Citation-capable video messages: overcoming time differences without losing interactivity.

Toshihiro Takada
Nippon Telegraph and Telephone Corp.
takada@t.onlab.ntt.co.jp

Yasunori Harada
Japan Science and Technology Corp.
hara@acm.org

Introduction

Communication networks like the Internet provide tools for surmounting space and time barriers. These barriers, namely distance and time differences, must be overcome to facilitate our activities that are fast taking on a global nature.

The telephone has been a tool for overcoming distance. We can also use videophones and tele-conferencing systems to surmount the space barrier. These tools, however, have not surmounted the time barrier. Participants of tele-conferences are required to be present at the same time. The problem of time differences has been a major obstacle in tele-conferencing and distance learning.

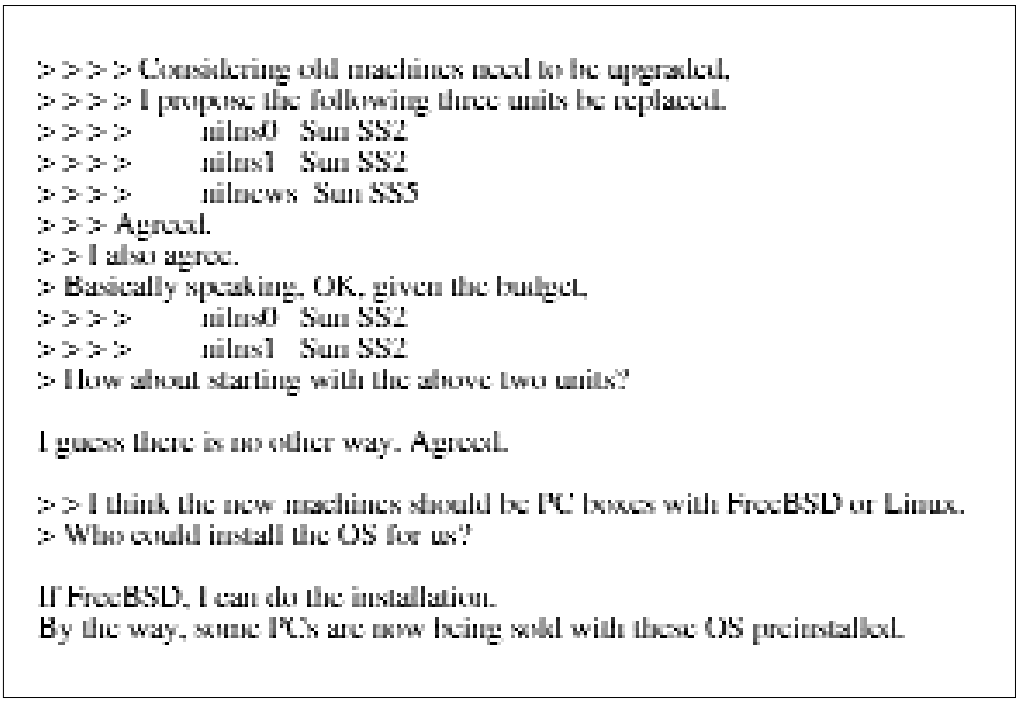
Commonly used tools for overcoming the time barrier are asynchronous communication systems such as E-mail, mailing lists, and Netnews (collectively referred to as "E-mail" below). Asynchronous communication systems enable each participant to read messages and write replies when convenient. There is no need for all participants to be present at the same time to achieve communications. Likewise, asynchronous audio/video communication systems as telephone answering machines, voice mail, and video mail (collectively referred to as "video message" below) enable dispersed groups to perform collaborative work without constraints caused by distance and time differences.

However, asynchronous communication systems suffer from poor interactivity. Interaction plays an extremely important role in communication between people. We often interact with others to express agreement with what the other person is saying, to bring up sudden questions and objections, or simply to interject comments. Since asynchronous communication systems lack simultaneity, participants cannot perform these kinds of interactions.

In this research, we focus on "citation-capable" messages as a means of introducing interaction into asynchronous communication systems. In the following sections, we describe what "citation-capable" means and how to achieve this feature in video messages. We also explore design and implementation of VideoPassage, a citation-capable video messaging system that has this capability. We then describe results of our experiments, discuss some design issues that the experiments raise, and conclude with a glimpse of possible future directions for this work.

