3. APPLICATION EXPERIMENT ON DESKTOP CONFERENCING

Desktop conferencing is a highly interactive application that involves real-time voice, video, and application sharing. The real-time performance and correct transmission are both important factors for usable desktop videoconferencing. In this experiment, each of the three components (voice, video, and application sharing) were analyzed for usability and stability. Additionally, a terrestrial baseline was established for evaluating changes in application performance when a satellite link was introduced.

3.1 Experiment Objectives

The experiments were intended to identify the performance issues that should be considered when a satellite link is included in the communications system. The objectives of these experiments were not intended to evaluate the COTS implementations.

3.2 Experiment Methods and Procedures

Intel Proshare® desktop-conferencing units, installed on 66-MHz/486 personal computers, were used for both proprietary and H.320 [3] experiments. In addition, one of the Proshare® units was replaced with a Picture Tel ® H.320 system to test for similar behavior and interoperability. These units represent typical COTS user equipment available and already widely distributed today. The NS/EP scenario discussed in Section 1 suggests that existing user equipment that has lost telecommunications connectivity may have its connectivity restored via a satellite link without modification to the user equipment. This user equipment most likely was not designed for use with a satellite nor the latest technology. However, the function it provides may be more than adequate for daily operations, and may be critical to daily business. The user equipment selected for these experiments is intended to model such an environment.

The experiments were conducted using the two equipment configurations illustrated in Figure 3.1. For terrestrial experiments, two desktop-conferencing units were connected to a local ISDN switch (Teleos ® Network Hub) with BRI ISDN. The ISDN switches were interconnected via PRI ISDN service from FTS2000. A bit-error simulator (Adtech SX-12) was connected between the FTS2000 PRI and the ISDN switch at the Boulder end of the circuit to inject errors into the data stream in both directions independently.

For the ACTS-based experiments, the same two desktop-conferencing units were connected to the same local ISDN switches with BRI ISDN service in Boulder (NTIA/ITS) and Gaithersburg (NIST). The Gaithersburg ISDN switch was connected via dedicated ISDN PRI service to an ISDN switch in Clarksburg (COMSAT). The Boulder and Clarksburg ISDN switches were interconnected via PRI ISDN service from the ACTS T1-VSATs in Boulder and Clarksburg. A bit-error simulator was connected between the T1-VSAT PRI and the ISDN switch at the Boulder end of the circuit to inject errors into the data stream in both directions independently.

The bit-error simulator was used to simulate error scenarios that might be observed over a link in actual operations. Two scenarios were investigated. The first was based on a "steady-state" environment. For this scenario, the statistics of the random bit errors were uncorrelated and Gaussian distributed for the duration of the experiment. This simulates the errors caused by noise from a variety of background sources. Thresholds for usability and application failure were determined. The second scenario was based on a transient error environment that causes severe error bursts for some period. These tests were intended to determine the stability/recoverability of the applications to transient error rates. A moderate error background was used with a periodic severe burst of random errors of a fixed duration. Thresholds for recovery/failure were determined by varying the burst length and intensity.

