COMMITTEE T1

CONTRIBUTION

Document Number: T1A1.5/94-107

****

STANDARDS PROJECT:	**Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service**
TITLE:	**Draft Standard**
ISSUE ADDRESSED:	**Changes from January 94 Meeting**
SOURCE:	**Editor - Stephen Wolf (NTIA/ITS)**
DATE:	**March 16, 1994**
DISTRIBUTION TO:	**T1A1.5**
KEYWORDS:	
DISCLAIMER:	

****

Summary of Changes

This document includes changes to the VTC/VT draft standard which were approved by the T1A1.5 Working Group at the January 94 meeting. These changes can be found in contribution T1A1.5/94-116, as modified by the T1A1.5 Working Group.

American National Standard for Telecommunications -

Digital Transport of Video Teleconferencing / Video Telephony Signals -

System M-NTSC Analog Interface Specifications and Performance Parameters

1 Scope, Purpose, and Application.

1.1 Scope.

This standard covers analog interface specifications in the system M-NTSC format and performance parameters of Video Teleconferencing/Video Telephony (VTC/VT) transmission service channels employing digital transport. VTC/VT signals created or transmitted in accordance with other standards or formats may not necessarily be compatible with the specifications of this standard.

This standard specifies the performance of transmission service channels employing digital transport provided to convey VTC/VT signals and their associated audio signals only. Performance values are for a single coding (refer to Figure 1) and are allocated by grades of service. Performance definitions and measurement methods are provided if appropriate. Interface specifications are provided to facilitate compatibility between end users, service providers, and carriers.

The performance characteristics identified within this standard apply to the transmission quality between the defined interfaces. Those interfaces are between VTC/VT transmission service providers and end users. This standard defines neither the interconnection nor the performance characteristics of specific apparatus or equipment.

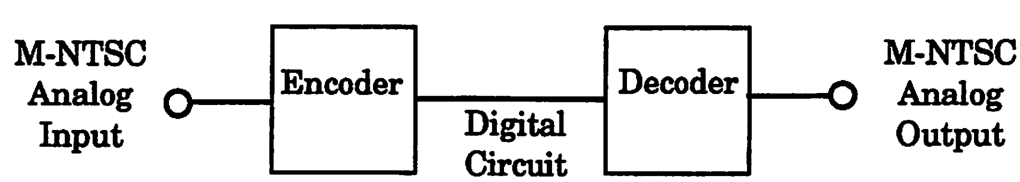


Figure 1 Digital Transmission Service Channel with Interfaces

1.2 Purpose.

The purpose of this standard is to assure the uniform application of standard values of transmission parameters for VTC/VT signals transported digitally by portions of the telecommunications network. It is intended to provide a common understanding by manufacturers, carriers, and their customers.

1.3 Application.

The primary applications of this standard are for specifying and evaluating the performance of a VTC/VT transmission service employing digital transport provided by common carriers. This service is used to transport the audio and video portions of the VTC/VT signals.

2 References and Related Standards.

2.1 American National Standards.

This standard is intended to be used in conjunction with the following American National Standards.

2.2 Other Related National Standards.

2.3 Other Related Standards.

2.4 Other Related Publications.

3 Definitions.

3.1 Special Word/Phrase Usage.

average picture level (APL): The average level of the picture signal during active scanning time integrated over a frame period and defined as a percentage of the range between blanking and reference white.

block distortion: Distortion of the received image characterized by the appearance of an underlying block encoding structure.

blurring: A global distortion over the entire image, characterized by reduced sharpness of edges and spatial detail.

coding: The digital encoding of an analog signal and decoding to an analog signal.

color errors: Distortion of all or a portion of the received image in which un-natural or unexpected hues appear.

digital transport: A portion of the telecommunication network using digital methods for the transmission of signals from one point to another to complete a transmission service channel. A transmission service channel may have one or more digital transport portion(s).

edge busyness: Distortion concentrated at the edge of objects, characterized by temporally varying sharpness or spatially varying noise.

error blocks: A form of block distortion where one or more blocks in the received image bear no resemblance to the current or previous scene and often contrast greatly with adjacent blocks.

jerkiness: Motion that was originally smooth and continuous is perceived as a series of distinct 'snapshots'.

longitudinally balanced: A circuit is perfectly longitudinally balanced if disturbing common mode voltages (V_c) result in zero differential voltage (V_d). For circuits that are not perfectly longitudinally balanced, the degree of longitudinal balance is reported in dB using:

$$\text{Longitudinal Balance} = 20 \cdot \log_{10}\left(\frac{V_c}{V_d}\right)$$

mosquito noise: Distortion typically seen around the edges of moving objects, characterized by moving artifacts around edges and/or blotchy noise patterns superimposed over the objects (e.g., a mosquito flying around a person's head and shoulders).

motion response degradation: The deterioration of motion video such that the received video imagery has suffered a loss of spatio-temporal resolution.

motion video: Video imagery that conveys movement.

object persistence: The persistence from earlier video frames of a piece or outline of an object which has subsequently moved (e.g., a piece of an object that has left the scene continues to appear in the received video imagery).

scene cut: Video imagery where adjacent frames are not highly correlated.

scene cut response: The perceived impairments associated with a scene cut.

smearing: A localized distortion over a subregion of the image, characterized by reduced sharpness of edges and spatial detail.

spatial application: An application needing high spatial resolution, possibly at the expense of reduced temporal positioning accuracy (or increased jerkiness). Example spatial applications include the ability to read small characters and see fine detail in still video or motion video which contains a very limited amount of motion.

spatial performance: A measure of the ability of a video transmission system to accurately reproduce still scenes.

still video: Video imagery that does not convey movement.

temporal application: An application needing high temporal resolution (or reduced jerkiness), possibly at the expense of reduced spatial resolution. Example temporal applications include the ability to accurately distinguish such items as facial expressions and lip movements in face to face or conference room settings - even though this may increase spatial distortion.

temporal performance: A measure of the ability of a video transmission system to accurately reproduce moving scenes.

tiling: See block distortion.

transmission service channel: A transmission service channel is the one-way transmission path between two designated points (analog in, analog out).

video: (1) The visually displayed images of video teleconferencing/video telephony. (2) Of or pertaining to the visually displayed images of video teleconferencing/video telephony.

video frame: A single frame of video composed of two interlaced fields.

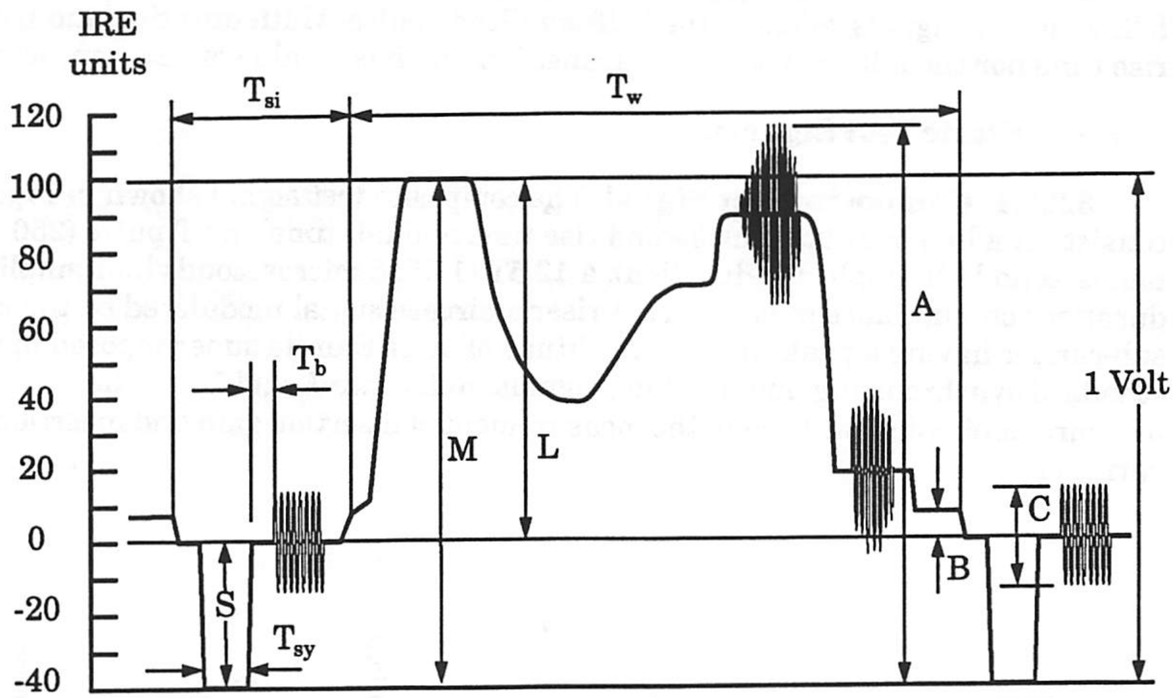
video imagery: A contiguous sequence of video frames.

video teleconferencing/video telephony motion artifacts: In a video teleconferencing/video telephony system, deteriorations of motion video due to image data compression that are observable by the viewer.

video teleconferencing/video telephony service (VTC/VT): The transmission of video signals capable of portraying motion and the accompanying audio signal(s) between two or more locations using digital transmission facilities. A typical example of this service is video teleconferencing between groups of personnel located at two or more locations.

3.2 Video Signals.

3.2.1 Video Signal Description. The waveform terminology used throughout the standard is in accordance with Figure 2, where the standard video signal waveform terminology is shown and measured in IRE units as shown in Figure 3.

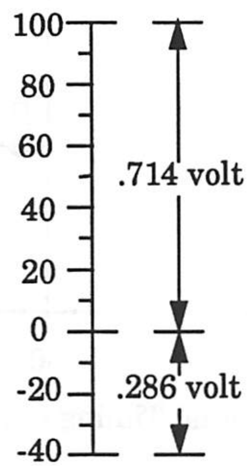


Waveform Terminology

A: The peak-to-peak amplitude of the composite color video signal
B: The difference between black level and blanking level (set-up)
C: The peak-to-peak amplitude of the color burst
L: Luminance signal - nominal value
M: Monochrome video signal peak-to-peak amplitude (M=L+S)
S: Synchronizing signal amplitude

T_b: Duration of breezeway
T_{si}: Duration of line blanking period
T_{sy}: Duration of line synchronizing pulse
T_w: Duration of active line period

Figure 2 Standard Video Signal General Waveform Terminology



(For a 1 V P-P composite signal)

Figure 3 IRE Unit Scale - Video

3.2.2 Test Signal Description. Reference to time (T) in the description of the following test signals refers to the half-amplitude pulse-width duration and not the rise time nor the fall time of a pulse transition and has a value of 125 nanoseconds.

3.2.3 Static Test Signals.

3.2.3.1 Composite Test Signal. The composite test signal shown in Figure 4 consists of a line bar (125 nanosecond rise time and fall time), a 2T pulse (250 nanosecond half-amplitude duration), a 12.5T (1.5625 microseconds half-amplitude duration) chrominance pulse, and a 5-riser staircase signal modulated by the color sub-carrier having a peak-to-peak amplitude of 40 IRE units superimposed upon standard synchronizing and blanking signals. Reference A and B are the measurement points utilized in the measurement of insertion gain and insertion-gain variation.

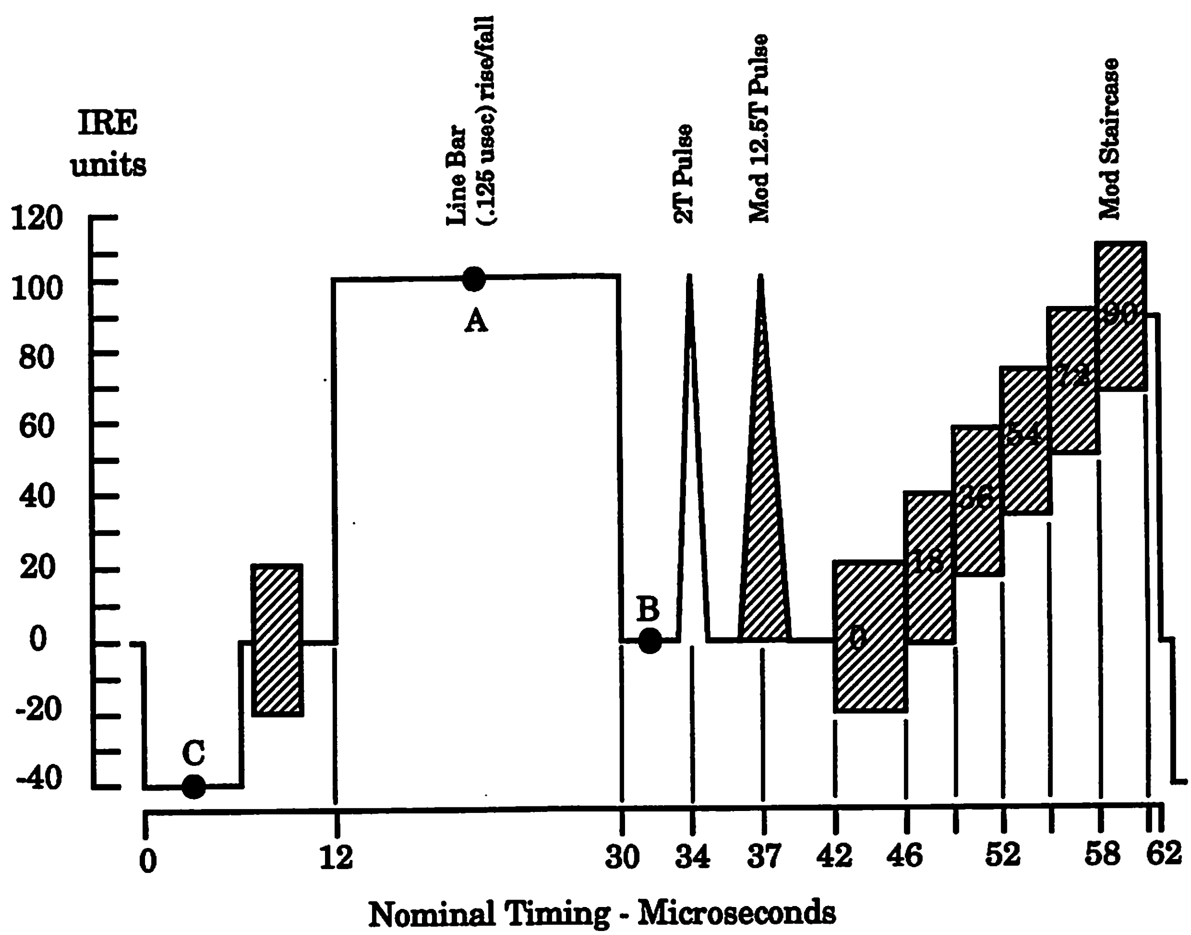


Figure 4 Composite Test Signal - Video

4 Baseband VTC/VT Interface Specifications.

4.1 Video Signal Electrical Interface Specifications.

4.1.1 Impedance.

4.1.1.1 Source Impedance.

4.1.1.1.1 Definition. The video source impedance of a transmission service channel, Z_s shown in Figure 5, is the impedance presented to the input terminals of a transmission service channel or other video baseband input point by the output terminals of the signal source. Proper source impedance is required for service channel evaluation.

