CONTRIBUTION TO T1 STANDARDS PROJECT

****

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

****

TITLE: Visual Channel Delay and Frame Rate Measurement - Results Using a Prototype Measurement System

****

AUTHOR: A. C. Morton

SOURCE: AT&T COMMUNICATIONS

CONTACT: A. C. Morton
AT&T Bell Laboratories
Room HO 3M-533 P.O.Box 3030
101 Crawfords Corner Road, Holmdel, NJ 07733-3030
908 - 949 - 2499

****

DATE: January 9, 1995

****

DISTRIBUTION: T1A1.5

****

ABSTRACT:

This contribution describes a prototype system designed to measure the transmission delay of a visual channel. The system differs from other methods by accurately quantifying delay through a series of measurements such that any delay variability will be revealed. The measurements can be treated as samples of a delay distribution, and summarized with the usual statistics. The prototype also determines the inter-arrival times of non-repeated frames, making calculation of an average frame rate possible. This method should be considered for inclusion in a standard addressing video performance measurements.

****

Visual Channel Delay and Frame Rate Measurement Results Using a Prototype Measurement System

1. Introduction

Knowledge of the transmission delay in a visual channel is essential when assessing the synchronization between video and audio, or when determining the suitability of a transmission system for interactive/conversational use. This contribution describes a prototype visual channel delay and frame-rate measurement system based on the method and considerations outlined previously in T1A1.5/94-126. The method employed here expects and measures the instantaneous delay variations that are frequently present in video conference systems and services today. The treatment of delay as a variable with a distribution of values dependent on several external conditions (e.g. bit rate, video test stimulus, etc.) sets this approach apart from earlier methods.

This paper is organized as follows: Section 2 gives an overview of the measurement method. Section 3 gives the results using the prototype. Section 4 gives the conclusions and areas for further work.

2. Overview of the Method

This method follows the steps listed below:

1. Collect a sequence of video frames at the receiver in response to a source sequence played into the transmitter.
2. Establish a common time scale between the received and source sequences.
3. Determine the frames in the received sequence that are Active (non-repeated frames), and continue processing with only these frames.
4. Determine the distribution of Active frame inter-arrival times (and subsequently calculate the average frame rate from this distribution).
5. Determine the distribution of transmission delays by matching each Active frame with its counterpart in the source sequence and calculating the time difference.

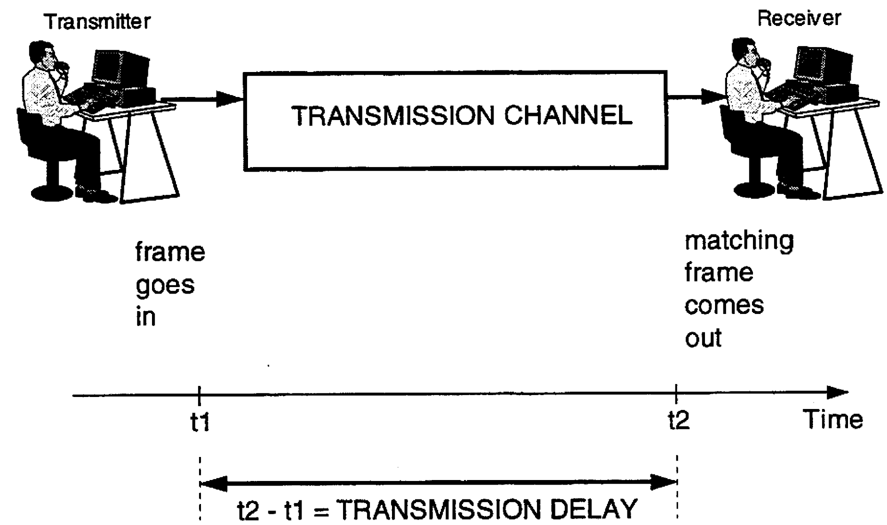

Figure 1 Visual Channel Transmission Delay

Figure 1 illustrates the simple calculation of transmission delay. The actual process of finding the best source frame - receive frame match is one of the complexities we will discuss in the next section. Another complication is the transmission delay variation, the value of t1-t2 changes over time - even from frame to frame. Finally, some video frames that enter the channel at the transmitter will never be displayed at the receiver because the channel bit rate is too low to allow all frames to be transmitted, and the compression system discards the frames.

3. Results Using the Prototype System

The subsections below follow the 5 steps in the overall procedure given in Section 2.

