Multimedia Systems 7: 214–221 (1999)

Multimedia Systems
© Springer-Verlag 1999

# TELEPORT – Towards immersive copresence

**Simon J. Gibbs, Constantin Arapis, Christian J. Breiteneder**

Visualization and Media Systems Design, Institute for Media Communication, GMD – German National Research Center for Information Technology, D-53754 Sankt Augustin, Germany; e-mail: {Simon.Gibbs, Costas.Arapis, Christian.Breiteneder}@gmd.de

**Abstract.** TELEPORT is an experimental teleconferencing system with the goal of enabling small groups of people, although geographically separated, to meet as if face to face. The innovative features of the system include the use of full-wall display surfaces, "merging" of real and virtual environments, viewer tracking, and real-time compositing of live video with synthetic backgrounds.

**Key words:** Copresence – Telepresence – Media spaces – 3D teleconferencing – Virtual reality

## 1 Introduction

Personal computers equipped with microphone, speakers, camera, and perhaps additional video monitors, are now widely used for desktop video conferencing. Conference participants appear in video windows, or on adjoining monitors, and may access shared applications shown simultaneously on each participant's screen. Several desktop video conferencing systems have been described in the literature; for example, see the work of Ahuja [1], Bly et al. [3], Gaver et al. [9], Mantei et al. [18] and Root [21], in addition several commercial products are now available.

This paper presents TELEPORT, an experimental teleconferencing system, where the goal is to provide greater realism than is found with conventional desktop video conferencing. The essential idea behind TELEPORT is what is termed *copresence* [4], the illusion that remote conference participants, although actually distant, are present in the local participant's physical space.

The system is based around special rooms, called *display rooms*, where one wall is a "view port" into a *virtual extension* (Plate I and Plate II). The geometry, surface characteristics, and lighting of the virtual extension are designed to closely match the real room to which it is attached. When a teleconferencing connection is established, video imagery of the remote participant (or participants) is composited with a rendered view of the virtual extension (see Plate III). The viewing position of the local participant is tracked, allowing imagery appearing on the wall display to be rendered from the participant's perspective. The combination of viewer tracking, a wall-sized display, and real-time rendering and compositing, give the illusion of the virtual extension being attached to the real room. The result is a natural and immersive teleconferencing environment, where real and virtual environments are merged without the need for head-mounted displays or other encumbering devices.

*Correspondence to*: S.J. Gibbs

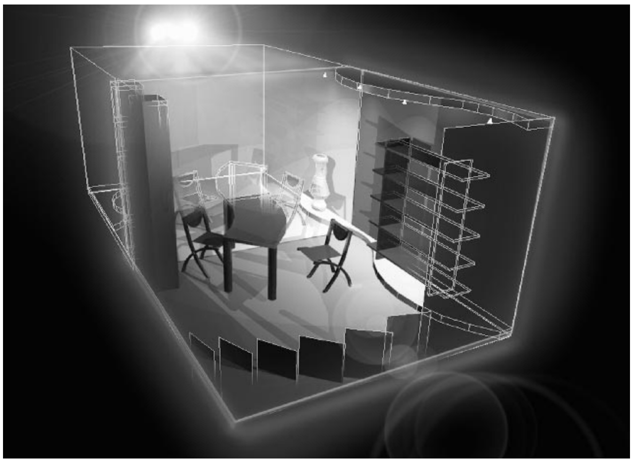

**Plate I.** TELEPORT Display Room (as designed)

From the above description, we view the contribution of TELEPORT to be that it integrates several new technologies and capabilities, and makes a step towards achieving true copresence. While each of the features used in TELEPORT has been studied previously, it is their integration which makes TELEPORT unique. Among the important elements of TELEPORT are

- a semi-real, semi-virtual meeting space;
- a wall-sized display supporting both mono and stereoscopic viewing;
- viewer tracking and real-time rendering;
- an image segmentation system capable of extracting foreground objects (participants) from a fixed background while operating at near video rates (10-25 fps);

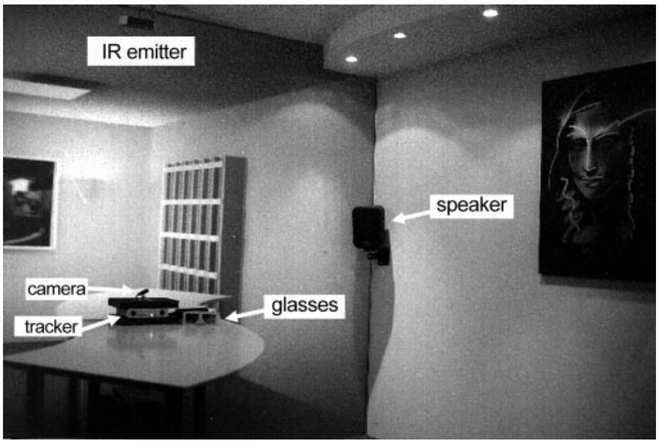


**Plate II.** TELEPORT Display Room (as constructed). Note: rendering view-point has been set to camera position

– compositing of video-textured surfaces within 3D geometric models.

