COMMITTEE T1
CONTRIBUTION

Document Number: T1A1.5/92-136

****

STANDARDS PROJECT: Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service and DS3 Television Project

****

TITLE: The Effect of Multiple Scenes on Objective Video Quality Assessment

****

ISSUE ADDRESSED: Objective Quality Assessment of VTC/VT

****

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

****

DATE: July 13, 1992

****

DISTRIBUTION TO: T1A1.5

****

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

****

The Effect of Multiple Scenes on Objective Video Quality Assessment

The Institute for Telecommunication Sciences, in support of T1A1.5, is conducting research to determine what objective measures of video impairment correlate well with human perception. This document uses the subjective and objective video quality data described in contribution T1A1.5/92-112 to demonstrate the effect of multiple scenes. In particular, we show that as the results from multiple scenes are averaged, prediction errors are reduced and objective results more closely approximate subjective results.

It is clear that the quality-of-service provided by a compressed digital video system is a function of the video signal that is being transmitted. This means that quality-of-service evaluations must be made using appropriate scenes. In many areas of science and engineering, the use of multiple measurements can lead to enhanced measurement accuracy. In the context of video quality measurements, this entails the use of an ensemble of scenes. If this ensemble of scenes provides a balanced sample of the video material that will actually be transmitted by the video service, then the mean of the quality measurements is a good indicator of how the service will actually perform. If the ensemble is representative, but not balanced, then a weighted mean can provide an approximation of the actual performance. This technique is equally applicable to subjective viewing measurements and objective electronic measurements.

As the number of relevant scenes in the ensemble of testing scenes is increased, the accuracy of the subjective measurements and objective measurements are both increased. If both measurements are converging to the same "true" value, it follows that the difference between them must go to zero. This convergence of subjective and objective measurements is a highly desirable situation. The following table and graphs demonstrate convergence between the subjective and objective measurements described in T1A1.5/92-112 as the scene ensemble size increases from one to five.

Contribution T1A1.5/92-112 describes tests using 36 scenes through 28 video systems, to generate a total of 128 impaired scenes. Of those 128 scenes, 64 were used to perform the linear regression that defines the objective measurement. The other 64 are referred to as "testing scenes". Since they are not used to determine the optimal linear measurement algorithm, they provide a fair test of that algorithm. Here we evaluate the performance of the algorithm on scene ensembles of size one, two and three, drawn from those 64 testing scenes. Due to the limited number of scenes, the characterization of the algorithm for ensemble sizes greater than 3 requires the use of some of the training scenes. We have done this for ensemble size 5 only.

The performance of the objective video quality assessment algorithm is shown in four scatter plots (Figure 1) and in Table 1. In the scatter plots, each data point has an abscissa equal to the mean of n subjective scores and an ordinate equal to the mean of the corresponding n objective scores, where n is the size of the scene ensemble. The members of each ensemble were randomly selected from the set of available scenes. As the number of scenes used to evaluate each video system is increased, the objective and subjective scores converge, and the scatter plots become tighter. Table 1 reports the RMS value of the error between the subjective and objective scores for the same four cases. From the third column of the table, it appears that the convergence has a tendency to follow the "one over root n" curve that one might expect.

We conclude that the objective and subjective measurements described in T1A1.5/92-112 converge as the size of scene ensembles grows. Assuming that the data used here is typical, Table 1 can be used as a design tool. It allows those who use these measurement techniques to select a scene ensemble size to attain a specified RMS prediction error.

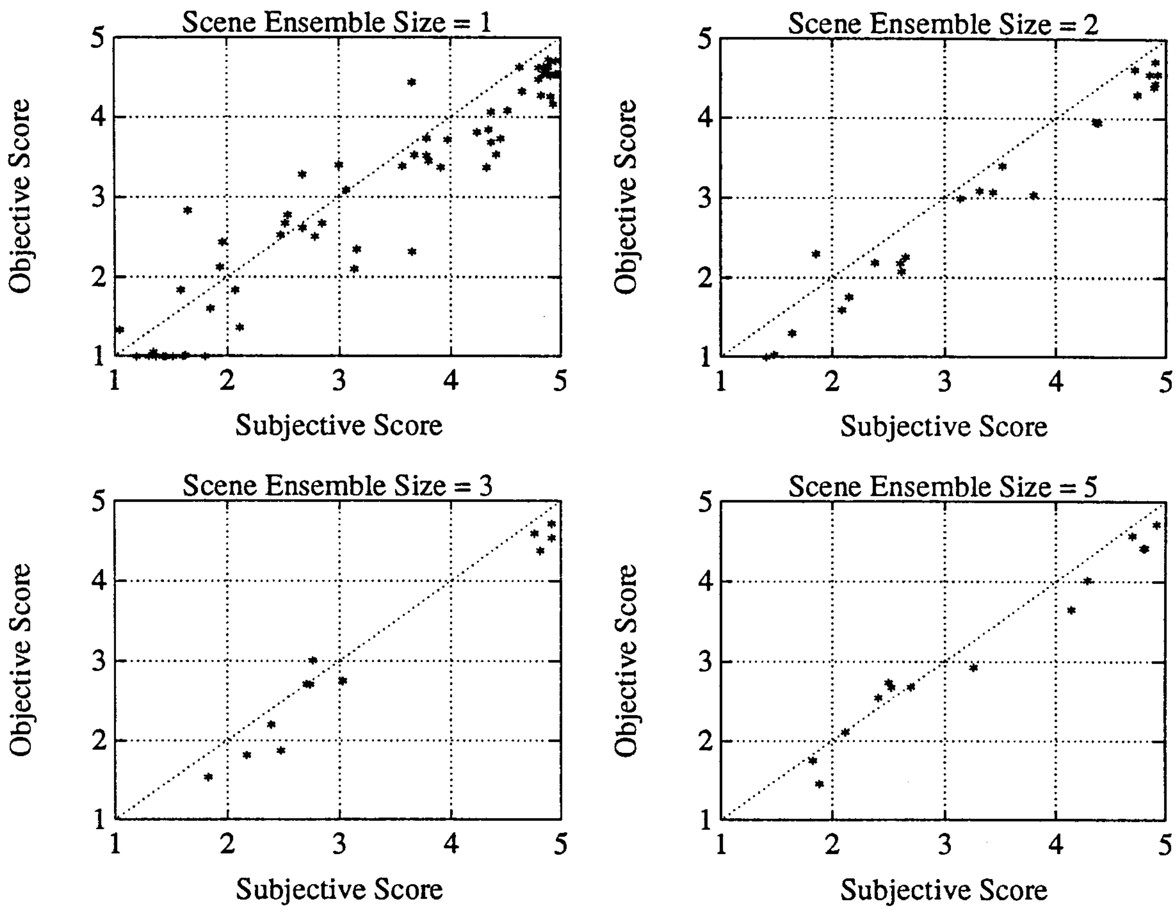

Figure 1

Scene Ensemble Size (n)	RMS Error	$e_1/n^{1/2}$	Number of Tests	Scene Distribution Testing / Training
1	.505=e_1	.505	64	100% / 0%
2	.397	.357	24	100% / 0%
3	.310	.291	12	100% / 0%
5	.274	.226	13	71% / 29%

Table 1