COMMITTEE T1

CONTRIBUTION

Document Number: T1A1.5/92-138

****

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

****

TITLE: An Automated Technique for Measuring Transmitted Frame Rate (TFR) and Average Frame Rate (AFR)

****

ISSUE ADDRESSED: Objective Video Performance Testing

****

SOURCE: NTIA/ITS, Stephen Wolf

****

DATE: July 13, 1992

****

DISTRIBUTION TO: T1A1.5

****

KEYWORDS: Video Performance Testing, Objective Quality, Subjective Quality

****

DISCLAIMER:

****

Introduction

The Institute for Telecommunication Sciences (ITS) presented contribution T1A1.5/92-112 entitled “Objective Measures of Video Impairment: Analysis of 128 Scenes” at the last T1A1 meeting. That contribution contained an International Broadcasting Convention (IBC) paper which discussed 3 objective parameters, (m_1, m_2, m_3), that have been found to be very useful for predicting subjective quality. Two of these parameters, m_2 and m_3, used statistics that were extracted from the first order motion difference images of the original and degraded video sequences. In this contribution, the same basic information is used to dynamically measure the Transmitted Frame Rate (TFR) and the Average Frame Rate (AFR) of the codec. The TFR measurement algorithm gives the TFR spectrum, so that in cases where the TFR is adaptive, each TFR being used by the codec is obtained. The AFR measurement algorithm gives the average number of frames per second that were transmitted by the codec during a selected time interval. The TFR and AFR measurements presented here can be used for in-service (section 6 of the VTC/VT draft standard, T1A1.5/92-107) as well as out-of-service (section 5 of the VTC/VT draft standard) testing of digital video systems.

TFR Spectrum Calculation

The standard deviation of the pixel values of the motion difference image ($std_{space}(F_n-F_{n-1})$), where F_n is the sampled video frame at time n, measures the flow of motion in a video scene. Useful temporal distortion measures (m_2 and m_3 in the IBC paper) have been derived from this quantity. When a codec is performing frame repetition (dropping input video frames and simply repeating a fixed output frame), this quantity drops to approximately zero. This is true because F_n is approximately equal to F_{n-1} at the codec output. When smooth motion is present in the input video sequence, and the codec is performing frame repetition, the standard deviation of the degraded motion difference image ($std_{space}(D_n-D_{n-1})$) will contain spikes that occur at the TFR. Since these spikes were not present in the standard deviation of the original motion difference image ($std_{space}(O_n-O_{n-1})$), temporal frequency components have been added. Thus, an analysis of the frequency components of the time histories of the $std_{space}(O_n-O_{n-1})$ and the $std_{space}(D_n-D_{n-1})$ will reveal the TFRs being used by the codec. In addition to frame repetition, other sources of added temporal frequency components include periodic and non-periodic noise.

Figure 1 illustrates the above discussion for two VTC scenes and two codecs. The two scenes have been described in ITS contribution T1A1.5/92-113 as “Steve1 - The Box” and “3 People in a Row (3inrow)”. Figure 1 (top) shows the time histories of $std_{space}(O_n-O_{n-1})$ for the original Steve1 and $std_{space}(D_n-D_{n-1})$ for Steve1 after passing through a 384 Kbps VTC codec. The TFR for this codec was a constant 15 frames per second (fps). Since every other codec output frame is repeated, the $std_{space}(D_n-D_{n-1})$ drops to approximately 0 for every other frame. When the codec updates the image, $std_{space}(D_n-D_{n-1})$ is a positive value indicating the image has changed. Thus, when one looks at the time waveform of $std_{space}(D_n-D_{n-1})$, which is sampled at 30 fps, one can see spikes occurring in $std_{space}(D_n-D_{n-1})$ at the rate of 15 fps. These spikes are not in the original. Figure 1 (bottom) shows the time histories of $std_{space}(O_n-O_{n-1})$ for the original 3inrow and $std_{space}(D_n-D_{n-1})$ for 3inrow after passing through a 56 Kbps VTC codec. In this case, one sees that the TFR of the codec is adaptive. The TFR varies from 10 - 15 fps for cases of little motion to as low as 2 - 3 fps for lots of motion (the time from frame 225 on includes a pan). For low motion, 56 Kbps provides adequate bandwidth for the codec to rapidly update the moving portion of the image. In cases of large motion, the codec is unable to rapidly update the motion portion of the image with the limited 56 Kbps bandwidth. Thus, the codec has resorted to updating the image only 2 or 3 times a second.