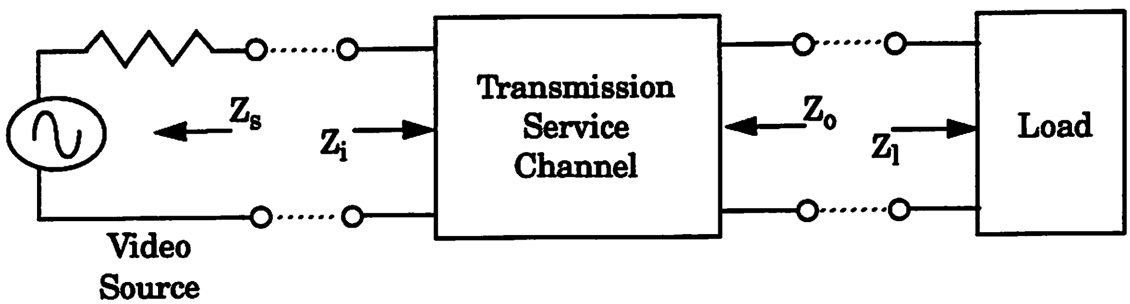


Figure 5 Impedance Reference - Unbalanced to Ground - Video

4.1.1.1.2 Standard Value. The standard value shall be 75 ohms, unbalanced to ground, with a return loss of at least x dB over the frequency range of 20 Hz - 4.2 MHz.

4.1.1.1.3 Method of Measurement. The impedance is measured by using impedance measurement equipment and the return loss is reported in dB using the following formula:

$$\text{Return Loss} = 20 \cdot \log_{10}\left|\frac{Z+Z_m}{Z-Z_m}\right|$$

Where

Z = specified standard impedance

Z_m = measured impedance

Alternately, the return loss may be measured using a return loss bridge.

4.1.1.2 Input Impedance.

4.1.1.2.1 Definition. The video input impedance of a transmission service channel, Z_i shown in Figure 5, is the impedance presented by the input terminals of a transmission service channel or other video baseband input point.

4.1.1.2.2 Standard Value. The standard value shall be 75 ohms, unbalanced to ground, with a return loss of at least x dB over the frequency range of 20 Hz - 4.2 MHz.

4.1.1.2.3 Method of Measurement. See section 4.1.1.1.3 on page 7.

4.1.1.3 Output Impedance.

4.1.1.3.1 Definition. The video output impedance of a transmission service channel, Z_0 shown in Figure 5, is the impedance presented by the output terminals of a transmission service channel or other baseband output point.

4.1.1.3.2 Value. The standard value shall be 75 ohms, unbalanced to ground, with a return loss of at least x dB over the frequency range of 20 Hz - 4.2 MHz.

4.1.1.3.3 Method of Measurement. See section 4.1.1.1.3 on page 7.

4.1.1.4 Load Impedance.

4.1.1.4.1 Definition. The video load impedance of a transmission service channel, Z_l shown in Figure 5, is the impedance presented by the input terminals of the device which will terminate the video baseband output of the transmission service channel. Proper load impedance is required for service channel evaluation.

4.1.1.4.2 Standard Value. The standard value shall be 75 ohms, unbalanced to ground, with a return loss of at least x dB over the frequency range of 20 Hz - 4.2 MHz.

4.1.1.4.3 Method of Measurement. See section 4.1.1.1.3 on page 7.

4.2 Video Signal.

4.2.1 Polarity of the Picture Signal.

4.2.1.1 Definition. The polarity of the picture signal of a transmission service channel is the sense of the potential of a portion of the signal representing a dark area of a scene relative to the potential of a portion of the signal representing a light area. Polarity is stated as "black positive" or "black negative". It is the polarity presented to the transmission service channel input terminals and presented by the transmission service channel output terminals.

4.2.1.2 Standard Value. The polarity of the picture signal is "black negative".

4.2.1.3 Method of Measurement. The polarity of the picture signal is determined by use of an oscilloscope or waveform monitor of known deflection polarity.

4.2.2 Input Signal.

4.2.2.1 Signal Level.

4.2.2.1.1 Definition. The input signal level of a transmission service channel is the difference in voltage between sync tip (-40 IRE units) and reference white (100 IRE units) of a composite video signal presented to the video baseband input terminals. It is expressed in volts.

4.2.2.1.2 Standard Value. The standard value shall be a nominal 1 volt peak-to-peak difference between sync tip and reference white, 140 IRE units (see Figure 2 on page 5 and Figure 3 on page 5).

4.2.2.1.3 Method of Measurement. The input signal level is measured by means of a properly calibrated and terminated oscilloscope or waveform monitor.

4.2.2.2 Time Base Error.

4.2.2.2.1 Definition. Time base error is defined as the difference between the instantaneous time base (which is the time between the 50% value of the leading edges of two successive horizontal sync pulses) and the time base averaged over one frame. The time base error defined here is typically that which is generated by a video tape machine.

4.2.2.2.2 Standard Value. The absolute value of the largest time base error shall be less than x microseconds.

4.2.2.2.3 Method of Measurement. The time base error is measured by applying the video signal to a time base error reading instrument.

4.2.2.3 Unweighted Signal to Noise Ratio.

4.2.2.3.1 Definition. The unweighted signal-to-noise ratio of the input signal is the ratio of the peak-to-peak luminance signal, blanking to reference white (nominally 0.714 volt = 100 IRE units), to the rms band limited noise level. The noise may be mixed random and quantizing noise. Synchronizing signals are not included in the signal measurement. It is measured at the output of the video source.

4.2.2.3.2 Standard Value. The standard value is greater than x dB. In addition, the input of the transmission service channel must tolerate noise that is out of band and could have voltage spikes several times that of the band limited noise.

4.2.2.3.3 Method of Measurement. The input signal of a transmission service channel is connected to the input of a rms reading meter through standard low pass filter x and high pass filter y. The video S/N is measured using a video line that is at a constant IRE level. The readings are averaged over a .4 second interval where synchronizing signals are excluded. This measurement is equivalent to the unweighted version of the signal to noise ratio measurement commonly known as NTC 7 3.16. The signal to noise ratio (SNR) is reported in dB using the following formula:

$$SNR = 20 \cdot \log_{10}\left|\frac{V_{p-p}}{RMS_{noise}}\right|$$

Where

V_{p-p} = peak-to-peak video signal amplitude

RMS_{noise} = RMS noise level measured as given in NTC 7 3.16

4.2.3 Output Signal.

4.2.3.1 Signal Level.

4.2.3.1.1 Definition. The output signal level of a transmission service channel is the difference in voltage between sync tip and reference white of a composite test signal presented by the video baseband output terminals. It is expressed in volts peak-to-peak.

4.2.3.1.2 Standard Value. The standard value shall be a nominal 1 volt peak-to-peak difference between sync tip and reference white, 140 IRE units (see Figure 2 on page 5 and Figure 3 on page 5).

4.2.3.1.3 Method of Measurement. Apply the composite test signal shown in Figure 4 on page 6 to the transmission service channel input. The output signal level is measured by means of a properly calibrated and terminated oscilloscope or waveform monitor.

4.2.3.2 Time Base Error.

4.2.3.2.1 Definition. Time base error is defined as the difference between the instantaneous time base (which is the time between the 50% value of the leading edges of two successive horizontal sync pulses) and the time base averaged over one frame.

4.2.3.2.2 Standard Value. The absolute value of the largest time base error shall be less than x microseconds.

4.2.3.2.3 Method of Measurement. See section 4.2.2.2.3 on page 9.

4.2.3.3 Non-Useful DC Component.

4.2.3.3.1 Definition. The non-useful DC component of the output signal is any DC component which is unrelated to the output signal. It will be present only as a result of the transmission equipment.

4.2.3.3.2 Standard Value. The standard value is ≤ x IRE units across a standard termination (75 ohms).

4.2.3.3.3 Method of Measurement. The non-useful DC component of the output signal is measured with a properly terminated DC coupled waveform monitor (WFM). Apply the composite test signal shown in Figure 4 on page 6 to the input terminals of the transmission service channel. With the WFM in a non-clamping mode, measure the signal at the output of the transmission service channel. Terminate the transmission service channel input in 75 ohm and measure the new position of the trace. The non-useful DC component is the difference in IRE units between the 0 and new trace position.

4.2.4 Input to Output Video Relationships.

4.2.4.1 Active Video Area.

4.2.4.1.1 Definition. The active video area is the portion of the input video signal that is not blanked by the transmission service channel.

4.2.4.1.2 Standard Value. The active video area shall be from x_1 usec to x_2 usec of video lines y_1 through y_2 inclusive.

4.2.4.1.3 Method of Measurement. Inject a video signal into the input of the transmission service channel with a marker at each of the following 4 locations: 1.) line y_1, x_1 usec, 2.) line y_1, x_2 usec, 3.) line y_2, x_1 usec, 4.) line y_2, x_2 usec. Check the output of the transmission service channel for the presence of the markers. If all 4 markers are observed, then the entire active video area is being passed by the transmission service channel.

4.2.4.2 Active Video Shift.

4.2.4.2.1 Definition. The active video shift is the amount of vertical and horizontal shift of the output active video area with respect to the input.

4.2.4.2.2 Standard Value. The maximum shift for any of the 4 markers shall be y lines, x usec.

4.2.4.2.3 Method of Measurement. Using the test signal of section 4.2.4.1.3, measure the locations (line, usecs) of each of the 4 markers at the output of the transmission service channel. Compute the absolute value of the shift for each of the 4 markers by comparing the output marker locations to the input marker locations.

4.3 Audio Signal Electrical Interface Specifications.

4.3.1 Balanced Audio Specifications. (The analog audio input to and output from the transmission service channel shall be balanced with respect to ground).

4.3.1.1 Input Common Mode Rejection Ratio (CMRR).

4.3.1.1.1 Definition. Input common mode rejection ratio indicates the degree to which unwanted signals coupled into both sides of a balanced line are rejected by the input of the transmission service channel.

4.3.1.1.2 Standard Value. The common mode rejection ratio for all audio input channels to the transmission service channel shall be greater than x dB within the respective service bandpass and at least y dB at 60 Hz. This specification must be met for common voltages up to x volts rms.

4.3.1.1.3 Method of Measurement. The audio input to the transmission service channel is connected to a center-taped 600 ohm resistive termination as shown in Figure 6. A common mode test signal with voltage V_c is applied between the center tap and ground. The voltage of the resulting signal at the differential output of the transmission service channel is then measured (V_d). The common mode rejection ratio (CMRR) is given by:

$$\mathrm{CMRR} = 20 \cdot \log_{10}\left(\frac{V_c}{V_d}\right)$$

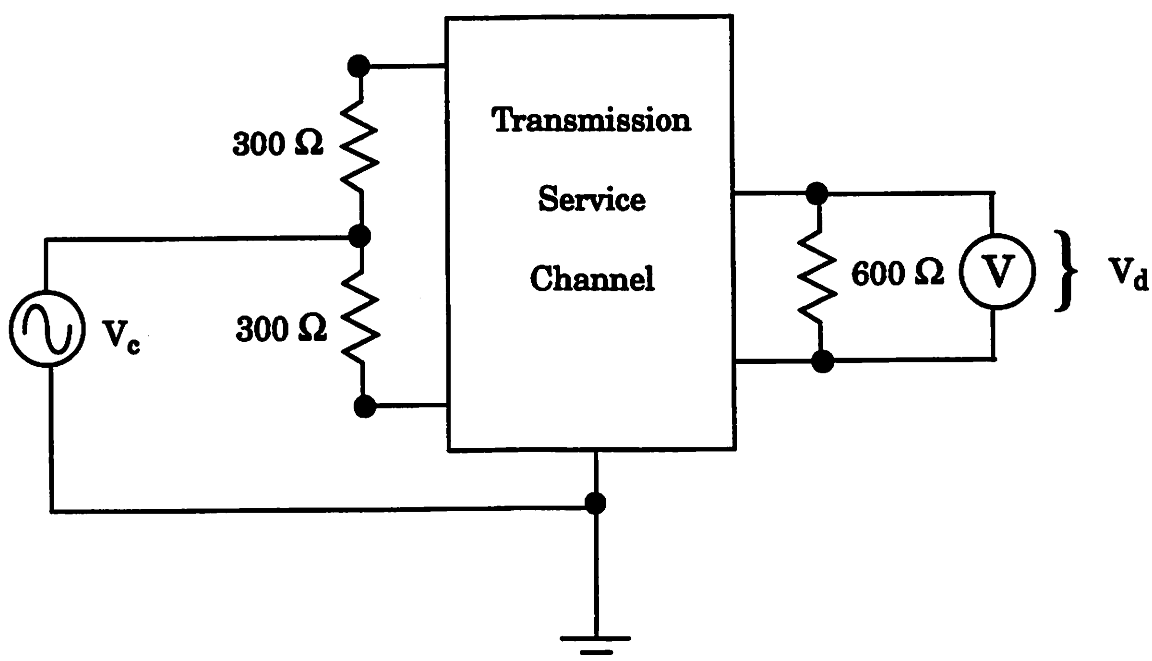


Figure 6 Measurement Set Up for CMRR

4.3.1.2 Output Common Mode Noise.

4.3.1.2.1 Definition. In a balanced audio system, any voltage that is common to both signal leads (measured with respect to ground) is common mode noise.

