Meetings in a Virtual Space: Creating a Digital Document

Lori Toomey, Lia Adams, and Elizabeth Churchill

FX Palo Alto Laboratory, Inc.; 3400 Hillview Ave., Bldg. 4; Palo Alto, CA 94304
{toomey, adams, churchill}@pal.xerox.com

Abstract

Improvements in computer network infrastructures and information utilities have led to an increase in the number of social and work interactions carried out 'virtually' by geographically separated group members [1, 5, 6, 7]. In this paper we describe the design and evaluation of a prototype system that supports synchronous and asynchronous collaboration between researchers separated by space and time. The system provides non-collocated team members with a digital, virtual space for information sharing and discussion. For synchronous interactions, our design prioritizes provision of shared context, real-time discourse, and real-time problem solving and negotiation between the team members. In the case of asynchronous interactions, we have prioritized the capture of team decision making and negotiation processes and the representation of these processes in a context-rich, hypertextual document of team problem solving and negotiation.

Introduction

In recent years, improvements in computer hardware, software and network infrastructures have increased digitally based interactions between geographically separated individuals. Recently published statistics state that the Internet reaches over 150 countries, is currently used by 30 million people and is growing at the rate of one million new users per year [6]. These developments are creating new ways of clustering people and the emergence of 'global cities' and 'virtual communities' [1, 2, 6, 7]. Digitally based interactions between members of these virtual communities include socializing, game playing, information sharing, sales and advertising, leisure pursuits, and collaborative work.

In this paper we consider the design of tools that support work interactions between non-collocated research teams, working across time zones. Specifically, our goals are to investigate the problems of effective collaboration within distributed teams incurred by spatial and temporal separation, and to offer some novel technological solutions to these problems.

Although communication media such as letters, the telephone, email, faxes and more recently video conferencing and media spaces have all played their part in supporting long distance communication, cooperation and collaboration, most of these have a common shortcoming: they fail to create sufficiently rich shared contexts for real-time collaborative work. As has been noted elsewhere [12], co-workers are more likely to collaborate if they are in the same place. This is because being physically collocated enables the easy construction of shared contexts for problem solving, negotiation and information sharing, and enables rich, multi-layered and multi-modal interaction through the use of speech, body movement, gesture and intonation. No technological medium to date has supported the creation of such rich interaction contexts.

Virtual spaces and distributed groups

Our current research centers on the use of virtual, digital meeting places and spaces in an attempt to address some of these shortcomings. Virtual spaces have been gaining popularity over recent years, being used mostly for social purposes in the form of 'chat spaces' [3]. Despite this initial use, we believe there is great potential for their use in business settings.

A number of points support this contention. First, central to the increasing popularity of such virtual spaces has been the fact that communication over large distances has sped up substantially in recent years. This speed-up is currently at a point where the time taken for a message to traverse the globe has reduced so much that the receiver and sender experience the interaction as being in "real-time", or synchronous. Thus these spaces offer advantages over letters, emails and faxes which do not so readily support fast, real time interaction. Secondly, virtual environments are highly flexible. Within one virtual space it is possible to represent text and graphics, and include streamed auditory and video information. Given the structure of the underlying architecture for most virtual spaces, it is possible to manipulate the incoming data streams while maintaining a coherent sense of location to a far greater degree than afforded by video conferencing. Finally, virtual environments represent a paradigm shift in the way in which individuals relate to the artifacts they

1060-3425/98 $10.00 © 1998 IEEE

share. Virtual environments literally provide a 'space to inhabit'; individuals share the space with other individuals (usually in the form of 'avatars' or graphical embodiments) and with the text and graphical information they are discussing. Media spaces and video conferencing still adhere to the metaphor of the window onto a shared context or window to view other individual's 'talking heads'. To a far greater extent than video conferencing or media spaces, therefore, virtual environments provide a sense of 'embodied action' upon shared artifacts and thus provision of shared context.

