

Submission to G-2.6.1

**Progress Report of Task Force
to define
A Unit of Measure and Means of Calibration
for
Video Quality Analysis**

Author: Leon Stanger
11 Mar 98

Task Force Members:

Alan Godber
Paul Jones
John Libert
Bruce Lilly
Al Morton
Mihir Ravel
Leon Stanger

The task force members have discussed several concepts via e-mail on the assigned topic. Although the task force is not making recommendations at this time, the following is a brief compilation of subjects discussed and viewpoints presented.

Goal:

To define scale or unit of measure and a calibration method for video quality impairments. The unit of measure must be defined in terms of human perception and calibrated by statistical analysis of viewer test results.

Application:

To apply the concept of a unit of measure, test sequences with varying degrees of impairment must be created. The range of test sequences must include common types of impairments found in compression systems. The test sequences can be stored on digital video tape which will be viewed and scored by analytical means to establish calibration benchmarks. Copies of tapes can be made available to interested parties for calibration verification of automated equipment. Reference tapes can also be used to establish the degree of picture degradation of compression/decompression equipment by simple visual comparison.

Unit of Measure:

One alternative is to use a multiple point scale indicating a continuum of picture degradation. The scale could use such terms as "just perceptible" through "objectional" indicating the degree of impairment observed by viewers.

Another approach is to use a singular unit of measure such as the just noticeable difference (JND) or just perceptible difference (JPD). Increasing degradation could be defined by multiple JNDs or JPDs. For example, a singular unit could be defined as the point where 75% of viewers prefer picture A over B. With monotonically increasing degradation, a second unit could be defined where 75% of viewers prefer B over C. Thus, from A to C is 2 units.

Both methods have benefits and pitfalls. The scale approach implies the degree of objectionableness which is more compatible with human perception. The difficulty with this method is defining the scale and units in quantifiable and statistically verifiable methods. For example, if the term “noticeable but not objectionable” were used, would that have the same meaning for different people?

The singular unit of measure method lends itself well to statistical validation for one unit since viewers are only concerned with preference (A or B). No subjectivity is required of the viewer so no interpretation of terms is needed. The difficulty lies in the labor required to find calibration points. For example, how many test runs are needed to find the point of degradation where exactly 75% of the viewers prefer A over B? A second drawback is that even though integer points can be defined and validated, a quantity between points such as 2.5 JNDs has no specific meaning.

Statistical Analysis:

Although no specific method has been discussed, it is generally agreed that observers should be trained or expert viewers. The viewing distance should be three screen heights so that artifacts can be seen easily.

Test Sequences and Anticipated Impairments:

Test sequences should expose impairments common to DCT based compression systems. Such sequences could include the following:

<u>Type of Test</u>	<u>Anticipated Impairment</u>
Uncorrelated motion (water or fire)	Blocking
Fast uniform motion (to stress search range)	Blocking
Complex scene with scrolling text	Mosquito noise
Nearly flat background with busy foreground	Contour lines
Slow pan of brick wall	Shimmering motion
Fades and Dissolves	Pixelization
Facial close-up with short movements	Pumping detail

Tape Storage Format:

An uncompressed CCIR 601 format such as D1 should be used. This will allow duplication of test material with minimal losses.