Face-to-Avatar: Augmented Face-to-Face Communication with Aerotop Telepresence System

Hiroaki Tobita
Sony Computer Science Laboratory Paris
Paris, France
tobby@csl.sony.fr

Takuya Kuzi
University of Electro-Communications
Tokyo, Japan
quzi@vogue.is.uec.ac.jp

ABSTRACT
We have developed a face-to-avatar system that integrates a blimp with a virtual avatar for a unique telepresence system. Our aerotop telepresence system has two advantages compared with conventional telepresence systems. One is to provide unique communication between user and physical blimp avatar. The blimp works as an avatar and contains several pieces of equipment, including a projector and a speaker. The user’s presence is dramatically enhanced compared to using conventional virtual avatars (e.g., CG and images) because the avatar is a physical object that can move freely in the real world. The other is that the user’s senses are augmented because the blimp detects dynamic information in the real world. For example, the camera provides the user with a special floating view, and the microphone catches a wide variety of sounds such as conversations and environmental noises. This paper describes our face-to-avatar concept and its implementation.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles, User-centered design

General Terms
Design, Experimentation, Human Factors

Keywords
Comic, Creation, Interactive System, Animation, Browsing, Communication, Visualization Technique.

1. INTRODUCTION
There are currently many systems in use for enhancing network communications. For example, in meetings, which are very important in terms of sharing information and discussing courses of action, computer graphics (CG) are frequently used to create virtual avatars or camera images and display them on a television screen. These avatars represent people who are not physically present in the meeting room, and thus the attendees can communicate with them through this avatar. This type of network communication system is useful not only in business but also for entertainment purposes such as gaming and chatting. Yet, although most network technologies have steadily improved in the last decade, telepresence systems have remained almost unchanged.

There have previously been many studies on telepresence systems for augmented reality (AR) and virtual reality (VR). However, these systems share two fundamental problems. One is a lack of versatility. Most systems restrict communication to displays and cameras that users must be physically near to. These displays and cameras are often fixed in place or cumbersome to move. Although these systems are useful in meeting rooms and classrooms, they cannot be effectively used for spontaneous meetings and chatting. The other problem is a lack of realism. A user is represented as a CG or camera image avatar, and while attendees can communicate with the virtual user and understand his or her feelings, the avatar itself does not have a realistic appearance. A sense of reality is extremely important for face-to-face communication, and we therefore believe it is necessary to find a new approach that will make avatars more realistic.

Therefore, we have developed a unique telepresence system for communications and entertainment whereby users can share visual and sound information and express their feelings and impressions more directly than when using conventional meeting systems. We mainly focused on two features to create a unique avatar—its presence in the real world and its ability to interact with people—and created a system based on a blimp. Blimps are physical, not virtual, so they can be used as avatars in the real world. We installed a projector as the output function inside the blimp so that our system can work as a display and express the user’s attributes. A camera and microphone mounted on the outside of the blimp provide the input function, which means the user can control the blimp from a distance through the network. The blimp extends the user’s senses because he or she can see through the camera and hear through the microphone. The proposed system makes unique network communications between face-to-avatars and humans a legitimate possibility. Moreover, we provide an automatic mode in which the floating blimp moves on its own. This mode is useful in areas such as media art installation.

Our main contribution is introducing a unique floating interface paradigm, named aerotop interface. The aerotop interface offers several advantages, such as being able to move through the air and above people, so the public can simultaneously share information obtained from these dirigibles. Our face-to-avatar is the application of the concept focusing on telepresence. Our face-to-avatar can interact with blimp systems both directly and indirectly because they move freely: users can move to the systems or the systems can move to the users. Our system enhances the robotic approach by floating, so it can avoid many obstacles on the ground compared with talking-head systems [9].
Figure 1: Face-to-avatar provides visitors with two experiences. One is unique communication between visitor and face-to-avatar (1). The other is that the user’s senses are augmented through video camera and microphone (2).

