t-Room: Next Generation Video Communication System

Keiji Hirata, Yasunori Harada, Toshihiro Takada, Shigemi Aoyagi, Yoshinari Shirai,
Naomi Yamashita, Katsuhiko Kaji, Junji Yamato, Kenji Nakazawa
NTT Communication Science Laboratories
hirata@brl.ntt.co.jp

SUMMARY

In this paper, we present t-Room, the next generation video communication system we are developing. Our approach is to build rooms with identical layouts, including walls of display panels on which users and physical or virtual objects are all shown at life-size. In this way, the user space enclosed by t-Room's surrounding displays can be shared as a common space at any other site. In other words, the enclosed spaces overlap each other. This configuration effectively provides symmetric reproduction of the audio-visual information surrounding local and remote users and objects. The feeling provided by t-Room is different from that by conventional videoconferencing systems, since there is no spatial barrier separating users such as the video screen of a conventional videoconferencing system. Furthermore, t-Room benefits in every way from Next Generation Network (NGN) technology: QoS, service productivity, and security. We view t-Room as a future form of telephone service.

INTRODUCTION

Recently, room-sized videoconferencing systems (VCS) have become available as high-end products (e.g., Polycom's RealPresence and Cisco's TelePresence) that are generally considered viable services in the Next Generation Network (NGN). The approach employed by these systems is to develop Hi-Fi audio-visual devices and high-performance networks based on the layout in Figure 1. As better system performance is achieved, users can see and hear others more clearly; this advance may lead to the ability to acquire audio-visual information on the users and surroundings and then reproduce this information in a remote room [2, 3]. Another approach to a room-sized VCS is to share a virtual room in which user images or avatars are placed [1]. Users look at a virtual room on the screens in front of them, which may give the impression that the users are all present in the virtual room.

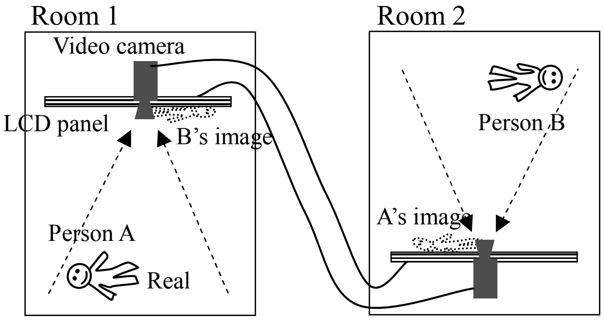

Figure 1: Display and camera layout of conventional videoconferencing system.

However, we believe it is difficult for these approaches to provide natural, efficient interaction similar to face-to-face communication [9, 8]. This is because they do not satisfy the critical property that face-to-face communication naturally satisfies: The *symmetric reproduction of audio-visual information*. This problem is known as the reciprocity of perspective [3] or presence disparity problem [10]. Achieving symmetric reproduction could allow each user to acquire audio-visual information as if he/she were actually in the same room used by remote partners.

In this paper, we present a video-mediated communication technique for properly arranging cameras and displays to facilitate symmetric reproduction of audio-visual information, and we describe a video communication system implementing the technique, our t-Room.

SYMMETRIC REPRODUCTION OF AUDIO-VISUAL INFORMATION

To achieve symmetric reproduction of audio-visual information, we use the simple approach of building rooms with identical layouts, including walls of display screens on which users and physical or virtual objects are all shown at life-size. Showing users and objects at life-size in the same positions held in a remote room allows users to immerse themselves directly into each other's physical space, and it also eliminates the need to wear any special equipment or input/output devices.

Figure 2 illustrates the basic mechanism for achieving the symmetric reproduction of audio-visual information. Since an LCD panel inherently emits polarized light, polarizing film placed over a video camera's lens enables the camera to capture the scene occurring *in front of* the opposite display without capturing the image shown *on* the display [11]. By using polarizing films, visual echo can be substantially cancelled. With the layout in Figure 2,

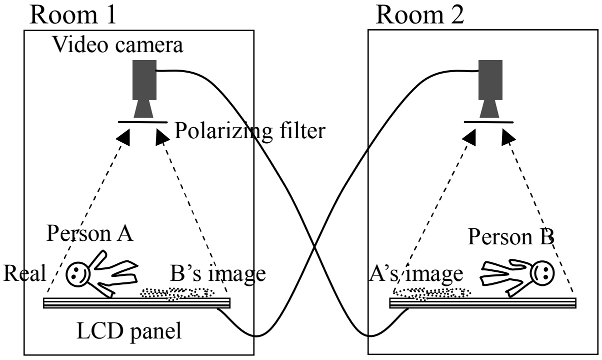

Figure 2: Basic layout for achieving symmetric reproduction.

although persons A and B are within a single display, they can recognize each other's gestures, give attention to peripheral cues, and point to real objects in either of the remote rooms. Accordingly, we use this basic layout to share audio-visual information around the surfaces of the LCD panels.