4.3.1.2.2 Standard Value. The common mode noise of all audio outputs from the transmission service channel shall be at least x dB below nominal operating level.

4.3.1.2.3 Method of Measurement. The audio output is terminated with a center-tapped 600 ohm load. The common mode noise is the voltage measured between the center tap of the 600 ohm load and ground.

4.3.2 Impedance.

4.3.2.1 Source Impedance.

4.3.2.1.1 Definition. The audio source impedance of a transmission service channel, Z_S, shown in Figure 7, is the impedance presented to the input terminals of a transmission service channel by the output terminals of the signal source. Proper source impedance is required for transmission service channel evaluation.

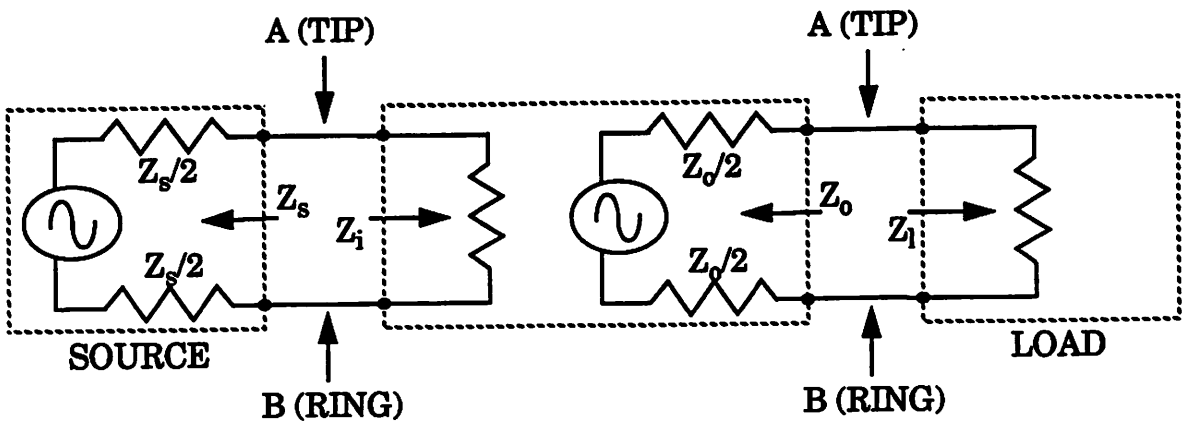


Figure 7 Impedance Reference - Balanced to Ground - Audio

4.3.2.1.2 Standard Value. The standard value shall be 600 ohms, nominally resistive, highly longitudinally balanced, and with a return loss of at least x dB over the frequency range of the respective service bandpass.

4.3.2.1.3 Method of Measurement. The impedance is measured by using impedance measurement equipment and the return loss is reported in dB using the following formula:

$$\text{Return Loss} = 20 \cdot \log_{10}\left|\frac{Z+Z_m}{Z-Z_m}\right|$$

Where

Z = specified standard impedance

Z_m = measured impedance

Alternately, the return loss may be measured using a return loss bridge.

4.3.2.2 Input Impedance.

4.3.2.2.1 Definition. The input impedance of a transmission service channel Z_i shown in Figure 7, is the impedance presented by the input terminals of a transmission service channel.

4.3.2.2.2 Standard Value. The standard value is 600 ohms, nominally resistive, highly longitudinally balanced, and with a return loss of at least x dB over the frequency range of the respective service bandpass.

4.3.2.2.3 Method of Measurement. See section 4.3.2.1.3 on page 13.

4.3.2.3 Output Impedance.

4.3.2.3.1 Definition. The output impedance of a transmission service channel, Z_0 shown in Figure 7, is the impedance presented by the output terminals of a transmission service channel.

4.3.2.3.2 Standard Value. The standard value is 600 ohms, nominally resistive, highly longitudinally balanced, and with a return loss of at least x dB over the frequency range of the respective service bandpass.

4.3.2.3.3 Method of Measurement. See section 4.3.2.1.3 on page 13.

4.3.2.4 Load Impedance.

4.3.2.4.1 Definition. The load impedance of a transmission service channel, Z_l shown in Figure 7, is the impedance presented by the input terminals of the device which will terminate the audio output of the transmission service channel. Proper load impedance is required for channel evaluation.

4.3.2.4.2 Standard Value. The standard value is 600 ohms, nominally resistive, highly longitudinally balanced, and with a return loss of at least x dB over the frequency range of the respective service bandpass.

4.3.2.4.3 Method of Measurement. See section 4.3.2.1.3 on page 13.

4.4 Audio Signal.

4.4.1 Input Signal.

4.4.1.1 Input Signal Level.

4.4.1.1.1 Definition. The input signal level to a transmission service channel is the signal level across the transmission service channel input impedance. When the signal is sinusoidal the input signal level is expressed in dBm.

4.4.1.1.2 Standard Value. The peak operating level of the input signal to the transmission service channel across the standard impedance is equal to the peaks of a sine-wave whose average power is 0 dBm. The nominal level is -16 dBm, and the minimum clip level is +9 dBm.

NOTE:

A. Clip level is defined as the level at which the total distortion exceeds 1%.

B. This 9 dBm level is only to be applied at the end users interface for program audio services as described in this document. This level is only used for an out-of-service test.

4.4.1.1.3 Method of Measurement. The audio input signal level is measured by properly terminated audio test equipment.

4.4.1.2 Non-Useful DC Component.

4.4.1.2.1 Definition. The non-useful DC component of the audio signal is any DC component which is unrelated to the audio signal and is also present when the audio signal is muted.

4.4.1.2.2 Loop Requirements. When the audio signal is muted, direct current flow shall be equal to or less than x mA through a zero ohm termination across the audio source output terminals.

4.4.1.2.3 Longitudinal Requirements. Direct current flow shall be equal to or less than x mA when the terminals are shorted together and the current flow is measured through a zero impedance to ground.

4.4.2 Output Signal.

4.4.2.1 Output Signal Level.

4.4.2.1.1 Definition. The output signal level of a transmission service channel is the signal level across the transmission service channel output impedance.

4.4.2.1.2 Standard Value. When a 1 KHz test tone at the nominal operating level of -16 dBm is applied to the input of the transmission service channel, the output level shall be -16 dBm plus or minus x dB.

4.4.2.1.3 Method of Measurement. See section 4.4.1.1.3 on page 15.

4.4.2.2 Non-Useful DC Component.

4.4.2.2.1 Definition. The non-useful DC component of the audio signal is any DC component which is unrelated to the audio signal and is also present when the audio signal is muted.

4.4.2.2.2 Loop Requirements. When the audio signal is muted, direct current flow shall be equal to or less than x mA through a zero ohm termination across the transmission service channel output terminals.

4.4.2.2.3 Longitudinal Requirements. Direct current flow shall be equal to or less than x mA when the terminals are shorted together and the current flow is measured through a zero impedance to ground.

4.4.3 Input to Output Audio Relationships.

4.4.3.1 Signal Polarity.

4.4.3.1.1 Definition. The polarity of the signal is the polarity sense of a nonsymmetrical audio transient signal on the A (tip) terminal with respect to the B (ring) terminal of the balanced pair (see Figure 7 on page 13).

4.4.3.1.2 Standard Value. The polarity sense of the audio signal at the output of the transmission service channel shall be the same as that at the input of the transmission service channel.

4.4.3.1.3 Method of Measurement. A nominal 400-Hz clipped sine wave (1/2 wave rectified, see Figure 8) is fed into the input of the transmission service channel. An oscilloscope of a known deflection is used at both ends of the transmission service channel to determine the polarity sense of the audio signal.

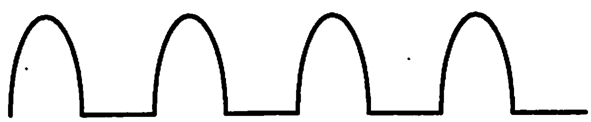


Figure 8 Nonsymmetrical Test Waveform - Audio

5 Out-of-Service Baseband VTC/VT Performance Specification.

This out-of-service VTC/VT performance specification gives the average performance of the transmission service channel for a representative group of VTC/VT video and audio signals. The user is cautioned that the performance for a particular input signal may vary from the average performance presented here. If the performance for a particular input signal is desired, the user of this standard should refer to section 6 on page 21. The block diagram for out-of-service testing is shown in Figure 9. As shown in the figure, a test signal and/or test scene generator is connected to the input of the transmission service channel. The test generator provides a source of known audio and video signals for testing the transmission service channel. Parameter measurement equipment is connected to the output of the transmission service channel. The parameter measurement equipment shown in Figure 9 determines the performance parameter values in this standard. Then, these parameter values are inserted into a set of performance calculations to obtain video and audio performance ratings for the transmission service channel. An optional provision may be made for sending telemetric data between the test generator and the parameter measurement equipment. Possible telemetric data includes parametric values that describe the test waveform or waveforms being used, commands required for automated testing, measured parameter values, and calculated performance ratings for the transmission service channel. Several options for transmitting the telemetric data include a separate data channel, and the use of the vertical blanking interval of the video signal (if transmitted by the transmission service channel).

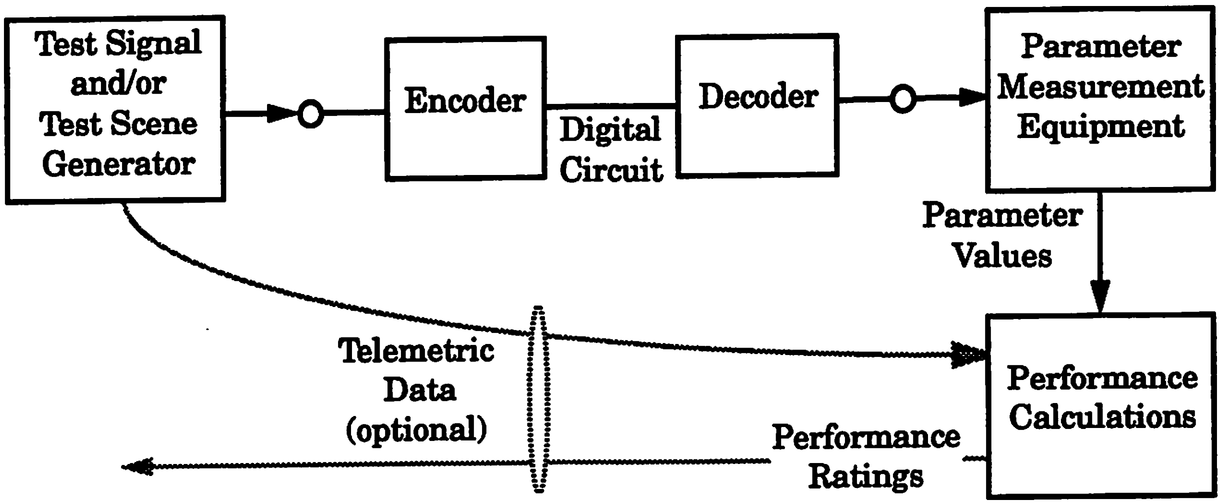


Figure 9 Out-of-Service Testing Block Diagram

5.1 Video Performance.

The video performance has been divided into two major areas; spatial performance and temporal performance. Spatial performance, a measure of the ability of a video transmission system to accurately reproduce still scenes, is a primary concern for graphical or spatial applications (see definition on page 3). Temporal performance, a measure of the ability of a video transmission system to accurately reproduce moving scenes, is a primary concern for temporal applications (see definition on page 3). A set of performance parameters for quantifying the spatial performance of a transmission service channel is given in section 5.1.1. The overall spatial performance rating for the transmission service channel is then obtained by inserting these parameter values into the algorithm given in section 5.1.2. A set of performance parameters for quantifying the temporal performance of a transmission service channel is given in section 5.1.3. The overall temporal performance rating for the transmission service channel is then obtained by inserting these parameter values into the algorithm given in section 5.1.4. The video performance application table in section 5.1.5 on page 18 can be used to specify the required spatial and temporal performance of the transmission service channel for a number of purposes (e.g., videophone, videoconference, etc.) and subject material (e.g., talking head, graphics, etc.).

5.1.1 Spatial Performance Measures. (This is the place to insert the useful traditional analog parameters as well as newer digital parameters that measure spatial distortions).

5.1.1.1 Parameter Name.

5.1.1.1.1 Definition.

5.1.1.1.2 Method of Measurement.

5.1.2 Spatial Performance Calculation. To compute the overall spatial performance rating for the transmission service channel (O_s), the measured parameter values in section 5.1.1 are inserted into the following equation: (This equation has been included as an example algorithm only, other options would include a table specifying the acceptable parameter values for each spatial performance level in section 5.1.5 on page 18)

$$O_s = \sum_{i=1}^{N} (c_i \cdot p_i) + c_0$$

Where p_i is the value of spatial parameter i, N is the total number of spatial parameters, and c_i (for i=0 to N) are constants determined according to the methods specified in Appendix A (i.e., the spatial performance algorithm shall accurately predict the subjective quality of spatial test scenes).