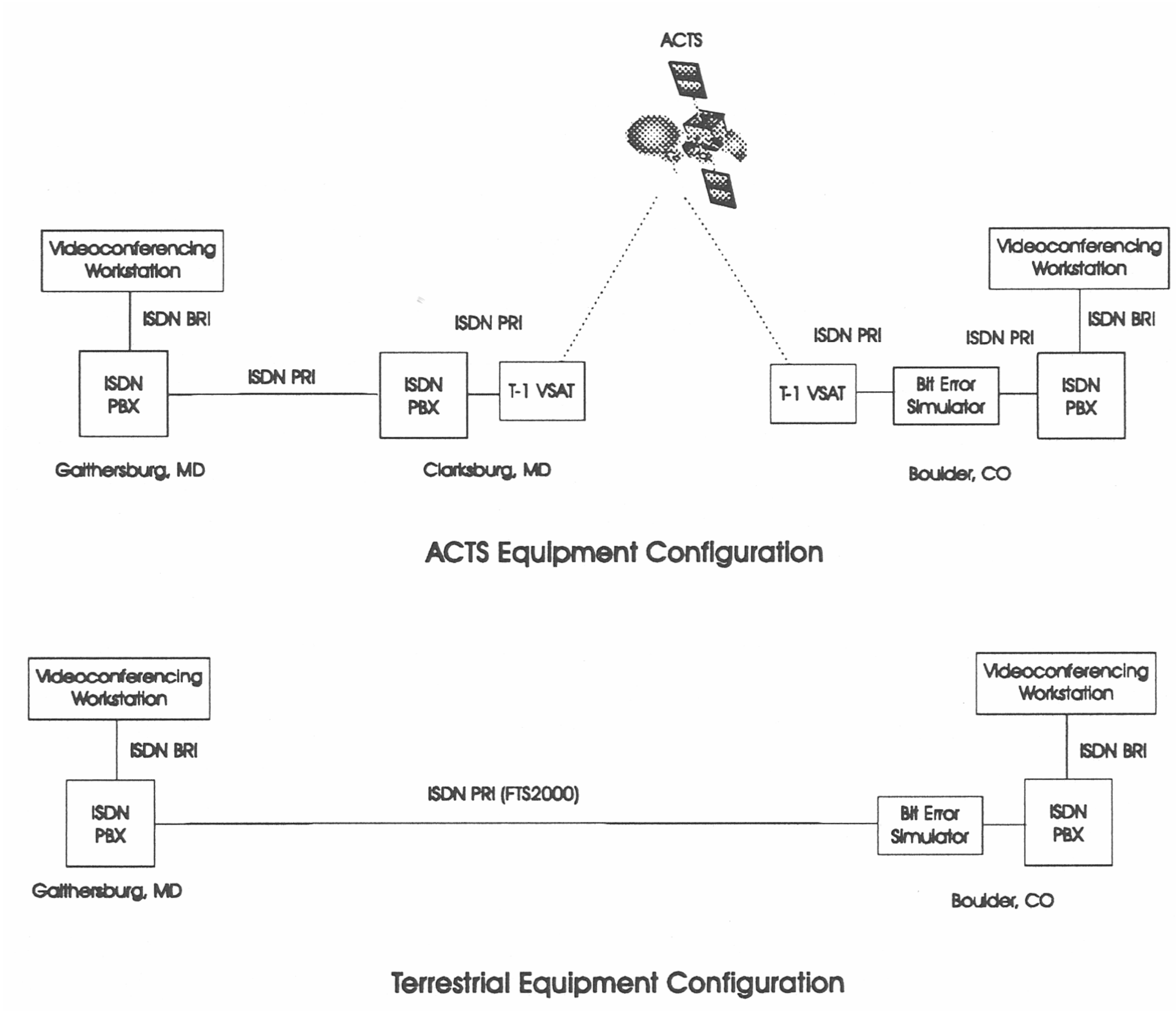

Figure 3.1. Equipment configurations for application experiment on desktop conferencing.

3.3 Metrics

Opinion Score: An opinion score (OS) was used to evaluate the subjective usability for the audio, video, and whiteboard components of the videoconferencing applications. The evaluator is assumed to be a person familiar with operation of the component being evaluated. In these experiments, the experiment operator was the only evaluator. The rating scale is defined in Table 3.1 (fractional scores were not excluded from use).

Table 3.1. Definition of Opinion Score Values

Opinion Score	Description
5 - Excellent	No difference between in-person and remote perception.
4 - Good	Subtle differences between in-person and remote perception.
3 - Fair	Obviously a remote reconstruction, but not disturbing to the user in most cases.
2 - Poor	Somewhat disturbing differences in reconstruction but still usable.
1 - Unusable	Use of system is not practical.

The following examples from everyday experience serve as guidelines for OS ratings:

- OS = 3 is the audio of a toll quality telephone connection;
- OS = 3.5 is the video quality from NTSC video (U.S. broadcast television); and
- OS = 5 is equivalent to two users sitting side-by-side editing a document on a personal computer.

Tracking response time: This is the time for the initiator of an event to observe the response of the recipient. This metric is an indicator of the ability of the system to support real-time interaction between two users.

Video error characterization: Descriptions of video errors were based on ANSI T1.801.01-1995. [4].

3.4 Experimental Procedure

The following procedure was followed:

1. Observe link signal quality via E_b/N_0 at Earth stations to verify low link BER. (A terrestrial link is assumed to have a BER less than 10^{-9}.)

2. Establish a desktop-conferencing session.

3. Set the BER for the experiment via the bit-error simulator (see Tables 3.2 and 3.3).

4. Assess video and audio quality via subjective OS, noting the nature of any audio and/or video distortions. Record OS and observations.

5. If available in the implementation, establish an application-sharing session (e.g., whiteboard). If not available, skip steps 6 and 7. (Note: H.320 does not provide for application sharing.)

6. User 1 quickly moves the cursor from one comer of the screen to the diagonally opposite comer. User 2 moves the cursor to attempt to follow the motion. User 1 records the time from User 1's initial motion to when User 2's cursor arrives at the final position via a stopwatch (tracking response time). Record subjective observations on behavior and an OS for usability.

7. User 1 continuously moves the cursor in a random pattern around the screen. User 2 attempts to follow. User 1 notes the ability of User 2 to follow a similar route. Record subjective observations on behavior and an OS for usability.

Table 3.2 defines the set of BER configurations used in step 3 of the above procedure for steady-state experiments.

Table 3.3 defines the set of BER configurations used in step 3 of the above procedure for burst BER's. These experiments characterize the transient behavior (e.g., failure/recovery) of the application in response to changes in the BER. The burst duration and the burst gap were periodic (i.e., not random). This permits viewing the experiment as a collection of burst events that could be observed independently to note transient behavior to a single burst event. The background BER is the bit error density between bursts. The burst BER is defined as the bit error density during an error burst event. The burst gap is defined as the interval between the end of one error burst and the beginning of the next. The burst interarrival time is the sum of the burst duration and the burst gap.

