12

AD A114120

DTIC
ELECTE
S MAY 5 1982 D
A

Prepared by
Bell-Northern Research Ltd.
Ottawa, Ontario

82 05 04 081

APRIL 1982

FINAL REPORT

A Virtual Space Video Conferencing Demonstration System

PREPARED FOR:

Defense Advanced Research Projects Agency

Contract No.: MDA903-81-C-0180
BNR Ref. No.: P379/TR6735

PREPARED BY:

Bell-Northenr Research Ltd.
P.O. Box 3511, Station C
Ottawa, Ontario
K1Y 4H7

SUMMARY

This report describes a system that was constructed to demonstrate and study the subjective effects of switching delay in a virtual space video conferencing system and presents some preliminary results of the subjective evaluations.

The concept of virtual space video teleconferencing provides geographically separated participants with a visual environment that very closely simulates a meeting room where all participants are physically present. At each site, TV camera/monitor pairs representing the remote participants are located in front of the physically present participant in approximately the same positions as these participants would occupy if they were seated around a meeting table in the same room. The local participant can at any time choose which monitor (participant) he wishes to watch and, by the head and eye positions as seen on the monitors, he can determine where the other participants are looking as well. A virtual space conference consisting of N single participant sites would require (N x (N-1)) TV camera/monitor pairs, and presenting a continuous video signal to each monitor would require as many dedicated video channels. By taking advantage of the fact that a participant can view only one monitor at a time, and utilizing a transmission scheme which dynamically allocates a single incoming video channel to each site to carry the signal for the monitor currently in view, the total video channel requirement can be reduced significantly. Using such a scheme, a non-trivial amount of time would be required for the system to detect the participant's change of viewing angle, effect the video switching at the signal sources, and ultimately deliver a live video signal to the new monitor.

An experimental system was constructed to demonstrate and study the switching transition period for virtual space teleconferencing over a dynamic channel allocation network. The system simulated a four site teleconference. A viewer room was equipped with three monitors providing either live or frozen images of three remote participants co-located in a nearby room. Only the monitor being watched by the viewer would receive a live camera signal. The remaining two monitors would present a frozen image. As the viewer's gaze switched to another monitor, a computer-controlled transition from frozen to live image would take place for that new monitor. The duration and visual attributes of the transition were programmable parameters and viewers were asked to express their subjective opinions concerning the video switching that took place as they participated in a mock meeting with the three remote participants.

Preliminary evaluation results indicate that a switching delay of one second is considered to be no handicap to the feeling of the virtual presence. Two seconds delay is considered acceptable although the delay is noticeable. Three seconds delay is considered unacceptable.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

The purpose of this component of the project was to construct a simulation facility to demonstrate the subjective effects of switching delay inherent in a video conferencing system utilizing dynamic allocation of bandwidth for the video channels. Through limited informal use of this demonstration system, a preliminary assessment of the tolerable delay time and signal switching characteristics was obtained.

2. BACKGROUND AND MOTIVATION

It has been postulated that the most effective form of video teleconferencing is one which preserves the interpersonal visual characteristics of the meeting room environment where all participants are physically present and share the same physical space. This shared ´physical space´ meeting environment is illustrated in figure 1 (for a four party meeting). In such an environment, each participant can at all times view the participant of his choice and can also determine where this participant is looking as well. When all participants are physically isolated, this physical space environment can be simulated reasonably well with a video teleconferencing setup as illustrated in figure 2 (again for a four party meeting). In this shared ´virtual space´ arrangement, only one participant is physically present in each room, the remaining meeting participants are virtually present as a TV camera / TV monitor pair. The positions of the camera/monitor pairs in the four locations are arranged to provide a consistent virtual seating arrangement pattern as perceived by all participants. When one