In the next sections we will discuss a number of experiments with our prototype system which we hope illustrate these points. First we discuss the design of the prototype, followed by several examples of the prototype in use. All of these examples illustrate the use of the prototype to support both synchronous and asynchronous collaborative and cooperative work activities. In each case, we discuss how the system provides support for these activities: synchronously by providing a greater sense of co-presence and shared context, and asynchronously through the provision of a rich multimedia log of all activities. This hypertextual log represents a context-rich document of all meeting activities to provide a rich resource for on-going collaboration. Before discussing the examples in detail, we will briefly consider the design of our prototype.

Designing the prototype: methodology

A set of meetings between physically collocated researchers was observed to identify a set of central features for successful meetings. Success was here defined simply to be that agenda items were discussed and action items noted.

From these initial explorations and observations, a number of meeting scenarios were developed from which our design requirements were specified. These design requirements specified a number of crucial features for meeting support. These will be discussed in turn.

Design scenarios

We began our design process by imagining the kinds of meeting scenarios we wanted to support. We focused on meetings of small groups (fewer than seven people), although we wanted to allow more than one group to use the space at a time. One meeting scenario was of a scheduled meeting such as a project status meeting to discuss a pre-arranged agenda. In addition, we wanted to be able to use the space for unscheduled discussions, analogous to the serendipitous consultations that happen in a hallway.

To support these meetings, we needed to support interactions including discussion, sharing of documents during a meeting, sketching, and markup of diagrams.

Features we identified

We identified features necessary for supporting these activities. The functions we deemed essential to provide were awareness, communication, sharing of artifacts, and a sense of place. Also, in order to support hallway-style interactions, we required that our prototype be low overhead to use, so that running it would not hinder doing other tasks between interactions in the space.

Next, we determined how to implement these features using the software we had available.

We chose the Palace software from The Palace Inc. [11]. The Palace provides a 2-D graphical MOO environment. Using the Palace, we implemented our features thus.

Awareness of the presence and state of users was implemented by graphical avatars. We originally created a few photographic avatars of each user's face to represent a user by an alert expression when they are participating in a discussion, and by a sleeping face when they are not participating. Users were free to create new avatars for themselves.

To convey information about the role being played by a particular user (such as being the presenter or moderator in a meeting), we planned to use the combination of spatial positioning and change of avatars.

For peripheral awareness of other users, we programmed some rooms to play sound cues to make users in the room aware that others were arriving in or leaving the room.

For communication and discussion, we used typed text. In the Palace, each user's typed words appear in a cartoon-style balloon coming from their avatar. This space does not support voice communication, but it permits users to play prerecorded sound files, which we anticipated would be used for exclamation or attention-getting.

To share documents, we could initially share only documents that could be viewed in a web browser. The Palace software allows users to direct their and others' web browsers to load a specific URL.

Since many types of business meetings we considered include frequent sketching on whiteboards, we included areas in the virtual space for drawing. In our implementation, room backgrounds can be used as canvases for drawing, and we created some rooms especially for this purpose.

The fairly compact footprint occupied by the Palace client, leaving screen space free for other tools, met our goal of low overhead use.

Implementation of prototype

The Palace's client-server architecture was appropriate for our prototype, because we wanted a central server that could be instrumented to capture all events from all clients, and assign them a sequential ordering.

Our space, the FXPALace, is a 2-dimensional graphical MOO, structured as a collection of rooms connected by virtual doorways. A room in the FXPALace consists of a set of background graphics. Users appear as customizable avatars and can carry props (graphical objects). Users visiting the same room can communicate textually and can draw in the same space. Users can play prerecorded sounds but cannot converse using live audio.

Rooms and users can each have attached scripts activated by movement to a spot in a room, in response to utterances, in response to arrivals and departures, or through other events.

Scripts can play sound effects, add props to a room, execute painting commands, and drive the user's web browser to a particular URL. Sharing web pages and other documents through independent client-side non-WYSIWIS web browsers and other access tools is the main way to share non-conversational information.

