Perceptual Effects of Noise in Digital Video Compression

Charles Fenimore and John Libert
National Institute of Standards and Technology
and
Stephen Wolf
Institute for Telecommunications Sciences

Abstract

We present results of subjective viewer assessment of video quality of MPEG-2 compressed video containing wide-band Gaussian noise. The video test sequences consisted of seven test clips (both classical and new materials) to which noise with a peak-signal-to-noise-ratio (PSNR) of from 28 dB to 47 dB was added. We used software encoding and decoding at five bit-rates ranging from 1.8 Mb/s to 13.9 Mb/s. Our panel of 32 viewers rated the difference between the noisy input and the compression-processed output. For low noise levels, the subjective data suggests that compression at higher bit-rates can actually improve the quality of the output, effectively acting like a low-pass filter. We define an objective and a subjective measure of scene criticality (the difficulty of compressing a clip) and find the two measures correlate for our data. For difficult-to-encode material (high criticality), the data suggest that the effects of compression may be less noticeable at mid-level noise, while for easy-to-encode video (low criticality), the addition of a moderate amount of noise to the input led to lower quality scores. This suggests that either the compression process may have reduced noise impairments or a form of masking may occur in scenes that have high levels of spatial detail.

1. Introduction

Digital video compression systems achieve bit-rate reduction by exploiting image information correlation within a single frame and between neighboring frames. The degree of correlation (and image compressibility) is reduced when noise is introduced. Examples of sources of noisy material include, archival material collected with low signal-to-noise tube cameras; modern digital, low-noise cameras operating in a low-light environment; and other degraded signal sources such as aging original film or video tape.[1]

In this study we investigated the effects of noise on an MPEG-2 compression system. The experimental setup for the measurements was based on ITU Rec. 500[2] as described in Section 0. The input test scenes were chosen for their variety, although they do not necessarily represent the full range of video of interest. Of the seven test clips used, one is in the public domain and available from NIST (*Wheels*) and two others are standard CCIR test materials (*Mobile and Calendar* and *Ballet Dancer*).

For noisy test scenes, the output of the MPEG-2 decoder can produce better subjective quality than its input since discrete cosine transform (DCT) filtering and higher-order coefficient truncation can behave like a low-pass filtering function. For this reason, a bipolar subjective quality scale was used where the quality of the input could be rated either higher or lower than the decoder output. Indeed, our data suggest that compression enhancement occurs, although the statistical significance of the effect is not especially high. We also find there is some ambiguity regarding the effects of increased noise on video quality and identify two possible mechanisms as sources of the ambiguity.

For some of our test materials the compression is nearly transparent, in a statistical sense. We find that the criticality (the difficulty of compression) of the video sequences has some predictive power for the bit-rate at which transparent coding occurs.[3,4] The Appendix to this paper details the basis for our definition of criticality.

2. Overview of the test plan

The primary purpose of the subjective experiment was to collect subjective viewer response data that could be used to construct an objective model of video quality for MPEG-2 video systems. For the purposes of this experiment, an MPEG-2 video system consisted of one pass through an MPEG-2 coder-decoder chain. The video input and output of this system conformed to ITU-R Recommendation BT.601.[5] In addition to examining the effect of bit-rate on perceived quality, another design factor that has been largely ignored in past experiments was included, namely, the effect on subjective quality of adding increasing amounts of noise to the input material. The range of added noise power was selected to produce a just-perceptible to slightly-perceptible change in video transmission quality. Viewers were given the task of rating the difference in quality between the input and output video. Figure 1 presents a conceptual block diagram of how each video clip pair (input, output) was generated for the subjective viewing experiment.

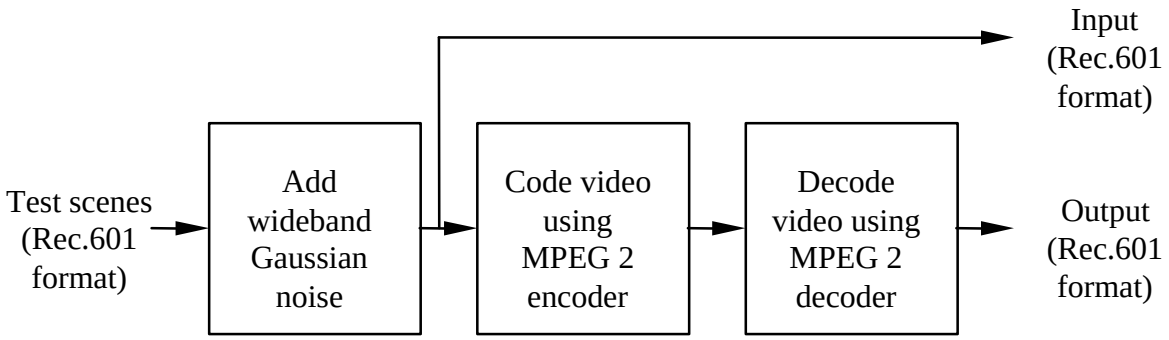

Figure 1. Block diagram for generating subjective test material.

2.1 Experimental variables

There were four experimental variables that contributed to the variability in the subjective data: (1) test scene, (2) noise level, (3) coding bit-rate, and (4) viewer.