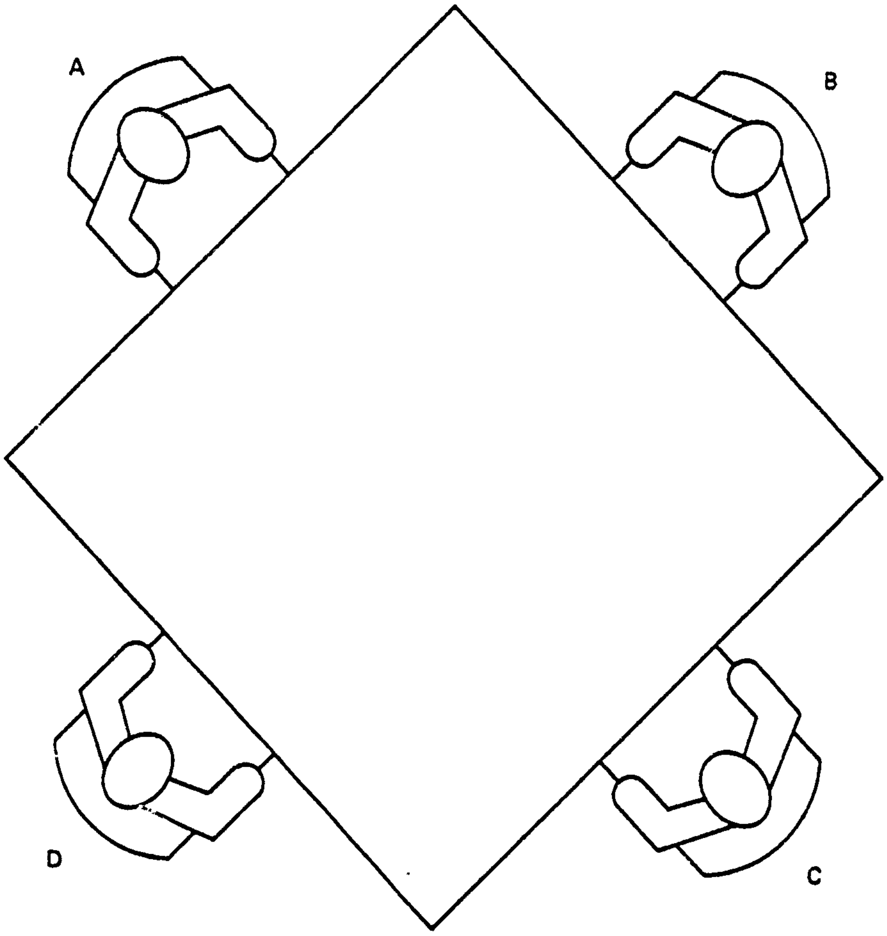


Figure 1 Shared 'Physical' Space Meeting

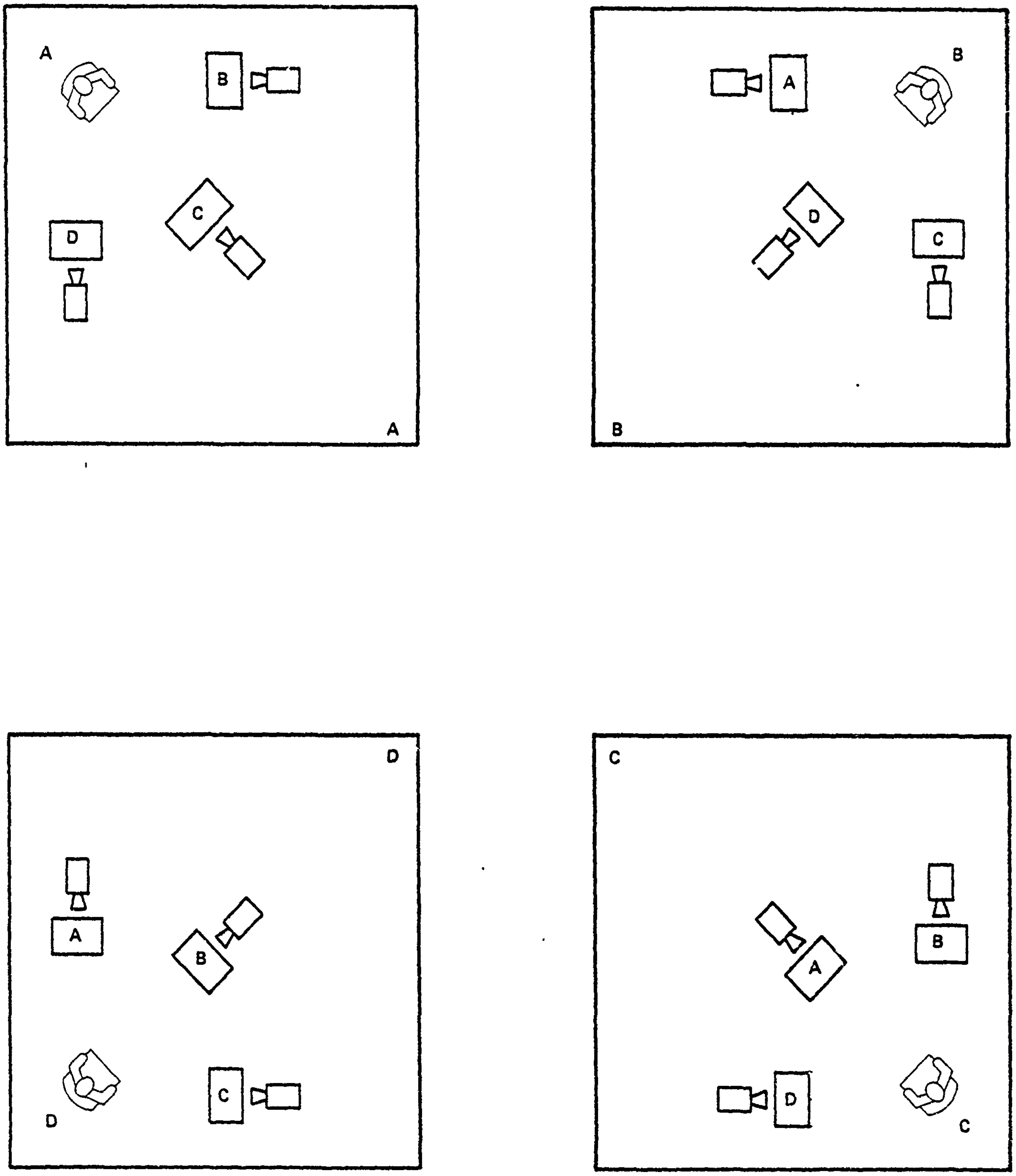


Figure 2 Shared "Virtual Space" Meeting

participant views another through the appropriate monitor at his location, he is presented with a viewing angle consistent with his virtual position in the remote room. Assuming one participant per physical location, for a meeting involving N conferees, this arrangement requires (N x (N-1)) cameras and monitors.