Table 3.2. Steady State BER Configurations

Measurement Number	BER
1	0
2	10^{-9}
3	10^{-6}
5	10^{-5}
5	10^{-4}
6	10^{-3}
7	10^{-2}

Table 3.3. Burst BER Configurations

Measurement	Background BER	Burst BER	Burst Duration (s)	Burst Gap (s)
1	10^{-6}	10^{-3}	0.5	10
2	10^{-6}	10^{-3}	1.0	10
3	10^{-6}	10^{-3}	5.0	10
4	10^{-6}	10^{-3}	10.0	10
5	10^{-5}	10^{-2}	0.5	10
6	10^{-5}	10^{-2}	1.0	10
7	10^{-5}	10^{-2}	5.0	10
8	10^{-5}	10^{-2}	10.0	10

3.5 Expected Results

Changes in picture quality due to introduction of a satellite link are likely to be manifested in various ways. Compression algorithms tend to have characteristic visible error signatures. For example, bit errors in a discrete cosine transform (DCT)-compressed picture result in distorted blocks in the picture. Lost frames produce a "jerky" picture that may be observed as a "frozen" frame. Errors in a system using differential frame updates may produce a portion of the frame with persistent distortions (object retention) until that portion is updated again. A group of several events may be recorded as a single "severe" event, as that would be the user's perception.

Satellite propagation delays may cause some difficulty in real-time whiteboard interactions. This difficulty arises from the time required for the parties to observe each other's actions. Also, the voice and video information is transmitted with higher priority than the data. This may cause actions on the whiteboard to lose synchronization with the voice and video information.

The experiments involving error bursts characterize the transient behavior (e.g., failure/recovery or interruption) of the application in response to changes in the BER. Longer burst durations are likely to cause application and/or link failure due to expiration of timers, while short bursts will probably make the application unusable (i.e., audio dropout, video frame loss, and suspended data transmission) during the burst, but recover. In the case of the error bursts with recovery, the experiment identifies the recovery behavior of the system.

3.6 Results

The graphs in Figures 3.2 and 3.3 illustrate the usability opinion scores collected for proprietary Proshare® videoconferencing over both terrestrial and satellite (ACTS) links, respectively. These scores were based on experiments with a BER defined as the mean of Gaussian bit error arrival times, which simulates errors from many independent noise sources. The statistics for these errors are time-invariant for the duration of the experiment. These experiments describe the steady-state application performance for a given environment.

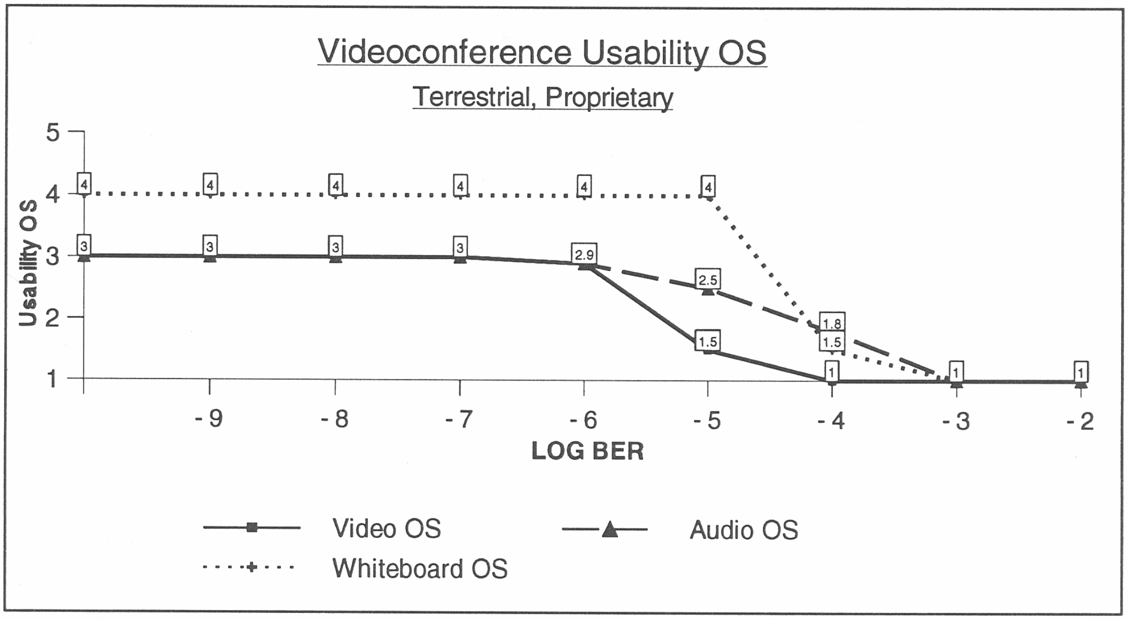

Figure 3.2. Terrestrial videoconference usability opinion scores.