The following sections motivate the role of systems such as TELEPORT. We describe related work, discuss our design objectives and the key technical components needed for TELEPORT's construction. We also report on initial usage trials and possible future extensions.

## 2 Motivation

With the advent of low-priced video codecs and greater network accessibility, desktop video conferencing is becoming available to large numbers of users. But, while desktop video conferencing has certainly been shown to be useful for a variety of tasks – design and planning activities [20, 25], medical cooperation [13] – and has many advantages when compared to earlier forms of video conferencing involving special meeting rooms, it is still recognized that there are many situations where desktop video conferencing is not appropriate [19, 23].

The limitations of desktop video conferencing are most clearly revealed when the nuances of body language and eye contact become important. Ideally, in such situations, we would like to provide conference participants with the sensation of being in the same room at the same time and meeting as if face to face. This is what is called "copresence."

Current desktop video conferencing systems provide only a limited degree of copresence. The reason is that they suffer at least three deficiencies: loss of shared physical context, differences of display size and body size, and difficulties in establishing eye contact and gaze awareness.

### *2.1 Shared physical context*

When a group of people meet in the same physical space they are implicitly aware of a consensus (or at least the potential for consensus) regarding what objects are present (furniture, paintings on the wall, etc.), the relative distances between the various participants, who is to the right of or in front of whom, and a score of other physical observations. Desktop video conferencing gives only a very limited form of such shared physical context: one may be aware of who is present, but there is no mutual physical reference frame so it becomes meaningless to speak of the relative placement of participants. In other words, rather than sharing a physical context, participants remain each in their own physical space.

### *2.2 Display size and body size*

Given typical monitor sizes and viewing distances, desktop systems simply do not have large enough displays to present a full-body life-sized, image of a conferencing participant. If the image is life-sized then at most the head and shoulders will be visible; if the image is full-body, then it appears doll-sized. In either case, the sense of copresence is minimal. Furthermore, desktop video conferencing sessions frequently opt for head shots, thus there is the additional disadvantage of a nearly complete failure to communicate body posture, hand gestures and many other physical mannerisms.

### *2.3 Eye contact and gaze awareness*

Admittedly, it is possible, through control over camera and subject placement, to provide for eye contact: i.e., when person A looks at the image of person B appearing in a window on A's screen, then A looks directly into a camera and so appears to make eye contact with B. Techniques allowing eye contact include mounting the camera behind a partially transparent mirror in front of the window [21], mounting over miniature monitors [23], and projecting the window onto a partially transparent screen behind which the camera is mounted [19]. However, in each case, eye contact is essentially a fortuitous consequence of the geometry of camera and subject placement. If participants move off center, it is possible to obtain both "false misses" and "false hits." For example, A may move to the side and although still looking at B, to B it appears as if A is looking elsewhere. Furthermore, these approaches, based on a correspondence between the local camera viewing direction and that of the remote participant, become unwieldy as the number of participants increase, since they require a separate local camera for each remote participant.

Now consider the more general issue of gaze awareness, i.e., the ability of one participant to determine what, or who, is being gazed at by another. Other than possible eye contact, desktop video conferencing provides only very coarse gaze awareness. For instance, consider a conferencing session with three participants: A, B and C. Suppose A has two windows, one showing B the other C. Even if the positioning of windows used by all participants is constrained, it will be difficult for A to determine, using purely visual cues, whether B is gazing at A, at C, at some object being held or pointed to by A or C, or in some other direction entirely.

We want to emphasize that the issues listed above are not always of importance to the effectiveness of teleconferencing. However, there are certainly situations where users may find it preferable to have a more natural and immersive interface than is provided by desktop video windows.

**Plate III.** A TELEPORT Session (*left*: virtual extension with raiosity lighting, *right*: virtual extension rendered in wireframe). Note: redenreing viewpont has been set to camera position rather than local participant position

### *2.4 Design goals*

The primary design goal of TELEPORT is to build a conferencing system that approximates, as closely as possible using current technology, the experience of meeting face to face. The unique contributions of TELEPORT are that it mimics, using 3D modelling and rendering, a shared physical context and that it provides life-sized display of remote participants placed within a virtual space. Eye contact and gaze awareness are still dependent on camera and viewer placement (however, since the display of participants is now part of a 3D rendering pipeline, we hope that by using new techniques such as view morphing [22], it may be possible to correct for the difference between local viewing perspective and remote camera perspective).

