Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®
Int. J. Human-Computer Studies 60 (2004) 421–454

International Journal of
Human-Computer
Studies

www.elsevier.com/locate/ijhcs

FreeWalk: a social interaction platform for group behaviour in a virtual space

Hideyuki Nakanishi*

Department of Social Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

Received 31 March 2002; received in revised form 5 September 2003; accepted 21 November 2003

Abstract

FreeWalk is a social interaction platform where people and agents can socially and spatially interact with one another. FreeWalk has evolved to support heterogeneous interaction styles including meetings, cross-cultural encounters, and evacuation drills. Each of them is usually supported by an individual virtual environment. This evolution extended the capability to control social interaction. The first prototype only provides people with an environment in which they can gather to talk with one another while the third prototype provides them with a whole situation to behave according to their assigned roles and tasks. FreeWalk1 is a spatial videoconferencing system. In this system, the positions of participants make spontaneous simultaneous conversations possible. Spatial movements are integrated with video-mediated communication. FreeWalk1 is able to make social interaction more casual and relaxed than telephone-like telecommunication media. In contrast to conventional videoconferencing systems, people formed concurrent multiple groups to greet and chat with others. In FreeWalk2, a social agent acts as an in-between of people to reduce the problem of the low social context in virtual spaces. When the agent notes an awkward pause in a conversation, it approaches those involved in the conversation with a suggestion for a new topic to talk about. We used this agent to support cross-cultural communication between Japan and US. Our agent strongly influenced people's impressions of their partners, and also, their stereotypes about their partner's nationality. FreeWalk3 is a virtual city simulator to conduct virtual evacuation drills. This system brings social interaction into crisis management simulation. People can join a virtual scene of a disaster at home. Social agents can also join to play their roles assigned by simulation designers. The system architecture has a split control interface to divide control of multiple agents into high-level instruction for them and simulation of their low-level actions. The interface helps simulation designers to control many agents efficiently.

*Corresponding author. Tel.: +81-75-753-4859; fax: +81-75-753-4820.
E-mail address: nuka@kuis.kyoto-u.ac.jp (H. Nakanishi).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2003.11.003

1. Introduction

Some virtual environments focus on communication support (Greenhalgh and Benford, 1995) while others focus on simulation (Macedonia et al., 1994). In the studies on the former type of virtual environments, awareness control and nonverbal communication support have been extensively discussed. In the studies on the latter type, the important topic has been how to control many distributed objects efficiently. To integrate these two research literatures, we have been developing a virtual space 'FreeWalk.' The system has evolved to support heterogeneous interaction styles including meetings, cross-cultural encounters, and evacuation drills.

FreeWalk1 is the first prototype and is designed to support casual meetings. The system provides only a virtual space where remote users can gather to talk with one another. The second prototype FreeWalk2 has the functionality to control to autonomous characters to conduct a social psychological experiment as well as facilitate casual meetings. FreeWalk3 is the latest prototype and simulates group behaviour including social interaction. People enter into an environment provided by this system to play a role assigned by the system.

We developed FreeWalk1 to support casual meetings. Casual meetings such as chatting during a coffee break or in a hallway occur daily. They maintain human relationships, and also play an important role in collaboration. Conventional videoconferencing systems, which simply transmit videos and voices like video-phone, cannot support casual meetings. FreeWalk1 allows people to change their positions dynamically during the meetings. Many systems use a 3D virtual space as a multiuser environment (Sugawara et al., 1994; Hagsand, 1996; Han and Smith, 1996; Waters and Barrus, 1997). Those systems usually aim to construct brilliant virtual worlds. The role of a 3D virtual space in FreeWalk1 is different from those systems, and similar to the spatial model of interaction in 'Massive' (Greenhalgh and Benford, 1995). However, the spatial position is much more tightly combined with video-mediated communication in FreeWalk1. Since earlier studies tried to compare face-to-face (FTF) communication with video-mediated communication (O'Conaill et al., 1993), various characteristics of conventional video communication became clear. However, the characteristics of the communication aided by a 3D virtual space remained unclear. We conducted an experiment to find the characteristics of 3D communication in FreeWalk1.

The second approach to facilitate casual meetings is employing an autonomous social character as a coordinator of meetings. Virtual spaces make it easy to have casual meetings between strangers from across town, or even across the world. Consequently, people have a difficulty in starting and advancing their conversation when they do not know what they should talk about. To diminish this difficulty, we developed a social agent that is a character in a virtual space FreeWalk2 (Isbister et al., 2000; Nakanishi et al., 2003). Our agent acts as a coordinator to maintain social interaction among people. In an experiment, the agent could help cross-cultural communication between Japanese and US students and influence their communication and their impressions of their partners.

The above approaches do not provide social context directly to people in a virtual space. FreeWalk3 provides a virtual scene of a disaster where many remote users are assigned their role and participate in an evacuation drill. The conventional crisis simulations ignore effects of social interaction. FreeWalk3 brings social interaction into crisis simulations. The differences of FreeWalk3 from conventional multiuser virtual environments are: (1) hundreds of agents participating in social interaction; (2) and the virtual space imitating the real space.

2. Evolutionary approach to the social interaction platform

FreeWalk has evolved toward a general-use platform for social interaction. FreeWalk1 is just a spatial videoconferencing system that enables social interaction among remote users. It was designed as a provider of communication environments where only human participants can gather to talk with one another. To FreeWalk1, FreeWalk2 added functionality of social agents that are autonomous characters as a part of participants. Social agents were designed to facilitate human participants to form their social context to communicate with each other. FreeWalk3, furthermore, equipped FreeWalk2 with a virtual city simulator that simulates a real-world city. It was designed to provide top-down social contexts for human participants and social agents. As described, the design of FreeWalk evolved and its function was extended along the direction in which the determinant of social interaction is transferred from human participants to virtual environments.

Social contexts, participants, and environments are major factors of social interaction. First, social interaction occurred based on its context. Next, social context is determined by identities and behaviours of participants in the social interaction. Finally, every participant exists in the environment shared by all the participants. The summary of these relations is that social interaction is controlled by its context, social context is controlled by its participants, and participants are controlled by the environment shared by themselves.

Virtual environments including FreeWalk can be classified based on the ratio of their focus on communication support to their focus on simulation. 3D-chat space and FreeWalk1 are examples of virtual environments that focus on communication support. Military training simulator (Macedonia et al., 1994) and FreeWalk3 are examples that focus on simulation. Pedagogical virtual environments and FreeWalk2 are intermediate between them. Criterion of this classification is how strongly virtual environments control social interaction. In the design of virtual environments focusing on communication support, every participant is a human user and only human participants form social context while, in the simulation-oriented design, virtual environments provide a social context and also a part of participants. The classification corresponds to how many factors of social interaction virtual environments can control.

