COMMITTEE T1
CONTRIBUTION

Document Number: T1Q1.5/90- 127

****

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

****

TITLE: PROGRESS TOWARD AUTOMATED VTC/VT QUALITY ASSESSMENT

****

ISSUE ADDRESSED: OBJECTIVE AND SUBJECTIVE QUALITY ASSESSMENT OF VTC/VT

****

SOURCE: NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION
INSTITUTE FOR TELECOMMUNICATION SCIENCES
(Stephen Voran and Stephen Wolf)

****

DATE: October 16, 1990

****

DISTRIBUTION TO: T1Q1.5

****

KEYWORDS: Video Teleconferencing, Video Telephony, Subjective Quality, Objective Quality, Levels of Service, Levels of Performance

****

1. Introduction

Further progress in T1Q1.5 toward validating objective VTC/VT performance measures will require accurate correlation of the objective measures with subjective test data. The Institute for Telecommunication Sciences (ITS) has made recent progress toward performing subjective and fully automated objective performance assessment of VTC/VT. This contribution provides an update on the progress of the ITS video quality measurement program. First, the contribution gives a description of the recently completed video viewing laboratory. Next, plans for conducting subjective video quality measurements are discussed. Then, a method for using these subjective measurements to validate the VTC/VT objective video quality measures is presented. Finally, a discussion on the relationship between ANSI quality objectives for international VTC/VT service (see ANSI contribution T1Q1.5/90-118 Rev 02, presented to CCITT as a proposed USA position) and the ITS approach to video quality measurement is given. To the extent possible, ITS will use the newly developed facilities to support T1Q1.5 efforts in validating objective VTC/VT performance measures.

2. Video Viewing Laboratory

The ITS video viewing laboratory was completed in September of this year. This laboratory provides ITS with a location for conducting subjective measurements of video quality in accordance with CCIR Recommendation 500-3. Recommendation 500-3 specifies a standard visual environment for conducting picture quality assessment. The ITS facility conforms to these recommendations in all respects except for the color temperature of room illumination. The 14 by 11 by 8 foot room is finished with full length white drapes on three sides and grey drapes on the fourth (rear) wall. Grey carpet completes the subdued visual environment and provides noise reduction. An offset stud wall reduces noise transfer from an adjacent office area and a specially designed air handling system operates with a minimum of noise.

The six light fixtures in the laboratory are independently steerable and dimable. This flexible lighting feature allows one to obtain the specified screen-to-background luminance ratio while simultaneously providing comfortable lighting conditions for note taking. Recommendation 500-3 indicates that a light source with a color temperature of 6500 degrees Kelvin is preferred, presumably to match NTSC white. Several problems (power requirements, heat dissipation, bulb life, fixed light output) seem to be intrinsically linked to such sources. As a practical matter, ITS chose to use tinted incandescent bulbs. These bulbs have a maximum color temperature of 4000 degrees Kelvin. This means that the "white" room illumination does not match NTSC white. Given the adaptive nature of human color perception, ITS expects the effects of this fixed background color error to be minimal.

Video scenes are displayed on a Sony BVM-1910 broadcast monitor. The monitor can accept component video signals and display them with a maximum of 900 lines of horizontal resolution on a 19 inch screen. The monitor was purchased with an option that allows repeatable monitor setup which is referenced to digitally stored values. This feature eliminates monitor setup as a potential source of variation in ITS video quality assessment tests. Comfortable seating for three viewers is provided at a distance of six picture heights from the monitor screen. By restricting the number of viewing locations to three, the viewers in the end locations are only 20 degrees off center and potential test variations due to viewer location are minimized. The laboratory is equipped with hidden studio quality speakers to allow for recorded voice instructions as well as future subjective video/voice quality assessment tests.

3. Plans

ITS will use the new video viewing laboratory to gather subjective video quality measurements for broadcast quality video and degraded video. Degradations in subjective video quality due to VTC/VT transmission channel impairments will be thoroughly investigated. The end goal is to provide automated, objective video quality assessment algorithms that correlate with subjective video quality judgements. Toward this end, ITS plans to combine the results of its subjective and objective video quality measurement programs and to provide the relevant methods and results to T1Q1.5.

3.1 Subjective Video Measurements

ITS will invite subjects selected at random from the site telephone directory to participate in the subjective quality video measurements. This gives a prospective pool of 1750 viewers. The pool of viewers includes maintenance workers, office workers, administrators, scientists, and engineers. The use of a large, random sample of viewers from this diverse pool should substantially reduce any occupational bias that might be present. By recording the occupation of each viewer, subjective quality measures within occupation can also be calculated.