2.1.1 Test scenes

Because the subjective perception of noise and the behavior of MPEG-2 systems are influenced by scene attributes such as spatial detail, amount and complexity of motion, brightness, and contrast, scenes that spanned a range of these attributes were selected. In addition to "natural" test scenes, we included one computer-generated test scene that was specifically designed to stress MPEG-2 systems. This computer-generated scene was specifically selected so that it was viewable (i.e., the range of motion and spatial detail was not excessive). Input material was selected to be of the highest quality that was readily available. The input material included some test scenes from the original Rec. 601 tests[6], scenes

produced with professional cameras and recorded onto ½-inch professional tape using a component format, and a computer-generated test pattern. Table 1 gives a description of the 7 scenes that were used for the experiment. Figures 2 and 3 display one frame from each of the clips used in this experiment, except for Ballet for which frames from both cuts are included. The length of each scene was 10 s, but the viewers only observed the center 9-second interval. The first and last 15 frames of each scene were eliminated to avoid possible coder transients at the beginning and ending of each clip.

Table 1 -- Description of Test Scenes Used in the Experiment

Scene Name (Abbreviation)	Description	Source
Mobile and calendar (Mobile)	Independent motion of many objects (e.g., red ball, toy train, calendar) against a highly detailed colorful background with a camera pan	Rec. 601 test material
Ballet dancer (Ballet)	Two ballet dancers against blue or brown backgrounds with camera pans and scene cuts	Rec. 601 test material from film
Grand Prix start (Start)	Start of a Grand Prix race -- colorful cars in foreground with detailed crowds in background and random camera motion	½-inch professional tape
Water bubbling (Water)	Ground level close-up of a bubbling stream in a forest with random camera motion	½-inch professional tape
One duck (Duck)	Close-up of a duck swimming and preening with scene cuts	½-inch professional tape
Taos boy with zoom (Boy)	Boy in Taos, NM in winter -- close-up shot with zoom-out to snow and blue sky	½-inch professional tape
Spinning color wheels (Wheels)	Three paddles in red, green, and blue form wheels which spin and move against a background with time-varying gray intensity levels	Computer-generated

2.1.2 Noise levels

Different levels of peak-signal-to-noise-ratio (PSNR) for the test scenes were achieved by adding wideband Gaussian noise to the Y (luminance) component channel. Noise was not added to the C_B and C_R chrominance channels in order to assess the increased MPEG-2 coding difficulty on the high-data-rate luminance component. The primary interest was to investigate the area where noise begins to produce perceptible, but slight, changes in MPEG 2 video system quality.

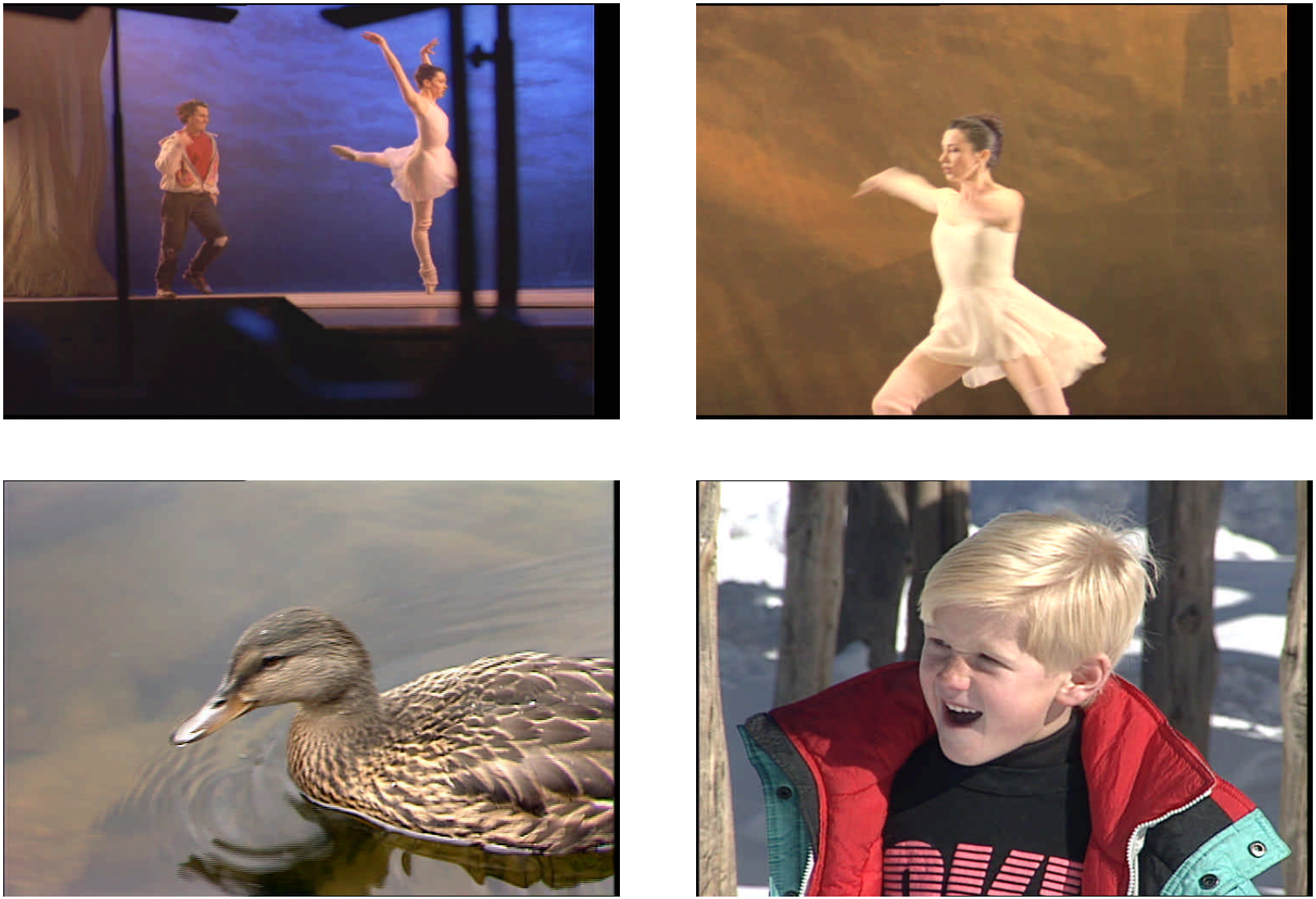

Figure 2. Two frames from the *Ballet* clip and single frames from *Duck* and *Boy*.