## 3 Related work

Work closely related to TELEPORT includes the wall-sized projection systems used in both CSCW [5, 11] and virtual-reality applications [5, 16], and prototypes in the field of video conferencing in general.

Wall-sized screens forming a cube onto which stereoscopic images can be projected are employed in the CAVE immersive virtual reality system [6]. Although TELEPORT uses a similar display technique, it differs in that the projection surface is part of a real room – this allows the rendered scene to be perceived as an extension of the real room.

The TELEPORT notion of a virtual extension of reality is similar to augmented reality [8]. However, rather than overlaying the real world with virtual objects, TELEPORT creates an adjoining virtual space.

At Bellcore, a system has been developed based on a window metaphor to connect two separated and remote rooms. The VideoWindow [9] system uses a very high aspect ratio (8:3) video display, allowing for the impression of talking to and seeing people in an adjacent room through a window. Like VideoWindow, TELEPORT is built around a large display surface; where it differs is in the use of video processing and real-time rendering to place the remote participants in a virtual extension.

Another related project is Virtual Window [10], where head-tracking is used to control a remote camera. In TELEPORT, tracking is used to control a rendering viewpoint rather than the position and orientation of a remote camera.

Panorama [27], a project within the European ACTS program, aims to establish 3D copresence amongst conference participants. Auto-stereoscopic displays will be developed that spatially separate left- and right-eye-view images, in order to avoid wearing glasses. In addition, the display system allows a user to move while observing 3D imagery. The goals of this work are similar to ours; however, Panorama focuses on developing new stereoscopic-display and image-processing hardware, while we make use of “off-the-shelf” components.

The GreenSpace project [17], funded by the Human Interface Technology Lab and the Fujitsu Research Institute, attempts establishing a “virtual common” for remote collaboration by the use of visual, aural and tactile cues. Currently, faces of participants are scanned and corresponding data sent to connected sites. The heads of participants are positioned in a virtual environment and updated according to head (for viewing) and hand (interaction) movements. TELEPORT differs in that live full-body video is placed within the virtual environment.

The goals of the Japanese project MAJIC (Multi-Attendant Joint Interface for Collaboration) [19] are the most similar to those of TELEPORT: multi-party meetings, eye contact between participants, gaze awareness of one participant to another, and life-sized images of participants. MAJIC projects the video imagery of remote participants onto a large curved screen positioned in front of the local participant; it is a good example of how eye contact and gaze awareness can be supported by the careful placement of participants, cameras and projectors. MAJIC’s limitations include difficulties in expanding to large groups (the total number of cameras and projectors increases with the square of the number of participants) and constraints on participant positioning. TELEPORT takes a different approach: rather than assigning distinct regions of the display surface to different participants, the remote signals are combined into a single composite image showing a shared virtual environment. The number of cameras and number of projectors now depend linearly on the number of participants (rather than quadratically), participants are free to move (although eye contact is lost with the current system), and can be positioned in the virtual environment under software control. Furthermore, the display surface is disguised as part of the local environment, thus increasing the sense of immersion.

PPS, the Bellcore Personal Presence System [4], bears some similarity to TELEPORT. Both systems use compositing techniques to combine the imagery of multiple participants. The main difference is that TELEPORT has an immersive interface, and so requires real-time 3D rendering, head tracking and stereo display. PPS does not address these issues.

## 4 Technical approach

In developing TELEPORT our overall technical goal was realism, i.e., we wanted the virtual extension of the display room, and the representations of remote participants within it, to appear as if they actually occupied physical space. This we identified with two basic technical requirements: to provide both an immersive interface and an "invisible interface."

*Immersive interface.* The section on motivation emphasized the importance of a shared physical context. However, since participants are not actually at the same place, the only way in which a shared physical context can be provided is via an immersive interface. Since immersive interfaces allow virtual objects (or representations of real objects) to be placed within the physical space perceived by each participant, it is possible to construct virtual environments which are consistent among the participants.