SURROUNDING-BACK-SCREEN METHOD

Since wider displays in general show a larger area for a user's life-size image, they enhance the symmetric reproduction. On the other hand, the eye contact recognition provided by the basic layout is inadequate. By arranging multiple LCD panels in a room so that they compose a cylinder (in reality, a partial polygon), we can create a shared space that enhances symmetric reproduction of audio-visual information and thus provides an environment for more satisfactory eye contact (Figure 3). The figure illustrates our method for reproducing face-to-face interaction among persons A, B, and C; we duplicate a space and project remote users' images to surrounding back screens [5].

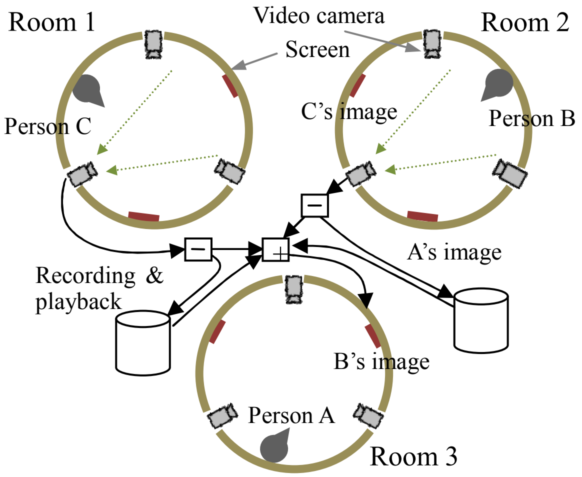

Figure 3: Duplicating space and projecting images to surrounding back screen

These screens and cameras play the same roles as the displays and cameras in Figure 2. For each room, we alternatively arrange three screens and three cameras to surround a user, who stands just in front of a screen. The preprocessing denoted by ⊟ and ⊞ in the figure is needed. The function of ⊟ is to extract only the light from real objects in front of the opposite screen and to cancel the light from the screen (visual echo canceller). That of ⊞ is to overlap or superimpose more than two images captured in Rooms 1 and 2 to correctly place images where they should be projected (overlapper). For example, images from the video camera capturing B are distributed to Rooms 1 and 3 (detailed illustration of the entire wiring is omitted for simplicity).

An advantage of this method is that it allows effective integration of recording and playback functions. This capability lets the system overcome temporal barriers and achieve asynchronous communication. For the recording and playback capabilities, the output of a visual echo canceller is stored. When later accessed, the stored data is put into an overlapper. Another advantage of this method is that it can be easily extended for connecting more than two rooms.

T-ROOM SYSTEM

To demonstrate and explore the symmetric reproduction described above, we have been working on the design of a prototype system called t-Room [5, 4, 12,13].

Hardware Configuration

Figures 4 and 5 show the hardware configuration of the current t-Room system. A single t-Room consists of eight modules (called Monoliths) arranged polygonally. With this setup, t-Room encloses a user space with surrounding displays showing life-size images. The enclosed space is shared with other enclosed spaces by the mechanism described above, and in this way the enclosed spaces overlap each other. As a result, users can freely come from and go into others' spaces, since there is no spatial barrier separating users such as the video screen in a conventional videoconferencing system. Consequently, the overlapping enclosed spaces can provide a feeling of mutual immersion.

We installed three nearly identical t-Rooms in our labs located in Atsugi City and Kyoto Prefecture (one in Atsugi and two in Kyoto). Atsugi is in the Tokyo area, and Kyoto

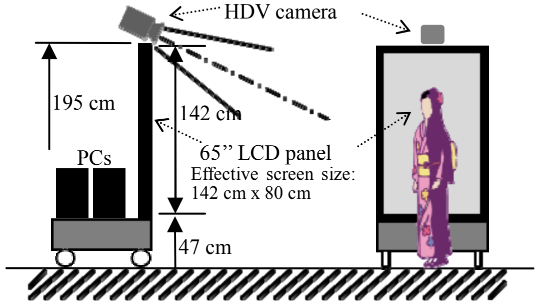

Figure 4: A "Monolith" assembly module: side view (left) and front view (right).