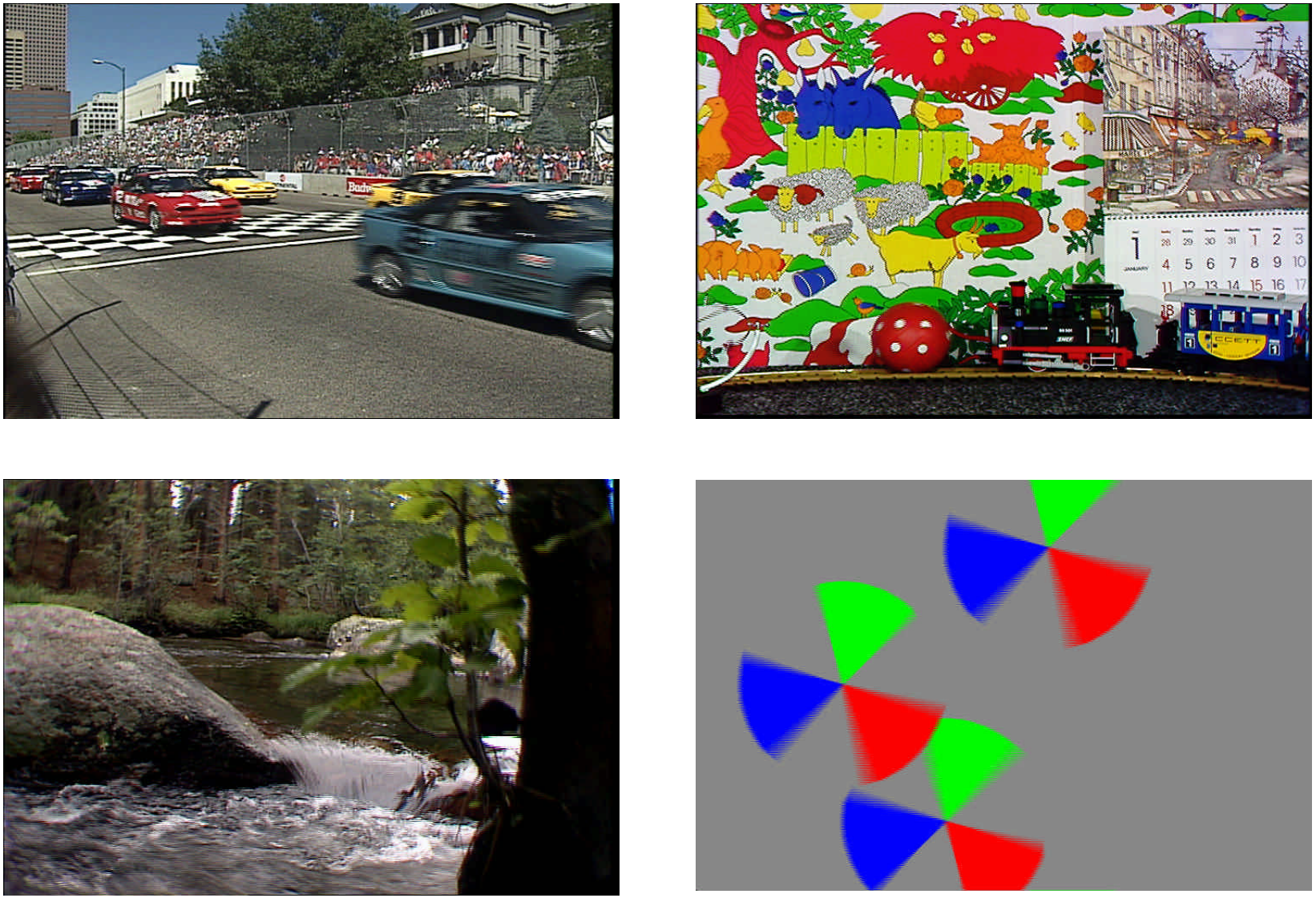

Figure 3. Single frames from test clips *Start*, *Mobile*, *Water*, and *Wheels*.

The direct method of Abramowitz and Stegun[7] was used to generate zero-mean Gaussian noise (i.e., N(0, σ^2)). The noise samples were added to the Y channel of the Rec. 601 video stream after conversion to floating point. The Rec. 601 format has some headroom (i.e., Y ranges from 16 to 235 for the 8-bit range [0, 255]) so small amounts of noise can be introduced without significant clipping effects.

Two independent Gaussian noise samples, n_1 and n_2, were generated from uniformly distributed [0, 1] noise samples u_1 and u_2 by:

$$n_1 = \sqrt{-2\sigma^2 \ln(u_1)}\cos(2\pi u_2)$$
$$n_2 = \sqrt{-2\sigma^2 \ln(u_1)}\sin(2\pi u_2)$$

The floating point Y-channel video samples with Gaussian noise added were rounded to the nearest integer and clipped at levels 1 and 254 (0 and 255 are reserved for synchronizing data in Rec. 601).