The entrance to our space models the entrance to the FXPAL building. Rooms in the space include lobbies, large and small public meeting rooms, casual spaces (such as our patio and idea room), and user-tailored spaces with personal appeal and less appearance of a workplace (for example, a beach and a waterfall). Some rooms are special-purpose places, such as rooms for collaborative discussion of math problems. Special-purpose rooms often have sets of links to web pages on a topic. For example, the meeting room has a scripted spot linked to the agenda of the current meeting.

Evaluating the prototype

In this section, we describe four kinds of meetings and discussions we observed in the FXPALace. For each meeting, we evaluate how the features we designed support the interactions.

1. Keeping in touch

One of our users was an FXPAL researcher who left on vacation early in the stay of a summer research intern working on his project. To keep in touch with her during his absence, he traveled with a notebook computer that could connect to the FXPALace and to his FXPAL email account through dial-up networking. Sporadically, he dialed into the FXPALace to ask the intern, who was usually logged in, if she had questions or problems.

Through short discussions in the FXPALace, often begun by email questions, they were able to help her get past a difficulty and continue her work. These meetings were short, unscheduled, and dedicated mostly to short problem-solving sessions.

The media they used were primarily text and audio cues by which the researcher signaled the intern that he had arrived in the space. Both participants in these meetings used cartoon avatars, but changed avatars very little. These can be seen in Figure 1. They did not use the drawing features of the space often. The missing feature they wanted most strongly was a shared editor for viewing and discussing source files, and referring to points in the code, since they were developing software together.

Figure 1: Intern and researcher

As there was no source code editor integrated into the space, the text logs that discussed locations of files, or places in code, were the only part of the meeting documents that were useful for reference after the meeting had ended. These meeting documents can be rendered and read simply as text, without losing much of the meeting experience. This example illustrates that the amount of information required for the re-creation of context depends substantially on how complex the current task is.

2. Co-design meetings

During July 1997, two high-school intern students used an FXPALace-based virtual space to collaboratively explore the design of on-line logic puzzles, and to design new FXPALace rooms and work presentations. They held daily meetings to discuss their summer project, discuss the layout of their poster presentations and more importantly to explore the design of the logic lessons they were transforming into interactive rooms. Text and drawings on whiteboards were used to communicate with one another, sounds were employed to get one another's attention (and to tease one another), and they used avatars extensively to express changes of mood and changes in the roles they felt they were playing. They kept the logs of all text they exchanged with one another (and with other consultants in the lab).

At the end of their project, the design rationale for the finalized products of the project (the rooms they created for the logic puzzles) can be re-created from their text

logs. These text logs provide a rich resource for this rationale reconstruction.

The design process rooms they developed for the co-design of the poster, by contrast, offer a document in themselves of the design's progress. The rooms contained drawings they had created using the whiteboard markup feature of the space. Each room contains a different stage of the design, offering a graphical log of their collaborative design process. Coupled with the text logs, the finalized product rooms and the design process rooms represent a very rich document of their collaborative activities.

A final multimedia meeting document capturing all of their summer activity would include the evolution of the drawings of their poster design, coordinated with the full text logs of their discussions and with the changes of avatar they used to signal changes of expression.

3. Daily use within lab

Since November 1996, members of the Collaborative Spaces research project at FXPAL have used the FXPALace almost daily. We have home computers running the client software, and so we can visit the space from home or from the office.

A number of observations can be made of the system. We tend to change avatars to express availability; that is, to indicate whether or not we are free to interact in the space. For example, we have changes of avatar appearance that indicate the presence of a visitor or a phone call, if we are away from a computer, traveling from home to work, or busy at another activity such as thinking or writing. In addition to these states of availability, we also use specific avatars to express moods such as tired, confused, frustrated, happy, smart, or inspired. Some of these are shown in Figure 2.

Figure 2: Common expressive avatars

All avatar choices are explicit; unlike video, which can capture and convey more than the user intends, avatar expressiveness is under the user's direct control. During a visit to Japan, one user made all of her FXPALace connections early in the morning or late at night. At these times, an avatar could make a professional appearance when a video image would have been completely inappropriate.

Within FXPAL, we use audio cues to get attention, initiate interaction and to punctuate conversation as exclamations.