What is "citation-capable?"

The one of the most outstanding feature of E-mail is that a received message can be cited. Generally, cited lines are duplicated with the character ">" placed at the front of each line. We can make a comment on, and ask a question about a subject on cited lines (see Figure 1). This citation function is a means of expressing agreement or objection and asking questions with respect to another party's messages in a discussion using E-mail, an asynchronous messaging system. We define a message with this sort of "citation" and "question/comment" functions as being "citation-capable."



> > > > Considering old machines need to be upgraded,
> > > > I propose the following three units be replaced.
> > > > nilus0 Sun SS2
> > > > nilus1 Sun SS2
> > > > nilnews Sun SS5
> > > Agreed.
> > I also agree.
> Basically speaking, OK, given the budget.
> > > > nilus0 Sun SS2
> > > > nilus1 Sun SS2
> How about starting with the above two units?

I guess there is no other way. Agreed.

> > I think the new machines should be PC boxes with FreeBSD or Linux.
> Who could install the OS for us?

If FreeBSD, I can do the installation.
By the way, some PCs are now being sold with these OS preinstalled.

Figure 1: Example of message citation in E-mail

The primary advantage of making messages citation-capable is that an asynchronous communication system becomes interactive. You can interact with others to express agreement, questions and objections, or comments, with this capability. Moreover, the use of citations eliminates the work and time to describe what one's own message is referring to. It also indicates the subject of discussion accurately and immediately.

This citation-capable function is also applicable to video messages. Incorporating citation and question/comment functions makes it possible to achieve an asynchronous video communication system that is interactive. An example application that we can consider here would be advice or instructions from remote sites in conjunction with presentation training or rehearsal for a performance. In this case, a video message that records the current state of a presentation or performance would be sent out. Then, a supervisor would directly cite and add comments on trainee's face, body or hand gestures, or even materials used in the presentation, as needed. In this way, appropriate guidance can be received on a trainee's actions or presentation materials.

Design scheme for citation-capable video messaging systems

Citation-capable video messaging systems must be built on an asynchronous remote videoconferencing system. To be more specific, a video message is first recorded in video and distributed to participants. Each participant other than the sender views the message. Then, to reply to it, a participant may create a new message while citing the original video and adding one's own material. Once a reply message is recorded and stored, it can be distributed to other participants. Exchanging video messages in this way essentially results in the production of a video-based dialog that begins with the original message.

"Editor" -- a unified input/output device

Some tools are necessary for carrying out the above tasks. As we ordinarily use text editors to perform the citation and question/comment functions in text-based E-mail, one of the tools clearly needs to be an "editor." The primary requirement of such an editor is that it combine output equipment (to display the other party's message) and input equipment (to enter one's own message) in a single device. This kind of unified input/output device must be provided for video-based messaging as well.

Continuous media like video has the both spatial expanse and temporal flow of information. To therefore achieve satisfactory citing in video messages, the following two requirements are crucial.

- Spatial citing -- Direct indication and correction of information presented by the other party
- Temporal citing -- Insertion of one's own message within that of the other party

In both, users also need to be able to clearly identify which parts of the message belong to which interlocutor.

A "shared board" model

In this research, we employ a "shared board" model as the unified input/output device to realise the video message editor. The following scenario illustrates our model. In this scenario, shown in Figure 2, the sender of the original message is called person "A," while the person replying to this message is person "B." These people continue to exchange messages as needed.