5.1.3 Temporal Performance Measures. (This is the place to insert the newer digital parameters that measure motion artifacts and temporal distortions).

5.1.3.1 Parameter Name.

5.1.3.1.1 Definition.

5.1.3.1.2 Method of Measurement.

5.1.4 Temporal Performance Calculation. To obtain the overall temporal performance rating for the transmission service channel (O_t), the measured parameter values in section 5.1.3 are inserted into the following equation: (This equation has been included as an example algorithm only, other options would include a table specifying the acceptable parameter values for each temporal performance level in section 5.1.5 on page 18)

$$O_t = \sum_{i=1}^{N} (c_i \cdot p_i) + c_0$$

Where p_i is the value of temporal parameter i, N is the total number of temporal parameters, and c_i (for i=0 to N) are constants determined according to the methods specified in Appendix A (i.e., the temporal performance algorithm shall accurately predict the subjective quality of temporal test scenes).

5.1.5 Video Performance Application Table. This table is provided as a guide to assist end-users and service-providers in specifying spatial and temporal performance levels. The spatial performance calculation (from section 5.1.2 on page 18) and the temporal performance calculation (from section 5.1.4 on page 18) are utilized as shown in Table 1. In the table, spatial performance increases as one moves from left to right and temporal performance increases as one moves down. Services within each block are example services for a given quality mix and are not intended to be exhaustive or exclusive.

	Spatial Performance			
Temporal Performance	Level 1 ($O_s>x_1$)	Level 2 ($O_s>x_2$)	Level 3 ($O_s>x_3$)	Level 4 ($O_s>x_4$)
Level 1 ($O_t>y_1$)		Desktop Graphics	Enhanced Graphics	For Further Study
Level 2 ($O_t>y_2$)	Head & Shoulders	People & Desktop Graphics	People & Enhanced Graphics	
Level 3 ($O_t>y_3$)	For Further Study	VCR	Training, Education	TV Studio

Table 1 Video Performance Specifications for Several Example Applications

5.2 Audio Performance.

A set of performance parameters for quantifying the audio performance of a transmission service channel is given in section 5.2.1. The overall audio performance rating for the transmission service channel is then obtained by inserting these parameter values into the algorithm given in section 5.2.2. The audio performance application table in section 5.2.3 on page 20 can be used to specify the required audio performance of the transmission service channel for a number of purposes (e.g., videophone, videoconference, etc.).

5.2.1 Audio Performance Measures. (This is the place to insert the useful traditional audio parameters as well as newer digital parameters that measure audio distortions, such as those being considered by the CCITT).

5.2.1.1 Parameter Name.

5.2.1.1.1 Definition.

5.2.1.1.2 Method of Measurement.

5.2.2 Audio Performance Calculation. To obtain the overall audio performance rating for the transmission service channel (O_a), the measured parameter values in section 5.2.1 are inserted into the following equation: (This equation has been included as an example algorithm only, other options would include a table specifying the acceptable parameter values for each performance level in section 5.2.3 on page 20)

$$O_a = \sum_{i=1}^{N} (c_i \cdot p_i) + c_0$$

Where p_i is the value of audio parameter i, N is the total number of audio parameters, and c_i (for i=0 to N) are constants determined according to methods similar to those specified in Appendix A (i.e., the audio performance algorithm shall accurately predict the subjective quality of audio signals).

5.2.3 Audio Performance Application Table. This table is provided as a guide to assist end-users and service-providers in specifying audio performance levels. The audio performance calculation (from section 5.2.2 on page 19) is utilized as shown in Table 2. In the table, audio performance increases as one moves from left to right. Audio performance levels within each block are examples and not intended to be exhaustive or exclusive.

Audio Performance			
Level 1 ($O_a>x_1$)	Level 2 ($O_a>x_2$)	Level 3 ($O_a>x_3$)	Level 4 ($O_a>x_4$)
G.711* & G.728*	G.722*	15 KHz of ANSI T1.502	20 KHz of ANSI T1.502

Table 2 Audio Performance Specifications for Several Example Applications

NOTE: * these document numbers shall be converted to ANSI numbers as appropriate

5.3 Audio-Visual Performance Measures.

Audio-visual performance measures quantify important attributes of the audio-visual signal. These performance measures fall into one of two categories: (1) parameters that quantify important attributes of both the audio signal and the video signal (2) parameters that quantify the interactions between the audio signal and the video signal, and hence require both the audio signal and the video signal for proper measurement.

5.3.1 Overall Path Delay.

5.3.1.1 Definition. Overall path delay is the greater of the one-way transmission delay of the audio signal, and the one-way transmission delay of the video signal, when only lips and eyes of the talking user (talking head) are moving.

5.3.1.2 Method of Measurement. (Under Study)

5.3.2 Audio-Visual Synchronization.

5.3.2.1 Definition. Audio-visual synchronization is the difference between the one-way transmission delays of the audio signal and the video signal when only lips and eyes of the talking user (talking head) are moving.

5.3.2.2 Method of Measurement. (Under Study)

6 In-Service Baseband VTC/VT Performance Specification.

In general, the performance of the transmission service channel depends upon the information content of the input video and audio signals. This in-service VTC/VT performance specification provides a non-intrusive method for measuring the performance of the transmission service channel for any input signal. The block diagram for in-service testing is shown in Figure 9. As shown in the figure, parameter measurement equipment is connected to both the input and output of the transmission service channel. The parameter measurement equipment is connected in a non-intrusive manner so that the video and audio performance of the transmission service channel are not effected. The provision is made for sending telemetric data from the parameter measurement equipment at the input to the parameter measurement equipment at the output. The exact requirements for the telemetric data have yet to be determined but will likely consist of extracted parameter values from the input signal. Several options for transmitting the telemetric data include a separate data channel, and the use of the vertical blanking interval of the video signal (if transmitted by the transmission service channel). The parameter measurement equipment determines the performance parameter values in this standard. Then, these parameter values are inserted into a set of performance calculations to obtain video and audio performance ratings for the transmission service channel. The final performance ratings may then be (optionally) transmitted back to the input of the transmission service channel. Performance ratings are calculated for each x second time interval of the transmission.

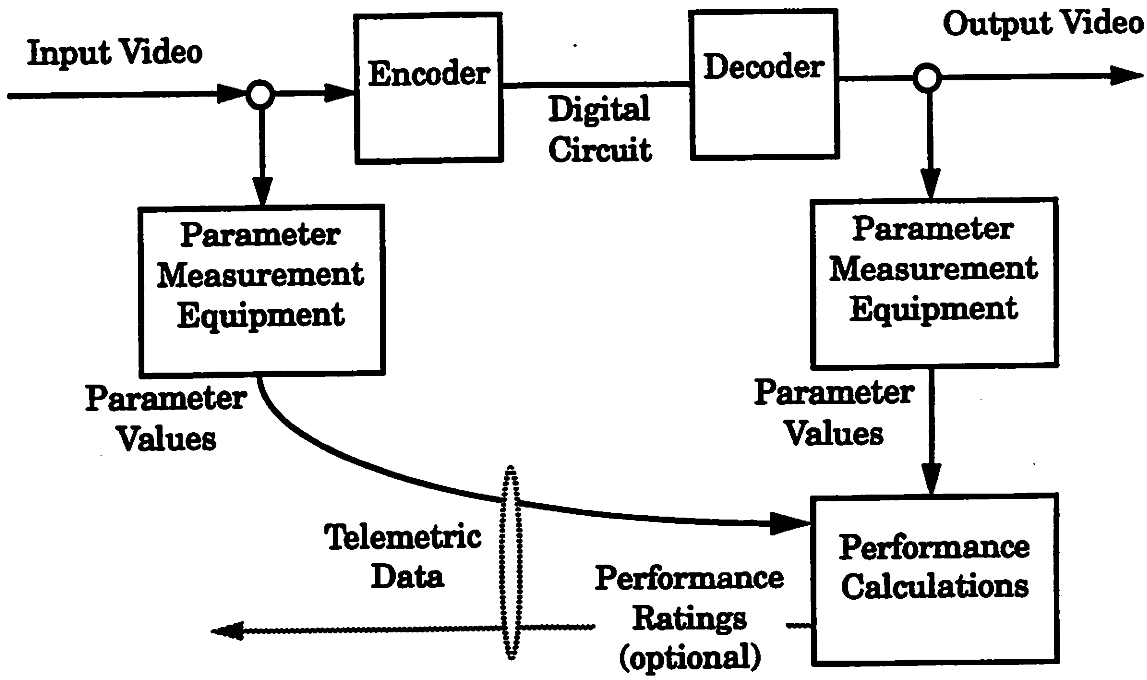


Figure 10 In-Service Testing Block Diagram

6.1 Video Performance.

A set of performance parameters for quantifying the video performance of a transmission service channel is given in section 6.1.1. The overall video performance rating for the transmission service channel is then obtained by inserting these parameter values into the algorithm given in section 6.1.2.

6.1.1 Video Performance Measures.

6.1.1.1 Parameter Name.

6.1.1.1.1 Definition.

6.1.1.1.2 Method of Measurement.

6.1.2 Video Performance Calculation. To compute the overall video performance rating for the transmission service channel (O_v), the measured parameter values in section 6.1.1 are inserted into the following equation: (This equation has been included as an example algorithm only)

$$O_v = \sum_{i=1}^{N} (c_i \cdot p_i) + c_0$$

Where p_i is the value of video parameter i, N is the total number of video parameters, and c_i (for i=0 to N) are constants determined according to the methods specified in

Appendix A (i.e., the video performance algorithm shall accurately predict the subjective quality of video test scenes).

6.2 Audio Performance.

A set of performance parameters for quantifying the audio performance of a transmission service channel is given in section 6.2.1. The overall audio performance rating for the transmission service channel is then obtained by inserting these parameter values into the algorithm given in section 6.2.2.

6.2.1 Audio Performance Measures.

6.2.1.1 Parameter Name.

6.2.1.1.1 Definition.

6.2.1.1.2 Method of Measurement.

6.2.2 Audio Performance Calculation. To obtain the overall audio performance rating for the transmission service channel (O_a), the measured parameter values in section 5.2.1 are inserted into the following equation: (This equation has been included as an example algorithm only)

$$O_a = \sum_{i=1}^{N} (c_i \cdot p_i) + c_0$$

Where p_i is the value of audio parameter i, N is the total number of audio parameters, and c_i (for i=0 to N) are constants determined according to methods similar to those specified in Appendix A (i.e., the audio performance algorithm shall accurately predict the subjective quality of audio signals).

6.3 Audio-Visual Performance Measures.

6.3.1 Overall Path Delay.

6.3.1.1 Definition. Overall path delay is the greater of the one-way transmission delay of the audio signal, and the one-way transmission delay of the video signal, when only lips and eyes of the talking user (talking head) are moving.

6.3.1.2 Method of Measurement. (Under Study)

6.3.2 Audio-Visual Synchronization.

6.3.2.1 Definition. Audio-visual synchronization is the difference between the one-way transmission delays of the audio signal and the video signal when only lips and eyes of the talking user (talking head) are moving.