PSNR is often used to specify the signal-to-noise ratio of a video signal. This method has the advantage of removing the signal power (which varies from scene to scene) from the signal-to-noise-ratio (SNR) calculation so that a given SNR is indicative of some fixed amount of noise power. We calculate PSNR according to the following formula:

$$PSNR = 20\log_{10}\left[\frac{V_{peak}}{\sigma}\right],$$

in which σ is the standard deviation of the added Gaussian noise and $V_{peak} = 235$ is "peak white," following the convention of ANSI T1.801.03-1996.[8] Other alternative formulas for calculating SNR use true signal power, maximum peak-to-peak signal amplitude (which for our case would be 235-16 = 219), and frequency-weighted noise. For SNRs based on weighted noise, the frequency-weighting function is normally based on the NTC7 filter.[9] While weighted noise is sometimes used because the human visual system is less sensitive to high-frequency noise than low-frequency noise, the PSNR figures presented in this paper are presented as unweighted numbers for simplicity.

A total of three noise levels were included in the subjective experiment. The maximum PSNR was limited by the 8-bit sampling of Rec. 601 and by the inherent noise level of the input scenes before digital sampling. Table 2 summarizes the three noise levels (σ's in the above equations) that were used.

Table 2 -- Noise σ s of the Test Scenes

Noise Condition	Noise σ (Rec. 601 units)	Unweighted PSNR (dB)
1 (original source)	1 (estimated)	47.4
2	3.0	37.9
3	9.0	28.3

2.1.3 Coding bit-rates

To generate the MPEG-2 impairments, the Test Model 5 (TM5) software encoder (main level, high profile, interlaced mode of operation) and the corresponding decoder provided by the MPEG Software Simulation Group was used. The MPEG-2 video target bit-rate was varied to generate five different MPEG-2 conditions: (1) 1.8 Mb/s, (2) 3.0 Mb/s, (3) 5.0 Mb/s, (4) 8.3 Mb/s, and (5) 13.9 Mb/s. These bit-rates were selected to concentrate more systems at the lower bit-rates (bit-rates above about 8 Mb/s were expected to produce nearly imperceptible impairments).

2.1.4 Viewers

A total of 32 viewers were randomly drawn from a pool of employees working at the U.S. Department of Commerce Boulder Laboratories site. This pool consisted of about 2000 scientists and engineers. Randomly selected viewers were pre-tested to verify that they had normal visual acuity and color vision.

2.2 Subjective testing

A full factorial design was used for the subjective experiment (i.e., all possible combinations of test scene, noise level, and coding bit-rate were rated by all the viewers). This yielded 7 x 3 x 5 = 105 conditions that were rated by each viewer. In addition, three test conditions were repeated to obtain a measure of session and viewer variability, for a total of 108 conditions.

2.2.1 Subjective rating scale

The goal of the subjective experiment was to measure the change in perceived quality between the input and the output as shown in Figure 1. This is equivalent to measuring video transmission quality, rather than the absolute video quality of the output. For noisy input test scenes, it was thought that the output of the MPEG-2 decoder might actually have better subjective quality than the input. This was because preprocessing and/or DCT filtering (e.g., higher order coefficient truncation) in the MPEG-2 encoder could behave like a low pass filter function and act to remove visible noise in the input. In view of this consideration, the quality comparison scale given in Table 5 of CCIR Rec. 500-5,[2] and reproduced in Table 3 was selected for the subjective experiment. With this scale, the viewers are shown two versions of each clip (first A, then B) and asked to rate the quality of the second version (B) using the first version (A) as a reference. A subjective rating that falls on the zero point or center of the scale represents the condition where the first and second presentations are perceived as being of identical video quality. To assure that the viewers made full use of both sides of the scale, the presentation ordering of the input and output were randomized so that the input appeared first 50% of the time and the output appeared first the other 50% of the time.

Table 3 -- Subjective rating scale

Score	-3	-2	-1	0	1	2	3
Subjective classification	Much worse	Worse	Slightly worse	The same	Slightly better	Better	Much better

2.2.2 Presentation ordering and scene length

Figure 4 details how the A-B clip pairs were shown to the viewers. To reduce clip ordering effects that might result from having all the viewers see the clips in the same order, two random orderings were used (a "red" randomization, R, and a "green" randomization, G). To reduce fatigue, the viewing of each randomization was further split into two half-hour sessions separated by a break. For this purpose, the R and G randomizations were each spread over two viewing tapes having 54 clips each, with three repeated test conditions appearing in both tape viewing sessions. The four tapes, called R1, R2, G1, and G2, provided four possible clip orderings that were shown to a particular viewer (R1R2, R2R1, G1G2, and G2G1). Each viewer was randomly assigned a particular clip ordering. For balance, eight of the 32 viewers saw each of the four possible clip orderings.

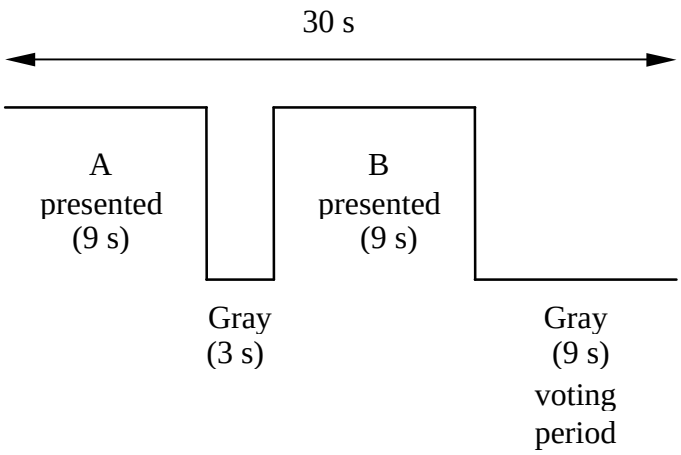

Figure 4. Layout of each clip pair on the video tape.