In accordance with CCIR Recommendation 500-3, each viewer will be given a visual acuity test and a color vision test. The results of these tests will become part of the viewer's data record. Groups of three viewers will then be seated in the video viewing laboratory for the subjective judgement session. The session is entirely contained on a Betacam SP (registered trademark of the Sony Corporation) cassette and lasts for approximately 20 minutes.

Each session will begin with an introduction that explains the purpose of the session. Next a training segment will expose the viewers to a random sequence of scenes which span the entire range of video quality levels. In the main portion of the test session, the viewers will be asked to rate the video quality of each scene by marking a seven (or more) point scale on a scoring form. Scene durations will be on the order of ten seconds, with a slightly shorter scoring time between scenes. The exact timing and rating scale for the ITS sessions are yet to be determined. ITS is currently conducting a pretest to determine what speeds and scales are, in the viewers' opinions, most likely to provide accurate responses.

ITS anticipates that each session will include roughly 40 scenes. To avoid training and fatigue effects, each viewer should participate in no more than four such sessions. This allows for the viewing of 160 scenes. Roughly 40 scenes will be repeated in order to provide intersession and intrasession redundancy. This leaves 120 distinct scenes. The likely format for these 120 scenes is ten versions of twelve source scenes, in random order. To provide for a reference point, one version will be chosen to be the undegraded broadcast quality source. The nine remaining versions will include codec degradations (several VTC/VT codecs at several channel rates), VHS degradations, and other digital/analog transmission channel impairments (digital errors, analog noise, etc.). The distribution of viewers scores, the number of viewers, and the number of times each scene is shown are data that will allow ITS to compute average subjective scores and confidence intervals in accordance with accepted statistical techniques.

3.2 Critical Issues for Scene Selection

The selection of source scenes is an important issue. In particular, the spatial and temporal information content of a scene are critical parameters. These parameters play a crucial role in determining the amount of video compression that is possible, and consequently, what level of video quality is attainable when the scene is transmitted over a fixed rate digital channel.

Video compression schemes attain various degrees of compression by removing the spatial and temporal statistical redundancies of the video signal. A highly detailed scene is said to have high spatial content. If the scale of the details is comparable to the scale of the spatial sampling grid and the details are hard to predict, then the video signal that describes this scene has very little spatial statistical redundancy. In this case, little compression can be gained by removing the small amount of spatial statistical redundancy. On the other hand, a scene with few details or with highly predictable details can often be greatly compressed. A parallel situation exists in the time domain. Scenes with large amounts of unpredictable motion have little redundancy in their temporal statistics and little temporal compression can be expected. Scenes with no motion or limited motion are highly redundant in their temporal statistics and can often be greatly compressed.

The preceding paragraphs describe the direct links between spatial and temporal information content, potential for compression, and potential for transmission at a given rate. In light of these links, fair and relevant video quality measurements must use video scenes that have spatial and temporal information content consistent with the user applications or video services being provided. As an example, video images of a soccer game may contain too much spatial and temporal information to be relevant test images for VTC/VT video quality measures.

ITS intends to perform measurements on scenes with widely varying spatial and temporal information content. Figure 1 diagrams the relative information content of some potential video scenes by locating them in a two-dimensional space. As one moves to the right in Figure 1, one encounters video scenes with increasing spatial information content. As one moves up in Figure 1, the video scenes have more and more temporal content.

