COMMITTEE T1

CONTRIBUTION

Document Number: T1A1.5/95-107

****

STANDARDS PROJECT: VTC/VT Performance Standard

****

TITLE: Draft Proposed American National Standard - Digital Transport of Video Teleconferencing/Video Telephony (VTC/VT) Signals - Parameters for Objective Performance Assessment

****

ISSUE ADDRESSED: Objective Performance Parameters for VTC/VT

****

SOURCE: NTIA/ITS - Stephen Wolf (Editor)

****

DATE: January 9, 1995

****

DISTRIBUTION TO: T1A1.5

****

KEYWORDS: Video Quality, Video teleconferencing

****

Introduction

The Objective Measurements Ad Hoc Group met by audio teleconference at 2:00 PM Eastern Time on November 4 and November 7, 1994. Members in attendance included

Sam Agin, Bellcore
Gregory Cermak, GTE Labs
Eric Hauch, OIR-Telecommunication
Bill Hughes, NTIA/ITS
Dan Klenke, Compression Labs, Inc.
Dwight Melcher, NTIA/ITS
Alfred Morton, AT&T Bell Labs
Margaret Pinson, NTIA/ITS
Neil Randall, Delta Information Systems
Marshall Schachtman, Consultant-PictureTel Corp.
Greg Sherrill, Bell Atlantic
Frank Taylor, Bell South Telecommunications
Arthur Webster, NTIA/ITS
Dan Wirth, Bellcore (Convener)
Stephen Wolf, NTIA/ITS (Chair, and Editor)

One contribution for the Forward section of the draft standard was received from Marshall Schachtman. Two contributions were received for the Scope, Purpose, and Application sections; one from Alfred Morton and the other from Stephen Wolf. Using these three contributions as a basis for discussion, the Ad Hoc group reached consensus on the Forward, Scope, Purpose, and Application sections of the standard.

Inputs were then solicited from the Ad Hoc Group members regarding the Objective Parameters section. Members of the Ad Hoc Group suggested (1) that a brief overview of the parameters be given up front to familiarize the reader with the parameters (Contribution T1A1.5/94-148 was mentioned as a good source for this overview material), (2) that the parameters be divided into orthogonal families, and (3) that the visual impairments measured by each parameter be discussed. The Ad Hoc Group Editor agreed to collect these inputs and produce a framework for the Objective Parameters section by December 1. On December 2, the entire document was faxed and E-mailed to the members of the Ad Hoc group for further comments (closing comments were due by December 15 so that the document could be mailed to members of T1A1.5).

This contribution contains the text of the draft standard as of December 2, 1994, plus several editorial changes that were suggested by Ad Hoc Group members during the December 2 to December 15 time period (these suggested editorial changes are given in *italics*, alongside the original December 2 text, which is ~~struck out~~).

FOREWORD

Intensive industry efforts have been and are continuing to develop highly efficient coding methods for transmitting and storing video images. Coding algorithms have been developed and standardized in ANSI T1 as well as in international standards organizations. To date, the coding algorithms and the overall transmission performance of the combination of encoders, transport facilities and decoders have been evaluated by subjective methods, drawing upon standard methods of the ITU-R and other domestic and international subjective testing experience. Traditional analog measurements will tend to either under-estimate or over-estimate the relative picture quality delivered. ~~As a result of indications from Industry that a more objective means (reflecting an understanding of coding techniques) to assess overall system performance would be desirable, especially for use in continuing maintenance of installed services, T1Q1, and its successor, T1A1, has actively pursued a project for objective assessment of Video teleconferencing/Video telephony performance.~~ *Industry has indicated that a more objective means to assess overall system performance would be desirable, especially for use in continuing maintenance of installed services. The objective means should reflect an understanding of coding techniques. Therefore, T1Q1, and its successor, T1A1, have actively pursued a project for objective assessment of Video teleconferencing/Video telephony performance.* This effort has required significant effort on two tracks: 1) Determination by careful subjective testing of the relative quality of video transmissions which included a wide range of digital video impairments on a variety of test sequences. 2) Postulation of analytical models based on proposed key parameters which can attempt to predict or characterize useful metrics which can have practical application in the provision and maintenance of services.