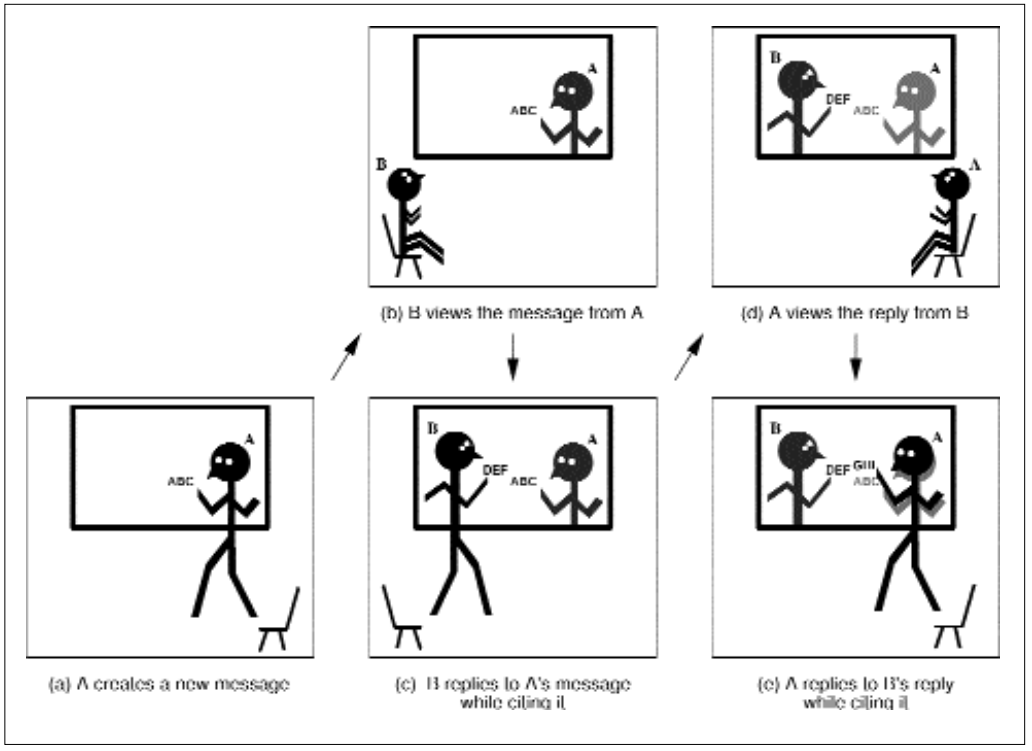


Figure 2: A "shared board" as the unified I/O device for video messages

To begin with, person A stands in front of a screen and speaks while being recorded to create the initial message [Figure 2(a)]. This video message can then be projected on a screen at person B's location [Figure 2(b)]. Person B, the receiver of this message, views the message by watching this projected video. Next, to reply to the message, person B stands in front of the screen. She or he then cites person A's projected message by overwriting text on or pointing to objects making up the message, such as person A's face or figure, text that person A may have written, etc. Recording the entire screen at this time creates the reply message [Figure 2(c)]. This reply includes (part of) person A's message on the screen and the video of person B standing in front of the screen and speaking while citing that message.

After receiving person B's reply, person A can view it by projecting the video message on a screen [Figure 2(d)]. To reply to this message, person A can stand in front of the screen in the same way that person B did. She or he then creates a new message while citing person B's message (which also contains part of person A's original message cited by person B). This situation is shown in Figure 2(e).

In this way, a screen, used for projecting a received message and creating a new message by having someone stand in front of it and speak, becomes the unified input/output device and the space on the screen becomes a virtual "shared board" between participants.

VideoPassage: experimental implementation of citation-capable video messaging system

System architecture

We have implemented an experimental citation-capable video messaging system called VideoPassage. The system architecture and external appearance of our system are shown in Figures 3 and 4, respectively. In this system, a whiteboard is used as a screen to play the role of the "shared board." A video projector to project video onto this screen and a camera to record new messages are connected to a personal computer with Microsoft Windows 98. A video capture board is used to store input video.

