COMMITTEE T1–TELECOMMUNICATIONS

STANDARDS CONTRIBUTION

STANDARDS PROJECT:	Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service
SUBJECT:	The Impact of Differential Delay on Video Conferencing Quality
SOURCE:	Bellcore
CONTACT(S):	Stephen Bersey Bellcore 331 Newman Springs Rd Red Bank, NJ 07701 (908) 758-3408 Bill Cotton Bellcore 331 Newman Springs Rd Red Bank, NJ 07701 (908) 758-2510
DATE:	March 8, 1996
DISTRIBUTION:	T1A1.5 Technical Subcommittee Working Group Members
KEYWORDS:	differential delay, video conferencing, video
ABSTRACT:	This is a proposed contribution to ITU-T Study Group 12 which addresses that portion of Question 22 dealing with global audio/video quality evaluation by subjective means. We are seeking review and comment from T1A1.5 prior to submitting it to ITU-T as a Bellcore contribution. A study was conducted to examine the impact of differential delay on video conferencing quality. The results of the study indicate that end user's perception of video conferencing service quality varies both as a function of pure delay and audio/video synchronization. This paper recommends that industry suppliers of video codecs provide audio/video synchronization and delay performance information for a suite of standard video conference clips as part of their product specifications.

1. INTRODUCTION

The loss of synchronization between the audio and video (i.e., differential delay or "lyp-sync") components during a video conference can be very disruptive to the end-user[1]. Although it has been found to have a negative influence on the video conference participants impression of the service quality, there is conflicting data on the acceptable limits of differential delay[4]. In 1995, Bellcore conducted a study which investigated the acceptable limits of differential delay[2,3]. The study concluded that a possible reason for the conflicting data could be attributed to the task that the participants were involved with during the experiments. A desirable task would promote a high level of speech activity and encourage eye-to-eye contact.

This paper reports on a new video conferencing differential delay study which uses a well established task to elicit high levels of both visual and verbal communication between subject pairs. This work is in support of ITU-T Study Group 12, Question 22/12, which focuses on the audiovisual quality in multimedia services. The objective of this contribution is to report the findings on this study to T1A1.5, and to provide an update of our continued work in this area.

2. METHOD AND PROCEDURE

2.1. Apparatus

In this study two video conferencing stations were configured using two test rooms in Bellcore's New Technology Performance Laboratory. These rooms are 17 by 12 feet with sound dampened walls. Ambient lighting, monitor settings, and sound levels were adjusted to levels which were agreed to be comfortable by the laboratory staff. A 6 by 3 foot conference table was centered in front of a conferencing system and the conferencing participants sat at a distance of 6 times the picture height of the monitor.

Each room was equipped with a oak console which housed a Sony DXC-151A CCD color video camera, a Sony PVM-2530 25" color monitor, a Panasonic WS-A10E-K speaker, and a Coherent 3000 echo cancelling system (see Figure 1). There was no video codec used in the experiment. The study used uncompressed video and audio channels throughout the testing. This configuration approximates the upper limit of the performance quality of a conferencing system employing compression and the external conferencing components listed above. The video output from the CCD camera and the audio output from the microphone were input to a Prime Image Pipeline video/audio delay generator located in the main laboratory. This apparatus inserts delay independently in both the audio and video channels. The delay generator was verified for accuracy by using Alfred C. Morton's procedure for measuring visual channel delay (T1A1.5/95-140 and T1A1.5/96-101R1)[2,3]. Next, the audio and video signals were routed from the delay generator to the experimenter's console where they were monitored. Finally, the signals were routed to the other conference room for display and audio amplification in the conferencing console. Participants used a Fujitsu handheld computer with a GUI interface and a stylus to provide service quality ratings. The experimenter was able to see and hear both participants during the sessions.

2.2. *Procedures*

2.2.1. Participants

A total of 24 people participated in this study. Participants were run in pairs, with one participant in each conference room. Each participant was paid $40.00 compensation for approximately 1 1/2 hours of their time.

2.2.2. Instructions

At the start of a session, each participant was seated in one of the test rooms. First, instructions were given on how to use the GUI based voting system. A practice screen was provided for selecting and entering votes by touching a stylus on the GUI screen. Once the subjects indicated that they were comfortable with the voting procedure, the experimenter read the instructions provided in Appendix A.