Historically in voice and data services, subjective and (in the case of data) error ratio tests were originally the basis for selection of encoding and decoding processes. However, conceptual frameworks including additional useful parameters were established and verified through extensive testing to gain Industry acceptance as appropriate measures for the operation of systems in service. A similar approach is the basis for this Initial Framework Standard for Objective Video Performance Assessment.

The Initial Framework provided in this Draft Standard is based on the initial results of the studies in T1A1 mentioned above. Initial Parameters have been identified as potentially useful in a proposed model for overall performance. The Subjective Testing that has taken place, has been extremely useful for assessing the potential suitability of the proposed framework, and for bounding suitability of key parameters. The results to date have sufficient promise that the experts of T1A1 believe that the Framework can be presented to Industry as suitable for carefully controlled initial application.

While the Subjective tests to date have been probably more extensive than any previously reported in this area, it is clear that any proposed Framework and supporting set of key Parameters will greatly benefit from initial results reported as a result of real-world application. This Framework must be applied extremely carefully; it is believed that its most effective initial use will be in the monitoring of changes in performance of installed systems. However, such applications must be critically followed to assess the degree to which results are: 1) are reproducible, 2) measure changes in systems that result in visible artifacts and/or distortions, and 3) are insensitive to service images different from those used in the T1A1 test program. Key to any assessment of usefulness would be failure to detect system changes that produce visible degradation and/or false indication of degradation where no visible changes are observed.

It is expected that work will continue to refine and improve the Framework and the specification of the key parameters. Therefore, it is expected that this draft Standard will be reissued from time-to-time. Based on initial experience with measurement techniques, the question of appropriate standard limits will need to be determined and are, *and is a topic* for further study.

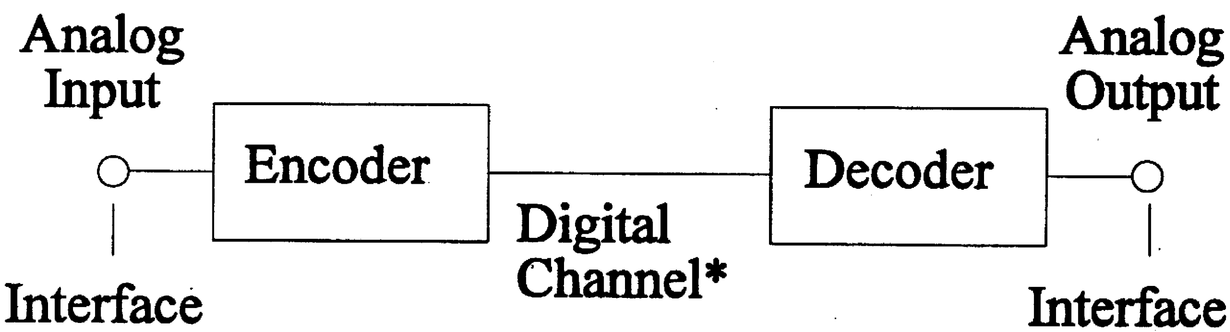
Global Change
"utilizing 525 line systems"

1. Scope, Purpose, and Application

1.1 Scope

This document covers the operational assessment of video teleconference/telephony systems on digital transport facilities utilizing M-NTSC and all related video formats and standards. It gives the measurement parameters which may be used to detect changes in the current status of a system when used in comparison with a set of reference measurements on the same system made under initial provisioning circumstances. Additionally, there are diagnostic parameters identified in this standard that may be utilized to characterize aspects of Video Teleconference/Video Telephony (VTC/VT).

This standard specifies methods of measurement for video performance parameters for VTC/VT transmission service channels that employ digital transport, as shown in Figure 1. The video performance parameters identified within this standard are defined for the end-to-end transmission quality between the interfaces shown. Those interfaces are between the VTC/VT transmission service providers and end users.