This classification is associated with how much social interaction is controlled by users who are not the participants of it. In pedagogical virtual environments, social agents play roles of instructors. In FreeWalk2, social agents helped cross-cultural

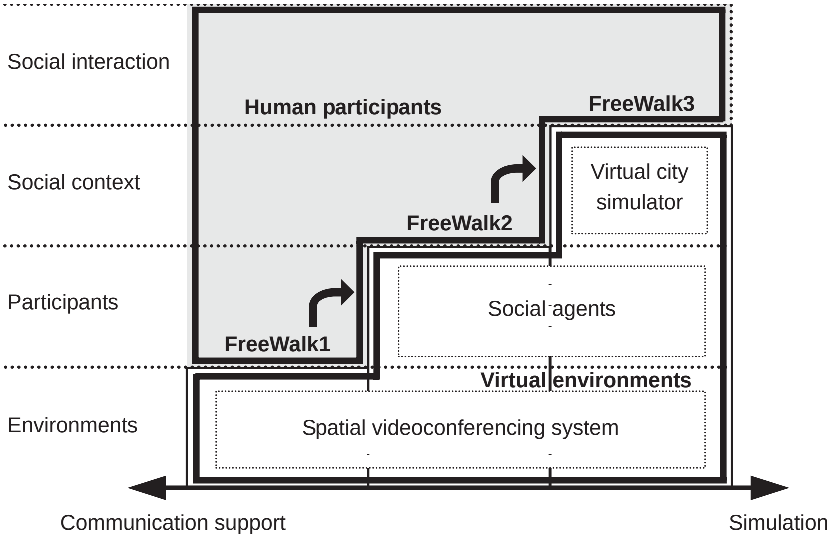

Fig. 1. Evolutionary design of FreeWalk.

communication of subjects in an experiment. In these virtual environments, designers of the social agents influence social interaction indirectly. In military training simulator and FreeWalk3, trainers design the details of training, which human participants and social agents follow. Consequently, designers of simulation influence social interaction directly.

In Fig. 1, each evolutionary stage of FreeWalk is classified. FreeWalk has evolved toward simulation-oriented design to control more degree of social interaction. Functional extension in the process of this evolution is presented together in the figure.

3. Freewalk1: spatial videoconferencing system

FreeWalk1 is a 3D virtual space as a social communication environment to facilitate casual meetings (Nakanishi et al., 1996, 1999). FreeWalk1 lets people meet casually in virtual spaces such as a park or a lobby. FreeWalk is just like a videoconferencing system that assigns a location and a view direction to a video window of each user. Users control their video windows in a virtual space.

In conventional videoconferencing systems, participants turn on the system when they start a meeting. The system displays the videos of all participants on their screens, which hinders free conversation. The system also lists the participants before the meeting starts, thereby prohibiting accidental encounters with other participants. Several videoconferencing systems have tried to extend their functions to support spontaneous conversations. 'Cruiser' developed by Root (1988) randomly selects some of the participants and displays their videos to other participants to simulate accidental encounters in a hallway. In contrast, FreeWalk1's approach provides a common virtual space for spontaneous conversations wherein participants can move and meet by themselves. It does not promote any system-directed encounters. The participants' videos are displayed on screens only when their embodiments meet.

In meetings such as parties, several tens of participants simultaneously exist in the same space. People gather physically to maximize the possibility of having spontaneous conversations. In the situation like that, it is almost impossible to use videoconferencing systems, since they try to display the videos of all participants at once. Plus, even if it were possible, it would be very hard for users to comprehend the situation. It is very troublesome for users to choose whom they observe. Furthermore, they may have to adjust each level of awareness such as the size of the video image, for each person they observe. Even if users do not mind doing that, it is almost impossible to have symmetrical awareness among participants. 'Vrooms' developed by Borning and Travers (1991) has virtual rooms, one of which is the window containing video images of participants. This design can divide all participants into several subgroups. Sharing the same room with others means that the user is observed by them as well as he/she can observe them. In this system, it is easy for users to form conversation groups. However, the user in a virtual room cannot observe other users in another room. In FreeWalk1, participants can freely change their locations and view directions to watch other participants in the virtual space. They can walk around before they talk to someone else. They do not need to see the videos of all participants at once. They can choose whom they observe and listen to with keeping symmetrical awareness among them in a simple manner.

3.1. *Design of social interaction*

This section describes the design of a communication environment for casual meetings. FreeWalk1 is a virtual meeting space where everyone can meet and talk with each other. In FreeWalk1, video and audio channels transmit nonverbal information such as facial expressions and paralanguage in the same way of videoconferencing systems, and spatial positions of participants make accidental simultaneous conversations possible. The spatial model is tightly combined with the video-mediated communication. The meeting style is more fluid and more dynamic than that in videoconferencing systems.

3.1.1. *Social interaction in a three-dimensional space*

Fig. 2 shows an image of a FreeWalk1 screen. FreeWalk1 provides a three-dimensional (3D) virtual space where people can meet. Participants move and turn freely in the space using their mouse. In this 3D space, a pyramid of 3D polygons represents each participant. The system maps live video of each participant on one rectangular plane of the pyramid, and the participant's viewpoint lies at the centre of this rectangle. The view of the space from a participant's particular viewpoint appears in the FreeWalk1 screen. Participants standing far away in the 3D environment appear smaller and those closer appear larger. FreeWalk1 does not display participants located beyond a predefined distance. The system also transfers voices under the same policy—that is, voice volume changes in proportion to the

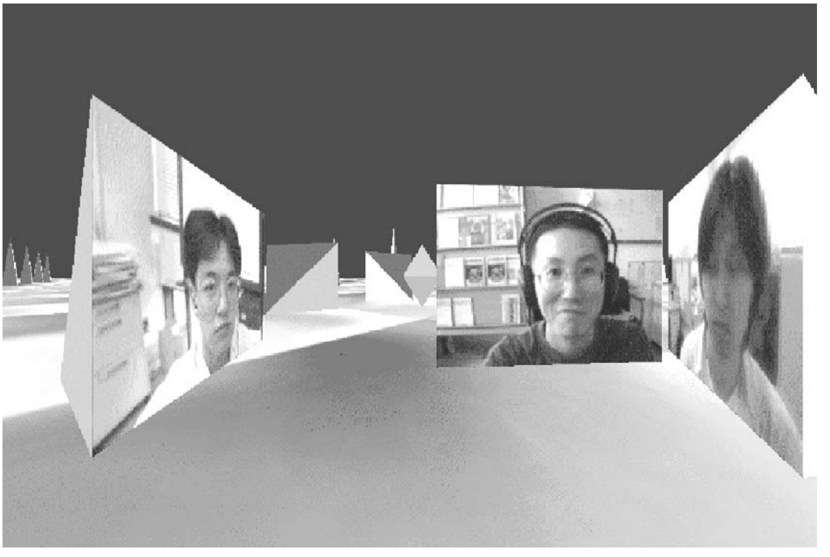

Fig. 2. FreeWalk: three-dimensional virtual meeting space.

distance between sender and receiver. Moreover, a participant hears others' voices in stereo so that he/she can easily recognize the speaker.

In FreeWalk1, meetings can start with an accidental encounter. Fig. 3 shows an example of an accidental encounter, where the user finds others on the radar screen displayed at the right bottom corner of the screen (see Fig. 3a), watches them to find out what they are talking about (see Fig. 3b), then joins them (see Fig. 3c). Since distance attenuates voice, a participant must approach the others in order to talk to them. On the other hand, not only can the participants in the conversation hear the speaker's voice, but also anyone in the neighbourhood can listen. This mechanism forces people to combine actions and conversations in the space. People can smoothly join the conversation that attracts their interest, since they can guess the topic by listening to the conversation beforehand. People can exit a conversation by leaving a group and join a conversation by approaching another group.