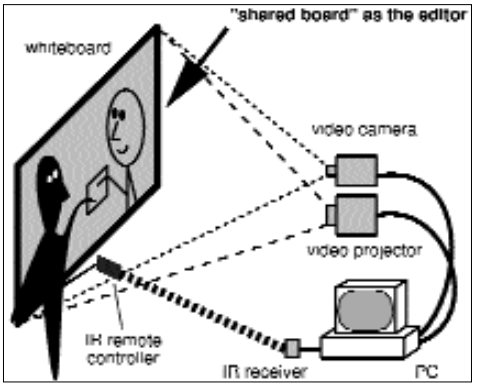


Figure 3: System architecture of VideoPassage

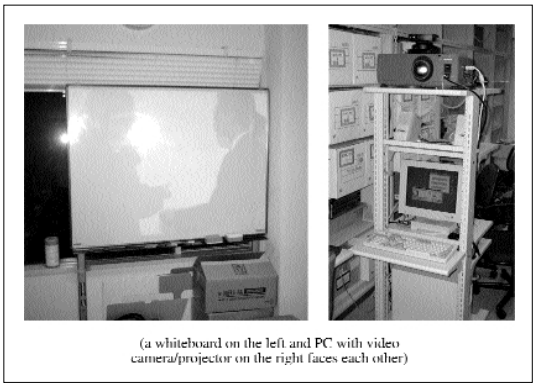


Figure 4: External appearance of VideoPassage

We have prepared controlling software to record new messages, display a list of stored messages, playback messages, and to reply to a received message. This software is written in Microsoft Visual C++, and consists of about 1900 lines. Video messages are recorded on a hard disk in Motion-JPEG format.

An infrared-type remote control is employed to operate the system, namely, to display a list of messages and to select, playback, and create messages. Video control operations while playing back messages, such as fast-forward, rewind, slow motion, and freeze, are similar to those of an ordinary VCR.

Examples of video messages

An example series of video messages created by our system are shown in Figure 5. Figure 5(a) shows the first created message. A trainee, the person on the right facing the screen, is rehearsing for a presentation. The trainee puts up a poster on the screen as presentation materials. Figure 5(b) and 5(c) show parts of the reply to this message. The former shows a supervisor, the person on the left, asking a question while pointing to the materials shown in the original message. The latter shows the supervisor giving some advice while citing the head of the person rehearsing on the right. (The supervisor advises that person not to look only at the screen but to look at the audience as well.) Figure 5(d) shows parts of the reply to the supervisor's message. It shows another person on the right making comment on the supervisor's question and the materials shown in the first message. For pointing to or citing the message, users draw marks on the whiteboard with standard whiteboard markers.

Figure 6 shows another example. In this series of messages, first, a person on the right draws the map on the board and describes the route to the destination. Then, another person on the left accordingly questions the sender about the route.