3.1 Collect Sequences

We collected a sequence of source frames similar to the actual sequence that was transmitted through a Hypothetical Reference Circuit (HRC). The prototype system captured its sequence of digital frames from a VHS copy of the source sequence. We captured only the luminance information from the even fields in a 320x240 pixel format. We edited the sequence down to 51 frames for further processing. The source sequence selected for this evaluation was "susie" and the HRC is number 15 (H.261 video compression using CIF at 62.4 kB/s).

The capture operation on the received sequence of frames (at the output of the HRC) was also collected from VHS tape, using the same conditions as above.

3.2 Common Time Scale

To measure delay, we must be able to align the transmitted and received sequences exactly as they appeared when entering and leaving the channel. There are two methods which will establish a common time scale between the frame sequences. The first requires both sequences be captured simultaneously or stored (on tape) with synchronous time codes. These time codes were not available on the VHS tape copies.

Figure 2 illustrates a second method that uses a video switcher to permit capture of a few frames of the source sequence using the same sequence numbers as the received frames. When the newly captured source frames are matched with corresponding frames in the original sequence, then the offset between the two arbitrary frame number sequences is found. Since this method was not employed during the taping of the received sequence, we will express our system's results only in terms of the delay variability and set the offset to zero.

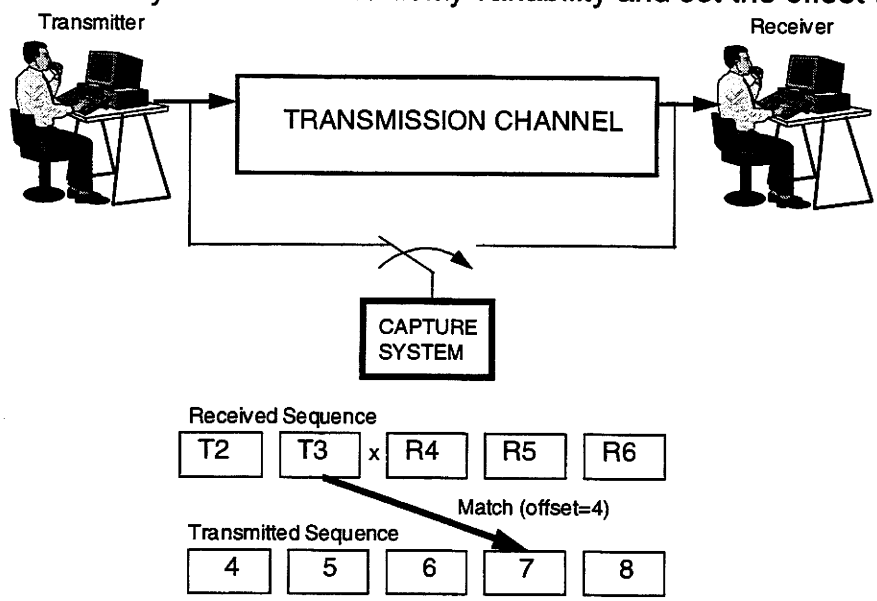

Figure 2 Method of Determining the Offset Between Captured Sequences

3.3 Find the Active Frames

In this step, we determine the frames in the received sequence that are Active frames, that is, frames which are not simply repetitions of previous frames. To find the Active frames, we calculate the pixel-by-pixel differences between a given frame in the sequence (N) and its preceding frame (N-1). We then summarize these differences over all pixels as a Mean Square Error (MSE) as shown below:

$$MSE = \frac{1}{K}\sum_{i=1}^{K}(N_i - [N-1]_i)^2$$

where N_i is the value of pixel i in frame N, $[N\text{-}1]_i$ is the value of pixel i in frame N-1, and K is the total number of pixels in a frame.

Recall that MSE is an important factor in the calculation of Peak Signal-to-Noise Ratio (PSNR), as described in T1A1.5/94-108, where:

$$PSNR = 10\log_{10}\left[\frac{S_p^2}{MSE}\right]dB$$

Figure 3 shows the frame-to-frame differences, in terms of Mean Square Error, for the captured sequences.