In order to provide a continuous video signal to every monitor in this system, (N x (N-1)) video channels would be required in what is called a 'fully connected' network (figure 3). By taking advantage of the fact that any participant can really only view one monitor at a time, a significant saving in the number of video channels can be realized through the implementation of a dynamic (channel) allocation network (figure 4). In this scheme, only one video signal is received at each site at any given time, the source of that signal being dependent upon which monitor is currently being viewed. As the participant changes his view from one monitor to another, the network must sense this change and route the appropriate new signal to his location. To be effective in terms of channel saving, this switching must take place at the signal source. This introduces the requirement for signalling channels between the sites, however this level of signalling would have an extremely low bandwidth requirement (100 bps or less) compared to video channels. The exact savings in channels is dependent upon the mode of transmission and the topology of the network. Perhaps the best case to illustrate the savings is through an implementation over a satellite network. In this case, for a four site conference, only four transponder channels would be required, one dedicated to feeding a single incoming video signal to each site. All sites would require the capability to transmit onto any or all of the three channels servicing the other sites. In the general case, for a

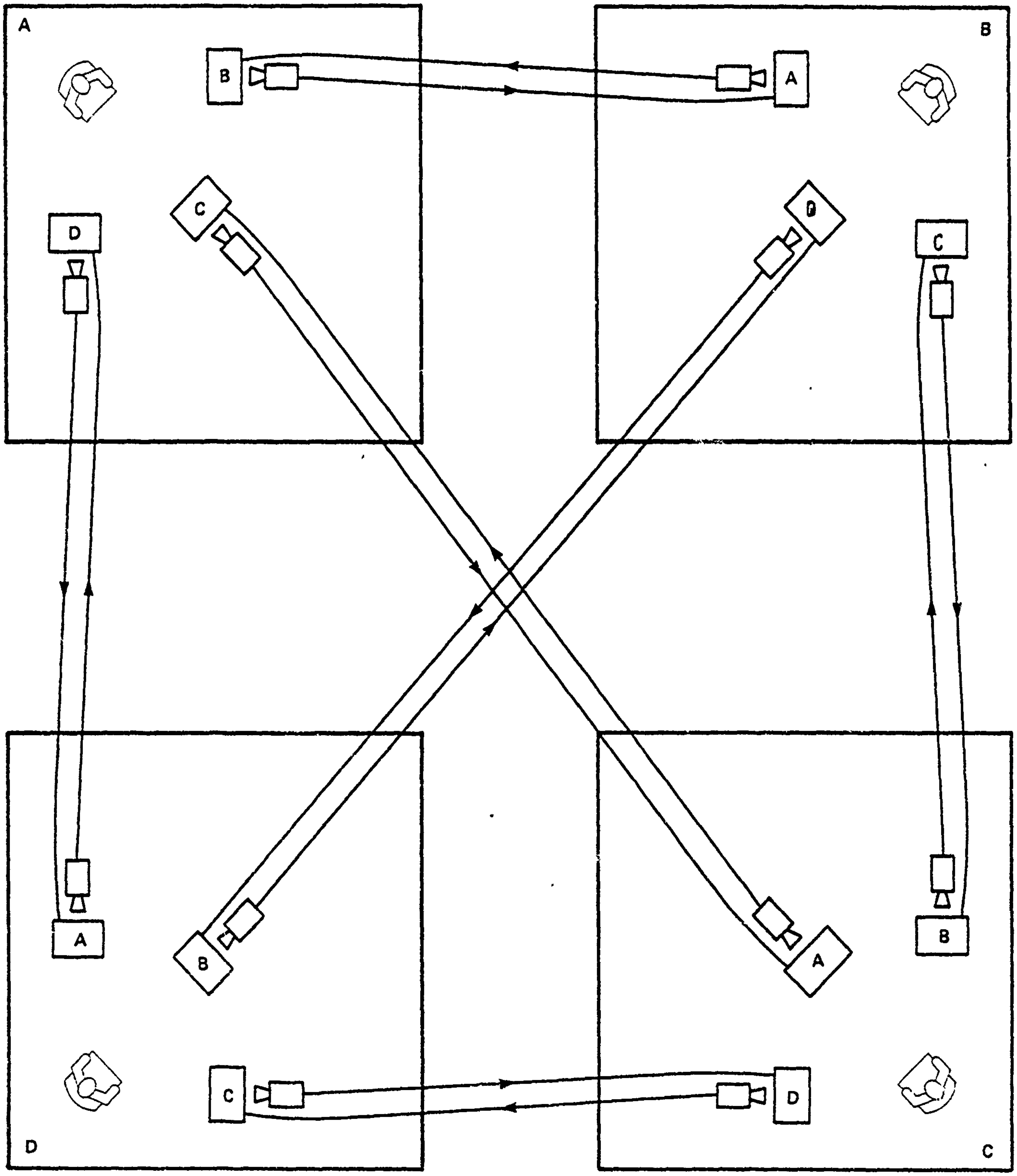


Figure 3 Fully Connected Network Maximum Channel Requirement

satellite-based dynamical allocation network connecting N single participant sites, the number of channels required would be N and the number of channels saved over a fully connected network would be (N x (N-2)).

Using the layout in figure 4 as an example, if participant A were speaking and looking at monitor (participant) B, and all other participants (B,C, and D) were watching monitor (participant) A at their respective sites, then three video signals from the three cameras at site A would be transmitted on video channels B,C, and D (to feed monitor A at sites B,C, and D respectively); and one video signal from camera A at site B would be transmitted on channel A (to feed monitor B at site A). If participant A changed his view to monitor (participant) C, then a control signal would be sent to site B to stop transmission from camera A on channel A, followed by a control signal to site C to start transmission from camera A onto channel A. This scheme of dynamic channel allocation will involve a non-trivial switching delay between times of steady state viewing. The overall delay consists of the following components - 1) the time to sense the participants' change of viewing position, 2) the time to transmit the STOP and START control signals to the remote sites (B and C), and 3) the time for the new signal (from camera A at site C) to reach its destination (monitor C at site A). Depending upon the responsiveness of the system to detecting viewer changes, the transmission path lengths for the control and video channels, and the type (if any) of video coding used, this delay could range anywhere from one half second to several seconds.