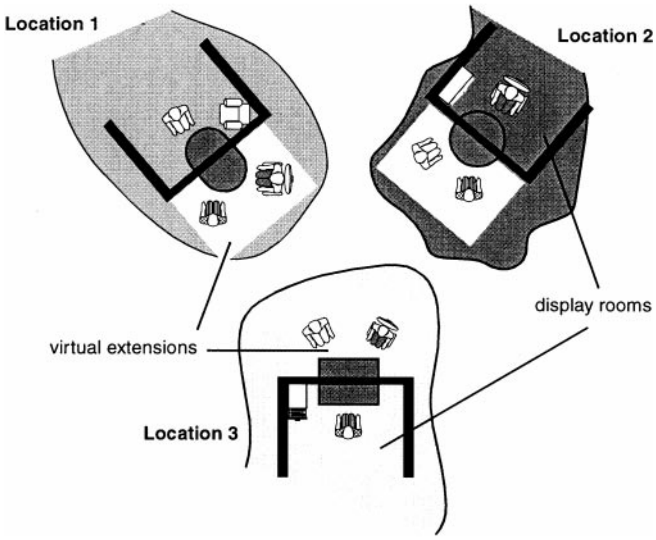


**Fig. 1.** A TELEPORT session between three locations. Each participant is located in a display room and views the remaining participants in its virtual extension

*Invisible interface.* One form of immersive conferencing requires donning head-mounted displays and entering a virtual environment. Currently, there are at least two drawbacks with this approach: first, the representations of participants are not life-like but some fairly simple visual approximation; second, it is hard to bring real-world objects, such as a cup of coffee or the papers on one's desk, into the virtual world. Rather than cut the user off from the real world, we seek an "invisible interface" – one which is disguised as and blends in with the real world. TELEPORT sessions take place in what resembles a typical meeting or office room, with normal lighting conditions and furniture. During a TELEPORT session, participants should be able to use objects located in the room in their customary manner (e.g., write on the desk, drink the coffee). Furthermore, we would like to avoid the use of encumbering equipment such as head-mounted displays or tracking devices with wire attachments.

### *4.1 General architecture*

A TELEPORT session involves a group of connected TELEPORT display rooms (see Fig. 1). Each room has a wall-sized display, video camera, microphone, speakers and network connection capable of audio/video transfer. In addition, each room is equipped with tracking and processing systems, the function of which is explained below.

### *4.2 Display room*

A TELEPORT display room consists of a full-wall display surface built into an office-sized room. The current system uses a $3 \times 2.25$ m rear-projected video wall attached to a $3 \times 3$ m room (see Plate III). The video wall is driven by a pair of high-luminosity video projectors; two projectors give more light (the video wall appears reasonably bright even with room lighting on) and allow for stereoscopic viewing with passive glasses. (Passive glasses, such as those made from cardboard and plastic commonly used for viewing 3D movies, require placing polarizing filters on each projector; the projectors can also be used for stereoscopic viewing using the more expensive liquid crystal shutter, or "active," glasses.) Both projectors can display either mid-resolution (PAL or NTSC) video signals and high-resolution RGB signals.

In constructing the surrounding walls and selecting furnishings, our main consideration has been to simplify matching of the virtual extension with the real room. In particular, we have chosen materials and furniture which are simple to match with textures and geometry in the 3D model of the virtual extension. To further enhance the illusion of a seamless extension, elements in the real room mirror or align with elements in the virtual extension. Thus, the real table and virtual table are similar in shape, real and virtual lighting are placed in a similar fashion, and floor and wall edges line up.

### *4.3 Camera*

The initial implementation of TELEPORT allows for a single camera in each display room. The camera is placed on a table or stand just in front of the video wall and set at approximate eye height (for a sitting person), the field of view should be wide enough to take in a full or upper body shot of the local participant (when seated near the middle of the display room).

### *4.4 Viewer tracking system*

A viewer tracking system determines the position of the local participant within the display room (from which their viewpoint can be derived). The tracking system must operate over a room-size volume with accuracy, latency and sampling rate adequate for low-lag low-jitter rendering. Additionally, the person being tracked should not need to wear any bulky apparatus or be attached to cables. We experimented with several possibilities, including acoustic and electromagnetic devices. Currently, we are using an optical system (*3rdEye* from DynaSight), the accuracy is sufficient and no cabling to the viewer is needed.

### *4.5 Segmentation*

In the context of TELEPORT, segmentation refers to the real-time identification of regions of a video signal where a participant appears. Currently, we are using two techniques: chroma keying and delta keying. The first requires the remote participant be placed in a uniformly colored and lit *blue room*. This is a test configuration used only to evaluate image quality in the display room. The image shown in Plate III was produced using chroma keying.

*Delta-keying* uses a "reference" video frame. For TELEPORT, reference frames are shots of empty display rooms. When a participant enters a display room, they are identified by essentially subtracting the current video frame from the reference video frame (see Plate IV). Variations of this technique have been described recently [10, 24]. The drawbacks are sensitivity to changes in shading (such as changes in ambient lighting, shadows and reflections) and sensitivity to noise in the input signal. Also edge quality and processing delay can be issues. Further details of the algorithm can be found at [2].