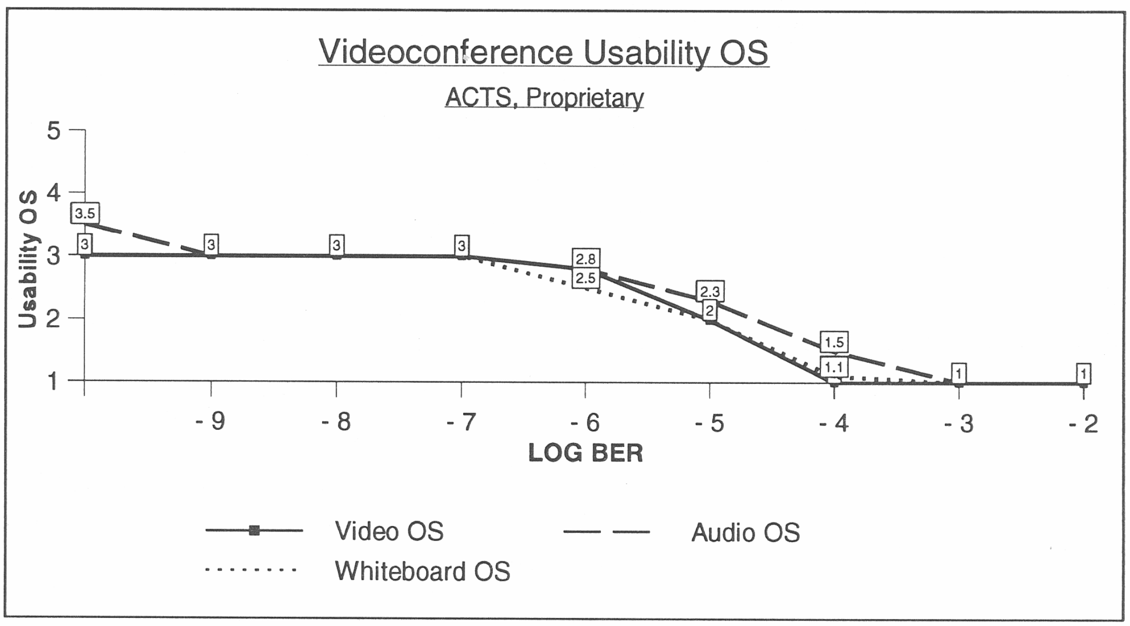

Figure 3.3. ACTS videoconference usability opinion scores.

Figures 3.4 and 3.5 illustrate the usability opinion scores collected for H.320 Proshare® videoconferencing over both terrestrial and satellite (ACTS) links, respectively. These scores were based on experiments with Gaussian bit errors that simulate errors resulting from many independent noise sources. The statistics for these errors are time-invariant for the duration of the experiment. These experiments describe the steady-state application performance for a given environment.

No whiteboard data was collected for the H.320 experiments as the H.320 recommendation does not support application sharing, it only supports audio and video.

A subset of these experiments also was performed between a Proshare® H.320 system and a Picture Tel® system to test for differences in performance with different implementations and interoperability. No differences were observed in behavior of either H.320 implementation from a performance or an interoperability perspective.

Table 3.4 summarizes the results of the simulated error burst experiments conducted with the proprietary system over ACTS. In all cases, the burst intensity was severe enough to render the application unusable during the burst. The recovery time indicates the time needed after return to the background BER for normal performance at the background BER. No whiteboard measurements were reported since the measurement intervals could not be determined reliably. However, the recovery of the whiteboard appeared to be similar to that of the video (i.e., updates occurred within a few seconds of the end of the noise burst).