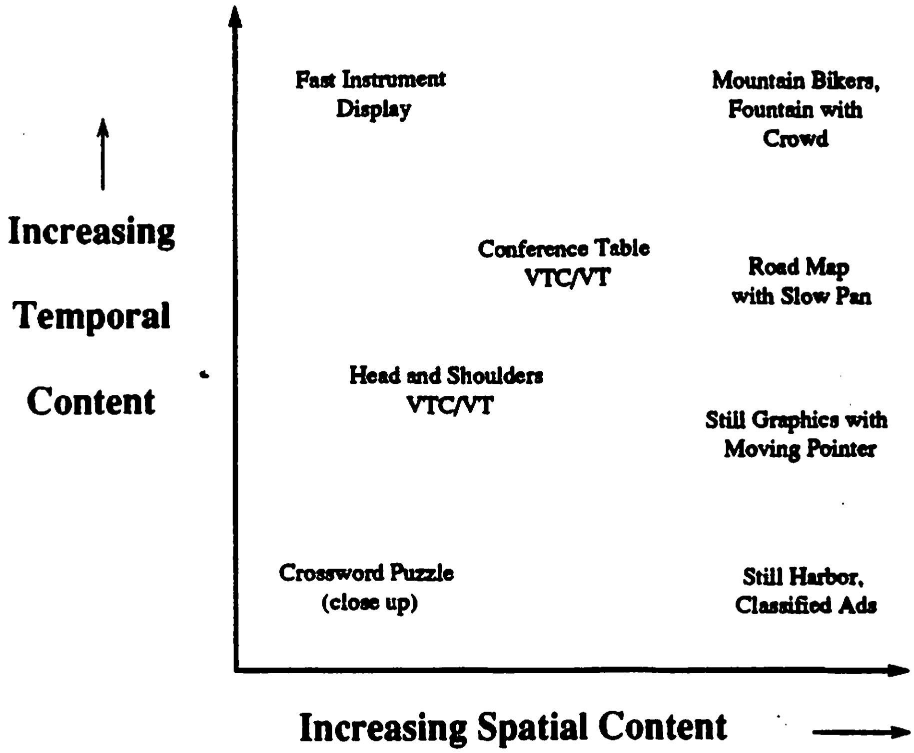

Figure 1. Video Scene Information Content

3.3 Integration with Objective Video Measurements

Once average subjective quality scores of acceptable variance are acquired, these scores will be compared with objective features, or parameters, that have been extracted from the same video scenes. The initial set of objective features are described in contribution T1Q1.5/90-123 entitled "Features for Automated Quality Assessment of Digitally Transmitted Video". Based on this comparison, ITS will derive a mapping from the multi-dimensional objective feature space to the one-dimensional subjective score space. Any objective measurements that prove to be redundant or not significant in the mapping will be discarded. By selecting a set of video scenes which span a wide range of quality levels and also span Figure 1, ITS hopes to generate a mapping that is accurate over a wide range of quality levels and scene types. This mapping, along with the automated video feature extraction hardware and software (nearly complete), will provide a fully automated estimate of perceived video quality.

4. Relationship to VTC/VT Levels of Performance

ANSI contribution T1Q1.5/90-118 Rev 02 (T1Q1/90-028), which proposes preliminary quality objectives for international VTC/VT service, was approved by Study Group A and forwarded to CCITT. In the appendix to that contribution, a levels of performance table specified three services (A, B, C):

A. Temporal Applications - For temporal applications, emphasis is placed on temporal positioning accuracy (or reduced jerkiness), possibly at the expense of reduced spatial resolution. This application group is concerned with the ability to accurately distinguish such items as facial expressions and lip movements in face to face and/or conference room settings.

B. Spatial Applications - For spatial applications, emphasis is placed on attaining high spatial resolution, possibly at the expense of reduced temporal positioning accuracy (or increased jerkiness). This application group is concerned with the ability to read small characters and see fine detail in still video and/or motion video which contains a very limited amount of motion.

C. Temporal/Spatial Applications - Both of the above.

For each of the services mentioned above (A, B, C), three levels of performance are available (Basic, Intermediate, and High). This gives a total of nine entries in the performance table. ANSI contribution T1Q1.5/90-118 also details two major VTC/VT performance characteristics; spatial resolution and temporal response. The placement of the nine performance entries within the two dimensional spatial resolution/temporal response space is yet to be decided. Ideally, the nine entries should be placed to accommodate the widest range of spatial and temporal applications.

For discussion purposes, Figure 2 suggests one such placement of the nine entries. The advantage of displaying the information in a two dimensional format is that the reader can more readily relate to the meaning of the nine entries when they appear in spatial resolution/temporal response coordinates. Note that there are potential "holes" in the two dimensional space. For example, some users desiring high resolution graphics presentation may want HDTV spatial resolution with a very low temporal response. Such a user must, as a minimum, purchase High B service, even though High B service may provide more temporal response than desired (and hence will probably also cost more than desired).

Considering the above discussion, T1Q1.5 should give much thought to where the nine performance entries are placed within Figure 2. Logically, the C service should have at least the temporal response of the A service and the spatial resolution of the B service. This condition is depicted graphically in Figure 2 with dashed lines.