For slightly asynchronous communication (short remarks not needing a real-time response or interaction), we often used delayed balloons that stay on the screen, to catch and direct a participant's attention, when it is convenient for the addressee. (E.g. "Elizabeth, what do you think of layout in the new Math Room?")

We used rooms to delimit conversation and avoid annoying others with sound cues; if all users spend their time in the same virtual room, they hear every attention-getting sound anyone in that room plays. Users often travel to a new room to converse privately. We found that the number of sounds people imported grew enormously; we currently have 19MB of WAV files.

Our meetings mixed synchronous use (conversational interaction), with some asynchronous use. Because the text, navigation, and sounds are logged, a person who has stepped out from an FXPALace session can quickly catch up by reviewing the log document. This useful feature of MOO logs has been noted in [4]. At present, we cannot review the media other than text, sounds, and navigation while a Palace session is going on. Since we keep the FXPALace connections active all day, the main use of documents generated by these daily meetings is to review discussion by other project members by members who didn't participate in real time.

4. Meeting with our Fuji Xerox colleagues

Our research group conducted a virtual meeting with some of our Fuji Xerox (FX) colleagues in Japan. Four researchers in Palo Alto met with a group of eight or so FX researchers.

This was the first time our FX colleagues had visited our virtual space. The purpose of the visit was (1) to meet one another and get acquainted, (2) to take a tour of the new room authored by the FX researchers, and (3) to vote on some technical issues we had been discussing through email. This agenda had been prepared before the meeting through email.

The meeting was held at 10:15AM Tokyo time (6:15 PM California time). Because of limited phone-lines, only two FX researchers could be in the space at the same time. The participants in Japan took turns using two laptop computers running client software; each laptop was on a separate floor of the building. Both laptops were in public areas, and so for politeness, the audio volume was turned down so low that it was practically inaudible.

The meeting lasted for about 40 minutes. The meeting included a lot of text conversation, some sound playing

(mostly by the FXPAL users already familiar with the space), navigation between rooms, and paint-style markup.

All of the available media were used in this meeting. Text was used for conversation and discussion. All text was typed in English; the client software does not have any provision for Asian character sets, and none of the FXPAL participants understands Japanese. The California participants used sounds for exclamations, but the new FX users were kept busy enough trying to respond to text typed by 4 or 5 co-workers that they never had time free to look at the sound-playing interface.

The meeting included a fair bit of room-to-room navigation, as one portion of the meeting was spent touring some newly added rooms and functionality. A new room shown in Figure 3 included a link to web pages describing an FX research project.

In this meeting, all participants used photographic avatars of their faces. Because several FX participants were sharing each laptop computer, the logged-in name of the client program was not enough to show the others who was currently driving the client. The use of photographic avatars let the FX researchers recognize some FXPAL people from pictures of previous visits, and also helped the FXPAL people recognize and welcome the FX visitors. Each time the user changed, he changed the current avatar to a picture of his own face.

We had further evidence that photographic avatars can assist with visual recognition and prompt introductory discussion. One of the FXPAL researchers, Lori Toomey, is Japanese-American, and the researchers from FX immediately recognized this when they saw her avatar. Just as FX researchers who physically visit Palo Alto have done upon meeting her, they asked if she was Japanese, whether she spoke the language, etc.

One unexpected behavior resulted from the use of photographic avatars to identify participants. Some of the FX participants found the real-time demands of typing in English to be difficult, and so one of the two groups spontaneously developed the practice of having the person whose turn it was to participate dictate their remarks to the best typist in the group. The good typist stayed seated at the laptop.

Near the end of the meeting, this person realized he could "spoof" his boss by typing his own flippant remarks while using his boss's avatar. Since we at FXPAL use the space as a workplace and tend to practice workplace behavior there, the California participants had no idea that the person typing was not the person whose avatar they were seeing.