### *4.6 Rendering*

The virtual extension is rendered from the viewpoint of a tracked participant located in the display room. Because this person is free to move within the display room, the virtual extension must be continuously re-rendered. Currently, we use an SGI Onyx with RealityEngine2 or InfiniteReality graphics hardware and can achieve rendering rates, with texturing and full anti-aliasing, of up to 50 frames per second (depending on 3D model complexity). Note that TELEPORT, as with CAVE systems, renders from a single viewpoint. Thus, if two or more people are present in a display room, at most one will see the virtual extension with the correct perspective.

### *4.7 Compositing*

By compositing we mean combining video imagery of remote participants with the rendered virtual extension. One approach is to take a remote signal, after segmentation, and simply overlay it (using video-mixing techniques) on that of the virtual extension. This is unsatisfactory for at least two reasons. First, tracking must be coupled to both the rendering viewpoint *and* to a perspective transformation for the overlay plane. Otherwise, there will be no parallax shift and the remote participant will appear to float and shift in space as the local participant changes viewing position. Second, with simple overlaying, all of the virtual extension appears behind the remote participant, i.e., it is not possible to place a virtual object in front of a remote participant (as occurs with the virtual table shown in the rightmost shot of Plate IV). TELEPORT uses a more versatile approach to compositing – by exploiting the capabilities of high-end graphics hardware, it is possible to create video textures. The video signal of a remote participant is texture-mapped onto a surface in the virtual meeting room. The video texture has a transparency channel (obtained from the segmentation stage), so only regions of the signal containing the remote participant are visible after texture mapping. Since the surface on which the video texture is applied is part of the 3D model of the virtual extension, it can be treated like any other piece of geometry. Thus, it is possible to scale, rotate and position (2D imagery of) the remote participant within the 3D virtual extension.

### *4.8 Audio*

Each meeting participant wears a small microphone. The audio signals from the microphones of remote participants are mixed together and sent to speakers mounted on either side of the video wall. Because of the video delay caused by segmentation, rendering and compositing (up to 200 ms), a matching audio delay is introduced to maintain lip sync. No additional audio processing is performed, although combining speaker position tracking with 3D audio localization hardware could improve spatial awareness and may be added in the future.

### *4.9 Transmission*

Because of the large display size and the video processing delay, TELEPORT has more demanding network "quality-of-service" requirements than window-based desktop video conferencing. Delay, in particular, is of concern, since the already significant delay from segmentation, rendering and compositing, when added to transmission and compression/decompression delays, could prove unacceptable. We have several transmission alternatives available within our laboratory; these include studio-quality uncompressed digital video over fiber, ATM, and satellite. In addition, we use the system with different video codecs. Currently, for local-area tests, we use uncompressed video; thus, the video quality is very high, the transmission delay is negligible, and there are no compression or decompression delays. For remote tests, we use ATM and motion-JPEG codecs (we will also test the system with MPEG-2 codecs in the near future).

The overall configuration of a single TELEPORT site is illustrated in Fig. 2. This configuration is based on that used in regular sessions between Bonn and Geneva, where we use motion-JPEG codecs (the Nemesys AVA/ATV 300) running at about 20 Mbps over ATM. However, as mentioned above, the transport method is entirely up to the user – the TELEPORT system just requires a bi-directional audio/video connection (for a two-party session). Looking at Fig. 2, the decoder produces an analog composite video signal, this is passed through an A/D converter, giving a 270-Mbps CCIR 601 4:2:2 serial digital video signal. The segmentation system, either a chroma keyer (Ultimatte System 7) or delta keyer (a program running on a multi-processor SGI Onyx)