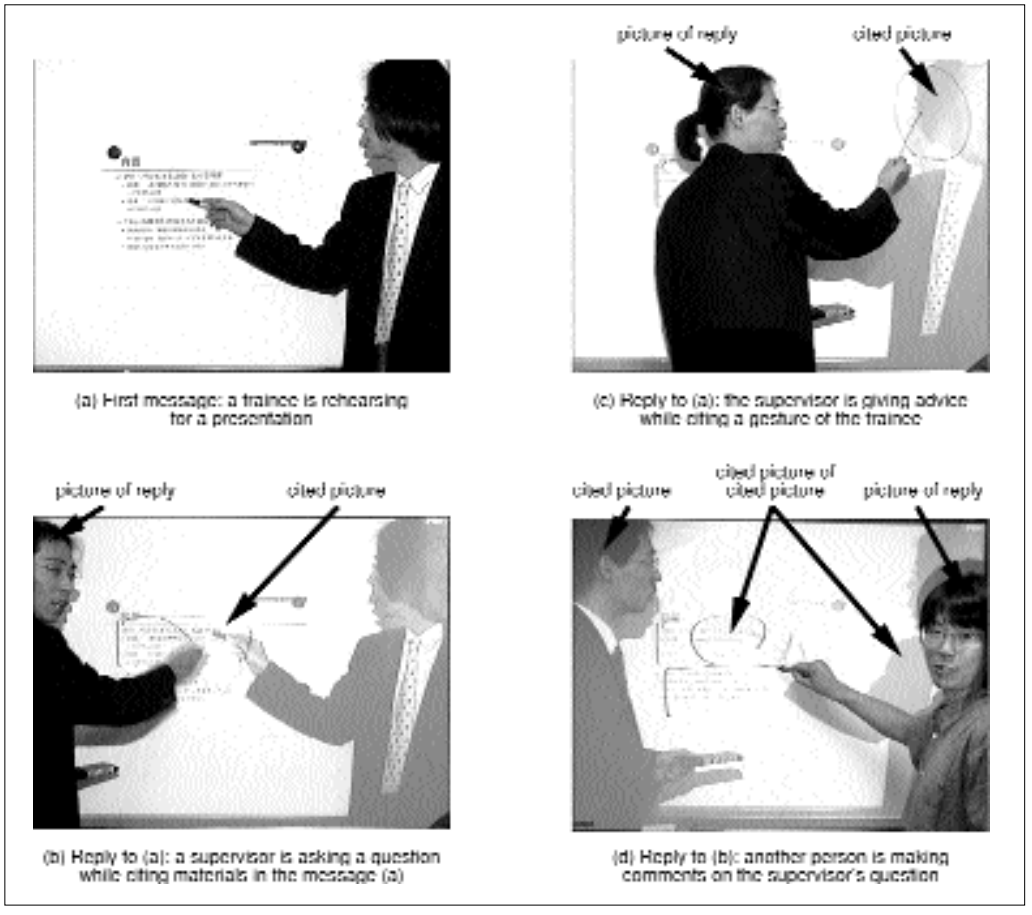


Figure 5: An example series of video messages

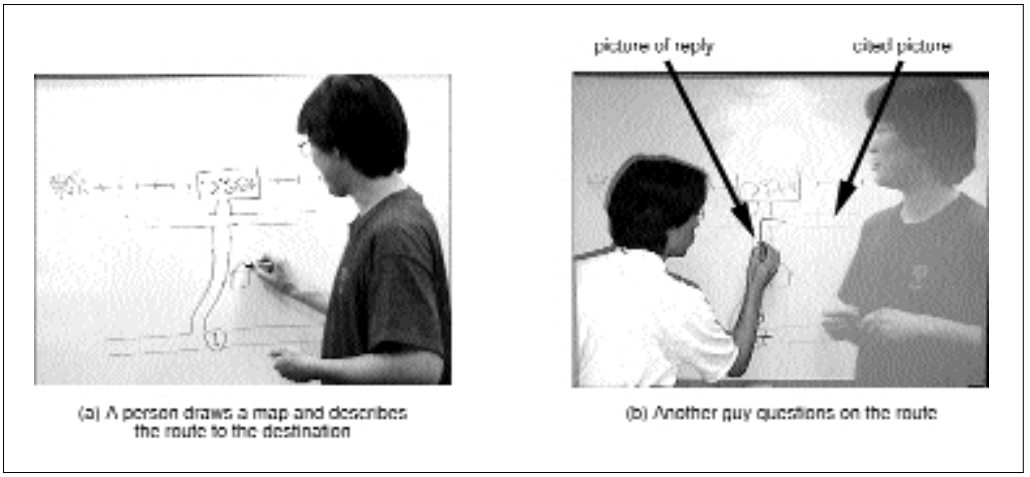


Figure 6: Another example series of video messages

Experimental results and discussion

Experiments with VideoPassage have shown that basic editor functions as described above can work using the shared board. First, in terms of spatial citing, diagrams drawn on the shared board and movements of the message sender pointing out such diagrams, for example, could be directly referenced, while corrections to these items could be indicated. Second, in terms of temporal citing, one's own message could be overlaid or inserted while playing back or freezing the received message.

In our system, playback control (like freeze, play, fast-forward, rewind, etc.) of the message to be cited was performed manually by remote control. During the course of our experiments, several cases were observed in which the responder became confused when having to make quick judgments, that is, judgments about when to jump in to add a message, or when to freeze the received message for making the reply. The importance of considering these problems in the design of a simple user interface was therefore highlighted. We should consider a user interface that would interpret gestures, speech, or drawings on the whiteboard as commands.

Furthermore, as the video to be cited and the new video to add comments were combined through simple optical overlaying, the resulting degradation in picture quality makes it possible to distinguish between cited pictures and comment pictures in the new message (See Figure 5 and 6). However, the degree of degradation in picture quality becomes progressively worse as the layer of citing deepens (it is similar to that, in E-mail citations, the number of ">" at each line increases). We have observed that, with this overlaying scheme, citation layers could be distinguished up to five layers. More than this number has been not practical due to the excessive image degradation .

To address this problem, it would be necessary to independently save the video data for each citation layer. In this scheme, we must synthesize the data for different layers by luminance keying, for example, at the time of playback. Furthermore, some video effects, such as dimming or fogging, would need to be used to let users distinguish citation levels.