Desktop videoconferencing systems provide various functions to support the organizational behaviour of participants, such as speaker selection. Although these functions let participants manage multiple conversation threads in parallel, they also damage the freedom we are aiming for. FreeWalk1 does not take this approach. Instead, it uses a common 3D space that promotes a casual feeling in communication. People form a group by standing close to each other to engage in conversation. Fig. 4 shows this situation. Since voice volume attenuates in proportion to the distance, people can have a confidential conversation by keeping away from others. If groups have enough distance between them, people in one group cannot hear people in other groups. Therefore, participants can form separate meeting groups and not bother each other. This feature makes FreeWalk1 an effective tool for holding a party.

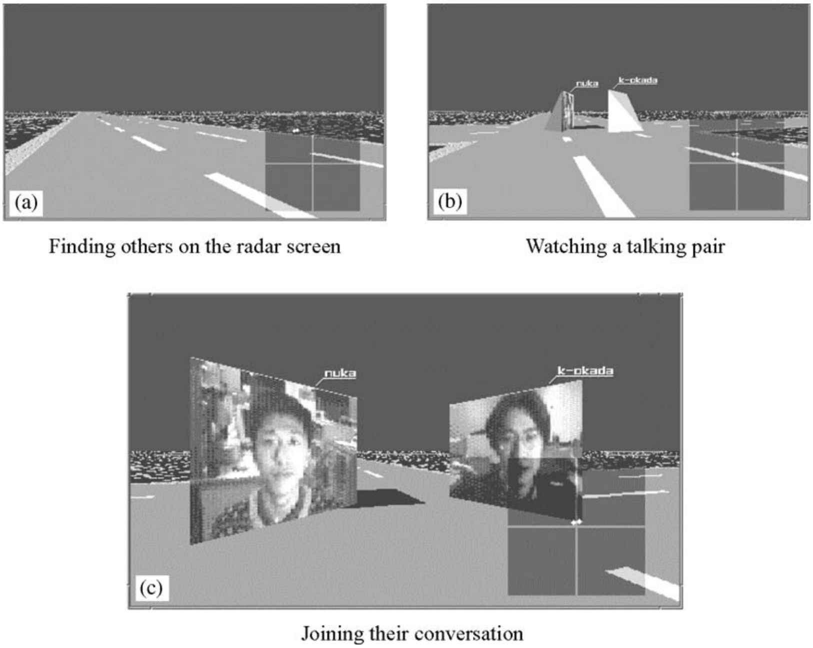

Fig. 3. Accidental encounter. (a) Finding others on the radar screen. (b) Watching a talking pair. (c) Joining their conversation.

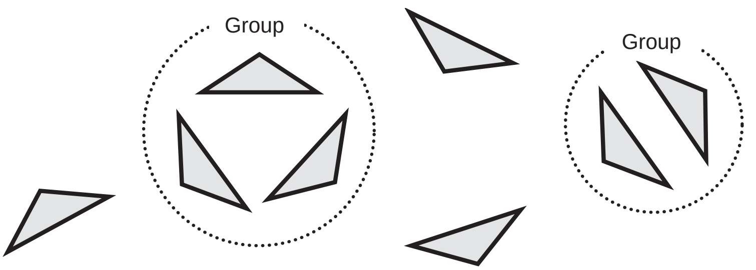

Fig. 4. Meeting organization.

3.1.2. Spatial nonverbal signals

Space provides a context for interaction (Duck, 1998). Spatial positioning is a nonverbal cue, which serves to communicate liking and disliking and attraction to a relationship. The orientation of a body and eye contact are used to start, sustain, and end interactions. Interaction is controlled by a behaviour that is changing body orientation, and communication becomes easy to do if this behaviour is judged correctly (Cranach, 1971). The behaviour that is turning directions of eyes, a head and a body is based on the structure of a human body. Turning behaviour reflects emotional attitude toward others. These nonverbal signals are important in casual meetings since they smooth out and regulate social behaviour.

In the virtual space of FreeWalk1, users can partially use these signals. The orientation of a pyramid represents the body orientation of a user. Since the participants' locations and view directions reflect a pyramid orientation, each participant can grasp the locations and view directions of other participants, and observe what other people are doing from a distance. Participants can also observe others around them by turning their body.

Since the volume and direction of one's voice is determined by its position, participants can easily identify the others in the same conversation group. Comprehending the correspondence between each face and each voice is not as difficult as that in conventional videoconferencing systems. The speaker can assume that the listeners recognize who is speaking. Furthermore, the speaker can turn to face the person he/she is talking to. These natures may transmit nonverbal information that makes it unnecessary to call the name of the person whom you talk to before you begin speaking.

3.1.3. Transmission control based on spatial locations

The FreeWalk1 system consists of a server and clients, each of which has vision and voice processes. Fig. 5 illustrates the interaction between the server and clients. When participants move in the 3D space using their mouse, the corresponding client calculates the new location and orientation, and sends them to the server. The server then compiles this information into a list of client locations in the space. The server

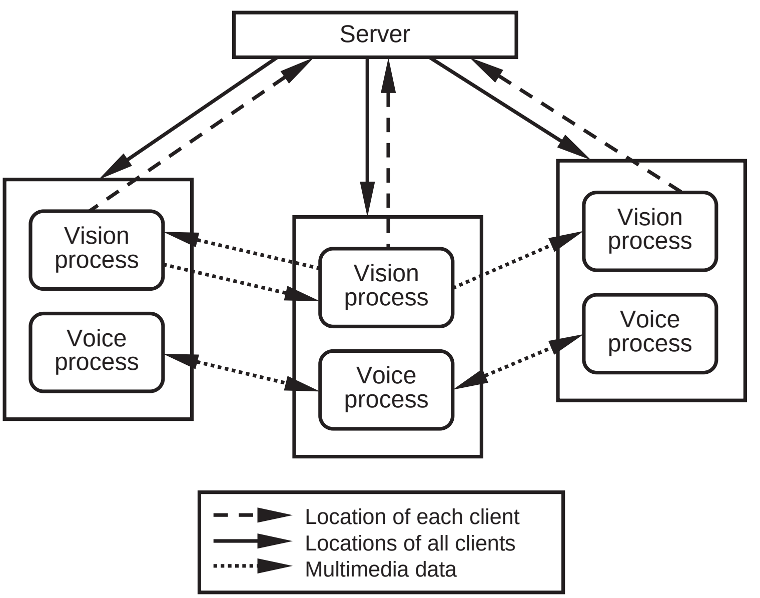

Fig. 5. FreeWalk system architecture.

finally sends the list back to each client for screen updating. Since only control information is transmitted between the server and the clients, the server can efficiently maintain ongoing activities in the space.