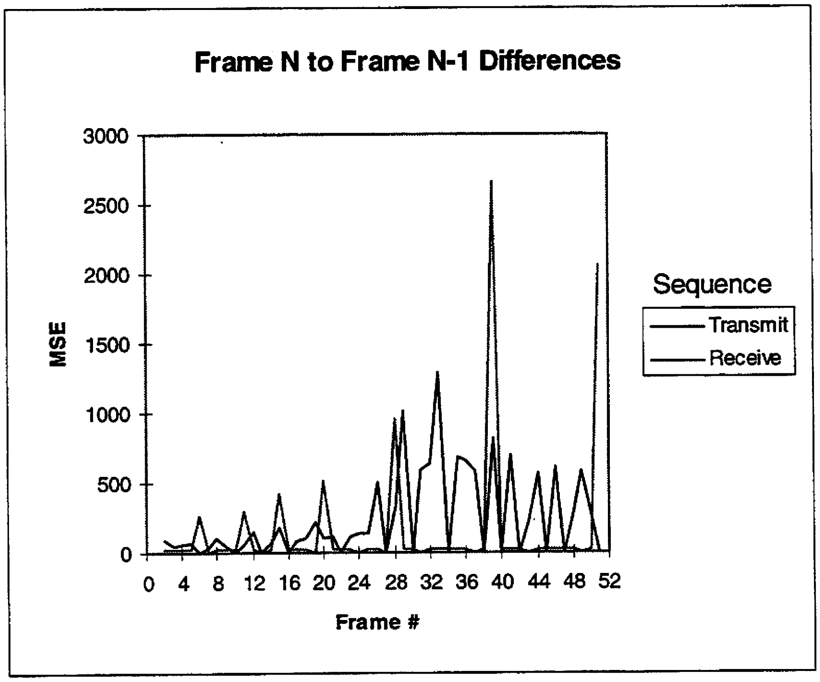

Figure 3 Frame-to-Frame Differences for the Transmit and Receive Sequences

Note that the received sequence has seven frames where differences from the previous frame are easy to distinguish from repeated frames, which differ only by the noise in the frame capture process. We declare these seven frames to be Active frames.

Active Frames	6	11	15	20	28	39	51

3.4 Active Frame Inter-Arrival Time and Average Frame Rate

We can now determine the distribution of inter-arrival times for the Active Frames. The distribution appears in Figure 4 below as a histogram.

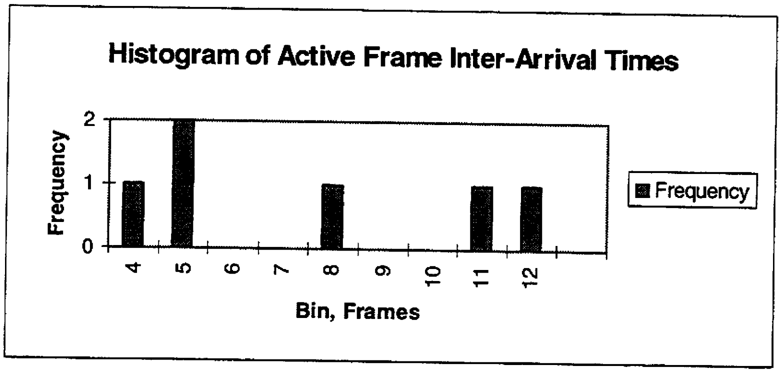

Figure 4 Inter-Arrival Times Measured by the Prototype System

There is an average of 7.5 frames between Active frames, with a standard deviation of 3.39 frames for the sample. With the measurement based on 30 frames per second, the average inter-frame time is 7.5 x 33 ms = 250 ms. This yields an average frame rate of 4 fps over this short sequence. The actual frame rate varies considerably over the sequence, depending on the amount of frame-to-frame differences, or motion in the source sequence.