**Figure 2:** Overview of face-to-avatar. The system is divided into two sections: blimp and gondola. The projector and camera are in the blimp (top) and several sensors are in the gondola (bottom). The system can communicate with local PC server for contents streaming through Wi-Fi.

**Figure 3:** Camera view: Users control blimp through network and find people to start communication.

**Figure 4:** Example of Face-to-Avatar communication. Users can communicate with each other from a distance through the blimp, so face-to-blimp communication is possible.

### 2. SYSTEM OVERVIEW

The system is active in two main areas: the blimp and the gondola. The blimp is filled with helium that enables it to float through the air, which is especially useful indoors (Fig. 2 (top)). Blimps are made of vinyl chloride, so they do not burst like rubber balloons. In addition, even though they are floating devices, they are quite safe: if a blimp collides with a user, he or she will sustain no injuries. The small projector and speaker are mounted at the center of the blimp, which supports rear projection. We used a balsa-wood plate to mount the projector and camera to the blimp because balsa is extremely light. Several modules were included in the gondola, including microprocessor and embedded Linux modules (Fig. 2 (bottom)). An avatar system is able to communicate with PC server and another avatar system through the Wi-Fi modules. The PC works for media server and uses for image processing such as face detection. Both the projector and the module used 3.7-V lithium ion batteries. Our current prototype can be operated for about two hours before the projector battery runs out.

### 2.1 Face-to-avatar Communication

Although typical net-meeting systems are an easy way to communicate, the attendees of such meetings must be physically close to the system. In contrast, our avatar can move freely, so it can approach the attendees directly. Moreover, attendees listen to the avatar blimp itself rather than to a speaker next to a TV screen. The position of the audio source moves along with the blimp.

Our face-to-avatar system has two modes: manual and automatic. In manual mode, the user controls the floating blimp through the network and in the automatic mode, the floating blimp moves by itself.

#### 2.1.1 Manual Mode

Users of our system can communicate with other users through the blimp. A blimp camera provides a special floating view of the user, so the user can control the position of the avatar and attributes via a PC connected to the network (Fig. 3). The user can change the floating view by controlling the position and angle of the avatar system. The PC camera is used to display user attributes on the blimp screen, so each user can recognize individual facial attributes. The avatar can express a reaction by displaying a wide variety of attributes just like the actual attendees (e.g., neutral, happy, surprised, sad, and angry). Members in a meeting can identify the speaker’s feelings by looking at the behavior of the avatar and the facial attributes. Figure 4 (left) shows a simple example of face-to-avatar communication between a person and a blimp that is geographically separated from each another. Figure 4 (right) shows a simple communication between several people and a face-to-avatar. The avatar is close to the ceiling, so the attendees can all see it and the associated facial attributes and gestures at the same time.
2.1.2 Automatic Mode
Since the face-to-avatar contains several sensors, including a camera and ultra-sonic sensors, users do not need to manually control the blimps because they can set the controls to automatic mode. For example, the blimps can detect obstacles (e.g., humans, walls, and desks) through their ultra-sonic sensors and automatically avoid them. The system uses the results of four ultra-sonic sensors (front, left, right, and bottom) to find spaces and avoid obstacles.

2.2 Augmented Face-to-Avatar
While our avatar is floating, the approach is similar to the related works. However, it is possible to enhance the two advantages by increasing the number of communication elements. As we mentioned, both our hardware and software was simply implemented. Thus, the hardware is easy to duplicate, and the software of both server and client applications are easy to update. Next, we mention how to enhance our basic concept that is enhanced face-to-face communication.

2.2.1 Avatar-to-Avatar Communication
Our first concept is to realize avatar-to-avatar communication between two remote users. Each user can use each avatar system through the network. Figure 5 shows examples of avatar-to-avatar communications. Two remote users use each blimp avatar and communicate each other, so they can chat in the air where normal people cannot come. Thus, they can use local space effectively. In our first concept, remote user communicates with a local person through network.