The algorithm for obtaining a direct readout of the TFR spectrum for the codec is simple and straightforward. One merely computes Fast Fourier Transforms (FFTs) of the two waveforms (original and degraded), and performs a bin by bin normalization of the resulting power spectrums. Bin by bin division of the power spectrum of $std_{space}(D_n-D_{n-1})$ by the power spectrum of $std_{space}(O_n-O_{n-1})$ results in values greater than one where spectral energy has been added, values equal to one where spectral energy has neither been added nor removed, and values less than one where spectral energy has been removed. Note that since the sampling rate of the two time histories is 30 fps, temporal components higher than 15 Hz will be aliased, or folded back, onto the 0 to 15 Hz spectrum. However, one can distinguish an aliased TFR of say 20 fps from a non-aliased TFR of 10 fps by observing the spikes in the original time waveform of $std_{space}(D_n-D_{n-1})$. In equation form, the above process may be written as:

$$FFTratio = \frac{|FFT\{std_{space}(D_1-D_0), std_{space}(D_2-D_1), \ldots, std_{space}(D_N-D_{N-1})\}|^2}{|FFT\{std_{space}(O_1-O_0), std_{space}(O_2-O_1), \ldots, std_{space}(O_N-O_{N-1})\}|^2}$$

where the magnitude squared and division operations are performed frequency bin by frequency bin on the two FFT outputs.

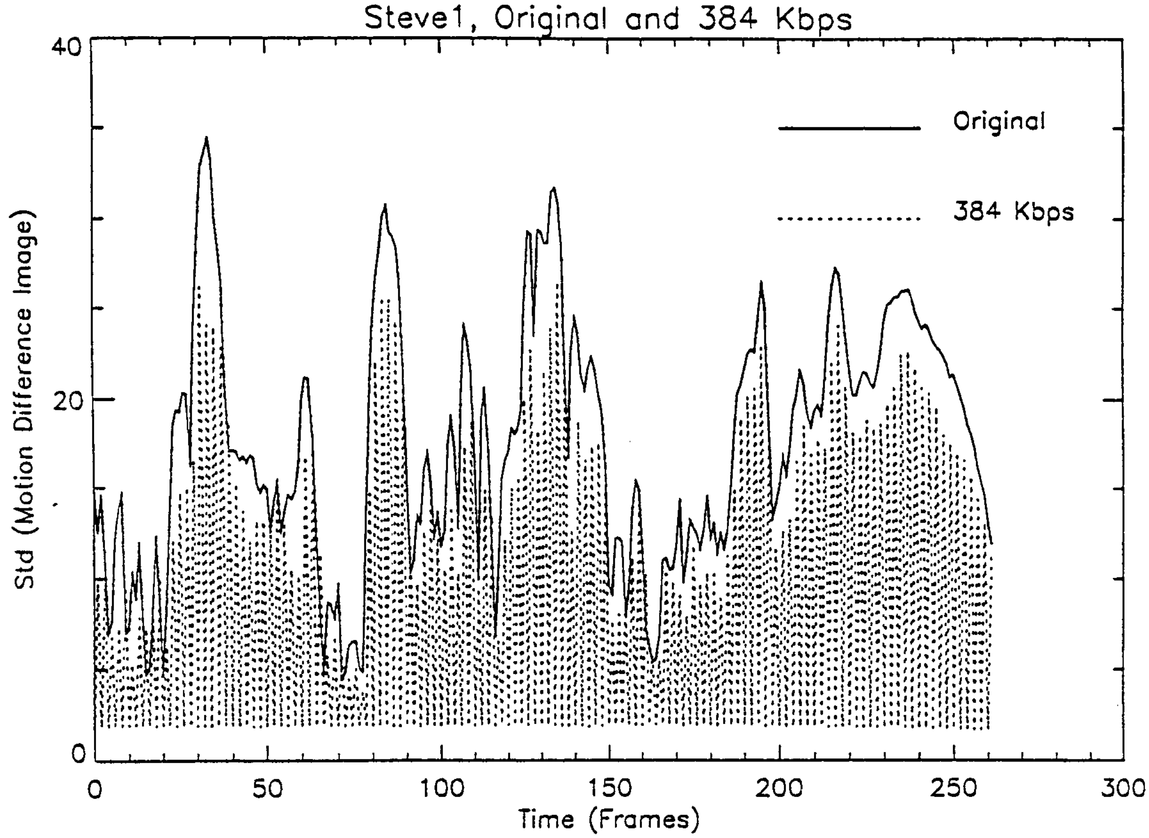
Figure 2 shows plots of the FFT ratio given by the above equation for the two cases depicted in Figure 1. Since the input is real, the FFT ratio is symmetrical about (N/2) + 1, and thus only the first (N/2) + 1 points are plotted in Figure 2. The horizontal axis has been scaled for 0 to 15 fps. Figure 2 (top) shows a very strong response at 15 fps for the codec that had a constant TFR of 15 fps. The adaptive TFR codec is shown in the bottom of Figure 2. It gives a large response at 6 fps and lesser responses at approximately 2.5 fps, and 8-15 fps. Thus, a readout of the entire TFR spectrum is obtained. The accuracy of the readout is determined by the width of the time window used for computation of the TFR.