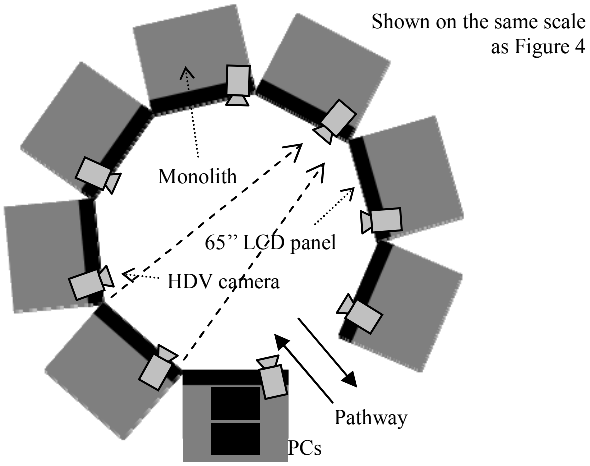

Figure 5: Top view of t-Room system composed of eight Monoliths arranged 8.5-gonally.

is approximately 400 km away from Tokyo. Currently, a commercially available 100-Mbps optical fiber line connects the three sites at Atsugi and Kyoto.

Figure 6 shows a partial view of a working t-Room system in which four people are standing and the others are displayed on the screen between them. At the local site, three spaces are overlapped: present Atsugi, present Kyoto, and the local site itself, also in Kyoto. Similarly, these spaces are overlapped at Atsugi and the other Kyoto location.

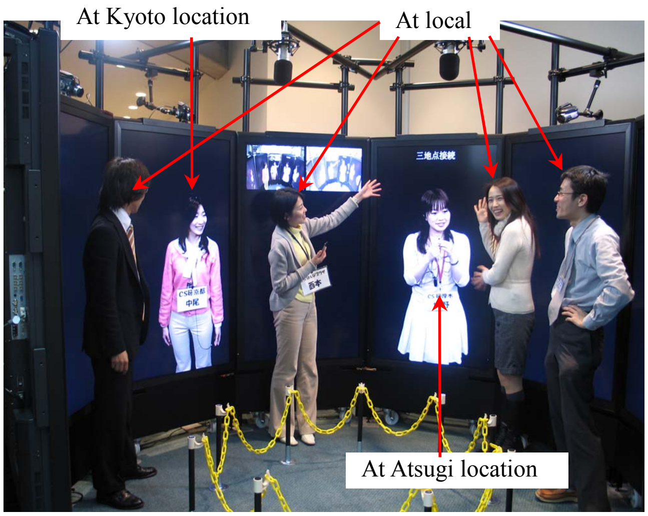

Figure 6: Demonstrating a t-Room made from eight Monoliths arranged decagonally, connecting three locations. Five central Monoliths are shown in this photo.

If t-Room overlapped a local enclosed space onto a remote site, the local users could feel as if they were actually in the remote place synchronously with the remote users; such a place might be an office, living room, classroom, restaurant, or place of scenic beauty and historic interest.

Moreover, using the recording and playback capabilities, t-Room can overlap a local enclosed space onto recorded past spaces that might include a past remote office, a past living room, and so on. Therefore, t-Room can convey a new communication feeling that is neither telepresence nor the reality evoked in conventional videoconferencing. Users do not need to use a metaphor to understand what is happening in t-Room, nor do they need to change their mental model at all. This is a reason why we believe that t-Room is a next-generation video communication system.

Specifications

A single t-Room as shown in Figures 4, 5, and 6 basically consists of eight 65-inch LCD panels (resolution of 1920 by 1080), eight HDV cameras, and four loudspeakers. In the current implementation, video data are transmitted in the Motion JPEG format over TCP/IP, and a single video camera transmits data at 1-2 M bytes per second (MB/s). Audio data are transmitted by PCM over UDP/IP, which requires a bandwidth of about 1 MB/s. Therefore, t-Room itself is considered a data stream source of max. 17 MB/s. The actual network delay for video data transmission between Atsugi and Kyoto is measured at around 0.7–0.8 seconds in normal operation.

NGN-Compliant SYSTEM

The Next Generation Network (NGN) is a packet-based network able to provide telecommunication services and to exploit multiple broadband and QoS-enabled transport technologies [6]. In NGN, service-related functions are independent from underlying transport-related technologies. NGN gives users unfettered access to networks and to competing service providers and services of their choice.

Furthermore, t-Room benefits in every way from NGN: QoS, service productivity, and security. Regarding QoS, t-Room systems consume a huge amount of network traffic, and the network traffic required for a single t-Room system increases in direct proportion to the number of overlapped spaces. In the case of Figure 6, three spaces are overlapped, and a single t-Room communicates with the other two. Thus, the Kyoto t-Room has to process data streams of max. 34 MB/s. Accordingly, when many t-Room systems become available, NGN can ensure high-performance, high-quality transfer of audio-visual information.