* A digital channel is implemented on a network composed of digital telecommunications components.

Figure 1 VTC/VT Transmission Service Channel

The measurements given here are intended to be used for the specific applications identified in the sections that follow. This document will be updated to include additional applications and measurements as research and development continues. Committee T1 will consider new methods and applications and add them here when there is industry consensus that they are appropriate. The scope of this standard does not include the following applications:

1) Measuring the following performance aspects of a VTC/VT system: audio, audio-visual interaction, and any implications of full duplex working (e.g. round-trip

delay that reduces the conversational spontaneity).

2) Discrimination between two or more systems beyond the accuracy of the objective measurements defined in this document. In such cases, subjective testing would be most appropriate. Definition of subjective test methods is beyond the scope of this document.

1.2 Purpose

The purpose of this standard is to assure the uniform application, provide a framework, and *provide* definitions of standard video performance parameters for VTC/VT video signals transported digitally by portions of the telecommunications network. This standard is intended to be especially useful as a basis for comparing the present operational readiness of a system with the same system's past performance. This standard is intended to provide a common understanding by manufacturers, carriers, and their customers.

1.3 Application

The primary applications for this standard are:

1) Detecting the continued operational readiness of VTC/VT systems on digital transport facilities utilizing M-NTSC and all related video formats and standards.

2. Objective Parameters

Two general categories of objective parameters have been identified:

(1) Parameters measured using artificial video signals.

(2) Parameters measured using natural video scenes.

Examples of category (1) parameters include signal-to-noise ratio measured with a constant IRE level test signal, and frequency response measured with a multi-burst or swept frequency test signal. Category (2) parameters that utilize natural video scenes are based on comparisons, just as the subjective measures are based on comparisons. A measurement is made on the input signal as it enters the transmission service channel shown in **Figure 1**, and an identical measurement is made on the output signal. Category (2) parameters are defined by comparing the input measurement with the corresponding output measurement according to explicit formulas. For category (2) parameters, each frame of each scene is captured and analyzed. The analyses can include measurements that quantify information in the time domain and/or the spatial domain.

2.1 Objective Parameters using Artificial Test Scenes

2.1.1 Gain

2.1.1.1 Definition. The gain is the ratio of the magnitude (amplitude) of a steady-state sinusoidal output relative to the causal input. The gain is separated into the following two factors: a proportional amplification which is frequency independent and associated with a dimensioned scale factor relating the units of input and output (denoted G_0), and a dimensionless factor which is frequency dependent (denoted G_f).

2.1.1.2 Method of Measurement.

2.1.2 Active Video Area

2.1.2.1 Definition. The active video area is the portion of the input active video area that is not blanked by the transmission service channel. The active video area is defined by the following four quantities: the uppermost active video line passed by the transmission service channel (denoted y_1, in standard NTSC line numbering), the lowermost active video line passed by the transmission service channel (denoted y_2, in standard NTSC line numbering), the left-most active video location passed by the transmission service channel (denoted x_1, in microseconds from the leading edge of line syncs, half-amplitude reference), and the right-most active video location passed by the transmission service channel (denoted x_2, in microseconds from the leading edge of line syncs, half-amplitude reference).

2.1.2.2 Method of Measurement.

2.1.3 Active Video Shift

2.1.3.1 Definition. The active video shift is the amount of vertical and horizontal shift of the output active video area with respect to the input. The active video shift is defined by the following two quantities: the vertical shift (denoted y_s, in lines where a positive shift indicates a shift downward), and the horizontal shift (denoted x_s, in microseconds where a positive shift indicates a shift to the right).

