COMMITTEE T1
CONTRIBUTION

Document Number: T1Q1.5/91-124

****

STANDARDS PROJECT:	Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service

****

TITLE:	CORRELATION BETWEEN ITS OBJECTIVE MEASURES AND SUBJECTIVE VIDEO QUALITY: PRELIMINARY RESULTS ON A SET OF 15 SCENES

****

ISSUE ADDRESSED:	Objective Quality Assessment of VTC/VT

****

SOURCE:	National Telecommunications and Information Administration Institute for Telecommunication Sciences (ITS) (Stephen Voran)

****

DATE:	July 9, 1991

****

DISTRIBUTION TO:	T1Q1.5

****

KEYWORDS:	Video Teleconferencing, Video Telephony, Subjective Quality, Objective Quality

****

Correlation Between ITS Objective Measures and Subjective Video Quality: Preliminary Results on a Set of 15 Scenes

This document is provided as a companion to document T1Q1.5/91-123. That document provides a general progress report on the ITS Video Quality Project as it relates to VTC/VT performance measures.[1] This document provides a more detailed description of a key stage of the measurement development process: evaluation of the correlation between proposed measures and subjective ratings of video quality.

ITS has planned a detailed and comprehensive set of subjective video quality tests. The implementation of these tests is under way, but a complete set of test results will not be available for several months. In order to proceed with the correlation step of the measurement development process, we have elected to utilize the results of an experiment conducted by Fish and Judd.[2] Their team selected 5 NTSC encoded test scenes. Each scene consists of a three second still followed by five seconds of full motion video. They created two impaired versions of each scene: a VHS record-play cycle and a simulated codec operating at the DS1 signaling rate. The resulting 15 scenes were shown to 45 viewers who rated each scene in terms of its "distance from ideal".

The researchers provided our lab with a copy of their test scenes and their subjective data set. We applied our family of objective measures to the test scenes to create a companion objective data set. This involved the processing of roughly four seconds from the motion part of each of the 15 scenes, resulting in approximately 120 values for each of 92 candidate measures. (One value for each frame of the four second sequence.) We then reduced each temporal sequence of 120 measurements to a single measure by selecting the median value of the sequence. The median operator was chosen over the mean operator because it generates results that are more closely correlated to subjective video quality. The result of these steps is an objective data set that contains 92 measures for each of 15 scenes.

Next we performed a correlation analysis between the objective and subjective data sets, and an analysis within the objective data set. Correlation analysis detects monotonic relationships between data sets. As relationships become more monotonic and closer to linear, the coefficient of correlation tends towards ± 1. A correlation coefficient (across the 15 scenes) was computed between the mean subjective impairment score and each of 92 candidate objective measures. We found absolute correlation coefficients larger than .8 for 43 of the measures, but many of the measures are highly correlated with each other, indicating that all of them cannot contribute unique information to the prediction problem. If we select a subset of these 43 measures by requiring that the absolute correlation coefficient between every possible pair of members of the subset be less than .9, we find five measures in the subset. Two of these five are described by the following equations:

$$p_{77} = -\text{mean\_}\{\text{still}(\text{Sobel}(S)) - \text{still}(\text{Sobel}(D))\},$$

$$p_{60} = \alpha \cdot 20 \cdot \left|\log_{10}\left\{\frac{\text{std}(\text{still}(\text{Sobel}(S)))}{\text{std}(\text{still}(\text{Sobel}(D)))}\right\}\right| + (1-\alpha) \cdot 20 \cdot \left|\log_{10}\left\{\frac{\text{std}(\text{motion}(\text{Sobel}(S)))}{\text{std}(\text{motion}(\text{Sobel}(D)))}\right\}\right|,$$

where:
S is the original video frame,
D is the distorted frame,
mean_ is a mean computed using only negative pixels,
still takes only the still parts of the frame,
motion takes only the motion parts of the frame,
std is the standard deviation of frame pixel values,
α = (number of still pixels)/(total number of pixels).

The measurement p_{77} has been named "Edge Fraction Gained, Still Portion", because it quantifies edges in the distorted frame that are not present in the original frame. The measurement is restricted to the still portion of each frame. The second measure, p_{60} is called "Absolute Edge Energy Difference, Motion-Still Weighted". Here a logarithmic energy difference measure is computed for both the motion and the still portions of each frame. The measures are passed through the absolute value operator and then combined using the weighting factors α and $(1-\alpha)$, which indicate the relative amounts of stillness and motion in the frame.

The coefficient of correlation between subjective score and "Edge Fraction Gained, Still Portion" (p_{77}) is .96. For "Absolute Edge Energy Difference, Motion-Still Weighted" (p_{60}) the correlation value is .94. These correlation values indicate that, for this set of scenes, a simple, first order linear predictor would do a respectable job of predicting mean subjective impairment values from either measurement. Figure 1 is a scatter plot of the subjective impairment score and the value of p_{77} for each of the 15 scenes, along with the best fitting (least squares) line. Like p_{77} and p_{60}, the remaining three measures in the set of five utilize the Sobel filter. One measure, "Edge Fraction Lost, Motion Part" is a direct complement to p_{77}. A fourth measure is formed from the ratio of edge energy gained to original edge energy. "Absolute Edge Energy Difference" rounds out this preliminary set of five measurements.