The participants were told that the purpose of the study was to determine how well people can use video conferencing to communicate and that they would be using different types of video conferencing services to see each other while conversing. They were told that they would be engaged in a problem solving task which would require them to converse for two minutes, then they would be asked to indicate their satisfaction with the different types of video conferencing services by using the palmtop to vote on a nine-point rating scale (see Appendix A).

2.2.3. Experimental Design

Pure delay was defined as the delay introduced in both the audio and video channels. Pure delays of 0 and 400 ms were used in the study. Differential delay was defined as the delay which was introduced to either the audio or video channel in addition to the pure delay. The study included differential delays of 100, 200, 400, and 600 ms. On half of the differential delay trials, the video information arrived after a pure delay interval, and then after a subsequent differential delay interval, the audio signal was heard. On the other half of the differential delay trials, the audio signal was presented and then followed by the video signal. For comparison purposes, two reference conditions were presented with pure delays of 0 and 400 ms, but no differential delay.

2.2.4. Experimental Procedure

A experimental session consisted of four practice trails and eighteen experimental trials. A different random order of test conditions was used in each session. At the start of a trial, the participants were given a choice of two items that might be of use to them if stranded in the desert. They were then instructed to converse about the two choices. After two minutes, the experimenter activated the laboratory microphone and requested their selection between the two items presented. The choice was then logged into the control computer, audio communication terminated, and color bars displayed on the participant's monitor. The experimenter then requested the participants to please rate their satisfaction with the video conferencing service. Once both participants responded, the next trial was presented. After 11 trails a five minute break was provided. During the break, the audio paths were disconnected and color bars were displayed.

After all the trials were presented, the participants were debriefed. Generally, they said that they had noticed either pure delay, differential delay, or both. Several participants reported that while they noticed the presence of delay during some conversational periods, and found it annoying, they did not necessarily indicate the degradation in service quality in their ratings.

3. Results

An Analysis of Variance[4] (ANOVA) was performed on the rating data. A significant main effect for pure delay was found (p<.001). Test conditions with no delay were rated higher than conditions with 400 ms of pure delay. A significant main effect for differential delay (100, 200, 400, or 600 ms) was also found (p<.001). Rated quality decreased as the amount of differential delay increased. The main effect for type of differential delay (i.e., video leading audio, or the reverse) was not significant (p>.31), nor were any of the interactions between these variables with one exception. The type of differential delay by amount of differential delay interaction was marginally significant (p<.05).

In the two reference conditions pure delay was presented with zero differential delay.
A dependent t-test was used to evaluate the effect of pure delay on service quality. This analysis found that 400 ms of pure delay also produced a significant reduction in quality (p<.001).

A descriptive statistic called Mean Opinion Score (MOS) was computed for each delay condition from the rating data. Figure 3 shows the MOS for each delay condition. The amount of differential delay for each test condition is shown on the abscissa. Separate curves are used to indicate the results for the two levels of pure delay (i.e., 0 vs 400 ms), the two types of differential delay (video leading, audio leading), and the amount of differential delay (100-600 ms). For example, the legend entry 400 V->A refers to test conditions for which the video signal arrived after 400 ms, and the audio signal arrived after 400 ms *and* an additional delay as indicated on the X-axis. Also shown on the Y-axis of the figure are the MOSs for the two reference conditions without differential delay.

Figure 4 shows the results of the study plotted as a function of the amount of pure and differential delay. From this figure it can be seen that while both pure and differential delay degrade quality, differential delay has a much more substantial impact. For example, when no differential delay was present, the addition of 400 ms of pure delay caused a reduction in rated quality of about 1.1 MOS units. On the other hand, regardless of the amount of pure delay, the addition of 400 ms of differential delay resulted in a reduction of about 2.5 MOS units.

The main effect for differential delay is shown in figure 5. For each participant, a mean rating score for each delay condition (in this case 0, 100, 200, 400, and 600) was computed by averaging the ratings obtained for the two pure delay and audio vs video leading differential delay conditions. As is to be expected, an ANOVA found a significant relationship between overall MOS and differential delay (p<.001). Tukey's HSD test was applied to the results of the ANOVA to test for significant differences between the overall mean scores. The differences between the means are shown in Table 1. Differences that are significant at .05 level are indicated by a single asterisk, and differences that are significant at the .01 level by double asterisks.