6.3.2.2 Method of Measurement. (Under Study)

6.3.3 Availability of Service. (Under Study)

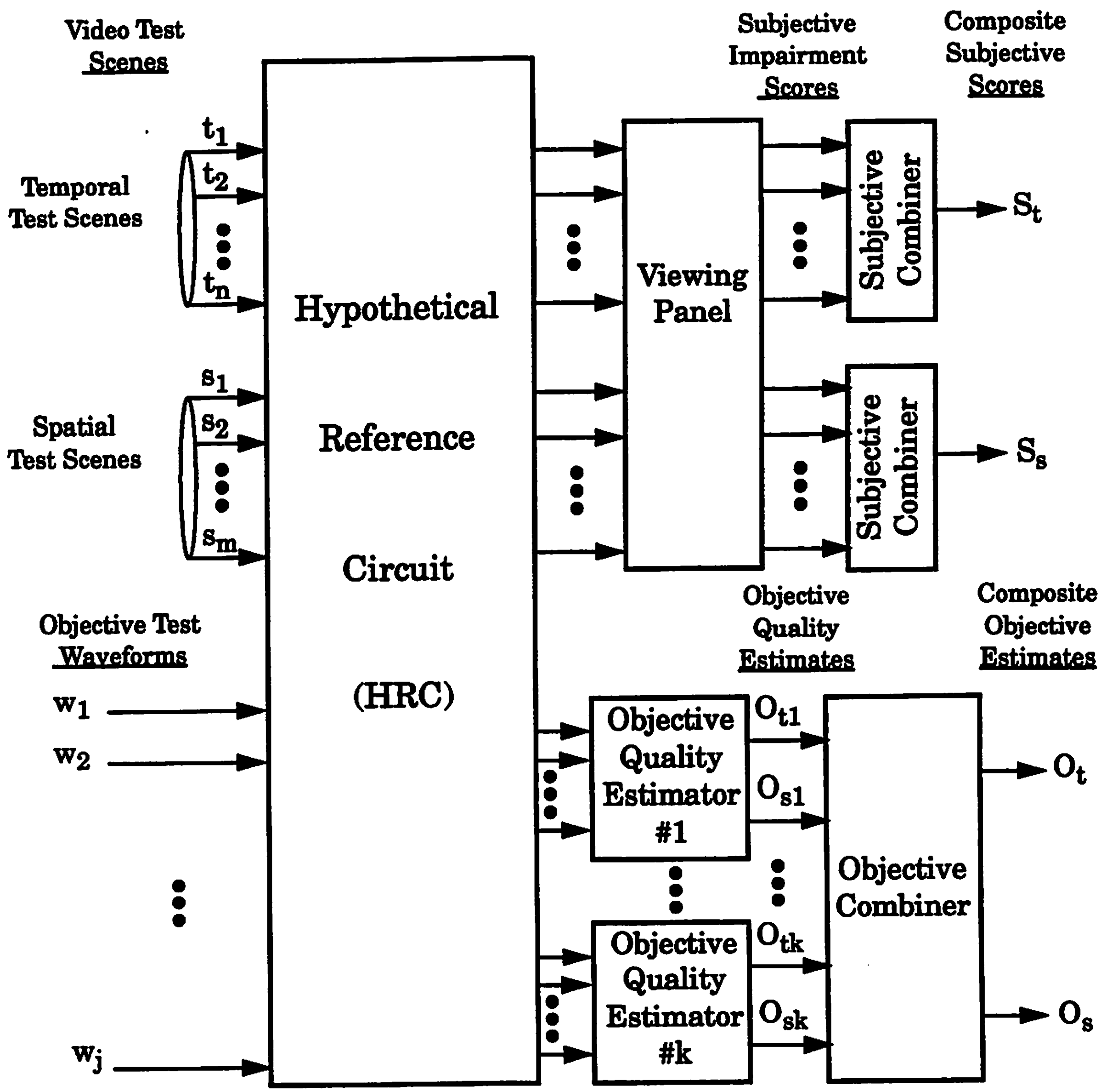
Appendix

Appendix A

Methodology Used for Evaluating Out-of-Service Performance Measures

The information in this appendix provides a historical record of the approach and the actual test results that were used to select the objective performance measures in this standard. The detailed methodology presented here can by used by others to verify the validity of this standard as well as to expand upon the standard so that it will be useful for measuring the performance of future, as of now unspecified, audio-visual services. This appendix describes the methodology used for developing the video performance measures. Extension of this methodology to include the audio performance measures should be straightforward. The extended methodology should consider the possibility of interactions between the audio performance and the video performance.

Overview of Approach. An overview of the approach that is used to develop and evaluate the objective quality estimators is shown in Figure A1 on page 26. A primary design goal for objective quality estimators is an output that accurately predicts the overall quality score as perceived by the end-user. Thus, subjective test results form a key criterion for the evaluation of acceptable objective quality estimators. The test scenes used for the subjective tests are divided into two groups; spatial test scenes (s_1, s_2, ..., s_m), and temporal test scenes (t_1, t_2, ..., t_n). The spatial test scenes are chosen to test the spatial performance of the Hypothetical Reference Circuit (HRC), while the temporal test scenes are chosen to test the temporal performance of the HRC. These subjective test scenes are input into a set of representative HRCs and the resulting output scenes are subjectively judged by viewing panels. Subjective impairment scores for these test scenes are then combined with the subjective combiners to produce two composite subjective scores for each HRC; a composite subjective score for the set of spatial test scenes (S_s), and a composite subjective score for the set of temporal scenes (S_t). Similarly, the objective test waveforms (w_1, w_2, ..., w_j) are also input into the HRCs and the proposed objective quality estimators (1 through k) each measure a set of objective parameters and produce objective quality estimates of the spatial performance (O_{s1}, O_{s2}, ..., O_{sk}) and temporal performance (O_{t1}, O_{t2}, ..., O_{tk}) of the HRC. The criterion for determining the accuracy of the objective quality estimates is minimization of the squared error differences between the objective and subjective estimates of performance averaged over all HRCs. These squared error differences are denoted by E_s^2 and E_t^2 in Figure A1, where E_s is the error in the spatial performance estimate and E_t is the error in the temporal performance estimate for a HRC. All of the outputs from the objective quality estimators can themselves be combined by an objective combiner. The final composite objective estimates from the objective combiner may provide improved estimates of the spatial performance (O_s) and temporal performance (O_t) of the HRC. This approach allows different organizations to independently develop objective quality estimators and enables the integration of the various methods into a coherent standard.



Minimize Over All HRCs

$$E_t^2 = (O_t - S_t)^2$$

$$E_s^2 = (O_s - S_s)^2$$

Figure A1 Approach Used To Evaluate Objective Quality Estimation Methods

Outline of Process Used to Validate Objective Measures. This section summarizes the validation process for proposed objective video performance measures and the time schedule for completion of the VTC/VT draft performance standard.

1. Determine how many laboratories would be willing to participate in the subjective tests that will be used for validating the objective performance measures. Each laboratory would volunteer to conduct subjective tests according to CCIR-500. This would involve collecting subjective ratings from approximately 30-40 people for 4, half-hour viewing sessions. For impairment testing, each viewer would rate no more than 2 clips per minute or 240 clips total. A clip is defined as a test scene - hypothetical reference circuit combination. A portion of these 240 clips would be used for training and inter-laboratory consistency checks (clips that are rated by all the laboratories to check for laboratory-to-laboratory consistency). Given that N laboratories agree to participate, then the product of the number of test scenes and hypothetical reference circuits can be no greater than 240 times N.

2. Select test scenes to use from the list of proposed test scenes. This involves selection of the scene and the particular time interval that the viewer will see (some of the test scenes are more than 10 seconds).

3. Select hypothetical reference circuits (i.e., video systems under test) to use from the list of proposed reference circuits.

4. Determine all of the objective test waveforms to include in the test.

5. Edit a master source tape that contains all of the test scenes and objective test waveforms. Editing should be performed using the highest quality source that is available.

6. Play the master source tape through each hypothetical reference circuit and record the output. This step produces one output tape for each hypothetical reference circuit.

(all the above steps have been completed)

7. Subjective Tests - (a) develop a subjective test plan and a subjective data analysis plan, (b) take the master source tape and the output tapes with the test clips (generated from step 6) and edit viewing tapes according to the subjective test plan, and (c) perform subjective tests according to test plan from step 7a.

8. Objective Tests - (a) distribute output objective test waveforms to participants (this step has been completed), (b) participants perform their own objective tests, (c) participants submit methods of measurement and their objective test data for other laboratory validation, and (d) other laboratories perform objective test validation.

9. Subjective to Objective Correlation Tests - (a) develop a subjective to objective correlation analysis plan, and (b) correlate the subjective data with the objective data.

10. Reach consensus on objective measures based upon step 9.

11. Editorial meeting - put objective measures from step 10 above into the draft standard.

12. Submit final draft for T1A1.5 approval.

Time Schedule. The following presents a time schedule for completion of the above process.

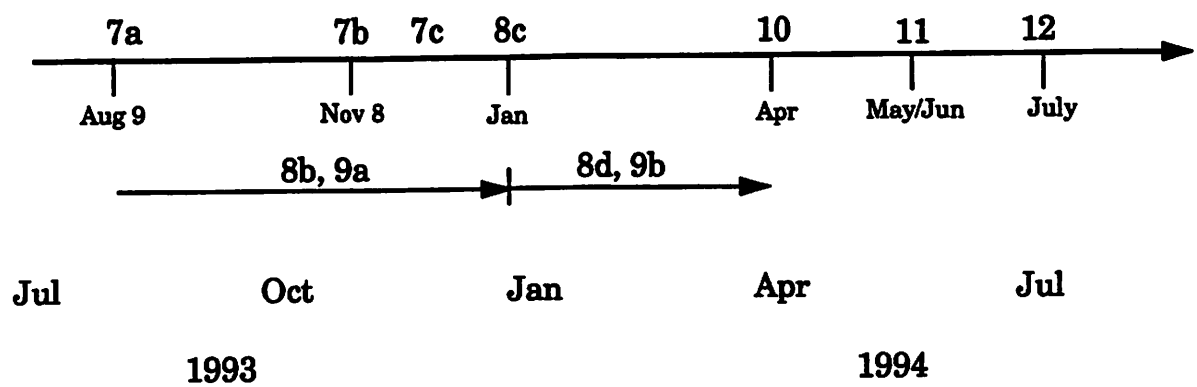


Figure A2 Time Schedule for Completion of Draft Standard

Test Scene Selection. The selection of test scenes is a very important issue. In particular, the spatial and temporal information content of the scenes are critical parameters. These parameters play a crucial role in determining the amount of video compression that is possible, and consequently, the level of impairment that is suffered when the scene is transmitted over a fixed-rate digital transmission service channel. Figure A3 shows the relative amounts of spatial and temporal information for some possible test scenes. Fair and relevant video test scenes must be chosen such that their spatial and temporal information content is consistent with the video services that the digital transmission service channel was intended to provide. As shown in Figure A1 on page 26, two groups of test scenes are proposed; spatial test scenes to test the spatial performance, and temporal test scenes to test the temporal performance. The set of test scenes should span the full range of spatial and temporal information content of interest to users of this standard. The specific set of video scenes to use for testing should be agreed upon by the VTC/VT sub-working group as soon as possible.

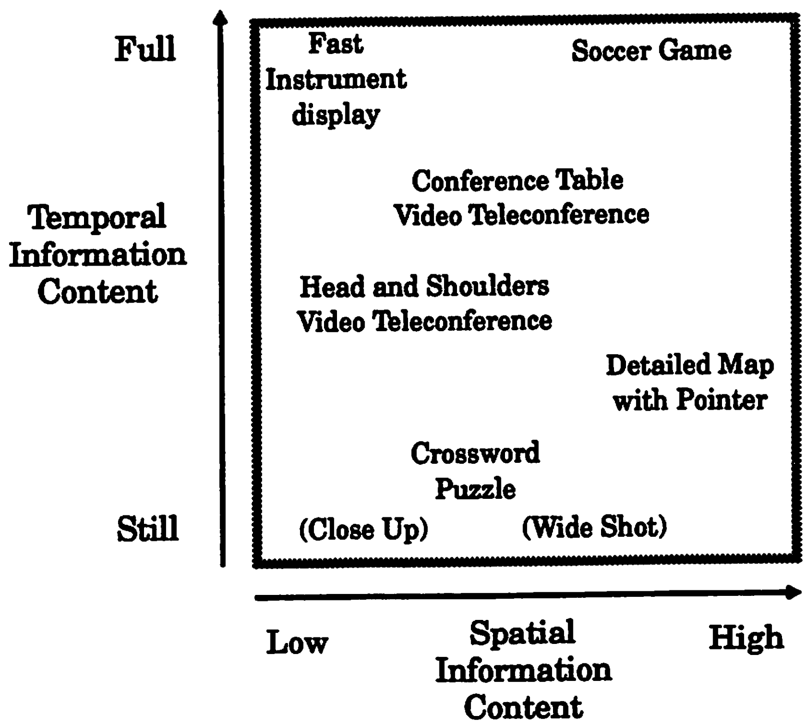


Figure A3 Information Content of Test Scenes

Spatial and Temporal Information Measures. This section presents methods for quantifying the spatial and temporal information content of test scenes. These methods for evaluating the spatial and temporal information content of test scenes are applicable to video quality testing - both now and in the future. The location of the video scene within the spatial-temporal matrix is important because the quality of a transmitted video scene (especially after passing through a low bit rate codec) is often highly dependent on this location. The spatial and temporal information measures presented here can be used to assure uniform coverage of the spatial-temporal matrix, as well as assuring that each pertinent performance block in the spatial-temporal grid shown Table 1 on page 19 is adequately stressed. For a video system to qualify at a given level, it must be tested with scenes that exercise that level of performance. A description of the spatial and temporal information measures is given first, followed by example results for several test scenes.

The Spatial Information content, SI, is based on the Sobel filter. Each video frame (luminance plane) at time n (F_n) is first filtered with the Sobel filter (Sobel(•)). The standard deviation over the pixels (std_{space}) in each Sobel-filtered frame is then computed. This operation is repeated for each frame in the video sequence and results in a time series of spatial information values. The maximum value in the time series (max_{time}) is chosen to represent the spatial information content of the scene. This process can be represented in equation form as,

$$SI = \max_{time}\{std_{space}[Sobel(F_n)]\}$$

The Temporal Information content, TI, is based upon the motion difference feature, $M_n(i,j)$, which is the difference between the pixel values (of the luminance plane) at the same location in space but at successive times or frames. $M_n(i,j)$ as a function of time (n) is defined as,

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j)$$

where $F_n(i,j)$ is the pixel at the i^{th} row and j^{th} column of the nth frame in time.

The measure of temporal information content, TI, is computed as the maximum over time (max_{time}) of the standard deviation over space (std_{space}) of $M_n(i,j)$ over all i and j.

$$TI = \max_{time}\{std_{space}[M_n(i,j)]\}$$

More motion in adjacent frames will result in higher values of TI. Note: For scenes that contain scene cuts, two values may be given: one where the scene cut is included in the temporal information measure, and one where it is excluded from the measurement.

Figure A4 on page 32, Figure A5 on page 33, and Figure A6 on page 34 show how three sets of test scenes can be placed on a spatial-temporal information matrix which has its origin at the upper left. The grid attempts to place scenes in boxes corresponding to the spatial-temporal matrix shown in Table 1 on page 19. Note that the scenes in Figure A4, Figure A5, and Figure A6 are divided into only 3 spatial levels instead of 4 as shown in Table 1 on page 19. This is due to the fact that Level 4 might be used for spatial performances beyond NTSC (such as HDTV).