For service productivity, we can efficiently develop new services for t-Room based on its service-related functions. A t-Room system is regarded as a simple input/output device to shared spaces and a transparent communication medium that is not dedicated to a specific purpose; in a sense, it can be considered a “thin client.” The NGN architecture is quite suitable for the combination of t-Room as a device and the various service-related functions provided through an application server-network interface (SNI).

Concerning security, since t-Room is an open, distributed system, the issue of security threats may arise for some users. The holistic protection supplied by NGN could support efforts to make the operation of t-Room secure.

CONCLUDING REMARKS

We have shown that a simple arrangement of cameras and displays can recreate a critical property in video-mediated communication: Symmetric reproduction of audio-visual information surrounding local and remote users.

Looking at the progress of ICT to date, there has been a tendency for technologies to make contents independent from their containers. Typical examples of the content and container pair include music and CD, text and paper document, and merchandise and real store. As a result, these contents have been produced and consumed at a much higher level than before. Content is more important than container. Here, a crucial point is that a t-Room system is merely a container that provides users a video-mediated face-to-face communication experience and video-mediated social interaction [7]. We think that what we must truly disseminate is the experience and the interaction provided by t-Room, rather than t-Room itself.

An obstacle to this dissemination among potential users is obviously the price of a t-Room system. However, the prices of PCs and LCDs account for the major portion of the total price, and a t-Room system is constructed from commodity hardware. Therefore, we are confident that the system's overall price will decrease rapidly in the near future. We hope that t-Room will foster a new, fruitful communication style and be widely disseminated.

We are considering how to apply t-Room as a future form of telephone service. Based on this technology, it will be possible to provide a wide variety of communication services, establish connections even to mobile phones and PCs, and give people completely new communication experiences. We believe there are endless possibilities for this technology.

REFERENCES

1. Baker, H., Bhatti, N. T., Tanguay, D., Sobel, I., Gelb, D., Goss, M. E., MacCormick, J., Yuasa, K., Culbertson, W. B., and Malzbender, T. Computation and performance issues in Coliseum: an immersive videoconferencing system, In *Proceedings of ACM Multimedia 2003*, 470-479.

2. Bly, S. A., Harrison, S. R., and Irwin, S. MediaSpaces: Bringing People Together in a Video, Audio, and Computing Environment, *Communications of the ACM*, 36, 1 (Jan., 1993), 28-47.

3. Heath, C. and Hindmarsh, J. Configuring Action in Objects: From Mutual Space to Media Space, *Mind, Culture, and Activity*, 7(1&2), 81-104 (2000).

4. Hirata, K., Harada, Y., Takada, T., Aoyagi, S., Shirai, Y., Yamashita, N., and Yamato, J. The t-Room: Toward the Future Phone, *NTT Technical Review*, Vol. 4, No. 12, pp. 26-33 (2006).

5. Hirata, K., Harada, Y., Takada, T., Yamashita, N., Aoyagi, S., Shirai, Y., Kaji, K., Yamato, J., and Nakazawa, K. Video Communication System Supporting Spatial Cues of Mobile Users. In *Proceedings of CollabTech 2008*, IPSJ.

6. ITU-T. http://www.itu.int/rec/T-REC-Y.2001-200412-I/en

7. Norman, D.A. *Emotional Design: Why We Love (or Hate) Everyday Things*, Basic Books (2005).

8. Sellen, A. J. Speech Patterns in Video-Mediated Conversations, In *Proceedings of CHI '92*, 49-59.

9. Short, J., Williams, E., and Christie, B. *The Social Psychology of Telecommunication*, John Wiley & Sons, 1976.

10. Tang, A. and Greenberg, S., "Supporting Awareness in Mixed Presence Groupware", In *Proceedings of ACM CHI Workshop on Awareness systems: Known Results, Theory, Concepts and Future Challenges* (2005).

11. Tang, J. C. and Minneman, S. L. VideoDraw: A Video Interface for Collaborative Drawing, In *Proceedings of CHI '90*, 313-320.

12. Yamashita, N., Hirata, K., Takada, T., Harada, Y., Shirai, Y. and Aoyagi, S. Effects of Room-sized Sharing on Remote Collaboration on Physical Tasks, *IPSJ Digital Courier*, Vol. 3, pp. 788-799 (2007).

13. Yamashita N., Hirata K., Aoyagi S., Kuzuoka H., and Harada Y. Impact of Seating Positions on Group Video Communication, In *Proceedings of CSCW 2008*.