When a client receives the list of other clients, the client's vision system sends its owner's picture to the other clients. On receiving pictures from other clients, the vision system redraws the display based on the information in the list and the received pictures. Since each client cannot see all the other clients, it is not necessary for each one to send its picture to all the others. Similarly, each client does not have to send full-size pictures to clients far away. FreeWalk1 uses these facts to optimize the bandwidth of video communication as follows: (1) the sender adjusts the size of the picture to the size the receiver needs; (2) the client sends its picture to others who can see the client. Fig. 6 shows an example of a video transmission in FreeWalk1. Since client A lies near client C, client C sends a large picture to client A. In contrast, client C sends a small picture to client B, because it is located far away. Voice communication occurs in the same manner. Clients do not send voice data to those clients located too far away to hear the participants' voices. The locations and view directions of participants in the space determine which pictures and voices get transmitted in FreeWalk1.

3.2. Evaluation of the virtual space

In this section we show the characteristics of 3D communication compared with FTF and conventional video-mediated communications.

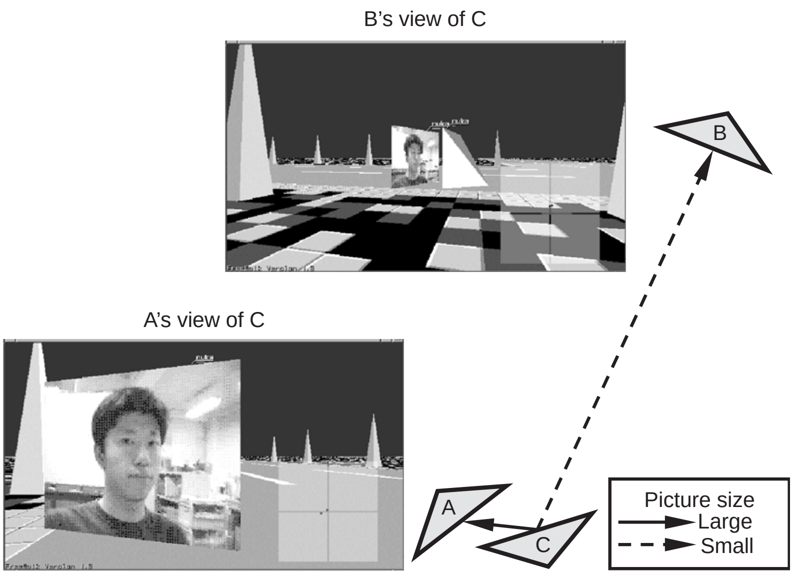

Fig. 6. Video transfer among clients.

There are many functional differences between 3D virtual meeting spaces and conventional videoconferencing systems. We compare these two kinds of communication environments to show the inherent advantages of virtual meeting spaces for casual meetings. We took Silicon Graphics' InPerson (see http://www.sgi.com/software/inperson/) as the example of a conventional videoconferencing system and FreeWalk1 as the example of a virtual meeting space. Table 1 shows the functional differences between these two environments.

Process of joining: In FreeWalk, the process of joining a meeting is just to enter the 3D virtual space. The protocol of InPerson inherits that of telephones: one calls the other to hold a meeting. If one wants to join the multiparty meeting, he/she needs to be called by one of the meeting participants. Newcomers cannot join the meeting freely.

Maximum number of participants: The maximum number of participants is seven in InPerson because of the limited size of a screen. FreeWalk does not limit the number of participants, though if the number increases too much, the performance of the system becomes intolerable.

Occurrence of conversation: In FreeWalk, a conversation may be started by an accidental encounter while the participants are walking around the 3D virtual space. A conversation is started by participants' contact of their own accord. In InPerson, conversation is started when the coordinator of a meeting turns on the system and contacts all participants.

Meeting group: In FreeWalk, participants approach one another to organize a meeting group. Participants can form multiple meeting groups simultaneously. In InPerson, however, participants always form a single meeting group since everyone faces the others and hears the voices of the others.

3.2.1. Experiment to evaluate FreeWalk1

We believe that 3D environments are more effective for casual communication than conventional video environment as follows:

(1) Participants using a conventional videoconferencing system tend to be strained and their conversations do not proceed smoothly. This is because all their faces are always displayed and the system keeps everyone facing the others. A 3D virtual space eliminates this strain by giving them locations and view directions.

Table 1
Functional comparison

	FreeWalk	InPerson
Process of joining	Enter a 3D virtual space voluntarily	Called by someone who has already joined
Maximum number of participants	Unspecified (practically, 20 or so)	7
Occurrence of conversation	Caused by participants' approach of their own accord	Caused by turning on the system by a coordinator
Meeting group	Multiple groups	Single group

(2) It is impossible to reproduce communication with moves like real life communication in a conventional video environment. A 3D virtual space reproduces communication with moves by enabling participants to move freely.

Sellen (1995) compared communication in two video conferencing systems, 'Hydra' and 'Picture-in-a-Picture (PIP),' and in the FTF environment. She found no differences among the three environments for conversation in terms of turns (transferring the initiative of speech), even though previous studies showed that more turns occurred in the FTF environment than in the videoconferencing environment. We expected that the number of turns might increase in casual meetings, so we analysed the number of turns in our experiment. Bowers et al. (1996) studied how the movement of avatars coordinated with conversation in a virtual environment. The results showed that the avatars' moves transferred the initiative of conversation. In 3D and FTF environments, the moves of people relate to their communication skills. In our experiment, we analysed the moves of people in meetings. Additionally, we counted the number of occurrences of chat and calculated the standard deviation of utterance. We thought a casual atmosphere might stimulate the occurrence of chat and change the amount of utterance of each participant.

Twenty-one undergraduate students participated in our 1-day experiment. We prepared three environments for conversation to compare FTF, conventional video, and 3D communications (see Fig. 7). We set up seven workstations connected by

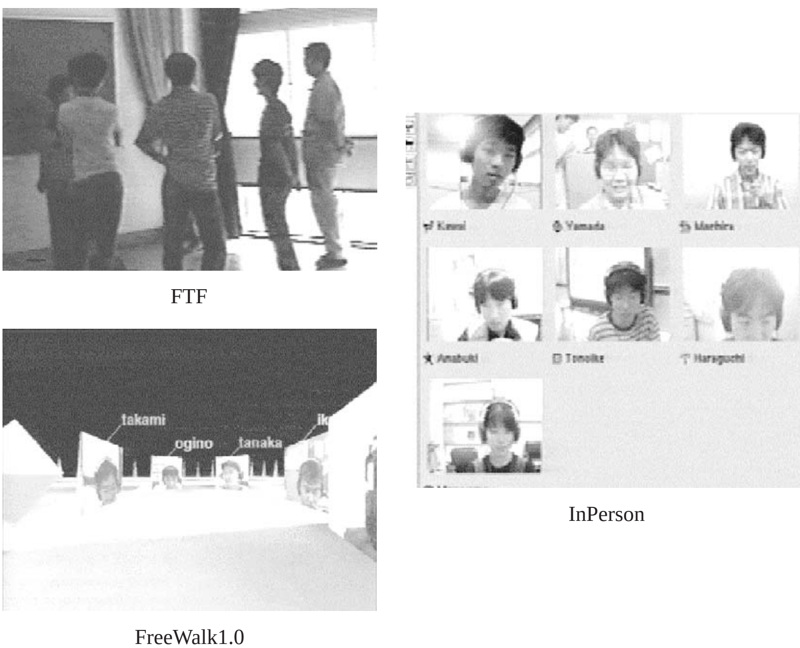

Fig. 7. Three different environments for conversation.