2.1.3.2 Method of Measurement.

2.2 Objective Parameters using Natural Test Scenes

2.2.1 Parameters Based on Analysis of Temporal Changes (Motion)

A set of parameters can be defined that are based on comparing the input and output statistics of temporal changes to the image pixels. These temporal statistics are indicators of the amount of temporal change, or motion in the video scene from one frame to the next. Computation of these temporal statistics for a video sequence is performed as follows. Each frame of the luminance signal is digitized, pixel-by-pixel according to CCIR Recommendation 601. This sequence of digitized frames will be denoted $Y(t_n)$, where t_n is ~~time~~ *the time at which frame n occurs.* The amplitude of a pixel in $Y(t_n)$ at column i and row j will be denoted $Y(i, j, t_n)$. To compute the temporal information (change) at time t_n for pixel $Y(i, j, t_n)$ (denoted generically as $TI(i, j, t_n)$), consider frame $Y(t_n)$ and one frame earlier in time, $Y(t_{n-1})$. Take the difference between those frames, pixel-by-pixel. Take the absolute value of those differences. This process for pixel $Y(i, j, t_n)$ is represented in equation form as

$$TI\,(i, j, t_n) = |Y\,(i, j, t_n) - Y\,(i, j, t_{n-1})|$$

To compute temporal information statistics for frame $Y(t_n)$, compute the mean and standard deviation of the $TI(i, j, t_n)$ values across all the pixels. This process is expressed mathematically as

$$TI_{mean}\,(t_n) = \frac{1}{N}\sum_i\sum_j\; TI\,(i, j, t_n)$$

$$TI_{stdev}\,(t_n) = \sqrt{\left[\frac{1}{N}\sum_i\sum_j\; TI\,(i, j, t_n)^2\right] - TI_{mean}^2}$$

Where N is the total number of pixels in the frame.

The above process is used to compute TI time histories for the input video scene and the corresponding output video scene from the VTC/VT transmission service channel in **Figure 1**. The objective parameters in this section are calculated by forming composite measures using two parallel streams of TI samples (one stream from the input and the corresponding stream from the output). These parameters can be viewed as indicating added or lost motion in the output scene compared to the original input scene. ~~Additions come~~ *Apparent added motion results* from impairments such as jerkiness, error blocks, and noise while ~~deletions come~~ *apparent lost motion results* from impairments like frame repetition (i.e., frozen motion).

2.2.2 Parameters Based on Analysis of Spatial Changes (Edges)

A set of parameters can be defined that are based on comparing the input and output statistics of spatial changes in the vicinity of image pixels. These spatial statistics are indicators of the amount of spatial change, or edges in the video scene. Computation of these spatial statistics for a video sequence is performed as follows. Each frame of the luminance signal is digitized, pixel-by-pixel according to CCIR Recommendation 601. This sequence of digitized frames will be denoted $Y(t_n)$, where t_n is time *the time at which frame n occurs*. The amplitude of a pixel in $Y(t_n)$ at column i and row j will be denoted $Y(i, j, t_n)$. To compute spatial information quantities for pixel $Y(i, j, t_n)$ (denoted generically as $SI(i, j, t_n)$) that are indicators of the amount of horizontal and vertical edges in the vicinity of that pixel, consider the two 3x3 edge enhancing filters shown in **Figure 2**.

-1	-2	-1
0	0	0
1	2	1

Horizontal

-1	0	1
-2	0	2
-1	0	1

Vertical

Figure 2 Horizontal and Vertical Edge Filters

The horizontal filter shown in **Figure 2** is used to enhance horizontal edges in the neighborhood of pixel $Y(i, j, t_n)$ while the vertical filter is used to enhance vertical edges. The output of the horizontal filter for pixel $Y(i, j, t_n)$ (denoted $SI_h(i, j, t_n)$) is obtained by centering the 3x3 horizontal filter over pixel $Y(i, j, t_n)$, multiplying the filter coefficients by the neighboring pixel values, and adding the nine values together. This is expressed mathematically as

$$\begin{aligned} SI_h(i, j, t_n) = {} & Y(i-1, j+1, t_n) - Y(i-1, j-1, t_n) \\ & + 2\,Y(i, j+1, t_n) - 2\,Y(i, j-1, t_n) \\ & + Y(i+1, j+1, t_n) - Y(i+1, j-1, t_n) \end{aligned}$$