While these preliminary results are encouraging, the data sets are much too small to draw any firm conclusions. On the other hand, the probability that a random measure will attain a correlation coefficient of .8 or greater across the 15 scenes is only .0003. This indicates that we are measuring and reporting real effects, not chance occurrences. Larger data sets present a greater measurement challenge and may yield lower correlation values. We are confident that we can expand and refine our preliminary set of measures to a create a set that provides a sufficient basis for characterizing video quality as perceived by viewers. We are in the process of developing prediction algorithms that utilize these measures to generate accurate, technology independent, objective video quality ratings.

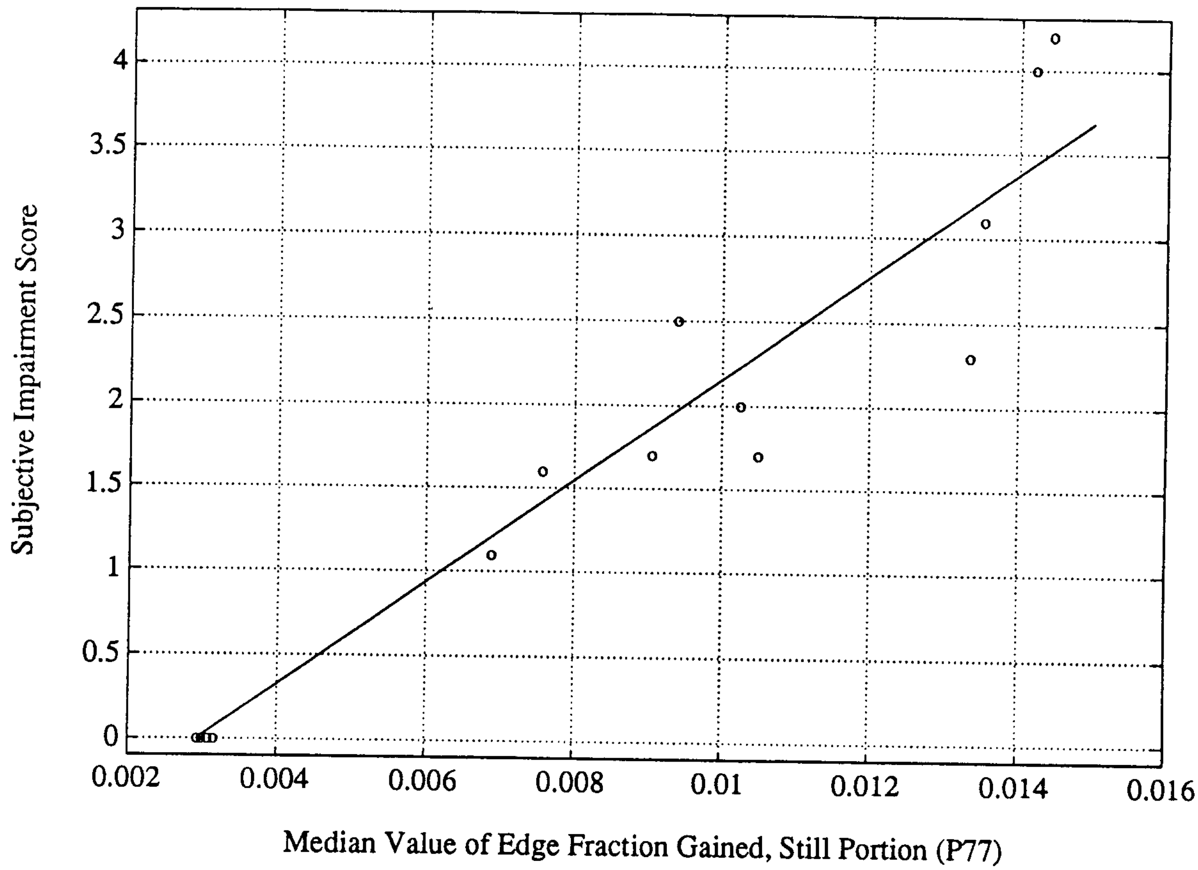

Figure 1.

Acknowledgements

Many persons have contributed directly and indirectly to the work described here. Major contributors to this research effort include Stephen Wolf, Margaret Pinson, and Arthur Webster. We express thanks to Robert Fish and Thomas Judd at Bellcore for providing us with video scenes and data from their experiment.

References

1. Institute for Telecommunication Sciences, Progress Report on Subjective and Objective Quality Assessment of VTC/VT Systems, Committee T1 contribution T1Q1.5/91-123.

2. R. Fish and T. Judd, A Subjective Visual Quality Comparison of NTSC, VHS, and Compressed DS1-Compatible Video, to appear in Proceedings of the Society for Information Display.