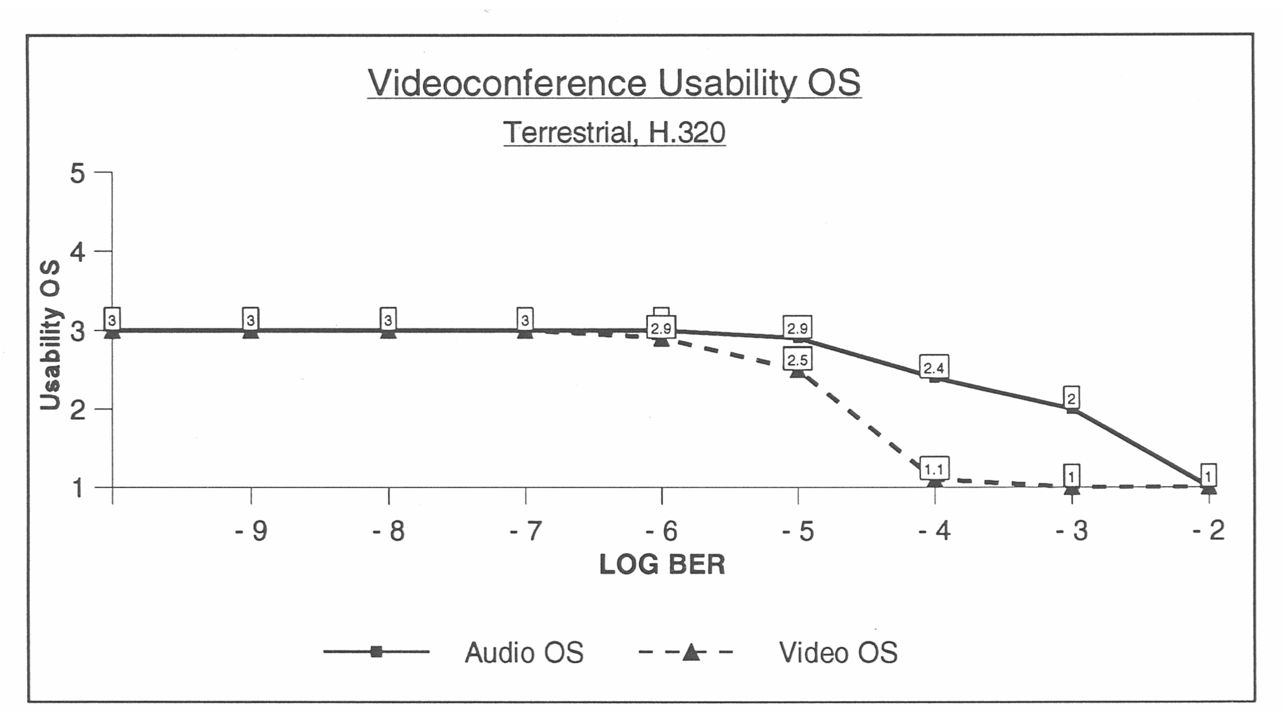

Figure 3.4. Terrestrial H.320 videoconference usability opinion scores.

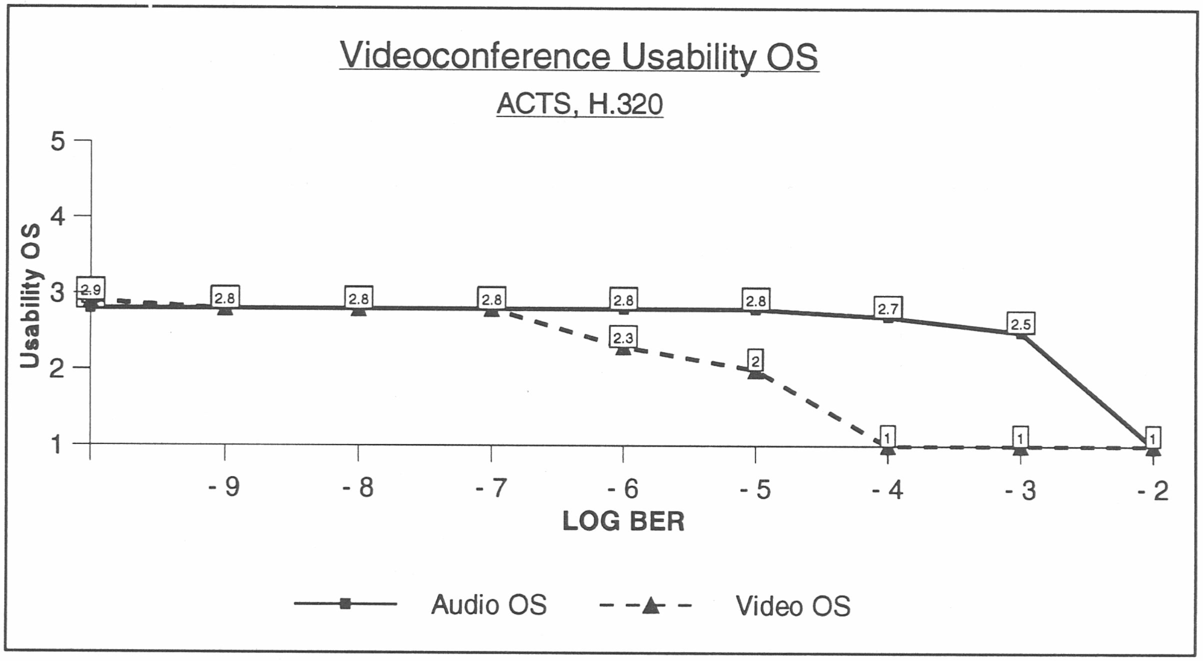

Figure 3.5. ACTS H.320 videoconference usability opinion scores.

Table 3.4. Recovery Response to Noise Burst (Proprietary, ACTS)

Background BER	Burst BER	Audio Recovery Time (s)	Video Recovery Time (s)	Burst Duration for Total Failure (s)
10^{-5}	10^{-2}	5	5	5
10^{-5}	10^{-3}	5	5	5
10^{-6}	10^{-2}	1	5	>10
10^{-6}	10^{-3}	1	5	10 (only video failed)

Table 3.5 summarizes the results of the simulated error burst experiments conducted with the proprietary system over a terrestrial connection.

Table 3.5. Recovery Response to Noise Burst (Proprietary, Terrestrial)

Background BER	Burst BER	Audio Recovery Time (s)	Video Recovery Time (s)	Burst Duration for Total Failure (s)
10^{-5}	10^{-2}	<1	2	0.5
10^{-5}	10^{-3}	<1	2	5
10^{-6}	10^{-2}	<1	2	>10
10^{-6}	10^{-3}	<1	2	10 (only video failed)

Table 3.6 summarizes the results of the simulated error burst experiments conducted with the H.320 system over an ACTS connection.