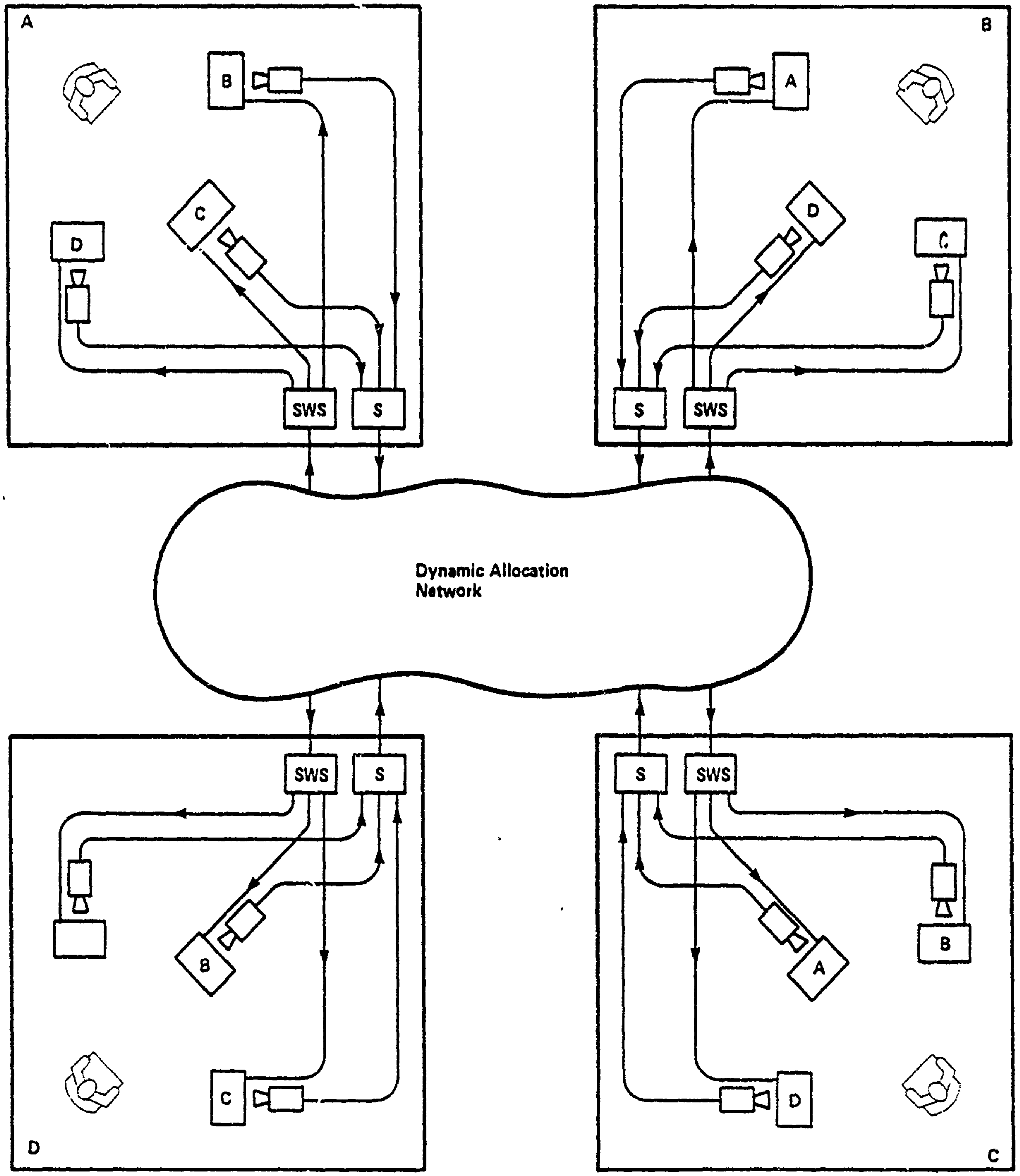
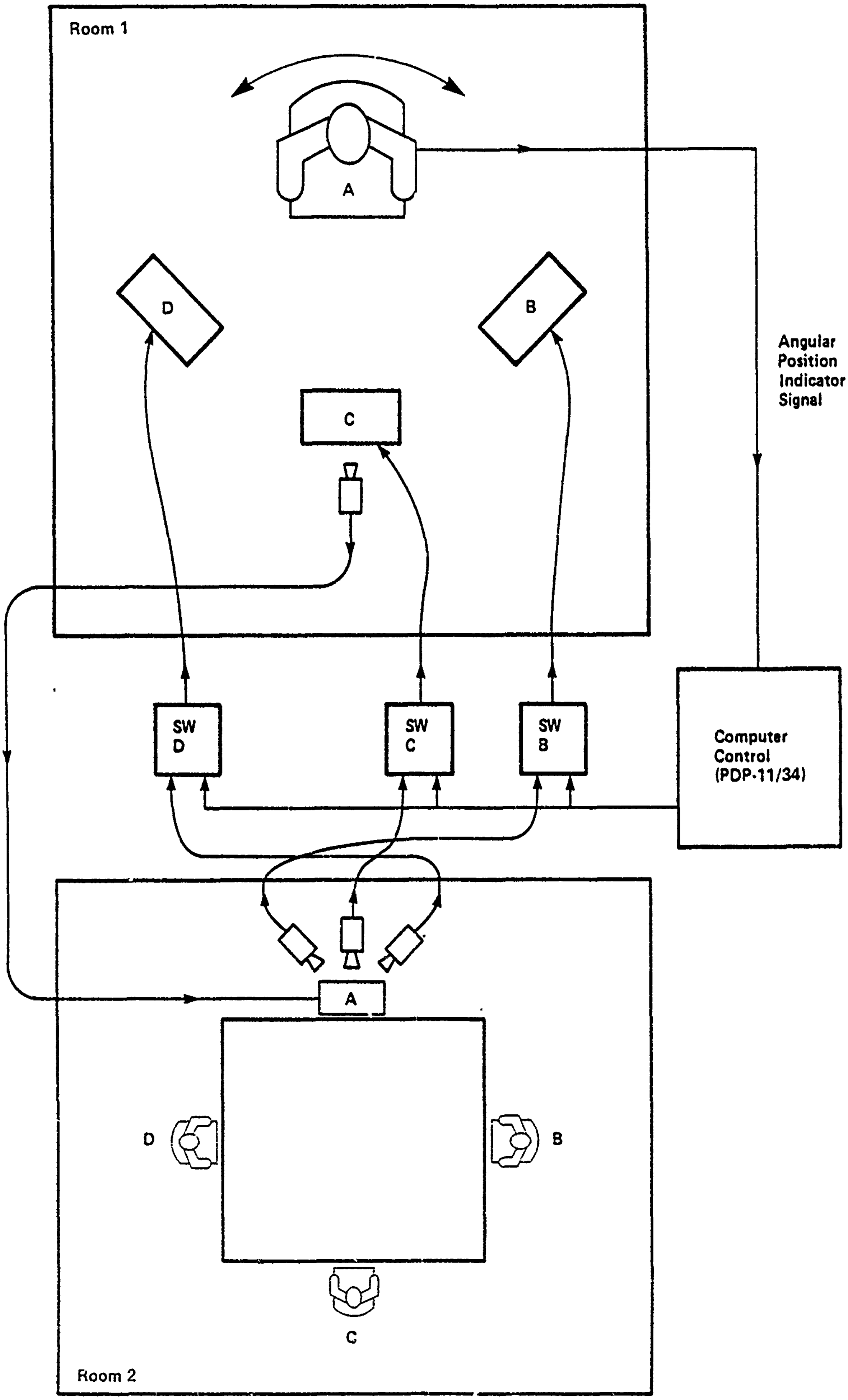