This meeting resulted in a multimedia document containing full text logs of the conversation, logs of the sounds played, captured images of the rooms and avatars, movement of "chip" props used to cast votes, and drawing

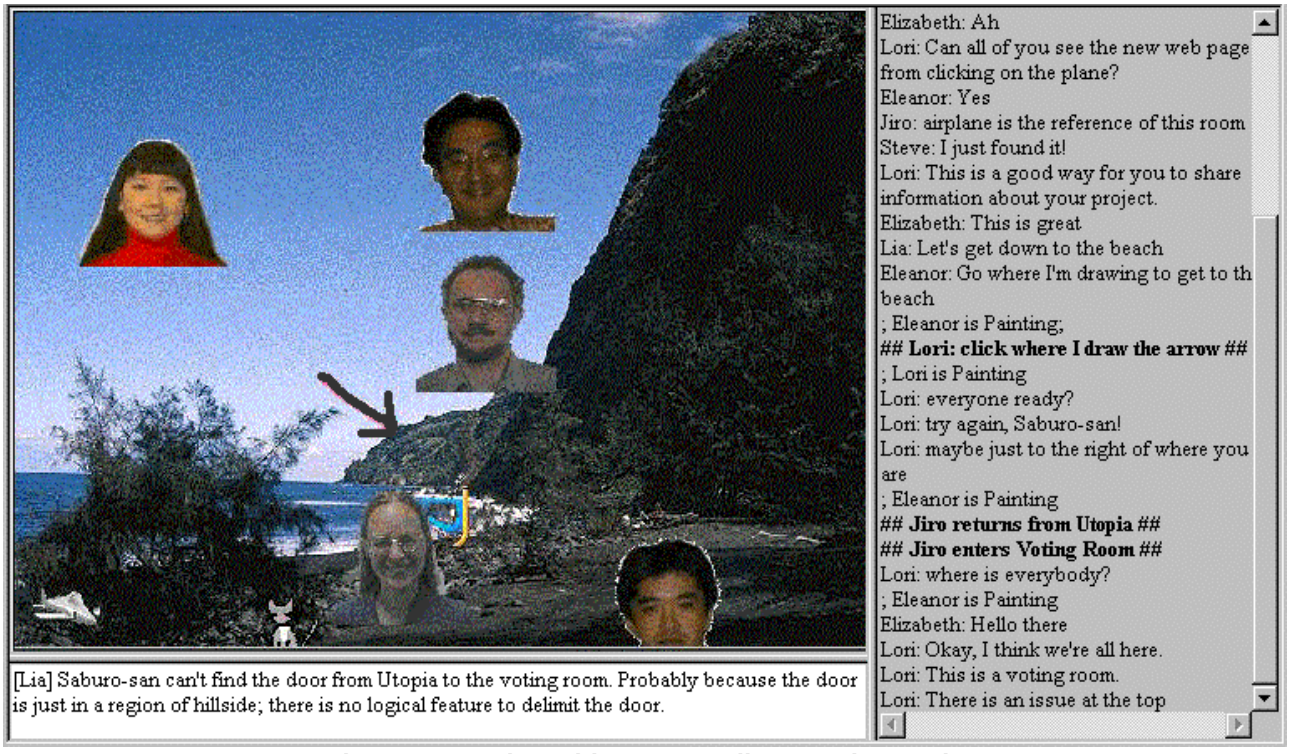

Figure 3: Meeting with our FX colleagues in Utopia

markups.

By instrumenting the space, all of these events can be correlated to one another by timestamps. With little or no manual intervention, this meeting document can be generated without a human agent summarizing or collating.

Figure 3 shows a sample of an FXPALace meeting as an HTML document. The page shown contains a frame showing the state of the graphics space (background, avatars, and props), and a frame showing the log of all text conversation, sounds played, and navigation information.

The document as generated affords reading straight through, for a linear replay of the meeting, as well as textual scanning or searching, for a rapid skim of the content discussion.

In addition to perusing a generated meeting document, a participant may wish to add further information after the meeting, through annotation. In the meeting described above, an activity that occurred directly after the meeting was the addition of annotations about the meeting. The researcher visiting Japan had been taking notes about our colleagues' experiences using the space, the problems they encountered, and any unexpected occurrences.