Table 1: Differences in mean scores as a function of amount of differential delay.

	MOS_0	MOS_{100}	MOS_{200}	MOS_{400}	MOS_{600}
MOS_0 = 6.56	-	0.10	1.17**	2.53**	3.40**
MOS_{100} = 6.46		-	1.06**	2.43**	3.29**
MOS_{200} = 5.40			-	1.36**	2.23**
MOS_{400} = 4.03				-	0.86*
MOS_{600} = 3.17					-

* $p < .05$
** $p < .01$

From the table it can be seen that the MOSs for all of the delay conditions except 0 and 100 ms are significantly different from each other. Increases in differential delay beyond 100 ms produce a significant degradation in video conference quality.

4. SUMMARY

The results of this study indicate that end-user acceptance of video conferencing services will vary as a function of video coding delay and audio/video synchronization. Of particular interest in this study was the characterization of the relationship between lack of audio/video synchronization and degradation in service quality. It was found that differential delays greater than 100 ms produce a substantial degradation in the quality of service. The impairment resulting from inherent coding delay (referred to as pure delay in this contribution) was also found to produce of significant degradation in quality. Due to the impact of these factors on user satisfaction with video conferencing quality, it is recommended that industry suppliers of video codecs provide audio/video synchronization and delay performance information as part of their product specifications. National and international standards bodies could facilitate this goal by identifying a suite of video clips for use in delay performance characterization.

References

1). K. Taylor & K. Tolly (1995). Desktop video conferencing: Not ready for prime time. Data Communications, pp. 64-80.

2). A. C. Morton, "Visual Channel Delay and Frame Rate Measurement - Initial Measurements using the Prototype System," contribution to the ANSI Accredited Standards Committee T1, Working Group T1A1.5, document number T1A1.5/95-140, 1995.

3). A. C. Morton, "Draft ANSI T1 Standard on Visual Channel Delay and Frame Rate Measurement," contribution to the ANSI Accredited Standards Committee T1, Working Group T1A1.5, document number T1A1.5/96-101R1, 1996.

4). L. A. Marascuilo & R.C. Serlin, Statistical Methods for the Social and Behavioral Sciences, W. H. freeman and Company, New York, 1988.

5). S. Bersey, "Teleconferencing Differential Delay," contribution to the ANSI Accredited Standards Committee T1, Working Group T1A1.5, document number T1A1.5/95-156, 1995.

6). N. Jayant (1993). "High Quality Networking of Audio-Visual Information," IEEE Communications Magazine, pp. 84-95.

7). T. Kurita, S. Iai, & N. Kitawaki (1993). "Effects of interaction between speech delay and video delay on communication quality," Nippon Telegraph and Telephone Research and Development Journal, Vol 42(10), pp. 1207-1214.