Figure 4 Dynamic Allocation Network, Reduced Channel Requirement
Remote Switching Control Required

In order to determine user acceptance of the dynamic channel allocation scheme, two issues relating to this switching delay must be explored. First, what would be the maximum acceptable delay, and secondly, what form of switching transition would minimize the negative subjective effects of the switch. With respect to the latter issue it could be assumed that viewers would prefer to see at least a frozen image of the participant on the monitor before the new live signal arrived. This could be a 'posed' photograph captured at the beginning of the meeting, or the last live frame transmitted when the participant was last viewing that particular monitor. In either case, it would require a local single frame store for each monitor at each site. Subjective preferences for the type of frozen imagery displayed on inactive monitors will be discussed later. The demonstration system described in the following section was constructed specifically to explore these two video switching issues.

3. DEMONSTRATION SYSTEM DESCRIPTION

The system constructed to demonstrate the effects of video switching in a dynamic channel allocation scheme is illustrated in figure 5. The viewing subject (A) sits in a high back swivel chair in Room 1 with three TV monitors representing the three 'virtually present' participants (B,C, and D). These three participants are physically located in Room 2 seated at three sides of a square table. A TV monitor which provides the virtual presence of the Room 1 participant is located on the fourth side of the table. Directly behind this monitor are three TV cameras angled to capture the images of the three participants in the room from approximately the same vantage point that the fourth


Room 1
A
D
B
C
Angular
Position
Indicator
Signal
SW
D
SW
C
SW
B
Computer
Control
(PDP-11/34)
A
D
B
C
Room 2

participant would have if he were physically present. The signals from these cameras pass through three video switching units (discussed in detail shortly) en route to the monitors in Room 1. A PDP 11/34 computer system controls the state of the video switches based upon the angular position of the viewer's chair in Room 1. For this reason, subjects were requested to rotate the chair to look from one monitor to another as opposed to turning their head or moving their eyes.

A schematic representation of the video switching unit is shown in figure 6. It consists of five functional components - a single frame digital memory, a composite black video signal generator, a video gating circuit, an analog video mixing circuit, and an analog fading circuit. The switching unit has four inputs - a composite NTSC colour video signal from one of the TV cameras in Room 2, a binary (0 or 1) gating control signal from the computer, an eight bit (0 - 255) fading control signal from the computer, and an eight bit (0 - 255) mixing control signal from the computer. The switch has a single output signal - a composite NTSC colour signal feeding one of the monitors in Room 1.

The frame store contains a frozen image of the participant upon whom the camera is focussed. When the gating control signal is true (1), the contents of the frame store are continually updated by the gating circuit which digitizes and stores new frames from the camera signal. When the gating control signal becomes false (0), the last frame digitized remains in the frame store.

The fading circuit is used to provide a controlled mix between the static frame image signal from the frame store and a composite black signal. The value of the fading control signal determines the mix. A value of 0 presents the full frame store signal at the output of the