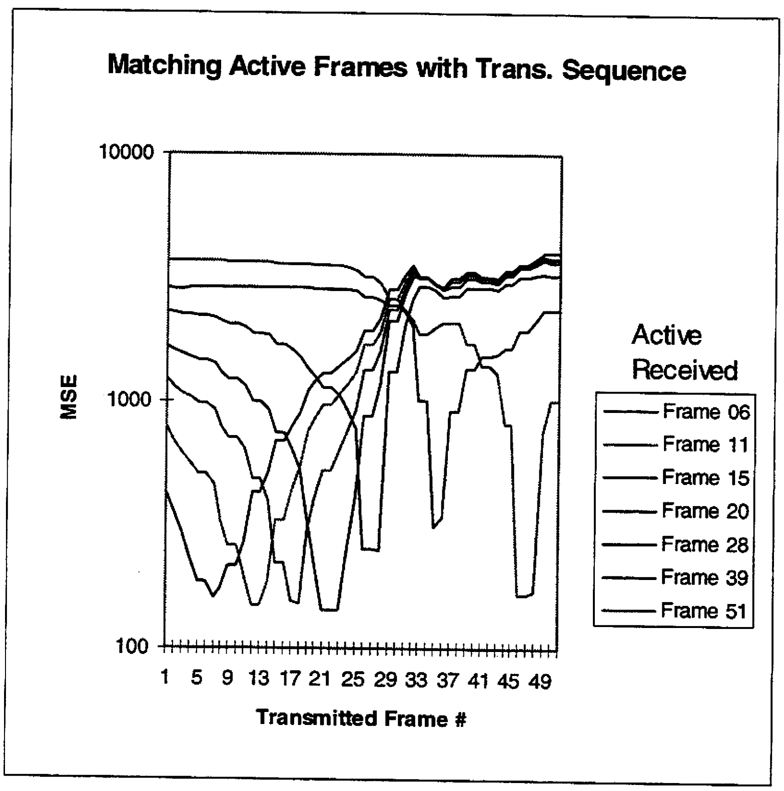

Figure 5 Results of Frame Matching Based on Mean Square Error

3.5 Frame Matching and Delay Distribution

We may now take the set of Active frames found in the received sequence and attempt to find their counterparts in the sequence submitted for transmission. To do this, we select the first Active frame and calculate its MSE when compared to all possible source frames. We take the source frame with the minimum MSE as the best match. Figure 5 shows the results of the MSE calculations for the seven Active frames.

The best matches for the Active frames are as follows:

Active Frame	6	11	15	20	28	39	51
Matching Frame	8	13	19	24	29	36	47
MSE	160.9	148.8	151.5	142.4	249.5	309.4	164.9
PSNR,dB	26.1	26.4	26.3	26.6	24.2	23.2	26.0

Frame repetition in the source sequence accounts for the close matches over several frames seen in Figure 5. Observe that the minimum MSE is higher during the last half of the sequence (with high motion) as expected, and PSNR for the matching frames is lower. PSNR varies over a 3 dB range for this short sequence, requiring an important choice among the various methods (e.g. average, minimum, etc.) to obtain an overall PSNR for the received sequence. The values obtained for PSNR are likely to be consistently low, owing to the use of a low resolution frame format (320x240), use of the entire frame in the calculation (excluding small borders would reduce the error), and use of the frames as captured (there was no attempt to align the sequences to remove any constant spatial offset).

We can conclude from these results that the matching process based on MSE was successful.

The next step is to calculate the delay distribution and observe its variability. We assume a common time scale between the sequences, and that the majority of the offset between them was removed from the transmit frame numbers. In practice, this offset could be 10 frames (330 ms) or more. We now plot the distribution of transmit-to-receive frame delays, normalized to the assumed offset.

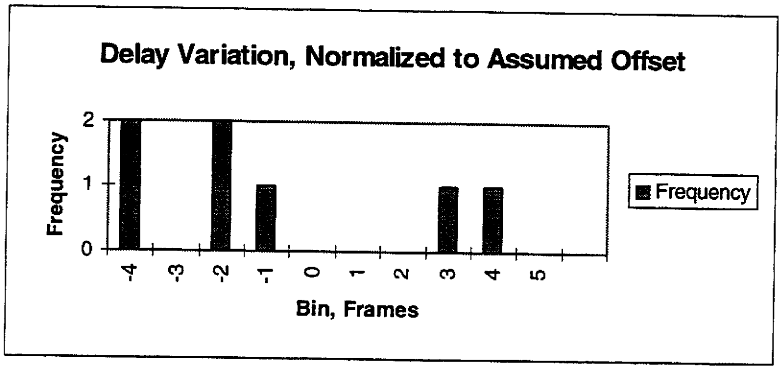

Figure 6 Delay Variation Measured by the Prototype System

We see that the delay variability is extensive, with ±132 ms range around the zero offset. It is important to measure transmission delay with a system capable of capturing and characterizing this delay variation.

4. Conclusions and Further Work

We have presented the measurements using a prototype system capable of measuring two important objective measurements of compressed digital video transmission systems, visual channel delay and frame rate. This system approaches the measurement problem with the expectation of short-term variation in transmission delay and frame rate measurements inherent in the devices under test, and characterizes the variation so that descriptive statistics can be applied to summarize the findings. The prototype has succeeded in making variable delay and frame rate measurements on its first outing, despite the distractions of VHS video storage and low resolution frame capture.

There must be further demonstrations of the prototype's measurements to seek practical refinements to the original approach and method (described in T1A1.5/94-126). For example, a method to establish a common time scale between the transmit and receive sequences must be demonstrated. Also, the video sequences must be collected without any frame repetition in the capture system. There may also be benefits from removing any spatial offset between the frame sequences, use of sub-image comparisons, and matching frames based on other measures, such as spatial correlation. Other suggestions for improvements are welcome. These areas will be addressed as time permits.