Figure 1: Differential Delay Configuration

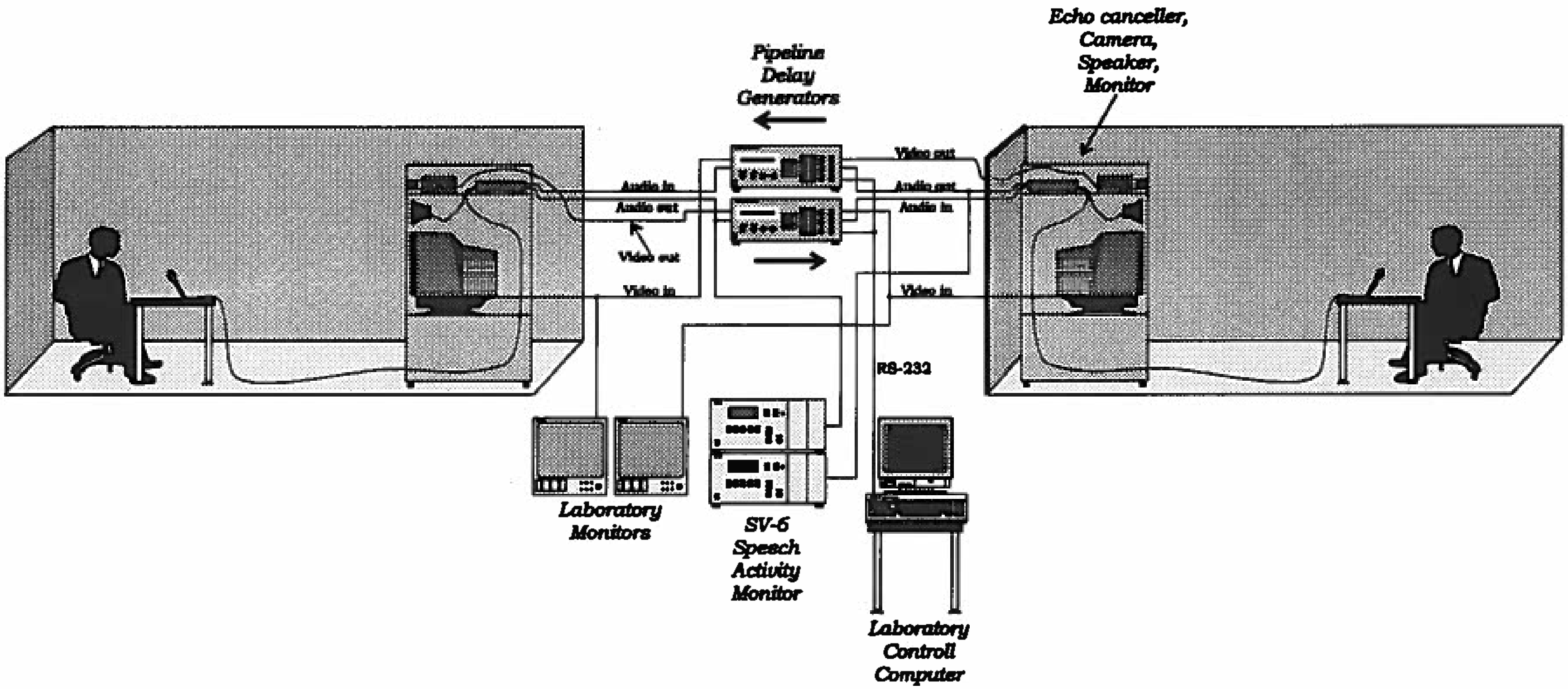


Figure 2: Handheld Computer GUI Screen

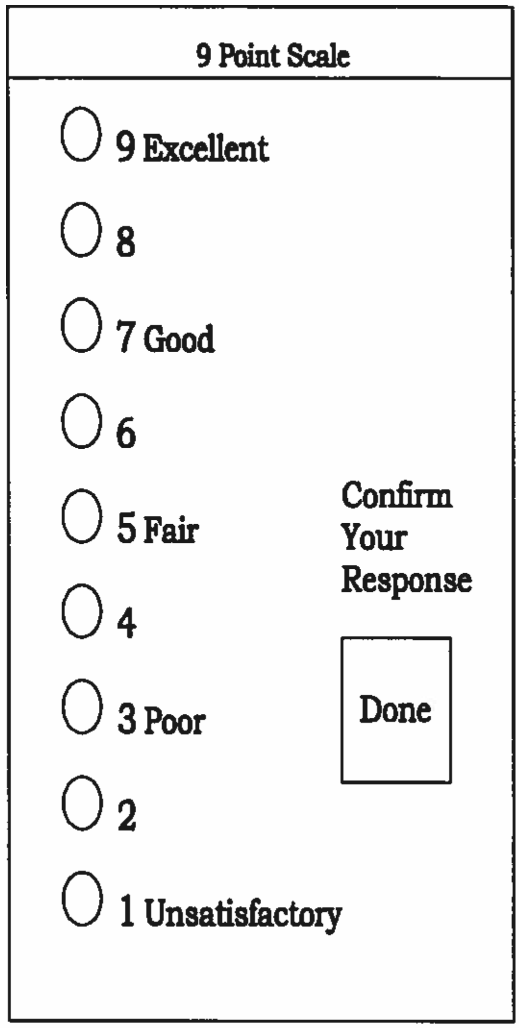


Figure 3. Mean Opinion Score as a Function of Pure Delay, and Amount and Type of Differential Delay