**Plate IV.** Delta-Keying Sequence (input video frame, key, composite)

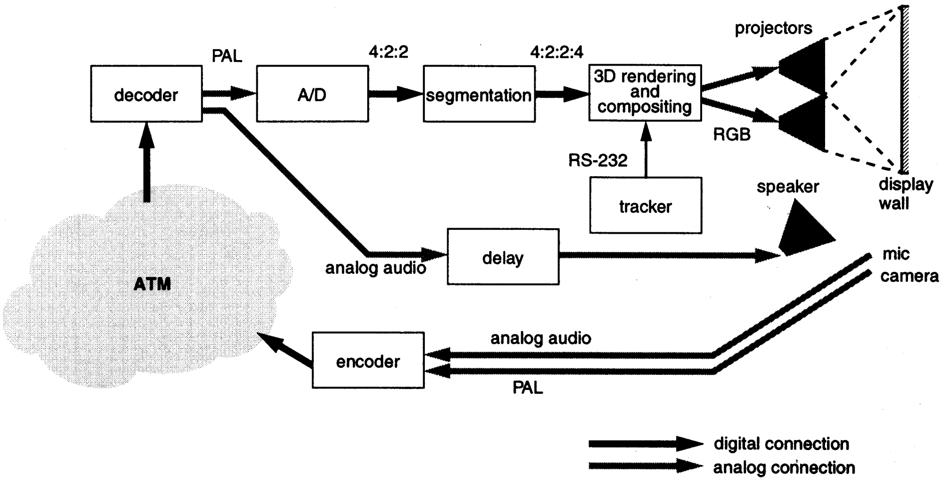


**Fig. 2.** TELEPORT system configuration

then gives a 4:2:2:4 video signal (i.e., the input signal plus accompanying "alpha" or "key" channel). The renderer, a process running on an Onyx with high-end graphics hardware, takes the 4:2:2:4 signal and produces left/right-eye views which are output as analog RGB signals to a pair of projectors. The additional delay introduced by segmentation, rendering and compositing ranges from about 2 frames (80 ms) to over 10 frames (400 ms), depending upon the choice of graphics hardware, segmentation technique, processing parameters (e.g., segmentation resolution) and geometric complexity of the virtual extension model. In order to maintain lip synchronization, the audio signal would be run through an audio delay device which could add delay in millisecond intervals.

## 5 Scenarios and trials

Scenarios for TELEPORT usage fall into two main groups: shared manipulation and visualization of 3D objects; and small group meetings. For cooperative 3D work, shared CAD applications are an example, the center of attention is more on the 3D model than the participants themselves. Here, the stereo capabilities of the display room are used (requiring participants to wear special glasses) and the system resembles a distributed (and vertical) Responsive Workbench [15].

The second group of scenarios, small group meetings, range from very focused and task-specific meetings to "generic" meetings with no particular topic or task. For the latter, TELEPORT simply offers an alternative to the telephone or face to face meeting. For task-specific scenarios, we are interested in situations where copresence is important to task completion.

We now list some of the usage trials for TELEPORT. These used one TELEPORT system located in Germany and a wall-sized video display system in Switzerland.

– *Remote consultation* represents a generic task. An ATM or satellite link from the University of Geneva offers graduate students the possibility to meet their professor in a virtual office.
– *Distributed rehearsal* falls into the task-specific category. Musicians in Geneva perform a rehearsal with a conductor in Germany [14]. These trials are part of DVP (Distributed Video Production), a project funded by the European Union's ACTS (Advanced Communication Technologies and Services) program, with the participation of several European broadcasters.

It should also be pointed out that it is possible to have "unsymmetrical" sessions, where some participants use TELEPORT display rooms and others use alternative display systems. This has been the configuration for our local tests: also, we may connect TELEPORT to a CAVE system located in our labs.

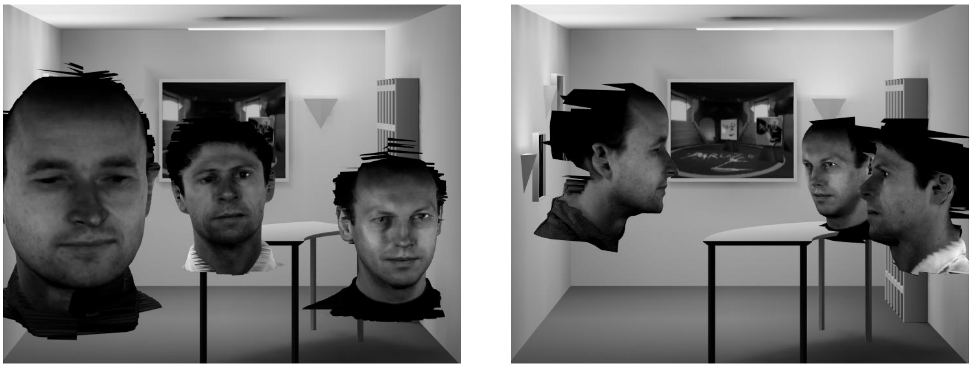

**Plate V.** A TELEPORT Session (*left*: virtual extension with raiosity lighting, *right*: virtual extension rendered in wire-frame). Note: redenreing viewpont has been set to camera position rather than local participant position

## 6 Observations

TELEPORT was designed to allow realistic meetings without the need for encumbering devices. Originally, we considered even stereo rendering to be of low priority, since it would require participants to wear special glasses. When the display room was installed, it became apparent that spatial perception of the virtual extension was very limited. Correct perspective continuation and a sense of depth could only be observed at a single point in the display room (and even then best spotted with one eye only). The result is similar to the perception of *trompe-l'œil* architecture or paintings, where the effect is typically quite flat unless viewed from the proper position.