2.2.3 Training

The viewers were given a brief training session (less than five minutes) at the beginning of the test, whose purpose was to expose them to the range of impairments in the test, and to allow them to gain familiarity with the scoring procedure. After the training session, the experimenter made sure that the test subjects understood the scoring procedure before beginning the actual test.

2.2.4 Test facilities

Testing was performed using quiet audio-visual testing rooms, meeting Noise Criteria 30 [10], and associated audio-visual test equipment. The rooms were finished in light gray and measured approximately 2.7 m by 3.0 m. The viewers sat in a chair centered in front of a video monitor and placed at a distance of 4 times the picture height of the monitor. Viewers were tested one at a time to avoid unwanted distractions. The illumination of the back wall was adjusted to be approximately 0.15 times the peak luminance of the picture. The 20 inch broadcast quality monitor that was used had SMPTE phosphors. The setting of the color temperature was D65 and the monitor was calibrated with a color analyzer probe and SMPTE color bar.

3. Data analysis

The analysis of the subjective data proceeded by determining the behavior of the data for each of the experimental variables: the viewer variability, the compressibility of the various scenes using a measure of scene criticality, the changes in quality as the compression bit-rate increases, and the effect of increasing noise level on the quality of the video. The analysis used the mean opinion score (MOS) averaged over the viewer responses and the half-width 95% confidence interval (2 standard deviations of the MOS), C_{95}, for each of the 108 test clips. The randomization of the order of the input and output clips dictated a reordering of the data.

3.1 Consistency of viewer ratings

With few exceptions, C_{95} varied from a low of 0.11 quality units to a high of 0.37. The average was 0.24 quality units. Corrections were applied in two cases. In the first, errors in writing one of the test tapes led to repeating of the first field in place of the second field on four clips. Cross-comparisons with viewer scores of the same clips on other test tapes found an average negative offset of about 1 quality unit. Therefore, an adjustment of +1 quality units was made to scores on those 4 clips for the 8 viewers who rated the output video worse than the input video. This adjustment affected less than 1.0 % of the data. None of our conclusions depends on the adjustment. In the second case, it appeared that a single viewer suffered momentary confusion and reversed the ordering of the pair of clips. Evidence of this was a single score deviating by 5.65 quality units from the MOS for the clip. Other deviations did not exceed 3.25 quality units. This viewer was retested for this scene. The new score was not an outlier and it was used in our data analysis. No other corrections were applied. The narrowness of the confidence bounds demonstrates a high degree of consistency across viewers in this subjective experiment.

3.2 Scene criticality and compressibility

Criticality is a measure of the difficulty of encoding a scene. We employed two measures of criticality. One was a subjective measure that was derived from the subjective data while the other was an objective measure that was derived from computer-based processing of the sampled video images. The objective measure of criticality (*o*) is detailed in the Appendix [3] and is given by

$$o = \log_{10}\left\{mean_{time}\left[SI(t_n) * TI(t_n)\right]\right\}$$

where *SI* measures spatial detail, *TI* measures frame-to-frame image changes, and t_n indexes the frames of the video clip. The objective measure of criticality (*o),* which was developed using a set of ANSI-standardized test scenes (see the Appendix) was evaluated using the set of MPEG-2 test scenes described in this paper. The subjective measure of criticality (*s*) was calculated by taking the absolute value of the averaged MOSs for each test scene with a noise level σ =1 (i.e., MOSs were averaged over bit-rates for each scene), namely,

$$s = \overline{MOS}\,.$$

Table 4 presents the comparison results of *s* and *o*. Here, higher *s* numbers indicate more impairment and hence scenes that are more difficult to code. Figure 5 presents a scatter plot of the results. The coefficient of correlation was -0.89, indicating a fairly strong correlation between *s* and *o*. Most of the remaining unexplained variance is due to a single outlier (*Duck*). The elimination of the scene *Duck* lowers the coefficient of correlation to -0.96. In this scene, the duck's feathers contain high spatial

information that is rapidly changing. However, the rapid motion that produces this change prevents the eye from tracking the spatial detail.

Table 4 -- Subjective and objective measures of scene criticality for seven MPEG-2 scenes

Scene	Mobile	Start	Water	Boy	Wheels	Duck	Ballet
Subjective criticality, *s*	-1.34	-1.21	-1.15	-0.84	-0.78	-0.76	-0.62
Objective measure, *o*	3.86	3.71	3.33	3.13	2.77	3.32	2.67

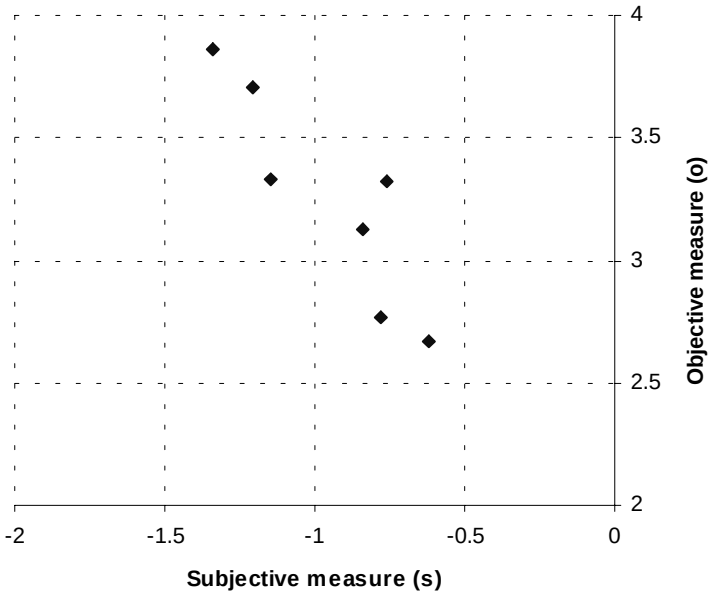

Figure 5. Plot of subjective and objective measures of scene criticality for 7 MPEG-2 scenes.