Table 3.6. Recovery Response to Noise Burst (H.320, ACTS)

Background BER	Burst BER	Audio Recovery Time (s)	Video Recovery Time (s)	Burst Duration for Total Failure (s)
10-5	10-2	static during burst only	2	0.5
10-5	10-3	static during burst only	2	5
10-6	10-2	static during burst only	2	>10
10-6	10-3	static during burst only	2	10 (only video failed)

Table 3.7 summarizes the results of the simulated error burst experiments conducted with the H.320 system over a terrestrial connection.

Table 3.7. Recovery Response to Noise Burst (H.320, Terrestrial)

Background BER	Burst BER	Audio Recovery Time (s)	Video Recovery Time (s)	Burst Duration for Total Failure (s)
10^{-5}	10^{-2}	I	2	>10
10^{-5}	10^{-3}	I	2	>10
10^{-6}	10^{-2}	I	2	1
10^{-6}	10^{-3}	1	2	>10

Figures 3.6 and 3.7 illustrate the tracking response time data collected for proprietary Proshare® videoconferencing over terrestrial and satellite (ACTS) links, respectively. These scores were based on experiments with Gaussian-distributed errors. Response times were not measured beyond 15 s as such times were considered far outside reasonable real-time interactive limits for such a primitive action.

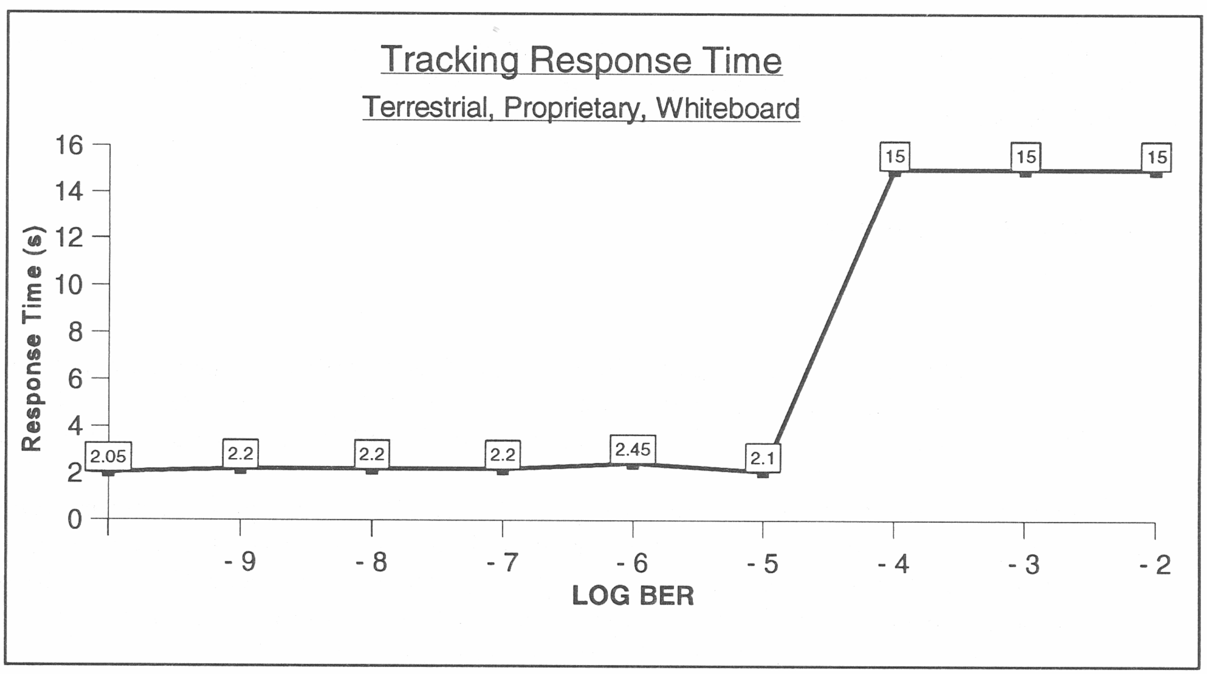

Figure 3.6. Terrestrial tracking response time.