AFR Calculation

The AFR may be directly calculated by counting the number of spikes in the codec output waveforms shown in Figure 1. Therefore, AFR = (number of spikes in codec output) / (time interval used to count number of spikes). Using the adaptive threshold method for detecting peaks that is given in contribution T1A1.5/92-139, the AFR of the 384 Kbps codec given in Figure 1 (top) was calculated to be 15 fps and the AFR of the 56 Kbps codec given in Figure 1 (bottom) was approximately 12 fps.

Conclusion

This contribution has presented simple methods for computing the TFR and AFR of a codec. These methods utilize information that has already been demonstrated to be useful for measuring the perceived quality of the video picture. Thus, these new measurements of TFR and AFR can be obtained with little additional computational complexity. The TFR and AFR measurements presented here can be used for in-service (section 6 of the VTC/VT draft standard, T1A1.5/92-107) as well as out-of-service (section 5 of the VTC/VT draft standard) testing of digital video systems.



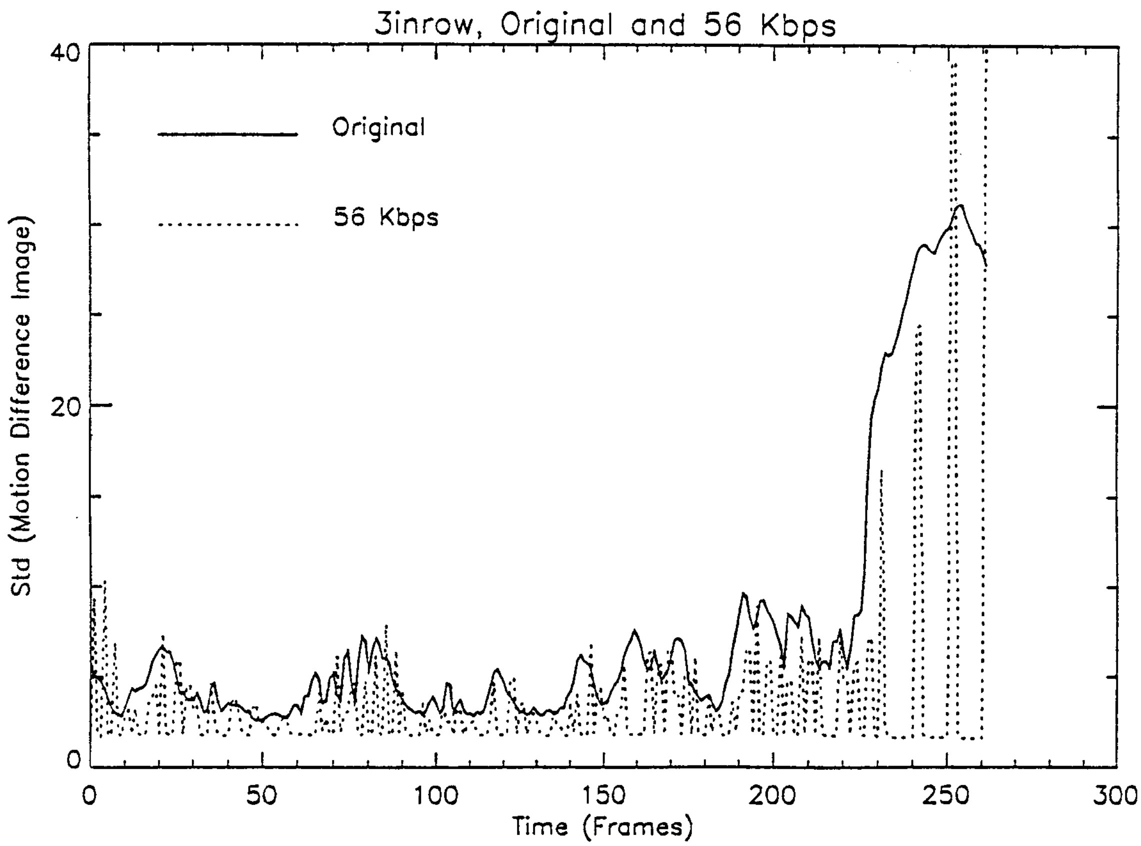


Figure 1 - Example Time Histories of $std_{space}(F_n - F_{n-1})$

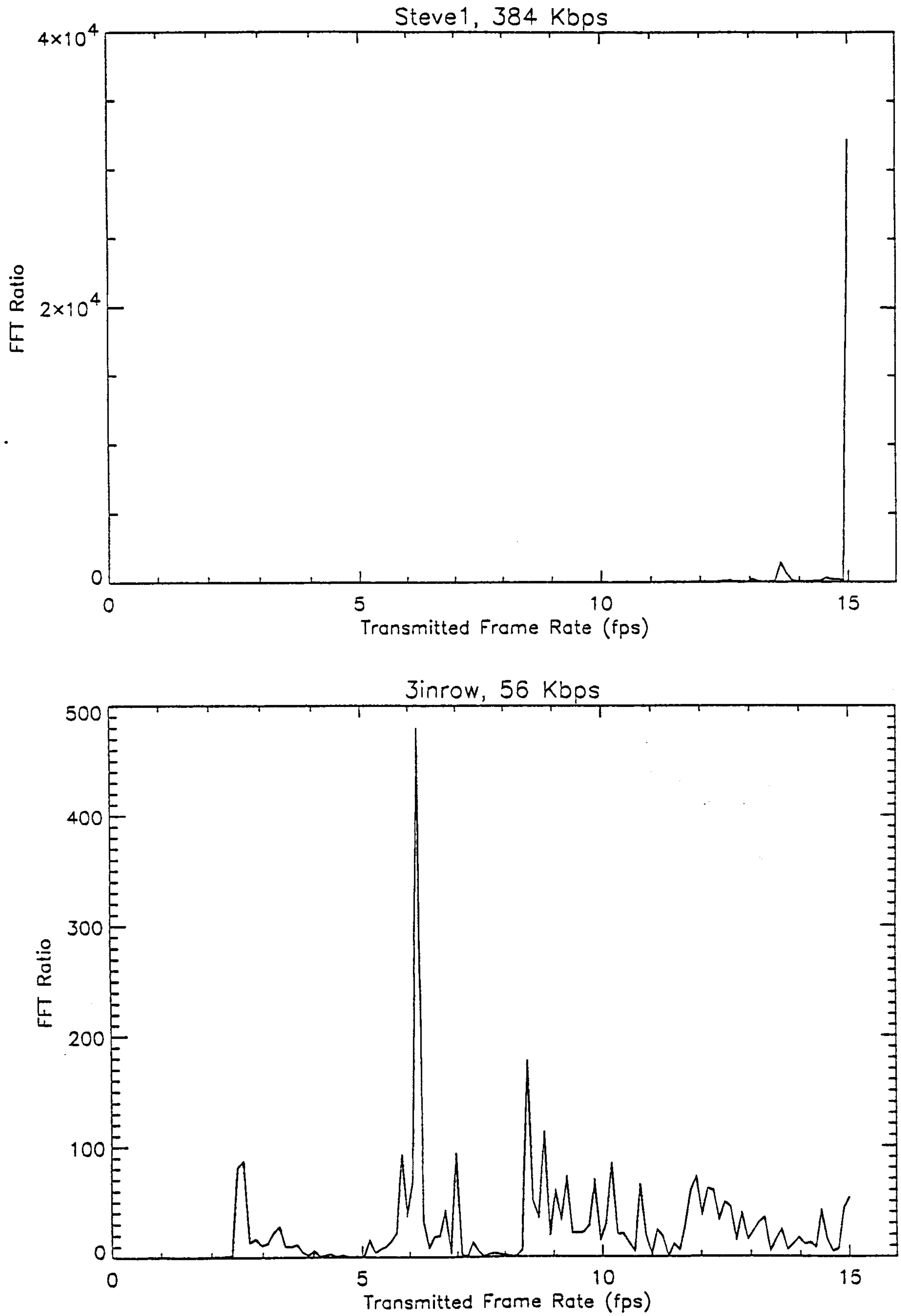


Figure 2 - FFT Ratio Plots of Data in Figure 1