3.2.1 Transparent coding bit-rates

Using the subjective mean opinion score, MOS, and the half-width 95% confidence interval, we determined the range in the encoding bit-rates for which the MOS first goes to zero (in the sense that the MOS is not statistically significantly different from 0 at the 5 % level.) The subjective measure of criticality properly orders the sequences with respect to this bit-rate as seen in Table 5. This measure is admittedly crude and might be shown to be less effective with more refined increments in the encoding bit-rate.

Table 5 -- Criticality ranking also ranks by transparent coding rate threshold for seven MPEG-2 scenes

Scene	Mobile	Start	Water	Boy	Wheels	Duck	Ballet
Subjective measure, s	-1.34	-1.21	-1.15	-0.84	-0.78	-0.76	-0.62
Transparent coding bit-rate (Mb/s)	>13.9	>13.9	5.0 to 8.3	5.0 to 8.3	5.0 to 8.3	5.0 to 8.3	3.0 to 5.0

3.2.2 Quality may be "improved" by compression

For low criticality scenes the data suggest that there may be improvements to the video by compression. Table 6 shows the MOS for each test clip in the low noise case, $\sigma = 1$. In the lower right hand corner, at higher bit rates, there are several positive MOSs, although the data do not support statistical significance at the 95% level. For scenes with high criticality, the MOS does not go to zero at these bit rates. The data support the use of the bipolar quality scale in these subjective quality measurements. Without positive going scores the MOS scores have a negative bias.

Table 6 -- Mean opinion scores (and half-width 95% confidence intervals) for each low noise test scene at the 5 bit rates. Shaded cells have non-negative MOS.

Bit rate (Mb/s)	1.8	3.0	5.0	8.3	13.9
Mobile	-2.78 (0.17)	-1.94 (0.29)	-1.16 (0.31)	-0.59 (0.24)	-0.25 (0.18)
Start	-2.69 (0.16)	-1.38 (0.30)	-0.84 (0.27)	-0.69 (0.24)	-0.44 (0.26)
Water	-2.84 (0.16)	-1.81 (0.27)	-0.69 (0.19)	-0.16 (0.21)	-0.25 (0.23)
Boy	-2.28 (0.35)	-1.56 (0.32)	-0.25 (0.20)	-0.09 (0.14)	0.00 (0.20)
Wheels	-2.75 (0.15)	-0.84 (0.37)	-0.19 (0.16)	-0.19 (0.26)	0.06 (0.26)
Duck	-2.34 (0.23)	-1.00 (0.22)	-0.22 (0.15)	-0.16 (0.18)	-0.06 (0.17)
Ballet	-2.66 (0.17)	-0.59 (0.26)	0.16 (0.22)	0.13 (0.29)	0.03 (0.14)

3.3 Effect of noise level

For some of the scenes there is a combination of high spatial detail and motion which leads to relatively high criticality, particularly for scene *Mobile and Calendar* and for scene *Start*. In these cases there is a suggestion of improvement in the bit-rate-averaged MOS as the level of the input noise is increased from $\sigma = 1$ to $\sigma = 3$ (see Table 7.) For scenes with lower criticality the only effect of increasing noise is to

decrease quality, particularly for scene *Ballet*. For this low criticality scene, the compression impairments generated by the addition of noise are very noticeable. This suggests that in high criticality scenes either noise is being reduced in the compression process or compression impairments are being masked. At the highest noise level the effects of compression were generally no less noticeable to the panel than at low noise.

Table 7 -- Subjective measures of scene quality (MOS) (and half-width 95% confidence intervals) averaged over 5 compression bit rates for 7 MPEG-2 scenes, shown at 3 noise levels

Scene	MOS and C_{95} (noise σ =1)	MOS and C_{95} (noise σ =3)	MOS and C_{95} (noise σ =9)
mobile	-1.34 (.07)	-1.23 (.09)	-1.29 (.09)
start	-1.21 (.10)	-1.03 (.09)	-1.33 (.09)
water	-1.15 (.08)	-1.18 (.08)	-1.12 (.08)
boy	-0.84 (.07)	-0.80 (.08)	-0.87 (.10)
wheels	-0.78 (.10)	-0.70 (.07)	-0.64 (.08)
duck	-0.76 (.07)	-0.81 (.06)	-0.79 (.07)
ballet	-0.59 (.08)	-0.76 (.10)	-1.07 (.11)

4. Conclusion

We have presented results that suggest the effect of noise on the perceived quality of compressed digital video is not described by a simple monotonic function. In some cases, the detail in an image masks the impairments introduced by the compression process. For the low-criticality scenes ($s > -1.0$) that we studied the *MOS* becomes positive for low noise at some of the higher bit-rates, although there is no single combination of scene, noise level, and bit-rate for which this effect is statistically significant at the 95% confidence level. The data suggests that for a larger range of test materials and bit-rates, one may find that the quality measurement process will rate compression "impaired" video as superior to the input material. In a practical sense, the subjective measurement process can detect this effect by employing a bipolar measurement scale such as that used in the experiment described here. If the effect is deemed to be significant, a more fundamental problem arises concerning objective measurement technology. It is common for such techniques to rate any image change an impairment while viewer preference may rate such change an improvement. This conflict will have to be addressed by new objective measurement techniques.