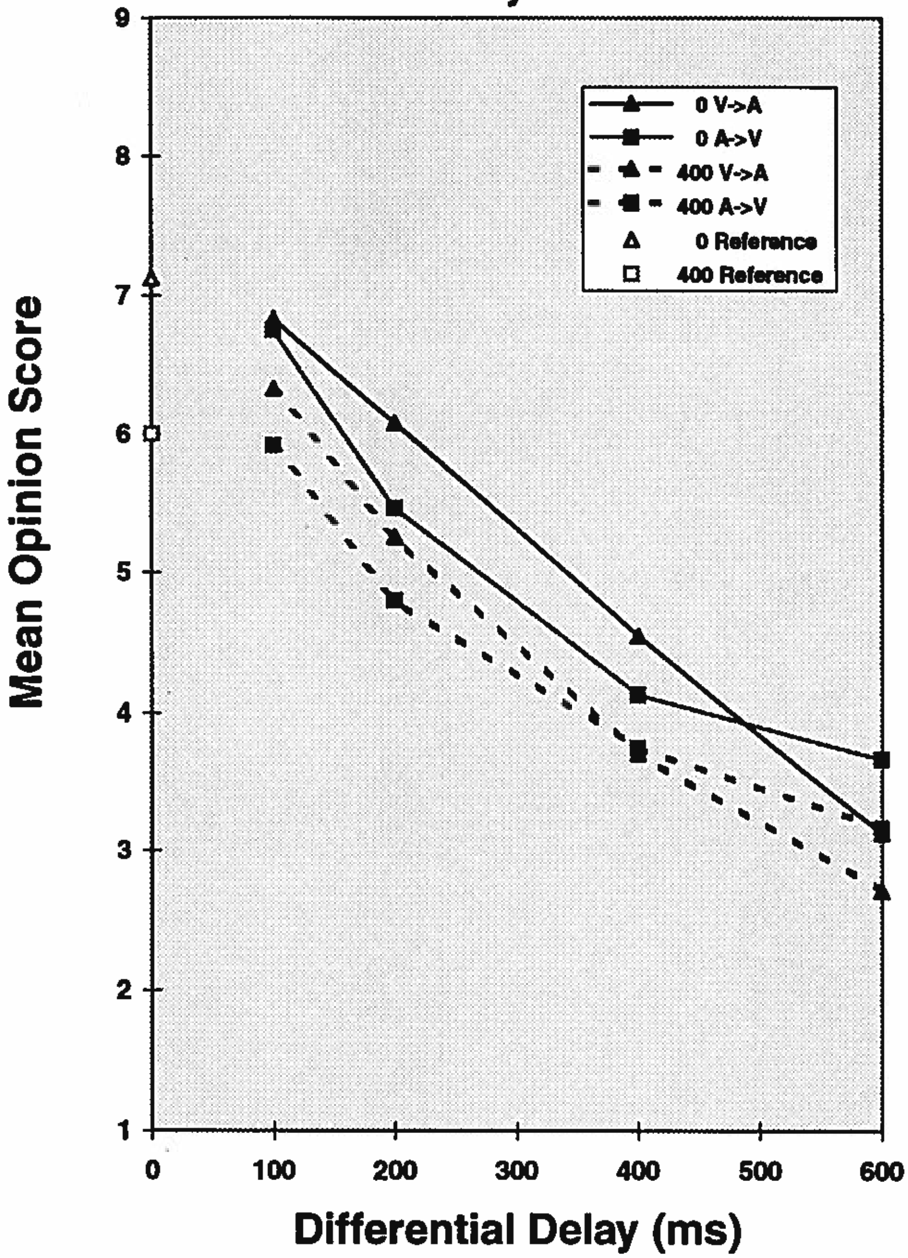


Figure 4. Mean Opinion Score as a Function of Pure and