As a consequence, we switched to stereo rendering. With stereo rendering, the impression of space is so strong that people wearing glasses often believe their position is tracked even when this is not the case (with no tracking, the virtual extension appears to "bend" in response to viewer movement). We now conclude that, for spatially immersive copresence, stereo display is a primary requirement.

While we have not attempted to measure the degree to which TELEPORT sessions approximate face to face meetings, it is probably safe to say that the system does allow for much more natural communication than with desktop conferencing systems. Since participants are rendered full scale and life-like, they are have some assurance that they will be perceived as they are normally perceived. In other words, they can be somewhat less concerned with one of the main hindrances of video conferencing – adapting ones position and posture to the camera rather than to the other participants.

In addition to face to face meetings between small groups of people, the situation for which the TELEPORT display room was designed, we have had much experience with using the room for tele-teaching sessions. In this configuration, the teacher sits in the display room and the class appears on the wall. The display room proved to be very comfortable for the teacher and was much preferred to using standard monitors. Although stereo and 3D were not used in this configuration, the large field of view provided by the wall display did enhance the sense of being in the remote class room with the students.

## 7 Future work

TELEPORT shows to what extent an immersive and invisible interface may be achieved using current technology; also, it points to some directions for further improvements in the realism of teleconferencing.

There are limitations of the TELEPORT system, the most noticeable involve eye contact and visual quality. First, despite the ability to place participants in a shared space, the establishment of eye contact is still dependent on camera and participant placement. Difficulty with eye contact is a symptom of the more general problem of *visual inconsistency* which arises from compositing material taken from different viewpoints. Solving this problem requires going beyond flat image-based representations of participants to 3D modeling. Efforts in this direction, such as warping images onto 3D facial models, are still lacking in realism, but we expect to see improvements from either combining tracking of facial features with video texture mapping or by using multiple cameras and view morphing. Plate V shows an attempt to extend TELEPORT in this direction. Here, static 3D facial models, serving as "avatars" of the remote participants, are placed with the virtual extension. By coupling avatar placement to a position- and orientation-tracking system, one approximates eye contact and gaze awareness.

Concerning visual quality, the two main problems are low image resolution and poor segmentation. The physical size of a TELEPORT video wall, and the close viewing distance, require a very high resolution signal – otherwise details blur and the scan line structure becomes visible. One possibility is to tile the wall with display modules [26]. However, this is expensive and creates slight seams in the image. The alternative is to wait for higher resolution display devices to become available. Visual quality also depends on the segmentation technique. Here, we hope to improve on delta keying by digital signal processing on a parallel machine.

In summary, this paper has concentrated on design and implementation issues of TELEPORT, since remote usage trials are just beginning. The main contribution of the system is as a concrete demonstration of the possibility of immersive teleconferencing without the need to leave the real world.

*Acknowledgements.* The authors would like to thank Dennis Tsichritzis for his support in creating the environment which made the TELEPORT project possible, August Tepper for helping realize the environment, Medien SL for designing and installing the display system, George Drettakis for his interest and many useful suggestions, Josef Speier and Jurgen Ziehm for their 3D modelling work, Karl Matthias Konke for the design of the display room and producing Plate I, and Chris Osland for helping improve the audio system.