Table A1 on page 32, Table A2 on page 33, and Table A3 on page 34 provide data as well as naming references for the plots in Figure A4, Figure A5, and Figure A6. Along the TI=0 axis (at the top edge of the plot) are found the still scenes and those with very limited motion. Along the SI=0 axis (at the left edge of the plot) are found scenes with minimal spatial detail. For the first data set shown in Figure A4, only the 9 second video segment given in Table A1 was processed. For the second and third data sets shown in Figure A5 and Figure A6, respectively, the entire video segment given in Table A2 and Table A3 was processed. These values of SI and TI can be compared to other test scenes measured using the above equations which have been spatially sampled at 4 times sub-carrier ($4f_{sc}$) or approximately 756 x 486 pixels, 30 frames per second (2 fields, no interpolation), and digitized with white set to 235 and black set to 16.

Note that the values are not normalized, although each plot is made to the same scale. Maxima for SI and TI are estimated to be 406 and 219 respectively. These values will not be encountered in typical video, although special test scenes can be constructed to yield these values.

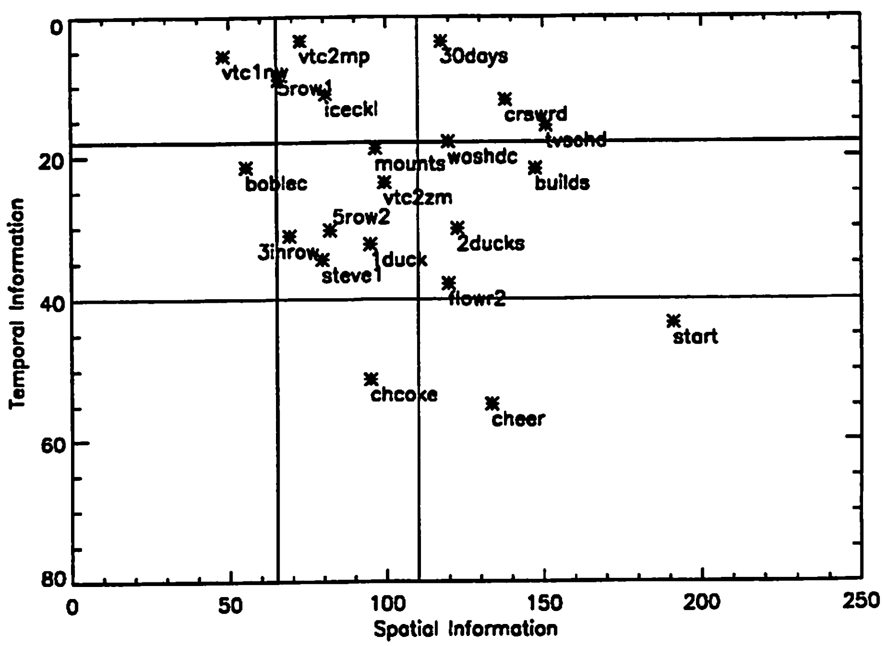


Figure A4 Spatial-Temporal Plot for Test Scene Set #1

Time Code mm:ss[ff] Begin - End	9sec seg Begin	Process Name	Length ss[ff]	Description	Spatial/Temp Matrix Location		Spatial Info	Temp Info	Audio Available
1:00 - 1:15	1:01[10]	vtc2mp	15[00]	VTC2 - MP + Map	2	1	72.7	3.5	Yes
1:15 - 1:30	1:20[29]	vtc2zm	15[00]	VTC2 - Zoom w/Pointer	2	2	99.4	23.6	Yes
1:30 - 2:00	1:32[00]	30days	23[25]	Calendar - 30 Days	3	1	117.5	2.7	No
2:00 - 2:30	2:19[23]	washdc	30[00]	Wash.,DC Map w/Pencil	3	1	119.8	17.9	No
2:30 - 3:00	2:36[00]	3inrow	30[00]	3 People - Two Pans	2	2	69.1	31.2	Yes
3:00 - 3:30	3:01[15]	boblec	30[00]	Bob's Lecture I	1	2	55.4	21.5	Yes
3:30 - 4:00	3:35[00]	steve1	30[00]	Steve1 - The Box	2	2	79.6	34.5	Yes
4:00 - 4:30	4:18[00]	builds	30[00]	Grand Prix -Buildings	3	2	147.4	21.7	Yes
4:30 - 5:00	4:45[00]	start	30[00]	Grand Prix -Start	3	3	191.0	43.4	Yes
5:00 - 5:30	5:17[00]	mounts	30[00]	Two Mountains	2	2	96.6	18.9	No
5:30 - 6:00	5:47[00]	flowr2	30[00]	Flowers II	3	2	119.8	37.9	No
6:00 - 6:30	6:02[00]	2ducks	30[00]	Two Ducks	3	2	122.7	30.2	No
6:30 - 7:00	6:32[00]	1duck	30[00]	One Duck (Close-up)	2	2	94.9	32.3	No
7:00 - 7:30	7:02[00]	iceckl	30[00]	Icecicle	2	1	80.7	11.2	No
7:30 - 8:00	7:44[00]	cheer	22[29]	Cheerleaders	3	3	133.5	51.3	No
8:00 - 8:30	8:05[16]	chcoke	15[18]	Cherry Coke on Mount.	2	3	94.9	54.8	No
8:30 - 9:00	8:32[12]	vtc1nw	30[00]	VTC1 - News Story	1	1	48.3	5.6	Yes
9:00 - 9:30	9:21[00]	crswrd	30[00]	Crossword Puzzle	3	1	138.0	12.0	No
9:30 - 10:00	9:51[00]	tvschd	30[00]	TV Schedule	3	1	150.8	15.7	No
10:00 - 10:30	10:19[20]	5row1	30[00]	5 People in a Row I	2	1	65.6	9.1	Yes
10:30 - 11:00	10:40[00]	5row2	30[00]	5 People in a Row II	2	2	82.1	30.3	Yes

Table A1 Spatial-Temporal Values for Test Scene Set #1

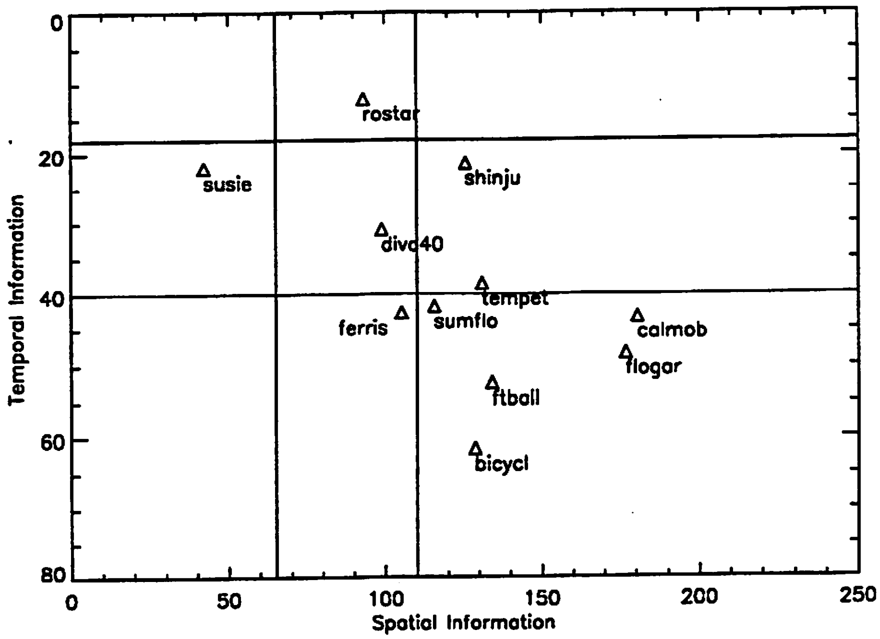


Figure A5 Spatial-Temporal Plot for Test Scene Set #2

Time Code mm:ss[ff] Begin - End	9sec seg Begin	Process Name	Length ss[ff]	Description	Spatial/Temp Matrix Location		Spatial Info	Temp Info	Audio Available
11:00 - 11:30		sumflo	20[07]	Summer Flowers	3	3	115.6	41.8	No
11:30 - 12:00		flogar	20[04]	Flower Garden	3	3	176.7	48.3	No
12:00 - 12:30		diva40	9[28]	Diva w/noise 40dB SNR	2	2	98.9	30.9	No
12:30 - 13:00		bicycl	19[27]	Bicycles	3	3	128.7	61.6	No
13:00 - 13:30		calmob	20[02]	Calendar & Mobile	3	3	180.5	43.2	No
13:30 - 14:00		tempet	20[03]	Tempete (Leaves)	3	2	131.0	38.5	No
14:00 - 14:30		shinju	20[00]	Shinjuku	3	2	125.7	21.4	No
14:30 - 15:00		rostar	13[00]	Rotating Star	2	1	93.1	12.1	No
15:00 - 15:30		ferris	20[00]	Ferris Wheel	2	3	105.1	42.7	No
15:30 - 16:00		ftball	19[29]	Football	3	3	134.1	52.5	No
16:00 - 16:30		susie	20[02]	Susie	1	2	42.2	22.0	No

Table A2 Spatial-Temporal Values for Test Scene Set #2

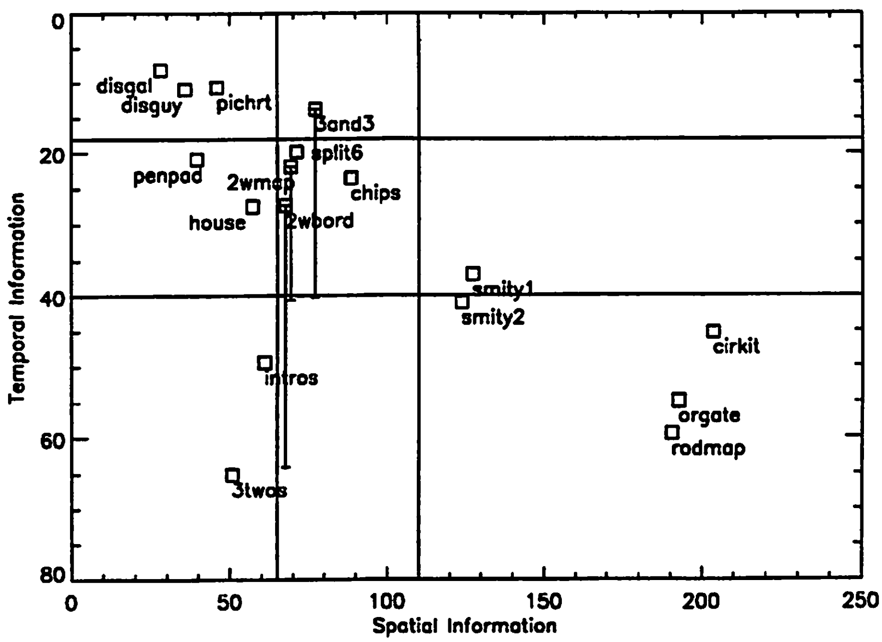


Figure A6 Spatial-Temporal Plot for Test Scene Set #3

Time Code mm:ss[ff] Begin - End	9sec seg Begin	Process Name	Length ss[ff]	Description	Spatial/Temp Matrix Location		Spatial Info	Temp Info	Audio Available
16:45[04] - 16:59[28]		A1 disguy	14[24]	Announcer	1	1	35.8	10.9	Yes
17:08[01] - 17:18[05]		A2 disgal	10[04]	Girl Close-up (Miss America)	1	1	28.1	8.1	No
17:24[23] - 18:15[12]		A3 smity1	50[19]	Salesman at Desk Equipment	3	2	127.2	37.1	Yes
18:20[28] - 19:17[18]		A4 smity2	56[20]	Salesman at Desk Close-up	3	3	123.8	41.0	Yes
19:28[00] - 20:00[21]		B1 orgate	32[21]	Circuit Diagram	3	3	192.6	54.9	No
20:04[12] - 20:29[05]		B2 house	25[23]	Grand House Poster	1	2	57.2	27.5	No
20:33[14] - 21:12[21]		B3 penpad	39[07]	Diag. Being Drawn	1	2	39.5	20.9	No
21:23[06] - 21:53[04]		C1 intros	29[28]	Intro of 6 Conferees	1	3	61.0	49.4	No
21:58[14] - 22:25[12]		C2 3and3	26[28]	2 Grps of 3 Conferees	1	3	77.1	40.2	No
22:30[00] - 23:09[11]		C3 3twos	39[11]	3 Grps of 2 Conferees	1	3	50.7	65.1	No
23:14[09] - 23:49[26]		C4 2wmap	35[17]	Discuss. at Wall Map	2	3	69.2	40.6	No
23:56[04] - 24:22[24]		C5 2wbord	26[20]	Argument at Board	2	3	67.7	64.0	No
24:29[24] - 25:14[10]		C6 pichrt	44[16]	Girl @Green Chalk Brd	1	1	46.0	10.6	No
25:20[00] - 25:48[14]		C7 split6	28[14]	2 Grps of 3 -Stacked	2	2	71.0	19.9	No
25:57[20] - 26:29[15]		D1 cirkit	31[25]	Circuit Board Layout	3	3	203.3	45.3	No
26:33[19] - 27:08[00]		D2 rodmap	34[11]	Map w/Pointer &Marker	3	3	190.4	59.5	No
27:17[26] - 27:47[13]		E1 chips	29[17]	Circuit Brd. Explanation	2	2	88.5	23.5	No

Table A3 Spatial-Temporal Values for Test Scene Set #3

HRC Selection. The objective performance measures and algorithms set forth in this standard should be tested using a representative set of HRCs that include all relevant technologies for which this standard applies. Preferably, the pool of HRCs should be sufficiently large to test the technology independence (i.e., not dependent on the coding algorithm or transport architecture) of the objective quality estimators.

Thus, selection of a HRC involves not only selection of the digital codec and bit rate, but also selection of the transport mechanism. The specific set of HRCs to use for testing should be agreed upon by the VTC/VT sub-working group as soon as possible.