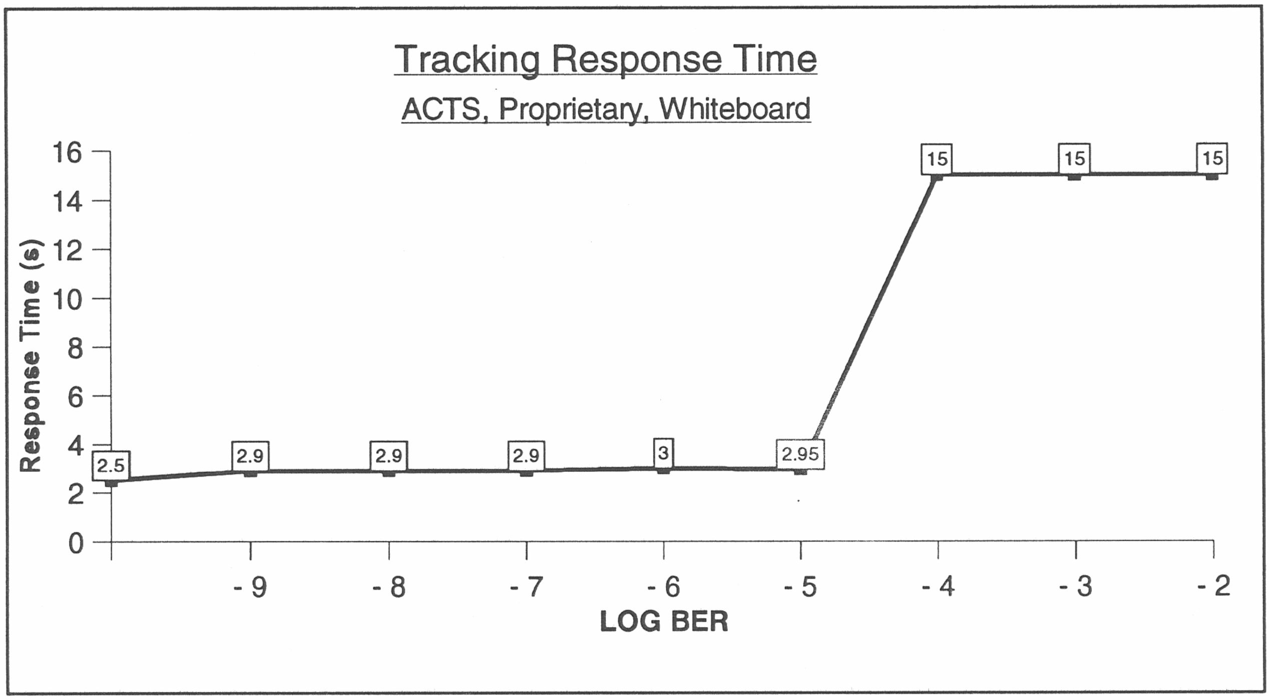

Figure 3.7. ACTS tracking response time.

Audio delays for both terrestrial and ACTS experiments caused conversation conflicts. An experiment conducted between two units in the same room revealed that processing and transmission, using only a local ISDN switch, resulted in a one-way delay of about 1s.

3.6.1 Characteristics of Errors

Video:

The proprietary video processing appeared to detect frames with errors and did not display video frames containing errors. This resulted in a "frame freeze" for the last correct frame. This resulted in jerkiness, a still frame, or no image until the first good frame was received.

A typical H.320 video frame is shown is Figure 3.8. The H.320 video presented frames with errors. Two classes of errors were noted: error blocks and motion-related artifacts. Motion-related artifacts resulted in incorrectly translated portions of the frame (Figure 3.10) or "ghosting" (Figure 3.11). The block errors resulted in block distortion or tiling in the image or blurring typical of errors in a DCT-compressed image (Figure 3.9).

Additionally, when noise bursts occurred with a high (10^{-5}) background BER, video recovery appeared to occur in multiples of 2 s. This was believed to be caused by background errors causing successive recovery cycles.

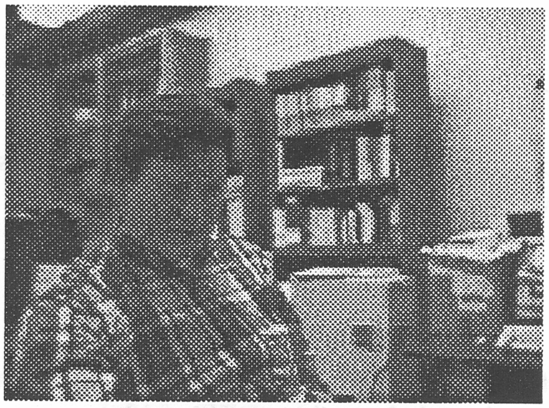

Figure 3.8. H.320 video frame with no errors.

Figure 3.9. H.320 video frame with block errors.

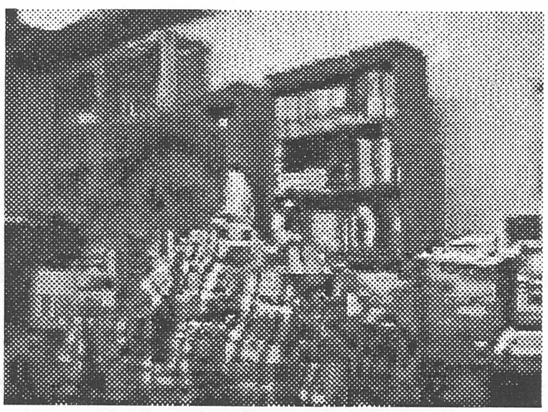

Figure 3.10. H.320 video frame with motion compensation distortion (translation).