Similarly, the output of the vertical filter for pixel $Y(i, j, t_n)$ (denoted $SI_v(i, j, t_n)$) is calculated as

$$
\begin{aligned}
SI_v(i, j, t_n) = {} & Y(i{+}1, j{-}1, t_n) - Y(i{-}1, j{-}1, t_n) \\
& + 2\,Y(i{+}1, j, t_n) - 2\,Y(i{-}1, j, t_n) \\
& + Y(i{+}1, j{+}1, t_n) - Y(i{-}1, j{+}1, t_n)
\end{aligned}
$$

The total spatial information for pixel $Y(i, j, t_n)$ (denoted $SI(i, j, t_n)$) is then defined as the summation of the absolute values of the vertical and horizontal filter outputs for that pixel, i.e.,

$$
\begin{aligned}
SI(i, j, t_n) &= \sqrt{SI_h^2(i, j, t_n) + SI_v^2(i, j, t_n)} \\
&\approx |SI_h(i, j, t_n)| + |SI_v(i, j, t_n)|
\end{aligned}
$$

To compute spatial information statistics for the entire frame $Y(t_n)$, compute the mean and standard deviation of the $SI(i, j, t_n)$ values across all the pixels. This process is expressed mathematically as:

$$
SI_{mean}(t_n) = \frac{1}{N} \sum_i \sum_j SI(i, j, t_n)
$$

$$
SI_{stdev}(t_n) = \sqrt{\left[\frac{1}{N} \sum_i \sum_j SI(i, j, t_n)^2\right] - SI_{mean}^2}
$$

Where N is the total number of pixels in the frame.

The above process is used to compute SI time histories for the input video scene and the corresponding output video scene from the transmission service channel in **Figure 1.** The objective parameters in this section are calculated by forming composite measures using these two parallel streams of SI samples. These parameters can be viewed as indicating added or lost edges in the output scene compared to the original input scene. ~~Additions come~~ *Added edges result* from impairments such as tiling, error blocks, and noise while ~~deletions come~~ *lost edges result* from impairments like blurring.

2.2.3 Parameters Based on Analysis of Scene Spatial Frequencies

A set of parameters can be defined that are based on comparing input and output statistics derived from the Fourier transform of frames in a video scene. Computation of these spatial frequency statistics for a video sequence is performed as follows. Each frame of the luminance signal is digitized, pixel-by-pixel according to CCIR Recommendation 601. This sequence of digitized frames will be denoted $Y(t_n)$, where t_n is ~~time~~ *the time at which frame n occurs*. (This introductory section to be completed)

3. Relationships Between Objective Parameters and Video Quality

3.1 Association of Impairments and Objective Parameters

Table 1 presents a list of impairment terms that are defined in ANSI T1.5MD (Multimedia/video Definitions) and an association of these impairment terms to objective parameters defined in this standard. Committee T1 considers that the objective parameters associated with a particular impairment may be useful for quantifying the perception of that impairment. Based on initial experience with the measurement techniques presented in this standard, the definition of explicit relationships between the parameters and perceived impairments will need to be determined and are for further study.

Table 1 Association of Impairments and Objective Parameters

Impairment	Objective Parameter(s)
Block Distortion, Tiling	Under Study
Blurring	Under Study
Color Errors	Under Study
Edge Busyness: Temporal Edge Noise Spatial Edge Noise	 Under Study Under Study
Mosquito Noise	Under Study
Error Blocks	Under Study
Jerkiness	Under Study
Object Persistence	Under Study
Object Retention	Under Study
Scene Cut Response	Under Study
Smearing	Under Study

3.2 Performance Model

The use of the objective parameters presented in this standard to model the subjective responses of a viewing panel is for further study. Committee T1 will consider models of performance for various applications and add them to this section when appropriate.