5. Acknowledgment

This research was supported by intramural funding from the Digital Video on Information Networks Program of the Advanced Technology Program at the National Institute of Standards and Technology.

6. References

1. J. Hamalainen, et al. “Facts and Fiction: Some Aspects Regarding the Design of Digital Television Cameras Using CCD Image Sensors,” *Intl J Imaging Sys Tech*, 5, 314-322, 1994.

2. CCIR Recommendation 500-5, “Method for the subjective assessment of the quality of television pictures,” Recommendations of the CCIR, September, 1992.

3. S. Wolf and A. Webster, “Subjective and Objective Measures of Scene Criticality,” ITU Experts Meeting on subjective and objective audiovisual quality assessment methods, SG 12 document number JRG010, Turin, October, 1997.

4. I. Yuyama, et al., "Objective Measurement of Compressed Digital Television Picture Quality," *SMPTE Journal*, 107, 348-352, 1998.

5. ITU-R Recommendation BT.601, “Encoding Parameters of Digital Television For Studios,” Recommendations of the ITU, Radiocommunication Sector.

6. ITU-R Recommendation BT.802-1, Test Pictures and sequences for subjective assessments of digital codecs conveying signals produced according to ITU-R Recommendation BT.601-4.

7. M. Abramowitz and I. Stegun, *Handbook of Mathematical Publications*, p.953, Dover [1972].

8. ANSI T1.801.03-1996, “American National Standard for Telecommunications - Digital Transport of One-Way Video Telephony Signals - Parameters for Objective Performance Assessment,” Alliance for Telecommunications Industry Solutions, 1200 G Street, N. W., Suite 500, Washington DC 20005.

9. E.B. Crutchfield, Ed., National Association of Broadcasters Engineering Handbook, Seventh Edition, Washington, DC, National Association of Broadcasters, pp. 4.1-38 to 4.1-40, [1985].

10. L. Beranek, *Noise and Vibration Control*, McGraw-Hill, [1971].

Appendix: Measures of Scene Criticality

The difficulty of coding a video scene to achieve a constant perceived quality level increases with the amount of spatial detail and motion. This Appendix describes preliminary results of an investigation to derive a combined spatial-temporal metric for estimating scene criticality, or coding difficulty. This objective metric of scene criticality has several potential uses, including use as a tool for systematically selecting an appropriate range of test material without unnecessary duplication, and as a method for performing dynamic bit-rate allocation in a "constant quality, variable bit rate", statistically-multiplexed transmission channel.

The emphasis for the investigation was to determine if scene criticality could at least be coarsely estimated from the set of low bandwidth spatial information (SI) and temporal information (TI) features.[1,2] The advantage of using these particular features is that they are simple to compute in real time and they can be readily transmitted or stored as digital side information due to their extremely low bandwidth and data storage requirements. They thus have uses for automatically controlling and monitoring the behavior of digital video transmission systems. The SI feature that was examined here is given by

$$SI(t_n) = \mathrm{rms}_{space}[\mathrm{Sobel}(F(t_n))],$$

while the TI feature that was examined is given by

$$TI(t_n) = \mathrm{rms}_{space}[F(t_n) - F(t_{n-1})],$$

where $F(t_n)$ is the luminance-only video frame at time t_n, Sobel is the Sobel filter [3], and rms_{space} is the root mean square function over the entire valid image subregion. Preliminary results presented here indicate that a *coarse* model of scene criticality can in fact be derived using these simple image features. Obvious refinements that can be made to improve this coarse model of scene criticality include the use of more localized estimates of SI and TI, scene-cut masking, object segmentation, and object motion tracking (including the randomness of the direction of motion) that emulates human perception.

1. Subjective measure of scene criticality

In 1995, ANSI accredited committee T1A1 undertook an extensive subjective experiment that involved the subjective evaluation of 25 test scenes [4] injected into 24 different digital video systems for a total of 600 scene-system combinations. Most of the digital video systems were video teleconferencing systems that included a range of bit rates from 64 kb/s to 1.5 Mb/s. VHS recorded scenes and 45 Mbit/s encoded scenes were also used as two reference conditions. To obtain a subjective estimate of the scene criticality, we averaged the subjective scores for each scene across all viewers and digital video systems that were used in the test. This computed average is referred to as the scene main effect by statisticians and provides a measure of the portion of the mean opinion score that is due solely to the test scene. Since a wide range of digital video systems were used in this test, the scene main effect should also provide an estimate of the scene criticality. Scenes that are the most difficult to code will have a lower scene main effect (or average mean opinion score—MOS) while scenes that are easy to code will have a higher scene main effect. Table A presents a summary of the subjective measure of scene criticality (s) for the 25 test scenes. Since the subjective scores were derived using an impairment scale that ranged from 1 to 5 (where, 5 = "imperceptible", 4 = "perceptible but not annoying", 3 = "slightly annoying", 2 = "annoying," and 1 = "very annoying"), we see from the table that the scene main effect varied from "annoying" to somewhere between "slightly annoying" and "perceptible but not annoying." As expected,

the football scene (*ftball*) was the most difficult to code while a head and shoulders scene (*disguy*) was the easiest to code. The 25 points in Figure A were used to develop the objective model of scene criticality that is presented in this paper.