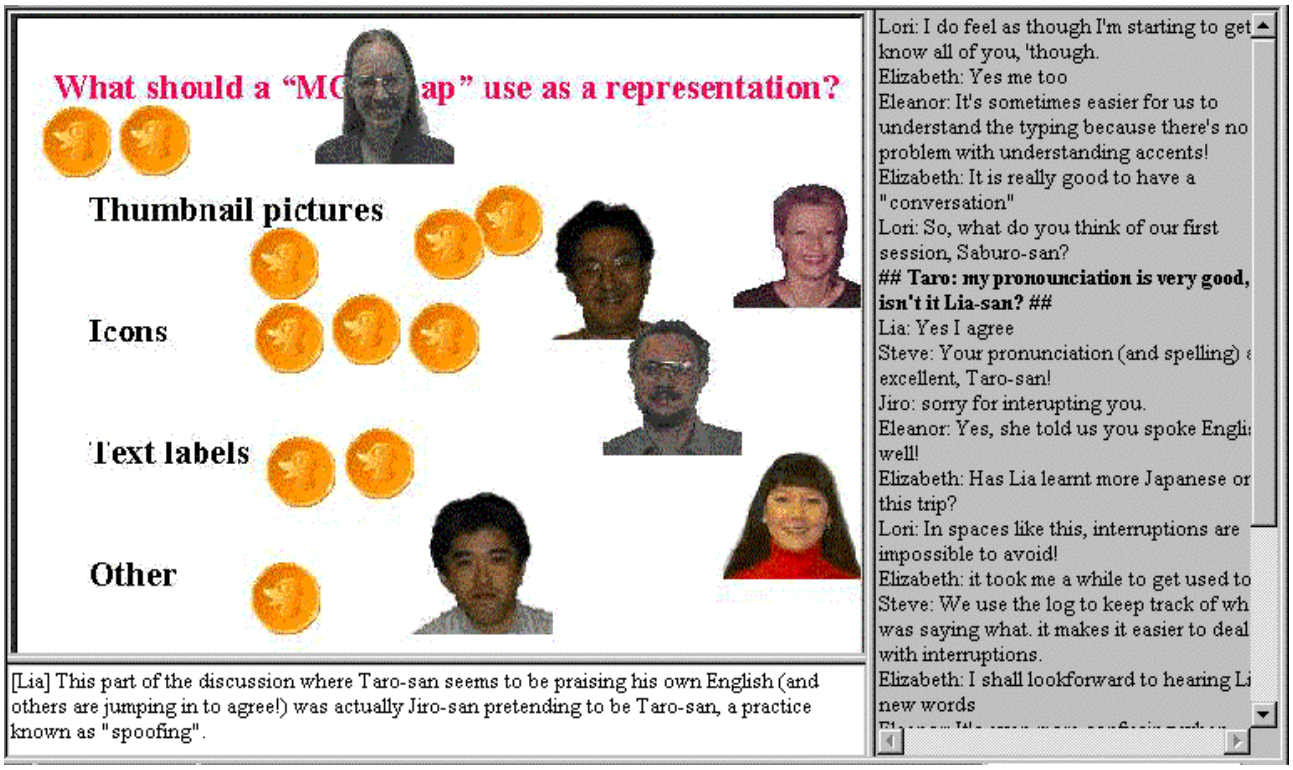

Figure 4: Document of our discussion in a voting room

Currently, the virtual space can be instrumented only on the client side. The meeting document shown above, therefore, shows only one thread of activity in the virtual space. The client captures only the events that occurred in the same room as the user running that instance of the client program. We plan build our next virtual space on a new platform in which we can instrument the server, allowing us to coordinate multi-room activities that occur in parallel. Representing the parallel activities within a single hypermedia document poses some challenges in how to structure the document.

Reading and annotating a meeting document

At the time, she wrote these notes as text on a notebook computer while she observed the meeting. She then reviewed the captured meeting document and inserted the annotations manually into the appropriate part of the time stream. In future, we would like to see annotation as one of the client-side applets running in the virtual space, so that the server can collate annotation events.