LAN for the video environment (InPerson) and the 3D environment (FreeWalk1). The meetings in the three environments consisted of three tasks. The first task (Task 1) is agreeing on a group travel destination. This was a decision-making task. We made the participants decide where they would travel a month later. The second (Task 2) is discussing social problems. This task was to shape ideas. The third (Task 3) is conversing freely. Participants had conversation without any guidelines. We chose these tasks to examine various types of communication comprehensively. For each task, we told participants to organize three groups of seven people. Thus nine types of meetings took place. Each meeting lasted for twenty minutes. We did not choose any chairpersons of the meetings in advance. Before performing the three tasks, the participants introduced themselves in each group so that they could memorize each other's faces and voices. They also practiced operating FreeWalk1. The independent variables of this experiment were the differences between the environments and the tasks.

We collected experimental data using videotape recordings. During the FreeWalk and InPerson meetings, we recorded the screen images of the workstations on videotape recorders. In FTF meetings, we recorded the scenes by video cameras. We reviewed the videotape pictures to record the start and end times of participants' utterances to create conversation records. In addition, we collected the logs of FreeWalk1 to find the pattern of moves in the 3D virtual space during meetings. The server stores logs in which it records locations and orientations of participants in the space. We developed a tool called 'SimWalk' to analyse participants' moves. SimWalk draws lines along the participants' moves and connecting their locations in sequence. It also can reproduce participants' moves by drawing animated triangles, each of which represents the location and view direction of each participant.

3.2.2. Results of the experiment

We present the analysis results of participants' conversations and moves. We organized the results of the conversations into number of turns, standard deviation of utterance, and occurrence of chat.

- Number of turns

  This value represents the number of events. Each event transfers the initiative of talking from a person to another. The turn occurs when someone starts talking immediately after or while another talks. We did not count cases in which someone stopped talking and started talking again after a brief silence.

  Fig. 8 shows the relation between the frequency of turns and environments. The frequency of turns equals the number of turns divided by the amount of utterances. The rankings of contributions of environments to the number of turns is characterized as:

  FreeWalk > FTFInPerson.

  The effect of the difference in environments showed that FreeWalk activated turns more often than InPerson and FTF.

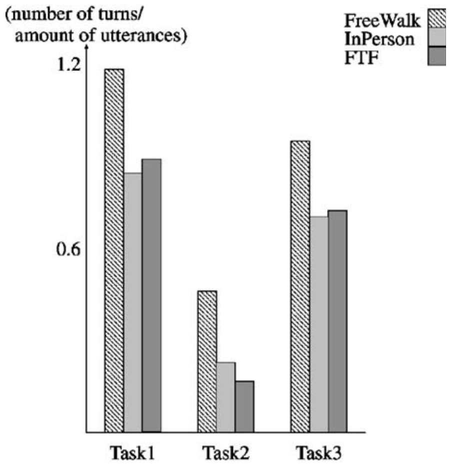

Fig. 8. Frequency of turns.

Table 2
Standard deviation of utterance

	Task 1	Task 2	Task 3
FTF	13.93	19.19	14.07
InPerson	12.31	15.97	17.25
FreeWalk	9.28	15.45	13.45

- Standard deviation of utterance

  This value represents the standard deviation of the ratio of the total time of utterances of each participant to the total time of all utterances of all participants. Table 2 summarizes the standard deviations of utterance. It provides the following ranking of environments for each task:

  Task 1 FTF > InPerson > FreeWalk,

  Task 2 FTF > InPersonFreeWalk,

  Task 3 InPerson > FTFFreeWalk.

  Interestingly, the deviation remained the smallest in FreeWalk for all tasks. This means that the number of utterances of each participant became equalized in FreeWalk.
- Occurrence of chat

  This value represents starting a conversation that does not contribute to accomplishing the task. Fig. 9 shows the occurrence of chat in Tasks 1 and 2 in each environment. In this figure, the horizontal axis represents time, and each mark represents the occurrence of chat. You can see that chat occurred more

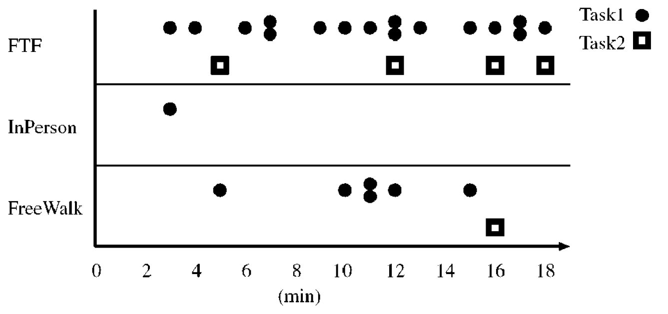

Fig. 9. Occurrence of chat.

actively in FTF than in FreeWalk, while it seldom occurred in InPerson. The rankings of the contributions of environments to the occurrence of chat follow:

FTF > FreeWalk > InPerson.

In FreeWalk, the atmosphere among participants might have been relaxed since they formed a circle to have a conversation, while in InPerson everyone faced the others.

- Participants' moves

  In FTF meetings, participants seldom moved after forming a circle to have a conversation. During InPerson meetings, everyone faced the others on the screen. Fig. 10 shows participants' moves during a 15-min period in each FreeWalk meeting. In Tasks 1 and 2, they seldom moved after forming a circle as in FTF. Unlike the other two tasks, they moved actively around the 3D virtual space in Task 3.

In Task 3—free conversation—we observed the following behaviours. At the beginning of the task, participants moved actively. For example, they rushed toward others. The occurrence of conversation was scarce. Fig. 11a presents snapshots of SimWalk, which reproduces the participant's moves. In the middle of the task, participants faced one another frequently to greet. The lengths of conversations were short. We noted that some participants blamed others for approaching them when they tried to whisper to each other. You can see the participants greeting each other in Fig. 11b. Toward the end of the task, all participants gathered to converse. We noted that a certain participant ran about trying to escape from the meeting place since he was unwilling to talk, while another participant looked for someone else who had gone elsewhere. This situation is represented in Fig. 11c.