Subjective Viewing Tests. The subjective viewing tests should be conducted in accordance with CCIR Recommendation 500. The 5-point impairment scale given in Recommendation 500 is useful when comparison of subjective results between laboratories is desired, and when the range of impairments will vary considerably. The subject is first presented with a scene and then with an impaired version of the same scene. The subjects are instructed to decide on and mark the level of impairment in the second scene, using the first scene as a reference. The five possible responses are: Imperceptible (IP), Perceptible but Not Annoying (P/NA), Slightly Annoying (SA), Annoying (A), and Very Annoying (VA). This 5-point impairment scale intentionally covers a very wide range of impairment levels in a non-linear fashion. By including reference scenes, impairment tests take advantage of the fact that the human eye excels at making comparisons.

Subjective Combiners. The subjective combiners shown in Figure A1 on page 26 are required to combine the subjective impairment scores into an overall composite subjective score for the HRC. Figure A7 on page 37 gives a block diagram for the functions of the subjective combiners. For every HRC, the subjective test results produce a histogram of subjective impairment scores for each of the spatial test scenes (s_1, s_2, ..., s_m), and temporal test scenes (t_1, t_2, ..., t_n). These histograms are depicted graphically in Figure A7 by the heights of lines occurring in each of the histogram bins (IP, P/NA, SA, A, VA). It is important to note that, even for a fixed HRC, the histograms of subjective scores could vary dramatically depending upon the scene input. Some output scenes from the HRC may be rated IP while others may be rated VA. The purpose of the subjective combiners are to produce overall composite subjective scores for the spatial performance (S_s) and temporal performance (S_t) of the HRC from the histograms of subjective impairment scores. The first function that must be performed is to produce a subjective score for each scene (S_{s1}, S_{s2}, ..., S_{sm}, S_{t1}, S_{t2}, ..., S_{tn}). This is shown in Figure A7 as a histogram combiner. There are many possible ways in which the histograms could be collapsed. Three possible methods are:

1. The total percentage of responses where people were not annoyed (%IP+%P/NA). This might be a good indicator of customer satisfaction since people that are annoyed tend to complain.

2. The total percentage of responses where people do not see an impairment (%IP). This might be a good indicator of high quality video and hence might be useful for people that desire contribution quality video.

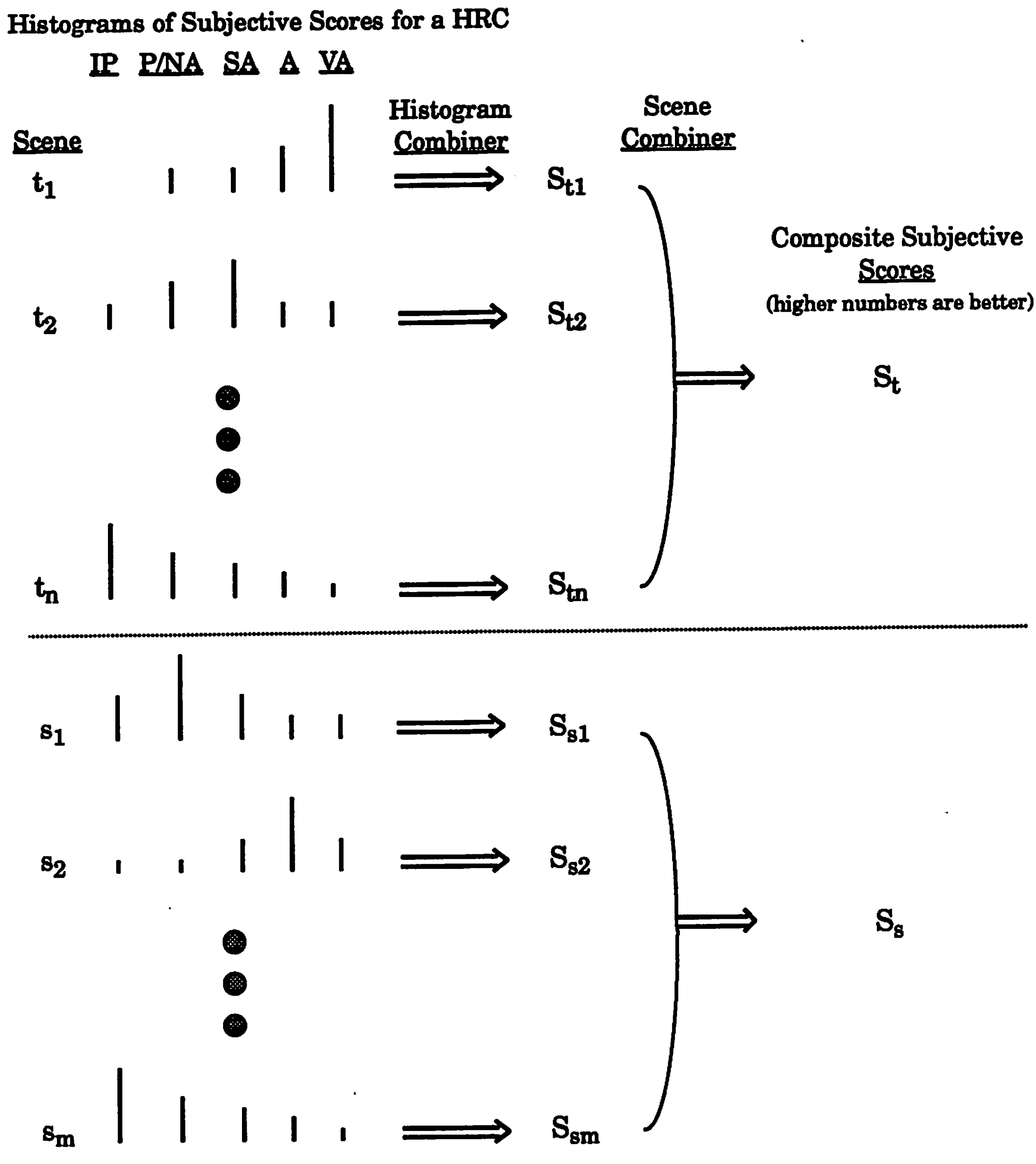
3. The weighted sum where each type of response is multiplied by a unique weight ($weight_1$(%VA) + $weight_2$(%A) + $weight_3$(%SA) + $weight_4$(%P/NA) + $weight_5$(%IP)). Here, the weight could vary in relationship to the importance of the particular histogram bin. A linear mean opinion score (MOS) can be generated from this equation if the weights are 1, 2, 3, 4, 5, respectively.

Next, the scene combiner shown in Figure A7 takes the subjective scores for each scene, as output by the histogram combiner (S_{s1}, S_{s2}, ..., S_{sm}, S_{t1}, S_{t2}, ..., S_{tn}), and produces the composite subjective scores for the spatial performance (S_s) and temporal performance (S_t) of the HRC. Once again, there are several possible methods to collapse the individual scene scores into overall composite scores for the HRC. Two possible methods are:

1. Assign equal weight (or importance) to each scene. Here, each spatial scene would have a weight of 1/m, and each temporal scene would have a weight of 1/n.

2. Assign zero weights to some of the test scenes depending upon the specific application as given in Table 1 on page 19. Then, assign equal weights to the remaining test scenes such that the sum of all spatial test scene weights is equal to one and the sum of all temporal test scene weights is equal to one. This method would allow the flexibility of tailoring the performance measurements to each of the applications given in Table 1 on page 19 (i.e., videophone - head & shoulders, videophone - desktop graphics, videoconference - people & enhanced graphics, etc.).

The specific subjective combiners should be agreed upon by the VTC/VT sub-working group as soon as possible.



NOTE:

The histograms of subjective scores for each HRC are a function of the scene input.

IP = Imperceptible
P/NA = Perceptible but Not Annoying
SA = Slightly Annoying
A = Annoying
VA = Very Annoying

Figure A7 Block Diagram of Subjective Combiners

Objective Test Waveforms. The objective test waveforms ($w_1, w_2, ..., w_j$) shown in Figure A1 on page 26 are analog video waveforms and include waveforms that are required to measure the traditional analog parameters as well as those that are required to measure any new objective parameters. After passing through the HRC, the output analog test waveforms are input into the proposed objective quality estimators (1 through k). Normally, a subset of the total waveforms will be utilized by each of the proposed objective quality estimators since some waveforms may have been specifically designed to work in conjunction with certain estimators.

Objective Quality Estimators. The proposed objective quality estimators (1 through k) shown in Figure A1 on page 26 each extract a set of objective performance parameters from the objective test waveforms, and combine these objective parameters to produce estimates of the spatial performance ($O_{s1}, O_{s2}, ..., O_{sk}$) and temporal performance ($O_{t1}, O_{t2}, ..., O_{tk}$) of the HRC. The designer of the objective quality estimator is free to choose the algorithm that produces quality estimates from the objective parameters. However, the criterion for determining the accuracy of the objective quality estimates is minimization of the averaged squared error difference between the objective and subjective estimates of performance over all HRCs. These squared error differences are denoted by E_s^2 and E_t^2 in Figure A1, where E_s is the error in the spatial performance estimate and E_t is the error in the temporal performance estimate for a HRC.

Objective Combiner. The objective combiner enables integration of the k objective quality estimates that are produced by the objective quality estimators. Composite estimates of spatial and temporal performance (O_s and O_t in Figure A1 on page 26) are produced by the objective combiner. The composite estimates will be at least as good as any of the individual estimates produced by the objective quality estimators. One could use linear predictors to predict O_s and O_t from the output of the objective quality estimators ($O_{s1}, O_{s2}, ..., O_{sk}$) and ($O_{t1}, O_{t2}, ..., O_{tk}$), respectively. Then, the prediction equations would take the form:

$$O_s = \left(\sum_{i=1}^{k} (\alpha_i \cdot O_{si})\right) + C_s$$

for the spatial performance and

$$O_t = \left(\sum_{i=1}^{k} (\beta_i \cdot O_{ti})\right) + C_t$$

for the temporal performance, where weights α_i and β_i are chosen such that the squared error differences between the objective and subjective estimates of performance are minimized over all HRCs. Constants C_s and C_t have been included to allow for the possibility that some of the individual estimates produced by the objective quality estimators are biased. One can look at the value of the individual weights α_i and β_i to determine if an objective quality estimator is not contributing to the overall composite score (the weights for that estimator will be small). In this case, the non-contributing estimator can be discarded without losing accuracy.

Test Results. This section presents detailed test results for the objective quality measurement system used in this standard. The objective quality estimates obtained by use of this standard are compared with actual subjective quality ratings. Close correlation of the objective quality estimates with the subjective scores demonstrates the validity of the objective video quality estimators used in this standard. (Insert results of tests).

Confidence Intervals for Correlation Coefficients. This section presents the results of a study done to evaluate the degree of confidence in the correlation coefficients calculated between subjective and objective video quality scores. This section also provides a clear, objective way to determine the number of hypothetical reference circuits (HRCs) that should be included in a performance test.

The confidence interval on the correlation coefficient, ρ, has been investigated by those interested in objective measures of speech quality. A useful equation has been found which closely approximates the true confidence interval for a large number (i.e., $N \geq 25$) of data points (See Quackenbush, S., Barnwell, T., and Clements, M., Objective Measures of SpeechQuality, Prentice Hall, 1988, p. 191 and Montgomery, D. and Peck, E., Introduction to Linear Regression Analysis, John Wiley, 1982, p. 49). For the proposed test methodology, each point corresponds to an (x,y) pair, where x is the subjective rating of the HRC and y is a proposed objective measurement. Correlation between subjective and objective scores is one method of comparing different objective measurements. It is this measure, ρ, which is addressed herein. The closer $|\rho|$ is to 1.0, the better the objective measure.

The 100(1-α) percent confidence interval (α=0.05 for a 95% confidence interval) for ρ is given in the above references as

$$\tanh\left[\operatorname{atanh}(\hat{\rho}) - \frac{Z_{\alpha/2}}{\sqrt{N-3}}\right] \leq \rho \leq \tanh\left[\operatorname{atanh}(\hat{\rho}) + \frac{Z_{\alpha/2}}{\sqrt{N-3}}\right]$$

where ρ is the true correlation coefficient, $\hat{\rho}$ is the correlation coefficient estimated from the data, $Z_{\alpha/2}$ is the normal(Gaussian) deviate, N is the number of points (in our case, the number of HRCs), **tanh** is the hyperbolic tangent, and **atanh** is the arc hyperbolic tangent.

A computer simulation was performed to validate the above equation and to extend the confidence interval results when the total number of HRCs is less than 25 ($N < 25$). In order to determine the confidence interval for a given correlation coefficient, four large (1,000,000 points) pairs of number sequences with known correlation coefficients of ρ = .85, .90, .95, and .99 were constructed. This was accomplished by utilizing the Unix-C random number generator **drand48** and constructing the pair of sequences according to the following equation:

$$x = \text{uniform}(0, 1)$$

$$z = \text{uniform}(0, 1)$$

$$y = \alpha x + \beta y$$

$$\rho_{xy} = \frac{\alpha}{\sqrt{\alpha^2 + \beta^2}}$$

The data from the simulation was collected by calculating $\hat{\rho}$ over 10,000 non-overlapping sequences of (x,y) pairs with each sequence containing N pairs. Each sequence was drawn from one of the large number of possible sequences, described by the above equation, that had known correlation coefficients. From these 10,000 estimates of ρ, the confidence intervals were obtained by finding the upper and lower boundaries within which 90%, 95%, and 99% of the estimates were found. N was varied from 5 to 100.

The twelve cases that were considered involved combinations of four correlation coefficients (.85, .90, .95, .99) and three confidence intervals (90%, 95%, 99%). For brevity, only five plots have been included here. Figure A8 shows the case where ρ=.95 and the confidence interval is 90%. Note that the simulation curve and the theoretical curve converge as N increases. Figure A9, Figure A10, Figure A11, and Figure A12 show the cases where ρ=.85, .90, .95, and .99 and the confidence interval is 90%.