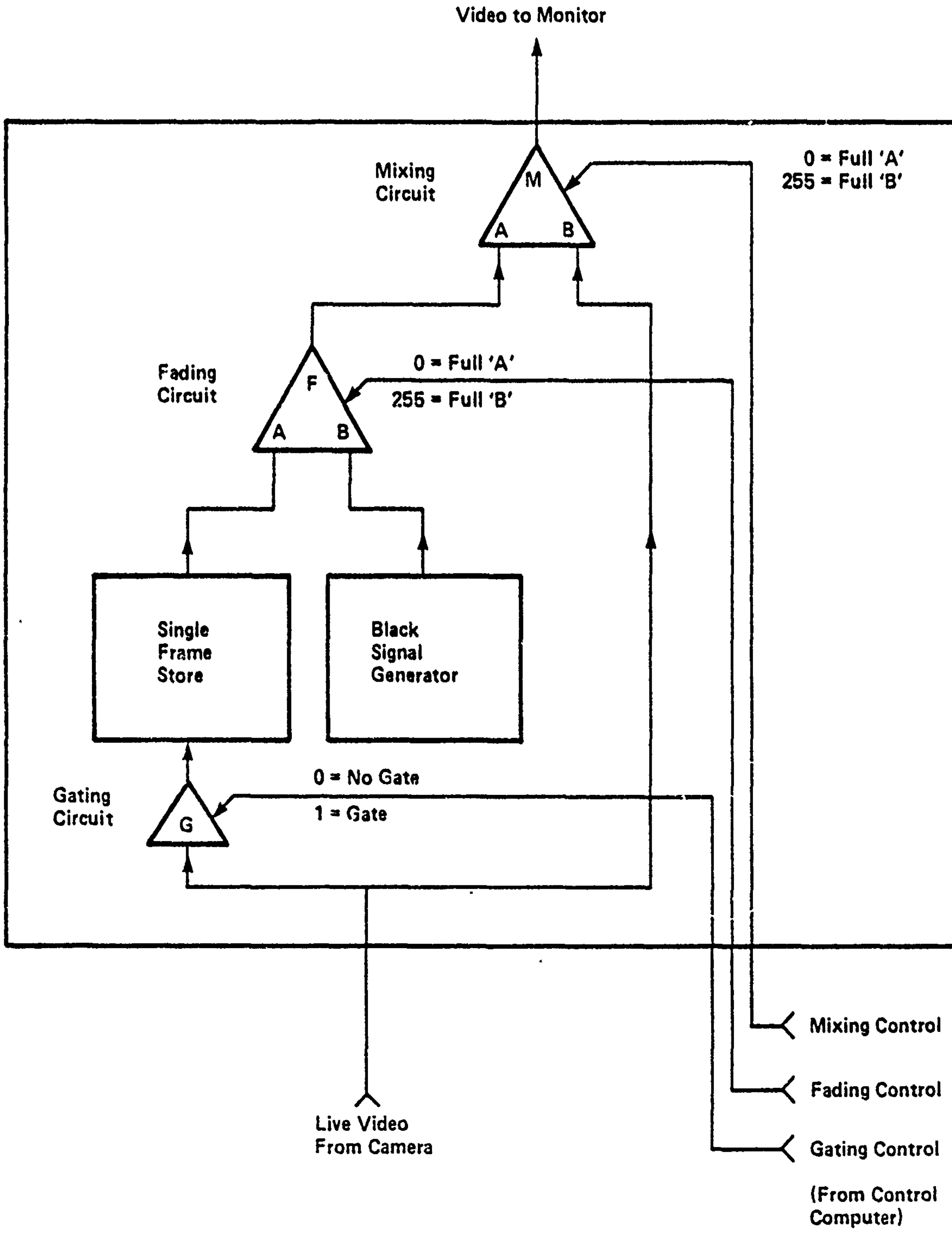


Figure 6 The Video Switch

fading circuit. A value of 255 presents the full black signal at the output. A gradual fade to black is achieved by successively incrementing the fading control signal from 0 to 255 over a specified period of time.

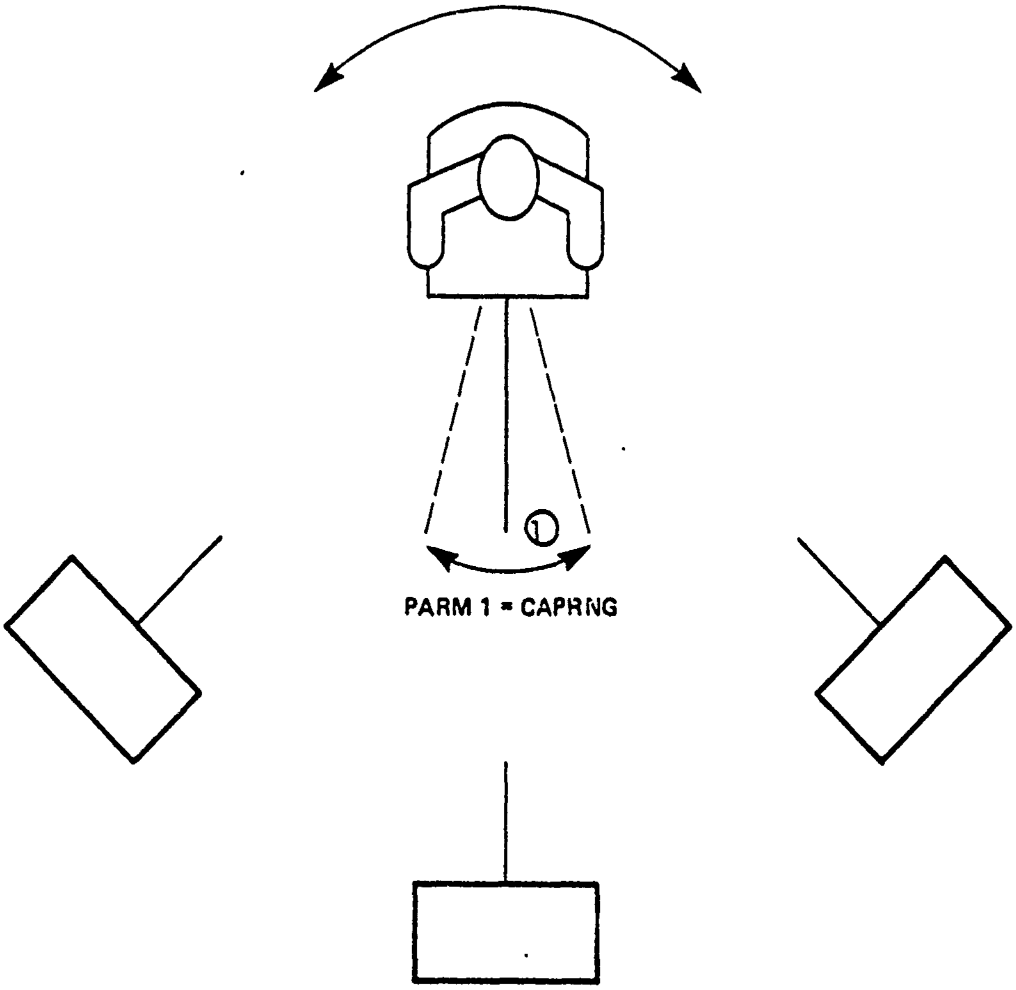
The mixing control circuit is used to provide a controlled mix between the output of the fading circuit and the live camera signal. A value of 0 presents the full output of the fading circuit to the TV monitor. A value of 255 presents the full live signal to the monitor. A gradual transition from frozen to live signal on the monitor is achieved by successively incrementing the mixing control signal from 0 to 255 over a specified period of time.