Differential Delay

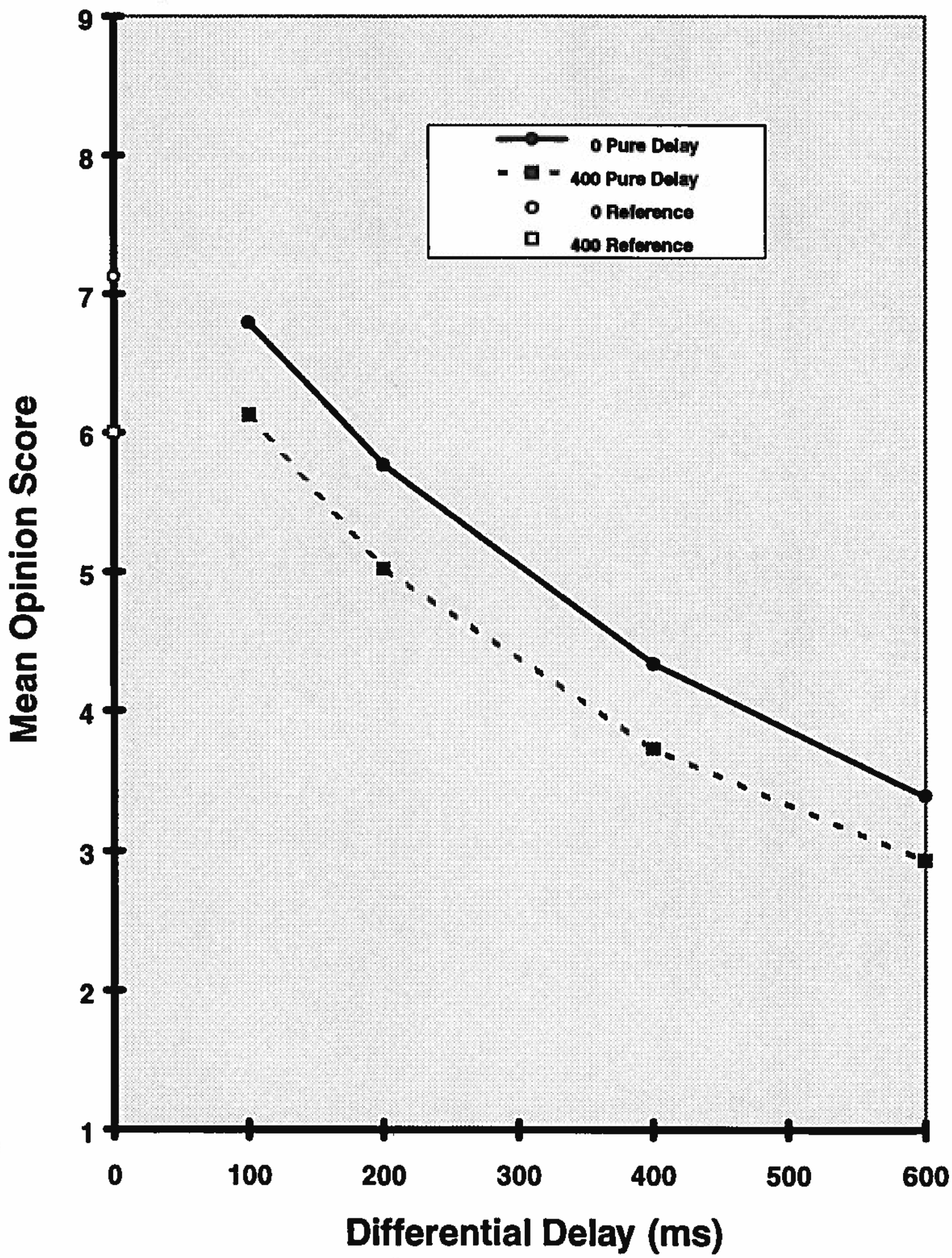
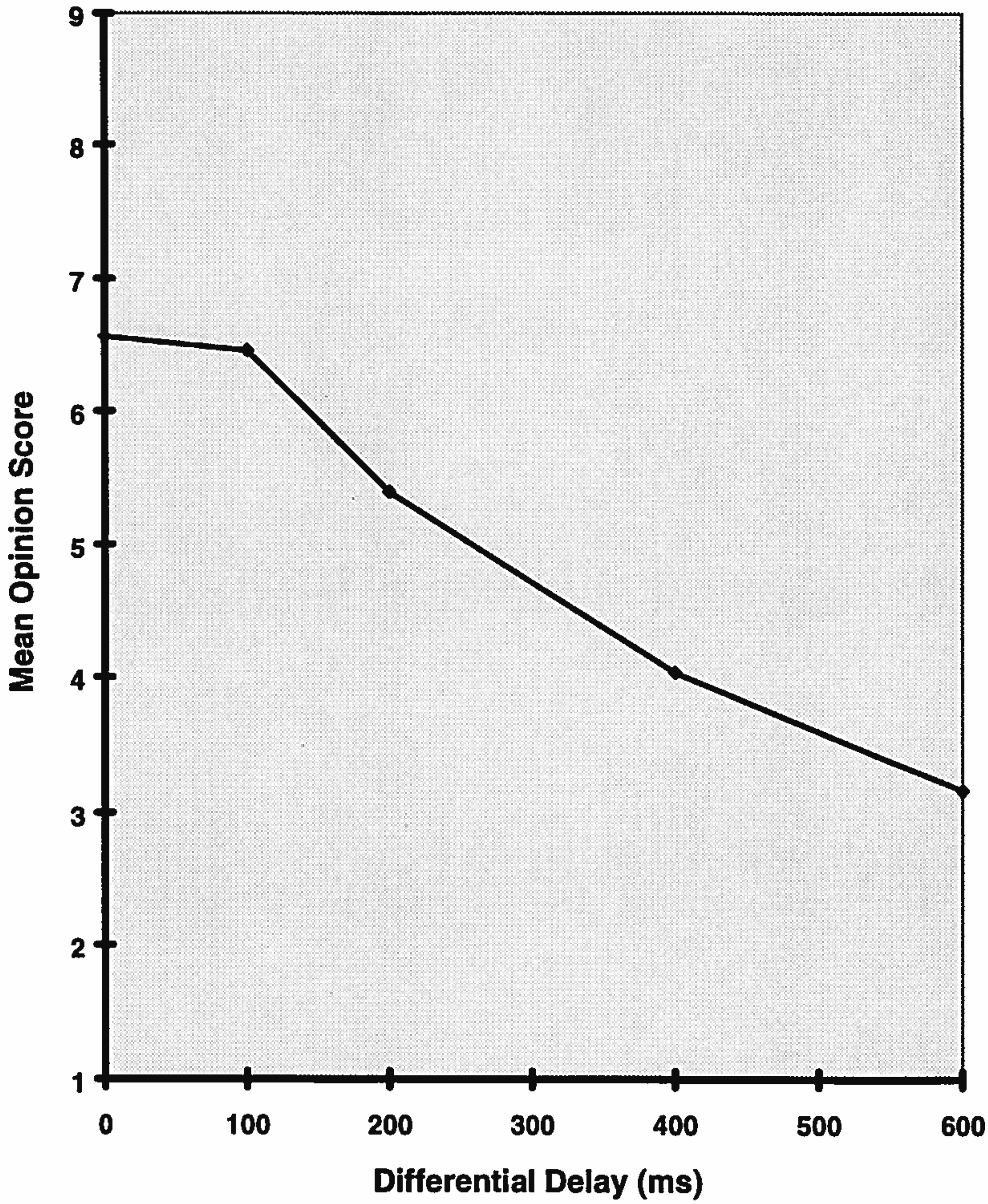