Related work

Several systems employ a "shared virtual whiteboard" scheme for tele-conferencing . The difference between these systems and our citation-capable video messaging system is that the former emphasize simultaneous behavior, such as gaze awareness, while we focus on asynchronous nature of communication systems and how to surmount the time barrier.

Some tele-conferencing systems transmit the views of people's faces and a common workspace separately. This technique makes it possible to effectively convey both the sender's facial expression, gaze, etc., and drawings, pictures, and other materials for presentation to participants . However, with this technique, when attempting to provide advice or instructions for presentation training, rehearsing for a performance, and the like, the sender's face or body cannot be directly cited and resent. SharedView is an example of previous work that emphasizes the usefulness of directly pointing out something on a received video with one's finger or other means and then returning that video.

These days, major E-mail client software, or MUAs (Mail User Agents) of electronic mail systems, allow users to attach text, graphics, sound, or video clips to messages . In addition, some multimedia messaging/mailing systems have been proposed . These systems let users create, send, and view audio and video messages. However, none of these systems is citation-capable, and is focused on introducing interaction into asynchronous communication systems.

Conclusion

In this paper, we have proposed the concept of "citation-capable" video messages that introduce interaction into asynchronous video communication. The space and time barriers, namely, distance and time differences, must be overcome to facilitate our global activities. "Citation-capable" asynchronous video communication system enables us to overcome both barriers.

We have also described design and implementation of VideoPassage, a citation-capable video messaging system. It employed a "shared board" model as a unified input/output device for video messages. We found that this model successfully allowed both spatial and temporal citing of video messages.

Acknowledgments

We would like to thank our colleagues who have been involved in fruitful discussions. In particular, we thank Satoshi Kurihara and Koji Sato who volunteered experimental use of our system. We are also grateful to Toshiharu Sugawara, for constant support and encouragement in completing this research.

References

Bly, S., Harrison, S., and Irwin, S. Media spaces: bringing people together in a video, audio, and computing environment, *Commun. ACM 36*, 1 (Jan. 1993), 28-47.

[Elrod, S., Brue, R., Gold, R., Goldberg, D., Halasz, F., Janssen, W., Lee, D., McCall, K., Pedersen, E., Pier, K., Tang, J., and Welch, B. LiveBoard: a large interactive display supporting group meetings, presentations and remote collaboration, *CHI '92 Conference Proceedings*, 1992, 599-607.

Freed, N., and Borenstein, N. Multipurpose Internet mail extensions (MIME) part one: format of Internet message bodies, *RFC 2045*, 1996.

Gaver, W., Sellen, A., Heath, C., and Luff, P. One is not enough: multiple views in a media space, *INTERCHI '93 Conference Proceedings*, 1993, 335-341.

Hess, C. K., Lin, D., and Nahrstedt, K. VistaMail: an integrated multimedia mailing system, *IEEE Multimedia 5*, 4 (Oct.-Dec. 1998), 13-23.

Ishii, H., Kobayashi, T., and Grudin, J. Integration of interpersonal space and shared workspace: ClearBoard design and experiments, *ACM Trans. Inf. Syst.* 11, 4 (Oct. 1993), 349-375.

Ishii, H., and Miyake, N. Toward an open shared workspace: computer and video fusion approach of TeamWorkStation, *Commun. ACM 34*, 12 (Dec. 1991), 37-50.

Kuzuoka, H. Spatial workspace collaboration: a SharedView video support system for remote collaboration capability, *CHI '92 Conference Proceedings*, 1992, 533-540.

Morikawa, O., and Maesako, T. HyperMirror: toward pleasant-to-use video mediated communication system, *CSCW '98 Conference Proceedings*, 1998, 149-158.

[Stafford-Fraser, Q., and Robinson, P. BrightBoard _ a video augmented environment, *CHI '96 Conference Proceedings*, 1996, 134-141.

Tang, J. C., and Minneman, S. L. VideoWhiteboard: video shadows to support remote collaboration, *CHI '91 Conference Proceedings*, 1991, 315-322.

Watts, L., and Monk, A. F. Remote Assistance: a view of the work and a view of the face?, *CHI '96 Conference Companion*, 1996, 101-102.