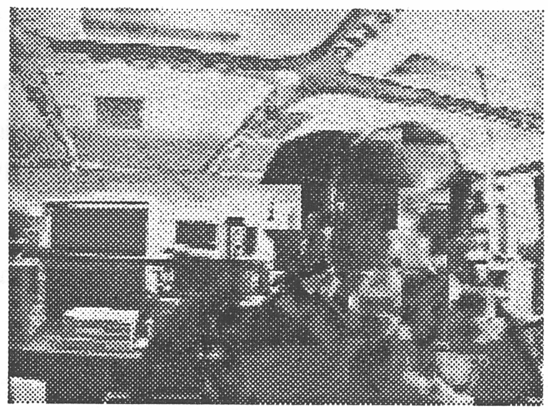

Figure 3.11. H.320 video frame with motion compensation distortion (ghosting).

Audio:

Errors in the audio data stream of the proprietary system generally resulted in audio blanking. No noticeable noise of any other type was observed other than very minor loss of fidelity at lower error rates ($<10^{-6}$).

In contrast, errors in the H.320 audio stream resulted in “static” which varied with the BER from an occasional “pop” or “crackle” beginning to be noticeable at error rates of 10^{-6} to a roar of white noise that overwhelmed the audio signal at high error rates ($>10^{-4}$).

The difference in the behavior appeared to be due to implementation choices, as both systems began to degrade at similar BER’s.

Whiteboard:

Errors in the whiteboard data stream caused loss of interactive responses. Display updates became irregular, and in severe cases ceased, with a simulated BER in excess of 10^{-5}.

General:

Service provided by the ISDN switch, the satellite signal processor, or FTS2000 failed when subjected to severe BER’s. Connections were consistently lost with a BER of 10^{-3}.

3.7 Analysis

The usability curves for proprietary and H.320 videoconferencing are very similar for both terrestrial and satellite configurations. For voice and video, both usability curves begin to degrade at a BER of 10^{-6}. However, the video degrades more quickly, becoming unusable at a BER of 10^{-4}. The voice remains usable at a BER of 10^{-2} or greater.

The behavior of the whiteboard, similar to that of the video, begins to degrade at a BER of 10^{-6} and becoming unusable at a BER of 10^{-4}.

The difference in the audio usability between the proprietary and the H.320 systems is due to the different implementation approaches used. The proprietary system blanks the audio signal when noise is detected, causing audio gaps but preventing loud blasts. The H.320 system does not blank errors; the signal simply gets increasingly noisy until it is unintelligible.

The initial difference in the usability of the whiteboard component between satellite and terrestrial configurations was attributed by the users to increases in the interactive response time. This additional delay caused a minor increase in annoyance when attempting to perform highly interactive tasks as tracking the other user’s cursor. However, the minimum interactive response time was 2 s. This is much greater than the increase in delay caused by satellite propagation (about 0.25 s).

Both proprietary and H.320 systems were extremely stable when subjected to severe simulated error bursts in either terrestrial or ACTS configurations. The videoconferencing systems rarely had fatal errors even when subjected to the most severe simulated errors (10^{-6} background BER, 10^{-1} burst BER for 10 s). In contrast, a burst sent into the terrestrial network with a BER of 10^{-3} for greater than 5 s consistently resulted in a lost connection.

3.8 Interpretations

1. BER, not delay, is the principal factor in videoconferencing usability. The data indicate that all components of videoconferencing began to degrade with a BER in excess of 10^{-6} and became essentially unusable with a BER of 10^{-3}. The usability scores for satellite versus terrestrial communications services were essentially identical, indicating that satellite delay was not a significant factor. This was partly because the delay added by a satellite was small compared to the processing delays of the applications themselves. Therefore, a user that is accustomed to the delays incurred with the terrestrial services is not disturbed much by the increased delay incurred by the use of a satellite. A maximum BER of 10^{-6} is recommended for reliable videoconferencing.

2. Bandwidth management of various data streams may cause synchronization problems in highly interactive activities. Voice and video take priority over data transfers in the Proshare® implementation and only BRI bandwidth is available for this application. Therefore, whiteboard activity tends to lag behind voice and video signals due to preemption of the whiteboard data. The voice and video signals must be synchronized to prevent very disturbing synchronization problems. The data are allowed to use any bandwidth that remains after voice and video requirements are satisfied. If major data transfers are required, the available data bandwidth is easily exceeded causing delayed updates in the data-based activities.

3. Additional delay due to the satellite is an important factor in highly interactive activities. Although the additional delay of a satellite link is only a fraction of the total transfer delay for this application, and therefore a minimal problem for most voice, video, and data uses, highly interactive uses (e.g., talking while using the cursor to point at objects with responses from the party at the other end) become increasingly difficult with increasing frequency of interaction.

4. Terrestrial communications links may fail when sustained, severe BER's are present. Data transmitted through the public terrestrial network with a BER in excess of 10^{-3} that persist for several seconds resulted in lost connections. However, the applications themselves were able to remain stable with higher error rates when the errors did not propagate through the network, but were generated locally. This behavior may be attributed to an administrative decision in the terrestrial network to terminate connections that have persistent errors.