Figures 3 and 4 show examples of hypertext annotations in meeting documents. The document presentation includes a replay window for the graphical state; a log window containing text, names of sounds played, and arrivals and departures; and a window for displaying annotations linked from elements in the log or picture.

An annotation in Figure 3 records a problem a user experienced in navigation. The newly authored room, Utopia, had an exit door that was hard to find. One new user couldn't find it, and another user painted an arrow to show him the location. Lacking telepointers, users paint and erase to point in the space.

The annotation shown in Figure 4 marks the lines typed by the person spoofing his manager. Readers of the meeting document need not be fooled by the spoofing as the original FXPAL participants were.

A third kind of annotation is not visible in the text log of the meeting. Because the name captured by the log is the login name entered at the beginning of the session, the turn-taking behavior of the FX participants is not reflected in the text log. So, later annotations on the login name can help a reader skimming of the meeting document know who is talking without having to use the graphical display.

Results

Our first somewhat surprising observation was that the FXPALace quickly became the default mode of daily communication in our research group. We are continuously connected to the FXPALace while at the office, as well as occasionally from our homes. It is interesting to note that there are times when we are unaware that one (or more) of us is working from home because our interactions are exactly the same as if we were all at the office.

As with physical workplaces, behavioral and social conventions are important. For example, we use avatars with "astral bodies" and closed eyes to indicate that we are not currently focused on the FXPALace, and audio cues such as a knock to get someone's attention.

Certain practices such as spoofing are not appropriate in all contexts, but rather than encode restrictions into the space, we rely instead on using place-based cues and social norms to indicate what sort of behavior is appropriate. For example, one rightly assumes that spoofing and using comic avatars is much more appropriate with peers in an informal discussion that takes place on the patio than it is during an on-line meeting with upper management in a meeting room setting. Harrison and Dourish refer to this as behavioural framing [8].

We were also somewhat surprised at the level of awareness and expressiveness avatars provided. Initially, we viewed avatars as an inexpensive, low-bandwidth alternative to video, hoping that the engaging nature of graphical virtual environments would counterbalance the loss of visual feedback provided by video. We rapidly discovered that a variety of avatars with different expressions and clothing, and/or with props such as question marks over the head, etc., could be used to provide visual feedback and to convey emotions to a remarkable degree.

Further, the use of avatars had two major advantages over video. First, users could explicitly control what they were expressing/conveying. Second, we avoided many of the concerns users of video-based systems have with regards to privacy, control of images, and general issues of camera shyness [9]. The primary disadvantage is the need for users to explicitly update the state of their avatars to avoid misleading others, e.g., changing from an "alert" avatar to a sleeping one when stepping out of the office to indicate they are no longer attentive.

In addition to the avatars, we found that simple, pre-recorded audio snippets were a powerful way to express emotions. FXPALace sessions frequently include the sounds of laughter, sobbing (to express frustration), applause, etc. There is a drawback, however, in that this use of audio can be inappropriate or intrusive in public settings. Interestingly, the text form of these sounds, i.e., the command one types in the FXPALace to play the sound, have spread to email communications within the group, used in much the same way as smiley faces or other emoticons.

Embedding links to web resources didn't seem to be as important in defining a place as we anticipated. A short text command provides access to arbitrary web pages from any room, so there didn't seem to be any real association of web pages with particular rooms.

It was clear that shared artifacts were critical for technical discussions. For a limited volume of text, e.g., URLs, file pathnames in a LAN, etc., it was easy to cut and paste into the normal text window used for discussion text. However, for more extensive text blocks, e.g., source code, a shared text editor is needed.

For simple mark-ups, we used the Palace drawing tools, but a more sophisticated whiteboard with operations such as copy, cut and paste is really needed for many tasks. We also used drawing to indicate items or areas of interest but realized telepointers would be much better.

During the course of the many meetings we held in the space, we realized that captured on-line sessions could be viewed as a multimedia document. Unlike summarization documents that are traditionally created after a meeting, the session is a document that also captures the context in which events occurred.