The demonstration system is programmed for the following operation. When the participant in Room 1 is not viewing the monitor controlled by a particular switching unit, the gating signal to that unit is false (0), the fading control signal is set to 0, and the mixing control signal is also set to 0. The monitor is presented with the stored signal from the frame store. This is called the INACTIVE state for the switching unit. Once the participant has been viewing a monitor for a period of time long enough for the switching transition to be complete, the gating signal is true (1), the fading control signal is set to 255, and the mixing control signal is set to 255. In this state, the frame store is continuously updated with new frames from the camera signal and the monitor is presented with the full live camera signal. This is called the ACTIVE state for the switching unit. When the participant turns away from the monitor, the switching unit reverts immediately to the INACTIVE state. When the system senses that the participant has turned towards an inactive monitor, the switching unit for that monitor enters the TRANSITIONAL state where the control signals to the switch are gradually changed from their INACTIVE values to their ACTIVE values by the computer according to several specified transition parameters. It is this transitional period that is the focus of the demonstration system. The control parameters which define its characteristics are described in the next section.

4. DEMONSTRATION SYSTEM PARAMETERS

The control program for the demonstration system has been designed to permit exhaustive experimentation with the video switching characteristics. At the beginning of a test session, the angular position of the three monitors with respect to the viewer's chair is recorded in the program for reference. The program then enters the test loop and a console operator is prompted to enter the six parameters which determine the visual characteristics of the switching transition period. These parameters are illustrated in figures 7 and 8 and described below. The program loop samples the chair postion 60 times a second and makes appropriate changes to the states of the video switching units and their software control variables based upon the chair position and the six transition parameters.

CAPRNG	Capture range of the viewer's chair. This is specified in degrees centered on a line perpendicular to the chair back. As the chair swivels, when the angular position of a monitor falls within the capture range of the chair, that monitor is considered ELIGIBLE FOR CAPTURE.
CAPDEL	Capture delay. This is specified in 1/160th second units and



CAPRNG = Capture Range of the Viewer's Chair
– Specified in Degrees Centred on a Line Perpendicular to the Chair Back