Table A -- Subjective and objective measures of scene criticality for 25 ANSI scenes

Scene Abbreviation	Scene Description	s (subjective measure)	o (objective measure)
Ftball	Football game	2.05	3.4
Cirkit	Circuit diagram, camera pan	2.16	3.75
2wbord	Two people at white board, scene cuts	2.33	2.69
Rodmap	Road map with hand and pen motion, camera pan	2.56	3.18
Smity2	Salesman at desk with magazine	2.56	3.43
Smity1	Salesman at deck with box	2.58	3.36
Flogar	Flower garden with windmill, camera pan	2.62	3.74
Washdc	Washington, DC, map with hand and pointer	2.63	2.82
Ysmite	Yosemite map & hand motion (intensity	2.73	2.77
Fredas	Fred Astaire tap dancing (black and white)	2.73	2.84
Split6	Split screen, six people	2.77	2.83
Intros	Introductions of people sitting at table, camera pans	2.8	2.69
Boblec	Bob’s lecture at chalkboard	2.86	2.59
3inrow	Men at table, camera pan	3.02	2.70
Vowels	Woman at whiteboard teaching vowels	3.1	2.85
Vtc2zm	Woman standing next to map with pointer, zoom and	3.14	2.88
Inspec	Woman at document camera	3.14	2.34
3twos	Two pairs of people, scene cuts	3.17	2.51
Susie	Susie on telephone	3.28	2.56
5row1	Five people in a row sitting at a table	3.37	2.44
Filter	Filter diagram on yellow pad with hand motion	3.51	2.43
Disgal	Female announcer	3.65	2.19
Vtc1nw	Woman sitting reading news story	3.66	2.13
Vtc2mp	Woman standing next to map	3.67	2.43
Disguy	Male announcer	3.68	2.16

2. Objective measure of scene criticality

Of several objective measures of scene criticality that were considered, the simplest that was developed, *o*, is given by the model

$$o = \log_{10}\left\{mean_{time}\left[SI(t_n) * TI(t_n)\right]\right\}$$

Values for this model were computed using a time window that was the same as the length of the video clips used in the subjective testing (nine seconds). The model measures the average value (over time) of the instantaneous frame-by-frame product of SI and TI. When a large amount of spatial-temporal gradient energy is present, the scene is difficult to code. The criticality number for this simple model is given in column *o* (objective measure) of Table A, while a plot of the performance of the model is given in Figure A. The coefficient of correlation between the objective and subjective measures is -0.82 (here, the objective model is negatively correlated to the subjective score since higher subjective scores indicate easier to code test scenes). Most of the remaining unexplained variance results from several outliers. Elimination of just one of these outliers (scene *2wbord*), which contains several scene cuts, lowers the coefficient of correlation to -0.87. The magnitude of the correlation achieved in the training phase is comparable to the correlation found in the test materials discussed in the body of this paper (0.89).The effect of scene cuts on coding difficulty cannot be explained by the simple objective model presented here.

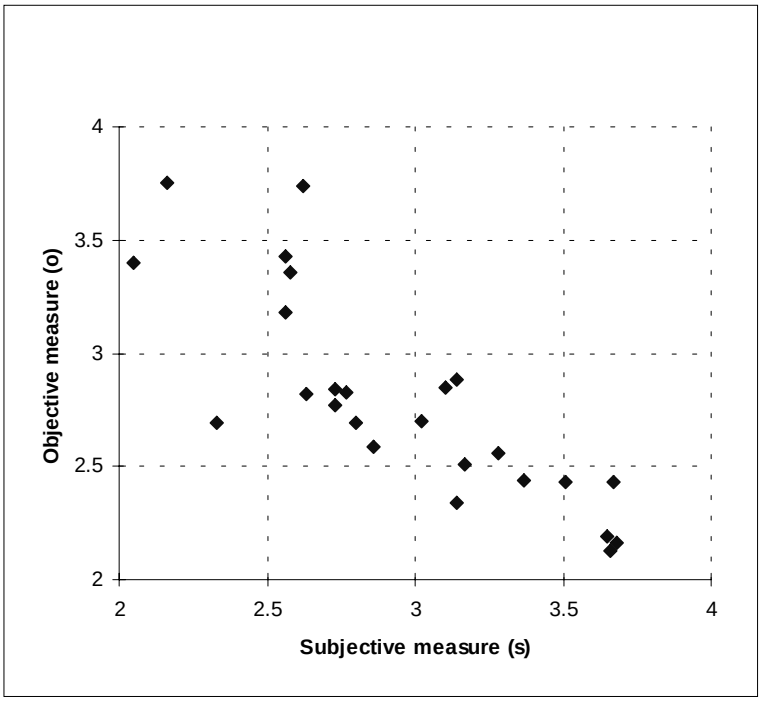

Figure A. Plot of subjective vs. objective measures of scene criticality for 25 ANSI scenes.

3. References

1. ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Recommendations of the ITU (Telecommunication Standardization Sector).

2. ANSI T1.801.03-1996, "American National Standard for Telecommunications - Digital Transport of

One-Way Video Telephony Signals - Parameters for Objective Performance Assessment," Alliance for Telecommunications Industry Solutions, 1200 G Street, N. W., Suite 500, Washington DC 20005.

3. R.C. Gonzalez and P. Wintz, *Digital Image Processing*, 2nd Ed., Addison-Wesley Publishing Co., Reading, Massachusetts, 1987.

4. ANSI T1.801.01-1995, "American National Standard for Telecommunications - Digital Transport of Video Teleconferencing/Video Telephony Signals - Video Test Scenes for Subjective and Objective Performance Assessment," Alliance for Telecommunications Industry Solutions, 1200 G Street, N. W., Suite 500, Washington DC 20005.