As a result of our analysis, we categorized the effects of a 3D virtual space into two types. In the first type, we observed that 3D communication resembles FTF communication. Two primary characteristics exist: frequency of chat and behaviour of participants. The second category, however, remains peculiar to 3D virtual spaces. These environments equalize the amount of utterances for each participant more than the other environments, increase the number of turns, and sometimes stimulate

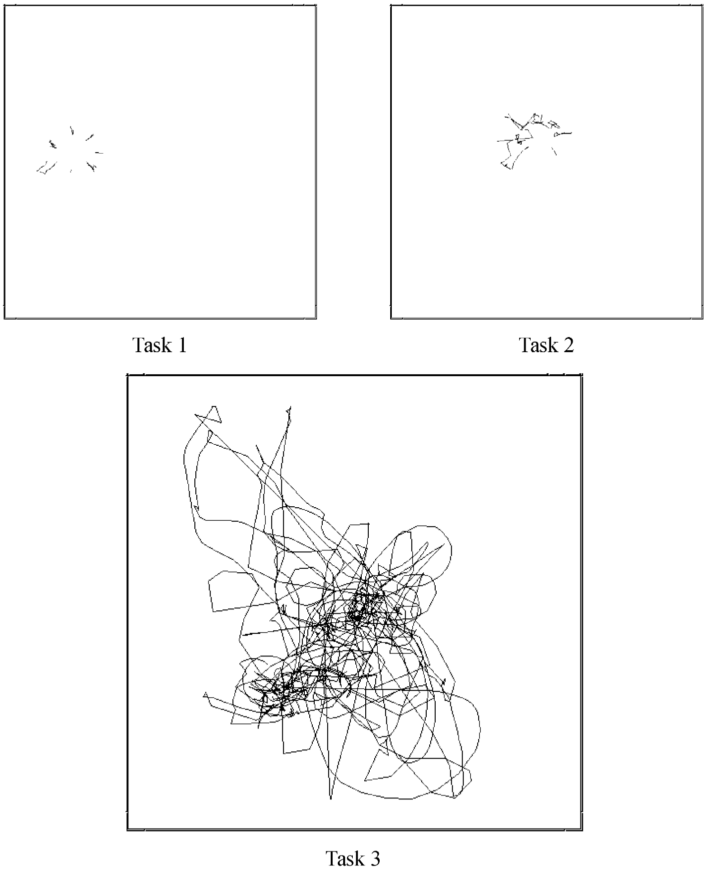

Fig. 10. Pattern of moves in a 3D virtual space.

participants to move around to converse freely. These results show the effectiveness of a 3D virtual space in casual meetings. The freedom of 3D virtual space lets participants enjoy their conversation, and its relaxed atmosphere stimulates participants into initiating conversations. On the other hand, participants in a 3D virtual space tend to concentrate less than participants in the other environments.

The difference of participants' moves in the three FreeWalk meetings gives us important implications about the advantages of virtual spaces. In Tasks 1 and 2, participants seldom moved after forming a circle, since they have common topics to discuss together. In this case, virtual spaces have no advantage except assigning positions, which provide directional voices and videos. Such kind of meetings as Task 1 or Task 2 can be held in conventional videoconferencing systems. However, participants formed multiple groups at once to greet and chat with others in Task 3, since they did not have any common topics to discuss. They cannot have similar conversations in conventional videoconferencing systems. Therefore, virtual spaces

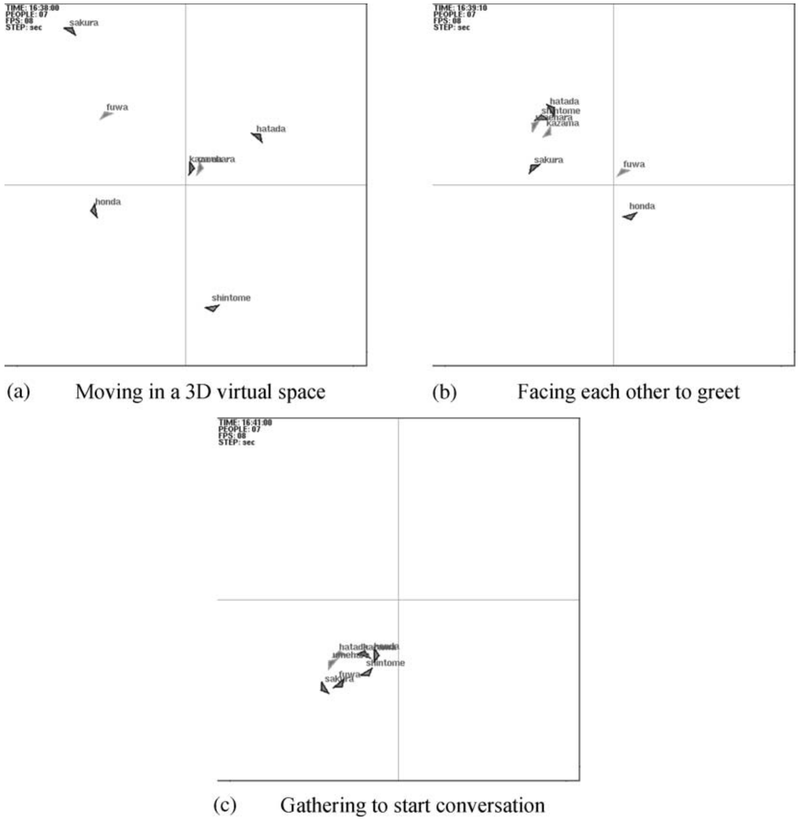

Fig. 11. Participants' moves in Task 3. (a) Moving in a 3D virtual space. (b) Facing each other to greet. (c) Gathering to start conversation.

have an incomparable advantage in free conversations. Virtual spaces can support the special style of casual communication as well as make video-mediated communication more casual.

4. FreeWalk2: social agents

As we described in the previous section, virtual meeting space enables people to meet accidentally and to have multiple conversations simultaneously. There are many virtual meeting spaces such as Community Place (Lea et al., 1997), InterSpace (Sugawara et al., 1994), Diamond Park (Waters and Barrus, 1997), DIVE (Hagsand, 1996), Massive (Greenhalgh and Benford, 1995) and CU-SeeMe VR (Han and Smith, 1996). Virtual meeting spaces make it easy to have casual meetings between strangers from across town, or even across the world. Unfortunately, virtual meeting spaces usually provide little socially meaningful context to use as a basis for finding common ground with each other. Since it is easy to arrive at a virtual meeting space from many entry points, it is often hard for visitors to assume much about one another's cultural backgrounds, group memberships, and other aspects of social identity. People need this sort of common context in order to build new human relationships (Clark, 1996).

We believe software agents could provide ongoing, in-context help in forming social relationships and building common ground between visitors to virtual meeting spaces. Software agents that communicate with human users in virtual worlds have been emerging. Some of them provide social services such as conversation in text-chat worlds, and these agents are becoming vital inhabitants of virtual worlds by playing a key role in forming on-line communities (Kim, 2000). We call these software agents social agents. In contrast to interface agents that support human–computer interaction, social agents support human–human interaction. We developed a social agent playing a role of party host in a virtual space. This agent was applied to support cross-cultural communication in our experiment.

4.1. Social agents

Previous studies have discussed and demonstrated some benefits of interface agents in one-on-one task settings, such as taking an educational tutorial, going on a tour (Isbister and Doyle, 1999), or looking at real estate. Lester et al. (1997) notes that the presence of an agent can lead to a strong positive effect on students' perception of their learning experience. Cassell et al. (1999) discusses the value of an embodied real estate agent with the proper human verbal and nonverbal communication skills. However, these findings concern task-support agents interacting with a single user.

There are projects, which have created agent-based social support through text-based conversation. Julia (Foner, 1997) plays a role of a guide in virtual worlds of MUD; the Extempo bartender agent converses with chat visitors, and is designed to enhance the social atmosphere (Isbister and Hayes-Roth, 1997); and there are bots for Web sites that answer questions and direct visitors in a friendly way (e.g. http://www.artificial-life.com). However, these agents are designed to engage in one-on-one social interactions, rather than facilitating human–human interaction. There are few studies about agents, which interact with multiple people. Takeuchi and Naito proposed a synthesized facial display as a conversational partner in multiparty meetings (Takeuchi and Naito, 1995).