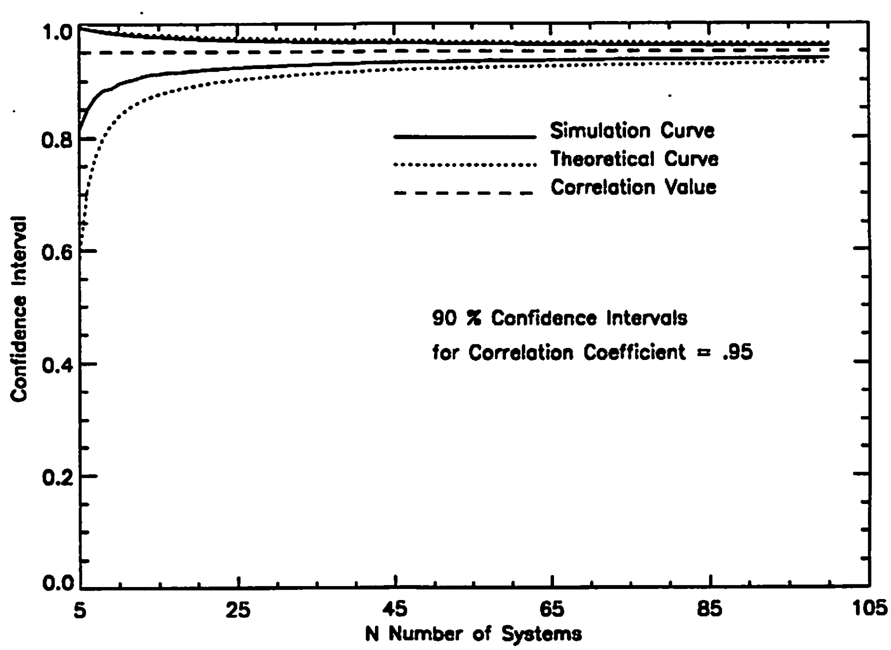


Figure A8 90% Confidence Interval for ρ = .95, N = 5 to 105

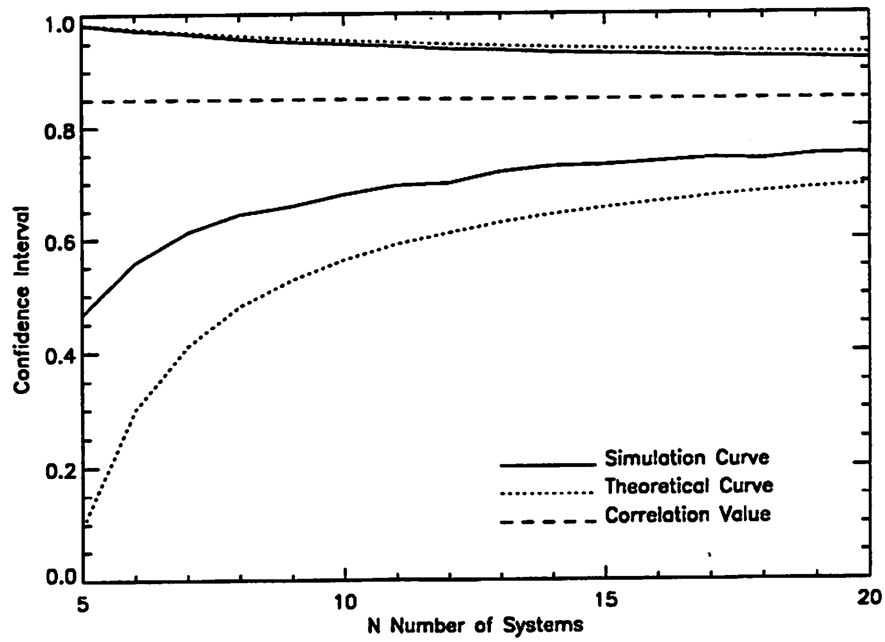


Figure A9 90% Confidence Interval for ρ = .85, N = 5 to 20

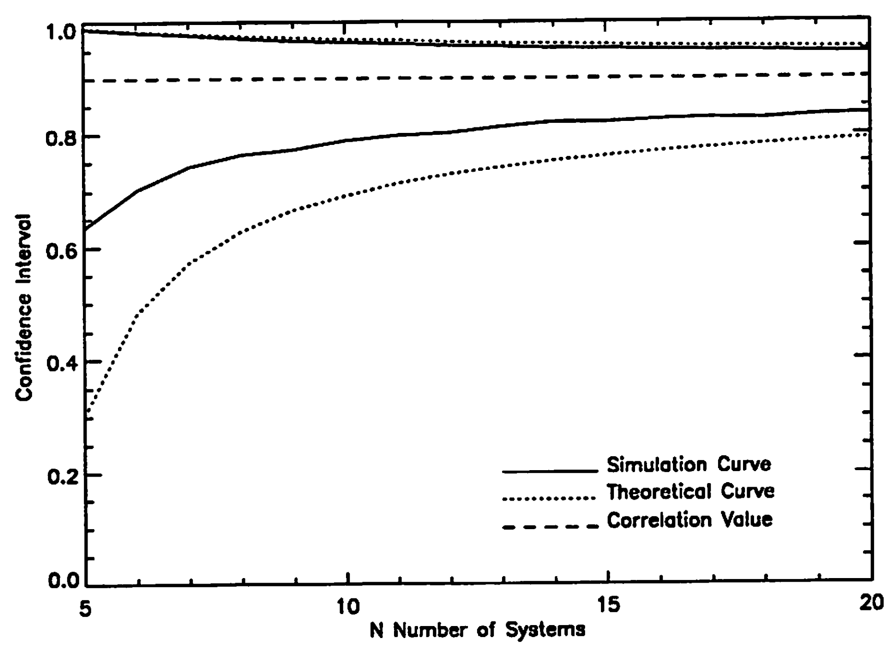


Figure A10 90% Confidence Interval for ρ = .90, N = 5 to 20

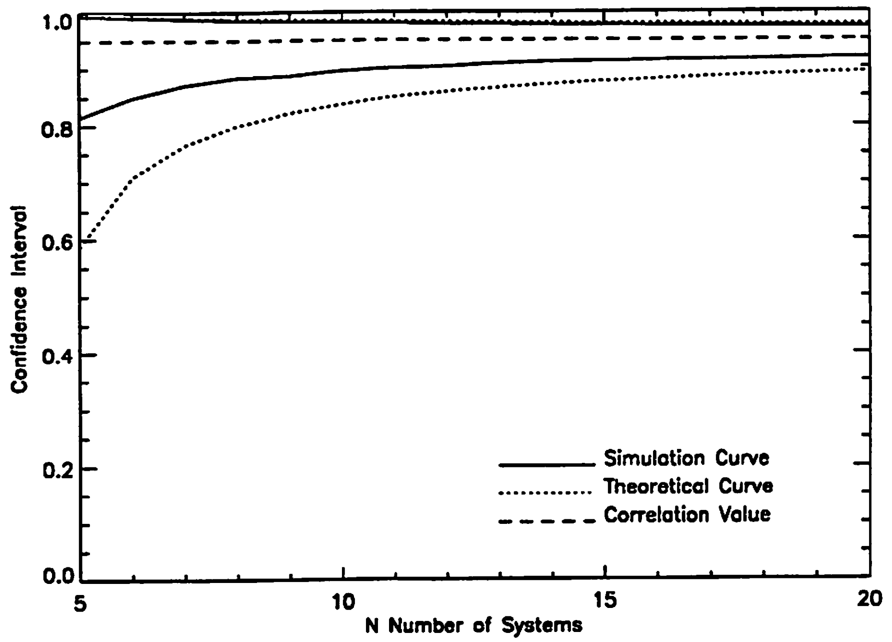


Figure A11 90% Confidence Interval for ρ = .95, N = 5 to 20

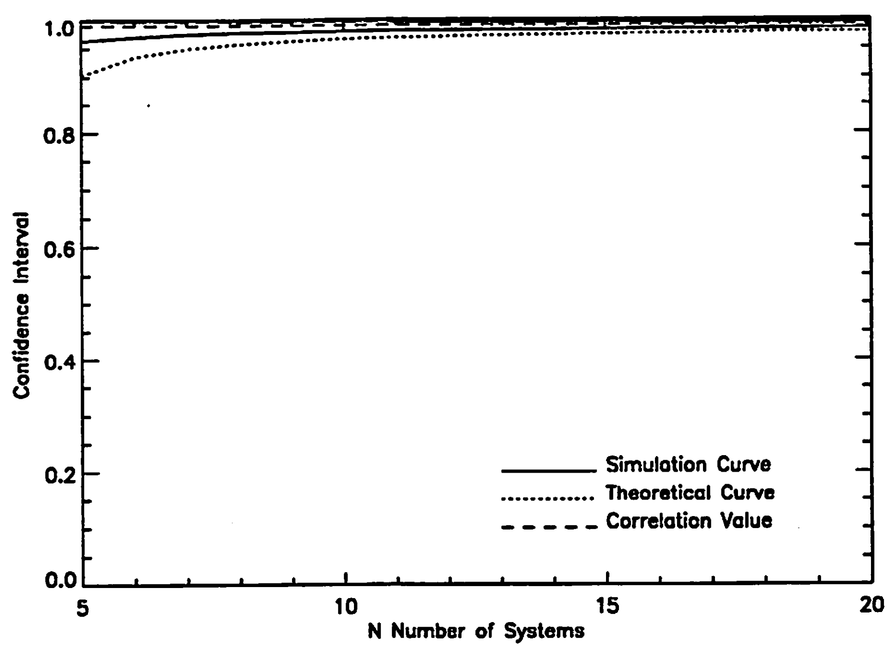


Figure A12 90% Confidence Interval for ρ = .99, N = 5 to 20

It is important to note that a correlation coefficient estimate is not statistically different from any other within its confidence interval. That means that to have a statistically significant result from a test a certain minimum number of systems must be used. For example, with a 90% confidence interval for a ρ of .90 and only 5 systems (see Figure A10), the true correlation coefficient may be as low as .63. One cannot claim that this is statistically different from another result which actually did give .63. Also, as can be seen from the plots, improvement in the confidence interval diminishes beyond N=15 or so. The knee of the curve is around N=10. Therefore, the minimum number of systems chosen should be at least 10 to 15.

If we assume that all test scenes will be used for each system chosen, then the number of clips to be subjectively rated will be (Number of Systems) x (Number of Scenes). For a single test, a reasonable number of clips would be between 150 and 250. The CCIR Recommendation 500-3 limits the length of time for a viewing session to half an hour and the number of sessions per viewer to 4. Therefore, the scene length must also be considered. Assuming that each test scene is about 10 seconds duration, a single laboratory should be able to test 15 systems with 10 to 16 test scenes.

Glossary

Exchange Carrier (EC).

The telecommunications common carrier franchised to provide telecommunications services within one or more exchanges. An EC may also provide exchange access service, intra-LATA long-distance service, and in some unusual cases, inter-LATA service.

Interexchange Carrier (IC).

A telecommunications common carrier authorized to provide telecommunications services between LATAs. An IC may also provide service within some LATAs.

Local Access and Transport Area (LATA).

A geographic area established for the provision and administration of telecommunications services. A LATA encompasses one or more exchanges that have been grouped to serve common social, economic, and other purposes.

Network Interface (NI).

The point of demarcation between the carrier's facilities and the customer's installation which establishes the technical interface and division of operational responsibility.

Point of Termination (POT).

The point of demarcation between carriers which establishes the technical interface and division of operational responsibility.

Items Under Study

Addendum

Standard Values

This addendum gives standard values for the video, audio, and audio-visual performance parameters. Standard values are given for each of the audio and video performance levels in Table 1 on page 19 and Table 2 on page 20, respectively.

Video Standard Values. Table 3 gives the standard values for each spatial-temporal performance level in Table 1 on page 19 (only spatial performance levels 1 through 3 have been included here). The performance parameters are defined in section 5.1 on page 17 and section 6.1 on page 22.

			(Spatial Level, Temporal Level)								
			(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
Out of Service Specifications	Parameter 1	Min									
		Max									
	Parameter 2	Min									
		Max									
	Parameter 3	Min									
		Max									
	etc.	Min									
		Max									
In Service Specifications	Parameter 1	Min									
		Max									
	Parameter 2	Min									
		Max									
	Parameter 3	Min									
		Max									
	etc.	Min									
		Max									

Table 3 Standard Values for Video Parameters

Audio Standard Values. Table 4 gives the standard values for each audio performance level in Table 2 on page 20. The performance parameters are defined in section 5.2 on page 19 and section 6.2 on page 23.

			Performance Level			
			1	2	3	4
Out of Service Specifications	Parameter 1	Min				
		Max				
	Parameter 2	Min				
		Max				
	Parameter 3	Min				
		Max				
	etc.	Min				
		Max				
In Service Specifications	Parameter 1	Min				
		Max				
	Parameter 2	Min				
		Max				
	Parameter 3	Min				
		Max				
	etc.	Min				
		Max				

Table 4 Standard Values for Audio Parameters

Audio-Visual Standard Values. Table 4 gives the standard values for each audio-visual performance level. The performance parameters are defined in section 5.3 on page 20 and section 6.3 on page 23.

Out of Service Specifications	Parameter 1	Min	
		Max	
	Parameter 2	Min	
		Max	
	Parameter 3	Min	
		Max	
	etc.	Min	
		Max	
In Service Specifications	Parameter 1	Min	
		Max	
	Parameter 2	Min	
		Max	
	Parameter 3	Min	
		Max	
	etc.	Min	
		Max	

Table 5 Standard Values for Audio-Visual Parameters