Also, rather than a raw recording and playback of meeting events, such as a videotape of a live meeting, flexible tools could be created for skimming and retrieving sections of the meeting document. This requires client software that enables users to manipulate the traditional meeting documents, and coordinates their creation and modification with other meeting events. The

Coral system [10] looked at doing something similar for technology-assisted face-to-face meetings.

Future work

There are a number of areas to pursue following our experiments with the FXPALace prototype. Efforts are underway to design and build tools for capturing, revisiting, navigating, and annotating meeting documents.

Another issue we are considering is the integration of additional media, which will require their own tools for capture and playback. Text chat has the advantage of generating an easily searchable log, but has other limitations. Thus, we are considering how voice might be included in future prototypes without losing this advantage. We are also considering alternative, less intrusive media for the audio cues we have found to be so useful.

Not surprisingly, our experience using the space to meet with our Japanese counterparts showed that the native speakers of English were typing much more quickly than the Japanese participants. The choice of ASCII text as a communication medium inherently biased the space towards our English-speaking users. We are eager to explore more of the cross-cultural issues that arise in holding business meetings whose participants are in multiple countries.

Finally, we are investigating how we can increase the information flow between our physical and virtual spaces. For example, using sensors to detect events such as a user being joined in her office by another person, or getting up and moving away from her computer. These events can then be used to trigger an automatic change in avatars to reflect what has happened.

Summary

We prototyped a virtual space to support collaboration between researchers and utilized it in a variety of settings. These experiences resulted in observations of how such a space may be used, indications of which features were valuable (or lacking), and an introduction to viewing meetings as documents. Future work will enable automated capture of on-line meetings as navigable multimedia documents, and the flow of information between the virtual and physical worlds.

References

[1] Amin A., PostFordism: A Reader, Blackwell, Oxford, 1994.

[2] Anderson, B., Imagined Communities, Verso, London, 1983.

[3] Curtis, P. and Nichols, D., "MUDs Grow Up: Social Virtual Reality in the Real World", Proceedings of the Third International Conference on Cyberspace, Austin, TX, 1993.

[4] Evard, R., "Collaborative Networked Communication: MUDs as Systems Tools", Proceedings of the Seventh Systems Administration Conference (LISA VII), Monterey, CA, 1993.

[5] Featherstone, M., Global Culture: Nationalism, Globalisation, and Modernity, Sage, London, 1990.

[6] Graham, S. and Marvin, S., Telecommunications and the City: Electronic Spaces, Urban Places. Routledge, London, 1996.

[7] Harasim, L., Global Networks: Computers and International Communication, MIT, London and Cambridge, Mass., 1993.

[8] Harrison, S. and Dourish, P., "Re-Place-ing Space: The Roles of Place and Space in Collaborative Systems", Proceedings of the 1996 Conference on Computer Supported Cooperative Work, Boston, 1996.

[9] Lee, A., Girgensohn, A., and Schlueter, K., "NYNEX Portholes: Initial User Reactions and Redesign Implications", Proceedings of ACM GROUP '97, Phoenix, 1997.

[10] Minneman, S., Harrison, S., Janssen, B., Kurtenbach, G., Moran, T., Smith, I. & van Melle, W., "A Confederation of Tools for Capturing and Accessing Collaborative Activity", Proceedings of the ACM Multimedia'95, San Francisco, 1995.

[11] The Palace Inc., "Virtual World Chat Software", http://www.thepalace.com,1997.

[12] Root, R. W., "Design of a Multi-Media Vehicle for Social Browsing", in Proceedings of the 1988 Conference on Computer Supported Cooperative Work, Portland, Oregon, 1988.

Acknowledgements

The authors would like to thank our colleagues in FX CRL in Japan, the Hypermedia Group and Les Nelson at FXPAL, and Vicki O'Day at Xerox PARC for participating in on-going evaluations of our prototypes. We would also like to thank researchers at Xerox PARC who are currently working on meeting capture and salvage for their useful and encouraging comments on our work to date.