The social agent we developed differs from the agents described above, which support specific tasks or play a role of a conversation partner. Our agent aims to work as an in-between of human–human interaction. Our agent is designed to conduct simple question and answer so that people whose conversation is faltering can find a common topic to talk about. Another possible solution for such an awkward situation is providing an information search tool to find a common topic based on the retrieved data about the social identities of conversation partners. However, that tool does not help the process to start a conversation. There is a gap between finding topics and beginning conversations. Through question and answer, people can share one another's answer to the same question. That is an opportunity to start a conversation based on the answers. Furthermore, it may be invasive for the participants' privacy to collect personal information about conversation partners.

4.2. Design of the social agent for virtual spaces

In FreeWalk2, our agent basically acts in the same way of a busy party host looking for clues that the guests' conversations are going badly. The agent tracks audio from a two-person conversation, to look for longer silences that will trigger its conversation aid. Pauses are a powerful cue for what is happening in a conversation (Clark, 1996). When the agent finds the pause, it approaches to the conversation pair. The agent then directs a series of yes/no questions to both conversation partners in turn, and uses their answers to guide its suggestion for a new topic to talk about. Then the agent retreats until it is needed again.

4.2.1. Nonverbal communication abilities

In FreeWalk2, our social agent is embodied the same way of users (see Fig. 12). This allowed us to take advantage of nonverbal cues in designing the agent, such as a spatial position and direction for turning to face users, and animated pictures to present facial expressions and gestures.

The agent approaches to the conversation pair to direct yes/no questions when it detects an awkward silence in their conversation. After concluding a suggestion

Fig. 12. Social agent in a virtual meeting space.

cycle, the agent departs from the conversation zone, and wandering at a distance, until it detects another awkward silence. This makes it easy for the conversation pair to know whether the agent joins their discussion (Hall, 1966). The agent orients its face toward the conversation partner it is addressing so that the pair can intuitively recognize whom the agent asks.

On the rectangle of the agent's embodiment, an animated dog is pasted. We chose a dog because we wanted users to think of the agent as subservient, friendly, and reasonably socially intelligent (Reeves and Nass, 1996). The agent has a set of animations for asking questions, reacting to affirmative or negative responses, and making suggestions. Each of them corresponds to a phase in the process of question and answer. We crafted these animations as a supplement to the agent's speech (Cassell et al., 1999).

4.2.2. Interaction mechanisms

Silence-detection: The agent decides there is silence when the sum of the voice volumes of both participants is below a fixed threshold value. When the agent detects a silence that lasts for more than a certain period of time, it decides the participants are in an awkward pause.

Positioning: The agent decides how to position itself, based on the location and orientation of each participant. The agent turns toward the participant that it is currently addressing. If the participants move while the agent is talking, the agent adjusts its location and orientation. The agent tries to pick a place where it can be seen well by both people, but also tries to avoid blocking the view between them. If it is hard to find an optimal position, the agent will stand so that it can at least be seen by the participant to whom it is addressing the question.

State-transitions: The agent has three states, which are idling, approaching, and talking. When idling, the agent strolls at the corner of the virtual space, further away than the normal conversation zone (Hall, 1966). When the agent detects an awkward pause in the participants' conversation, it begins an approach. Upon reaching the participants, the agent goes into the talking state. However, if the participants start talking again before the agent reaches them, it stops the approach and goes back to idling. This behaviour is strikingly similar to the actions of a hesitant subordinate trying to approach a superior, who is engaged in a conversation with another dominant person. The agent will also remain in idling state if the participants are standing far apart from each other (out of conversation range), or are not facing each other. If the participants turn away from each other during the agent's approach, or while it is talking, it will return to idling state, as well.

4.2.3. Interaction design

The user interface for communicating with the agent is very easy to learn. The agent presents questions to the participants in a text-balloon above its head. We did not use synthesized voice because we were afraid that unnatural utterance may affect participants, and participants may fail to catch what the agent says. The participant indicates 'yes' or 'no' by clicking the mouse on his/her answer displayed under the question in the text-balloon. We did not use natural language as an input interface to prevent participants from expecting too much intelligence of the agent, since they might be frustrated by not smooth conversation with the agent. Both participants see all questions, but only the addressed person sees the Yes/No options. When the person answers the question, his/her answer is displayed in a text-balloon above his/her own embodiment (see Fig. 13).

Each topic has a tree structure, with nodes that are: a question for a participant, possible answers by participants, agent's reply to each answer, and flags indicating whether the agent will address its next question to the other person or to the same

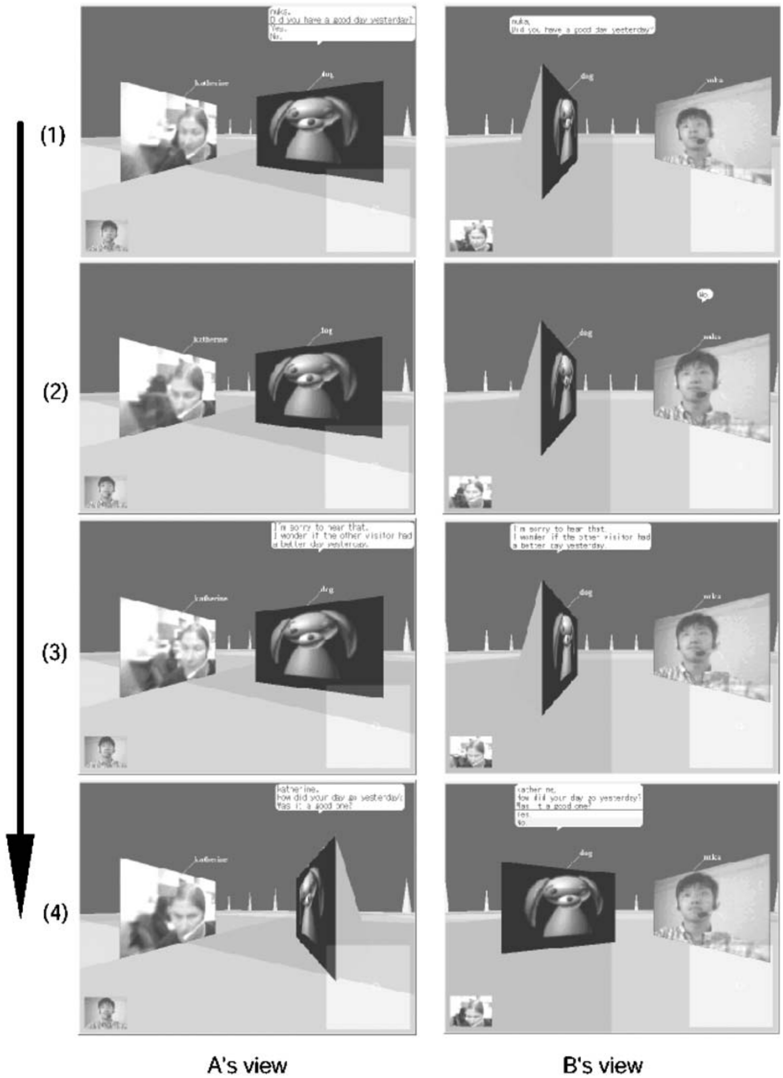

Fig. 13. Conversation from both participant's point-of-view.