2.2.2 Multi Avatars
In next example shown in Figure 6, a user users two avatars at the same time. The user can control two avatars with two different PC. It is not so difficult, because people usually multi computers for their desk or workspace. Thus, they are familiar with using multi computers. Also, the user can use one avatar with the manual mode and the other avatar with the automatic mode, so he/she can use them mixing between manual and automatic modes. For example, until a user can find a person to start communication, the user uses automatic mode.
4. DISCUSSION
In this section, we discuss our demonstration of the system through reactions from users and our observations.

The main purpose of the face-to-avatar is to achieve a unique telepresence to enhance communications and entertainment. Our concept that integrates a blimp as a physical avatar was well accepted by the visitors. We observed many users who smoothly controlled the face-to-avatar system just as they would a flying simulator. Because of the payload of the blimp, the system is initially weightless. The propeller in the gondola area can move slowly to move entire system. Also, the blimp is useful to reduce damage even if the system crashes obstacles. Thus, users could control our system easily and smoothly. We observed some people walking over to and touching the floating blimp, which highlights one of the physical advantages of the system: users can communicate with the blimps not only verbally but also physically. Besides of cost and time, our system also uses space effectively. Most of our system is helium and vinyl sheet, so users can make it compact after finishing demonstration.

About installation, we received both positive and negative reactions from the visitors. Our system looks really fragile, so visitors quickly understood the feature and touched softly. In Exhibition “Singing Balloon”, the installation was not interactive. However, our face-to-avatar system had a positive impact on them. Our system definitely attracts casual users through its novel appearance, which is significant because impact and impression are key elements for entertainment purposes. Although we used human face images in our demonstrations, we believe 2D and 3D characters from TV animations and comic books are also useful for entertainment purposes. Moreover, the system configuration is so simple that even inexperienced gallery staff could set up the system. In Siggraph demonstration, our system did not work well at first. One of the serious problems is the weight, because the payload is related with temperature. Thus, the payload is really sensitive to the environment. For example, if the temperature in the room is changed by air conditioner, the payload is also changed. Thus, we had to adjust the system configuration to the environment.

We received several questions about our prototype system, particularly relating to the face image and gestures. The image of a talking face depends on the webcam position, so the image is sometimes awkwardly cropped. However, this is a problem that conventional camera-based net-meeting systems have, too. We feel it could be partly solved by using a wide-angle camera and face detection technology. Moreover, it is possible to enhance both remote and local user’s face. Although we used face image without any effects, our system supports face recognition to the input image from the remote and local cameras, so the system can overlay virtual information onto the face area.

5. RELATED WORK
We used blimps to achieve our concept. There are several other projects that have also focused on computational blimps [1, 2, 3]. Users of autonomous light air vessels (ALAVs) [1] systems can detect obstacles because they contain sensors. Moreover, since all the blimps are connected to a network, they can communicate with one another and work as a group. Unmanned aerial vehicle (UAV) systems [2] also allow users to control blimps through portable devices, so people can use them for personal tasks.

Several systems to enhance communication through networks have been developed. The EPS system [4] is a telepresence device for real-time communication. A remote user is physically represented by the system and communicates through a video camera. In the Hydra system [5], each remote participant’s image is placed in one quadrant of the screen of a single monitor. The Porta-Person system [6] is a telepresence device that sits in a conference room or near the table. Remote users of this system can use video images or animations to represent their attributes. ClearBoard [7] is designed to integrate interpersonal space and shared workspace seamlessly. ChatScap [8] supports spontaneous meetings by using short texts and images. Facemasks are traditionally used for storytelling because they effectively show a character’s emotions and to audiences. HyperMask [9] provides a unique interface for displaying emotions. Yotaro [11] also provides several emotions by allowing users to control projected images.

6. CONCLUSION
We described our face-to-avatar system, which integrates blimps with virtual avatars to create unique communication and entertainment environments. We first discussed our prototype system focusing on the implementation and then demonstrated our current prototype for meetings. We also discussed our system based on user comments and reactions.

7. REFERENCES
[9] Talking head