Figure 5. Mean Opinion Score as a Function of Differential Delay



Appendix A. Instructions to Participants

In this study we are investigating how well people can use video conferencing to communicate.. You will use different types of video conferencing services see each other while conversing.

In the study you will talk to each other for about two minutes. During the conversational period you will engage in a problem solving task used by such groups as the Explorer scouts. Then you will be asked to indicate your satisfaction with the different types video conferencing services using a nine-point rating scale. As demonstrated a few moments ago, you will use the stylus and response box on the table in front of you to provide your rating..

In the problem solving task, you should pretend that the two of you are the only survivors of a light plane crash in the Sonora Desert in the southwestern United States. Although neither of you are injured, shortly before the crash the pilot indicated that the plane was 65 miles off course, and 70 miles north of the nearest town. This is cause for concern since the weather report indicated daytime temperatures of 110 degrees, which means ground temperatures in excess of 130 degrees. The immediate area is quite flat, and except for some occasional cactus, appears to be rather barren. You are dressed in light weight clothing -- short sleeved shirts, pants, socks and street shoes.

You should assume that you are the actual people in the situation, and that you have decided to stick together. In the first two minute period you will decide whether the wisest decision is to stay with the plane and wait for possible rescue, or attempt to walk to the town. Then in the following two minute periods you will be given a choice of two items that might be of use to you. Use the two minutes to decide which of the two items would be of most use in increasing the chances of your survival.

If you agree on a decision before the two minutes is up, consider whether the item that was not selected might be of use in some manner that you have not considered. If you do not converse with each other for the full two minute period, you may find it difficult to rate your satisfaction with each service..