## References


1. Ahuja SR (1988) The Rapport Multimedia Conferencing System. Proc. of ACM Conf. on Office Information Systems, 1998, pp 1–8
2. Arapis C, Gibbs S, Breiteneder B (1996) Experiences with a Multiprocessor Implementation of a Real-Time Segmentation Technique for Video Streams (submitted for publication)
3. Bly SA, Harrison SR, Irwin S (1993) MediaSpaces: Bringing People Together in Video, Audio, and Computing Environment. Commun ACM 36(1): 29–47
4. Boyer DG, Lukacs ME (1994) The Personal Presence System – A Wide Area Network Resource for the Real Time Composition of Multipoint Multimedia Communications. Proc. ACM Multimedia 1994, pp 453–460
5. Buxton W (1992) Telepresence: integrating shared task and person spaces. Proc. Graphics Interface 1992, pp 123–129
6. Cruz-Neira C, Sandin DJ, DeFanti TA, Kenyon RV, Hart JC (1992) The CAVE: Audio Visual Experience Automatic Virtual Environment. Commun ACM 35(6): 65–72
7. Elrod S, Bruce R, Gold R, Goldber D, Halasz F, Janssen W, Lee D, McCall K, Pedersen E, Pier K, Tang J, Welch B (1992) Liveboard: A Large Interactive Display Supporting Group Meetings, Presentations and remote Collaboration. Proc. CHI, 1992, pp 599–607
8. Feiner S, MacIntyre B, Seligmann D (1993) Knowledge-Based Augmented Reality. Commun ACM 36(7): 52–62
9. Fish RS, Kraut RE, Chalfonte BL (1990) The VideoWindow System in Informal Communications. Proc. CSCW, 1990, pp 1–11
10. Gaver W, Moran T, MacLean A, Lövstrand L, Dourish P, Carter K, Buxton W (1992) Realizing a Video Environment: Europarc's Rave System. Proc. CHI, 1992, pp 27–35
11. Gaver W, Smets G, Overbeeke K (1995) A Virtual Window on Media Space. Proc. CHI, 1995, pp 257–264
12. Ishii H, Kobayashi M, Grudin J (1992) Integration of Inter-Personal Space and Shared Workspace: ClearBoard Design and Experiments. Proc. CSCW, 1992, pp 33–42
13. Kleinholz L, Ohly M (1994) Multimedia Medical Conferencing: Design and Experience in the BERMED Project. Proc. of the International Conference on Multimedia Computing and Systems, 1994, pp 255–264
14. Konstantas D, Orlarey Y, Gibbs S, Carbonnel O (1997) Distributed Musical Rehearsal. Proc. Intl. Computer Music Conf, 1997
15. Krüger W, Bohn C-A, Frölich B, Schüth H, Strauss W, Wesche G (1995) The Responsive Workbench: A Virtual Work Environment. IEEE Computer 28(7): 42–48
16. Leigh J, Johnson AE (1996) Supporting Transcontinental Collaborative Work in Persistent Virtual Environments. IEEE Computer Graphics 16(4): 47–57
17. Mandeville J, Furness T, Kawahata M, Campbell D, Danset P, Dahl A, Dauner J, Davidson J, Kandie K, Schwartz P (1995) GreenSpace: Creating a Distributed Virtual Environment for Global Applications. Proc. IEEE Networked Virtual Reality Workshop, 1995
18. Mantei MM, Baecker RM, Sellen AJ, Buxton W, Milligan, T (1991) Experiences in the Use of a Media Space. Proc. CHI, 1991, pp 203–208
19. Okada K, Maeda F, Ichikawaa Y, Matsushita Y (1994) Multiparty Videoconferencing at Virtual Social Distance: MAJIC Design. Proc. CSCW, 1994, pp 385–393
20. Olson JS, Olson GM, Meader DK (1995) What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work? Proc. CHI, 1995, pp 362–368
21. Root RW (1988) Design of a Multi-media Vehicle for Social Browsing. Proc. CSCW, 1988, pp 25–38
22. Seitz SM, Dyer CR (1996) View Morphing. Proc. SIGGRAPH, 1996, pp 21–30
23. Sellen A, Buxton W, Arnott J (1992) Using Spatial Cues to Improve Videoconferencing. Proc. CHI, 1992, pp 651–652
24. Takeuchi A, Naito T (1995) Situated Facial Displays: Towards Social Interaction. Proc. CHI, 1995, pp 450–455
25. Watabe K, Sakata S, Maeno K, Fukuoka H, Ohmori T (1990) Distributed Multiparty Desktop Conferencing System: MERMAID. Proc. CSCW, 1990, pp 27–38
26. Woodward PR (1993) Interactive Scientific Visualization of Fluid Flow. IEEE Computer 26(10): 13–25
27. http://www.tnt.uni-hannover.de/tnt/project/eu/panorama/overview.html


Simon Gibbs is a staff engineer with the Sony Distributed Systems Laboratory in Santa Clara. Before joining Sony he was a senior scientist at GMD – the German National Research Center for Information Technology – where he was responsible for the Digital Media Lab within the Institute for Media Communication. Gibbs graduated with a PhD in Computer Science from the University of Toronto in 1983. He has publications in several areas including office information systems, CSCW, object-oriented programming, and multimedia systems.

Constantin Arapis completed his computer science studies at the University of Geneva and obtained his Master thesis in 1986. He received his PhD thesis in computer science from the same university in 1992. From 1994 to 1996 he was research scientist at GMD the (German National Research Center for Information Technology). Since September 1996 he is "Maitre d'enseignement et de Recherche" at the University of Geneva. Current research interests include teleeducation applications, teleconferencing systems and Internet multimedia applications.

Christian Breiteneder is an Associate Professor in the Institute of Applied Computer Science and Information Systems at the University of Vienna, Austria. His current interests include multimedia information systems, media processing systems and telepresence systems.