Figure 7 Demonstration System Parameters

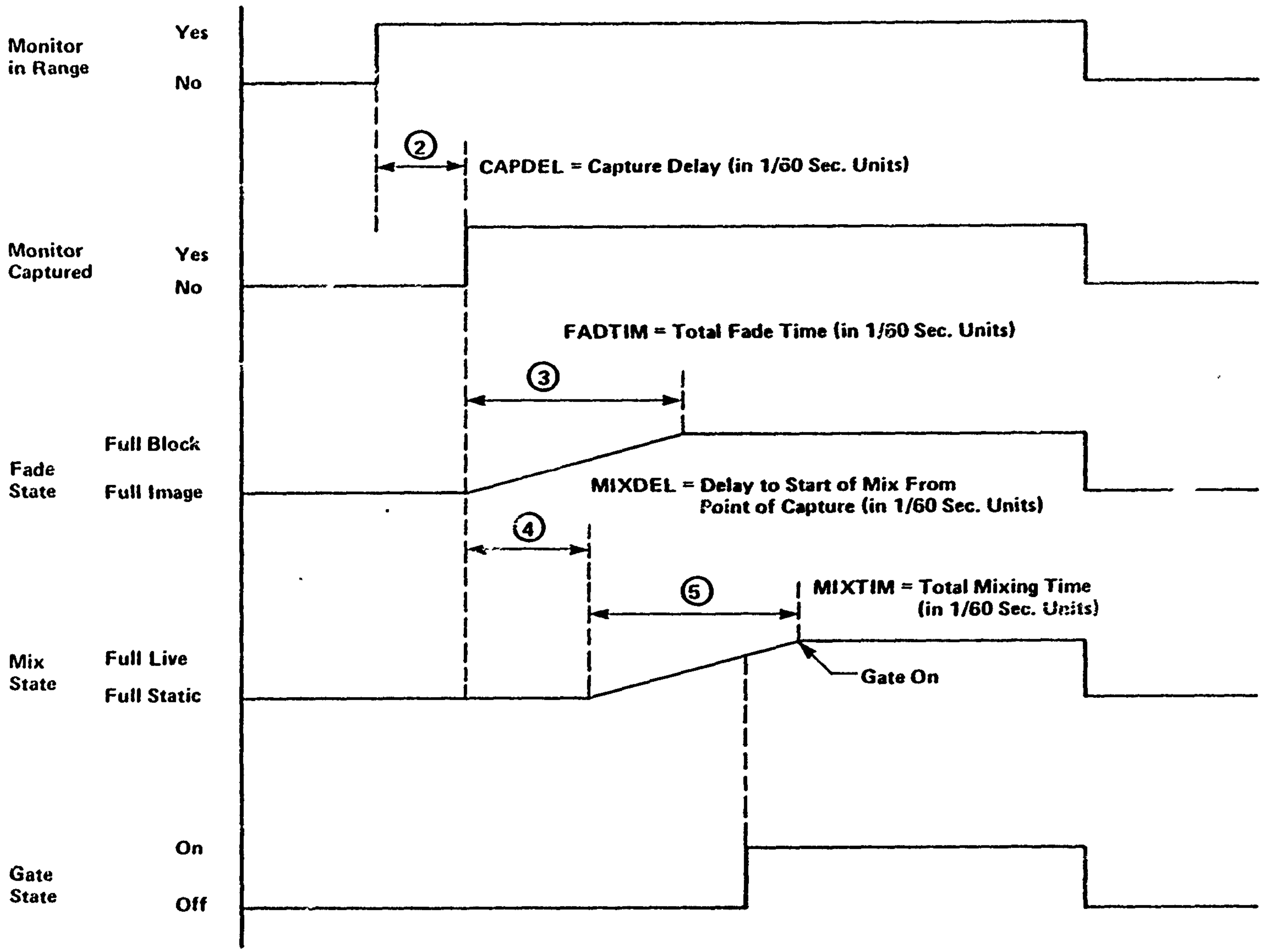


Figure 8 Demonstration System Parameters (Cont'd)

	determines how long a monitor must wait in the ELIGIBLE FOR CAPTURE state before being considered CAPTURED. Only after the monitor is captured does the switching transition period start. If a monitor is in the ELIGIBLE FOR CAPTURE countdown and the chair swivels so that it becomes out of range, the monitor looses its ELIGIBLE FOR CAPTURE status and its progress in the countdown. This parameter was included to prevent transient capture of the middle monitor as the participant shifted his view from the left monitor to the right monitor, passing momentarily through the capture region of the middle monitor.
FADTIM	Fadng time. This is specified in 1/160th second units and determines the transition time from full static image to full black in the fading circuit. The fading transition begins as soon as the monitor is CAPTURED. This is the beginning of the TRANSITIONAL period.
MIXDEL	Mixing delay. This is specified in 1/160th second units and determines the time delay to starting the static-to-live mixing transition in the mixing circuit. If 0 is specified, the fading and mixing transition start concurrently.
MIXTIM	Mixing time. This is specified in 1/160th second units and determines the transition time from full fading signal to full live signal in the mixing circuit. This mixing transition begins as soon as the mixing delay countdown has expired. When the mixing transition has completed, the TRANSITIONAL period is over and the monitor and video switching unit are in the ACTIVE steady state.
GATENB	Gating circuit enable. This is specified as 1 (true) or 0 (false) and is used to determine whether or not the gating

circuit is enable once the monitor enters the ACTIVE state. If this is specified as false, the contents of the frame store will not be updated and a ´posed´ image of the remote participant will appear in that monitor whenever the monitor returns to the INACTIVE state.

5. DEMONSTRATION RESULTS

The demonstration system was operational for about two weeks and was used for a limited number of four party meetings culminating with a several hour session on Friday, January 14, 1982. During this last session, Dr. Clinton Kelly from DARPA participated from the viewer´s chair in Room 1. Several BNR and INRS personnel took turns participating in Room 2.

For most trials, adjacent monitors in Room 1 were placed approximately 45 degrees apart with respect to the viewer´s chair and the capture range used was 40 degrees which means that adjacent monitors could not be simultaneously captured, but the ´dead zone´ between monitors (all monitors frozen) was almost never encountered and caused no irritation to the viewer. Overall switching delay (CAPDEL + MIXDEL + MIXTIM) was set for 1 second, 2 seconds, and 3 seconds. Switching transition characteristics were varied in a similar manner for all three delay settings. The following comments summarize the personal opinions of those participants, including Dr. Kelly, who spent time in the viewer´s chair in Room 1.

An overall switching delay of 1 second was considered to be no handicap at all to the feeling of the virtual presence of the three participants in Room 2. Two seconds was considered acceptable although the delay was noticeable. Three seconds was considered not acceptable. However three second delay was only tried after a significant period of usage at lower delay times and this may have influenced the viewer´s judgement.

Regardless of the delay time, viewer´s preferred to have the frozen image on the monitor start fading as soon as they started to look at the monitor (that occurs when CAPDEL is very small, i.e. no more than 1/12th second before the fading of the frozen image started). Also, viewers preferred a direct transition from frozen to live signal without the monitor going dark in between. This was accomplished by choosing a fading time (FADTIM) at least twice the value of the mixing time (MIXTIM) with a very small mixing delay (MIXDEL) period. For example, a two second overall switching time consisting of:

CAPDEL = 5 (1/12 sec.)
FADTIM = 240 (4 sec.)
MIXDEL = 1 (1/60th sec.)
MIXTIM = 120 (2 sec.)

seemed to produce the most favourable switching transition in the two second category.

As a final note, some viewers preferred a ´posed´ frozen image to be presented on the non-active monitors while others preferred the frozen image to be grabbed from the live signal the last time that monitor was viewed. Participants, whose frozen image would be appearing on the viewers monitor seemed to prefer the posed image option for fear of being captured with an 'unflattering' expression, etc.

6. CONCLUSIONS

The limited experience with the demonstration system suggests that the sense of remote participant presence in a virtual space video teleconferencing system is preserved when sites are interconnected using a dynamic channel allocation network as long as channel switching delays are no greater than two to three seconds in duration. By maintaining a local frozen image of participants and performing a smooth transition between frozen and live signals, the negative subjective effects of the switching delay can be minimized.

Although the conclusions reached from this initial experience are favorable, it is recommended that more exhaustive subjective tests be performed to verify these preliminary conclusions and to determine more precisely the tolerable delay period and the most optimum switching characteristics.