The ITS video quality measurement system extracts objective measurements directly from the test scenes described in Figure 1. For any placement of the nine performance entries in Figure 2, relevant test scenes (that contain the proper mix of spatial and temporal content) can be selected. Using subjective test data to validate the objective measurements will insure that the objectively measured performance matches the perceived video quality.

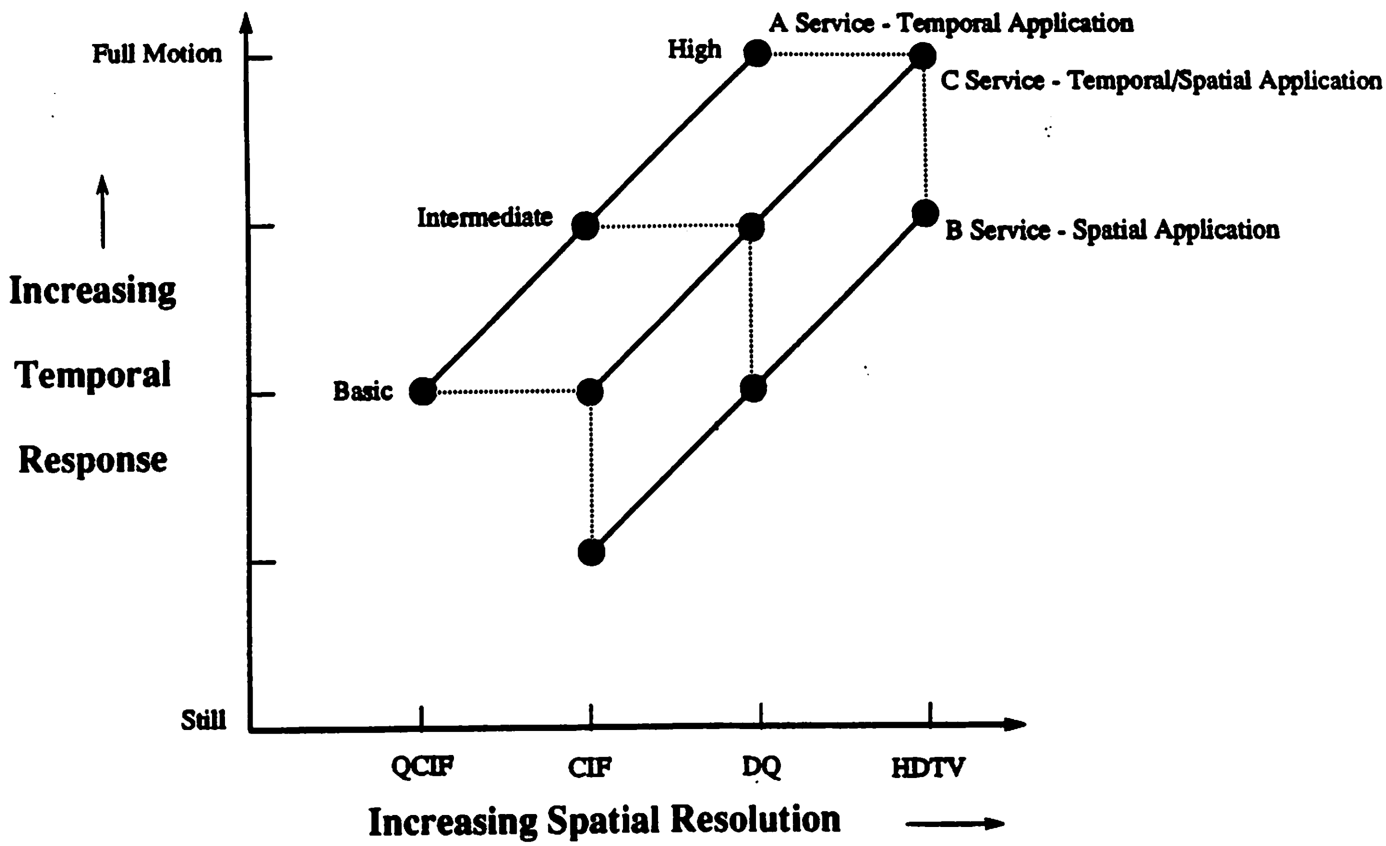

Figure 2. Levels of Performance Space

<u>Note:</u>

QCIF = Quarter (1/4) Common Intermediate Format
CIF = Common Intermediate Format
DQ = Distribution Quality component and composite signals
HDTV = High Definition TeleVision