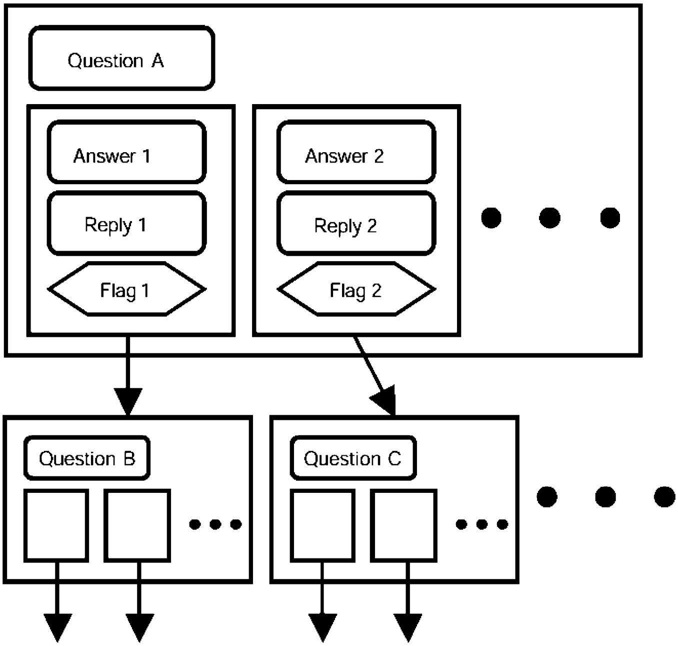

Fig. 14. Tree structure of topic data.

person (see Fig. 14). Topics were designed to draw participants into a dialogue, so each turn is tailored for this purpose. Basically, the agent asks both participants the same question to draw shared or conflicted points from the interaction. The cycle always concludes with a recommendation for how the participants could make use of the particular topic area, given their own answers to the agent.

When the agent approaches to start a cycle, it selects a topic from its repertoire of topics randomly, out of those that have not yet been used. Then it randomly chooses one of the two participants as the target for the first question. Let us call this person A. When A answers, the agent replies to A's answer. Based on what A answered, the agent then chooses a follow-up question. This question might be directed at A or at B. If it is directed at B, the agent turns to B to pose the question. When B answers, the agent makes a general comment that is meant to guide the participants into using this topic. This general comment is selected based upon the previous answers from the participants. Fig. 13 shows a part of this cycle from both participants' point-of-view. In this figure, (1) person A is asked the first question (2) and responds, (3) then the agent comments. (4) Next person B is asked a question. As we described above, the agent faces the person it is addressing.

After making this comment, the agent departs. If at any time a user does not respond to the agent's question, the agent will wait for an interval, and then go back into idling mode, without trying to continue its question cycle. This makes it clear that the agent is an in-between only to lead a question cycle. Therefore, participants can intuitively understand they do not need to include the agent in their conversation.

4.3. Evaluation of the social agents

We conducted an experiment to evaluate our social agent. The experiment was collaboration among NTT, Kyoto University and Stanford University. We used a dedicated line provided by NTT to connect both PCs in the two universities.

For testing our agent, we focused on an extreme case of low social context in a virtual meeting space: strangers from different national cultures, meeting for the first time. Even when people can use a common language with reasonable fluency, they do not necessarily have a common context for their conversation. Different cultures have different notions of how to begin and develop conversations. What is a safe topic that is unlikely to harm the conversation and destroy the relationship in one culture, may be very unsafe in another culture. For example, in some cultures it is appropriate to ask about family members right away; whereas in other cultures this is private (Hall and Hall, 1990; Clark, 1996). Since it is very hard to establish a common ground in this sort of meetings, we thought we could find the clear effect of our agent's assistance in conversations. We focused on conversations between Japanese and Americans. These two national groups are known to have very different interaction styles and cultural norms (Hall and Hall, 1990), and so we felt this was a good test case.

We gathered safe and unsafe topics for the first time meeting, using a Web survey, which university students from Japan and the United States filled out. We used the collected pool of topics to select common safe and unsafe topics for people from both countries. From these topics, we crafted a set of questions that the agent could ask in the question and answer process. Safe topics included: movies, music, the weather, sports, and what you did yesterday. Unsafe topics included: money, politics, and religion.

4.3.1. Design of the experiment

Our initial expectation is that the safe-topic agent would create a more satisfying experience, than if there were no agent. Participants would feel they were more similar, would be happier with the interaction and conversation partner, and would form more positive impressions of one another's nationality. We designed a three-condition experiment using pairs of students who were located in the United States and in Japan. Pairs either interacted one-on-one, or had the help of the safe-topic or unsafe-topic agent. We divided the 20-min conversation session into five segments, and forced the agent to display a topic within each 4-min segment. The agent looked for an awkward pause during a minute in each time segment. The agent introduced topics immediately if it could not find a pause. Thus, in the safe-agent condition, the agent introduced all five safe topics in random order. In the unsafe-agent condition, the agent introduced all five unsafe topics in random order.

The Stanford students were a part of an undergraduate class, which required participation in experiments for credit. The Japanese students were undergraduates from Kyoto University and other nearby universities, who were paid for their participation. In total we had ninety participating students. Students were assigned randomly to same-gender pairs. Each pair was randomly assigned to one of the three conditions. Students were told that they would be testing out a new communication environment with a student from the other country. They were asked to talk about anything they liked. They were trained in how to use the system, and then left alone to talk for 20 min. After their conversation, participants filled out a survey. The questionnaire included questions about the conversation, their conversation partner, and the agent (in agent conditions). We also asked them to make assessments of themselves, and the typical person of both participants' cultures on some commonly used stereotypic adjectives.

4.3.2. Results

The statistical analysis result of questionnaire data shows our agent strongly influenced subjects' impressions of the agent, their partners, and stereotypes about their partner's nationality.

In the experiment, the safe agent had positive effects for American students. Their opinions of their own behaviour, their partner, and the typical Japanese person were higher. On the other hand, it had negative effects for Japanese students. Their opinions of the experience, their own behaviour, and the typical American person were lower. But simultaneously it made them think their partner was more similar to themselves. One reason of these different effects may be that the agent's questions were implemented in English. It is possible that Japanese subjects felt it was a two-against-one situation. Another reason may be that Japanese subjects disliked the sudden interruptions by the agent that failed to find an awkward pause. Most of Japanese subjects seemed to be interested in talking with Americans.

In the unsafe agent condition, both Japanese and American students thought their conversations were more interesting, and Japanese students acted more American. This result indicates that it may be possible to mold user behaviour with the choices one makes about how the agent behaves and what it talks. Safe/unsafe agents were perceived differently by Japanese and Americans. Americans preferred the safe agent while Japanese preferred the unsafe agent.

4.3.3. Extension of the agent

In the experiment described above, our agent provided a topic chosen randomly from the topics that we had prepared beforehand, when an awkward pause occurred in the subjects' conversation. Hence, the agent often provided a topic which had been already talked, or which is not related to their ongoing conversation at all. We describe the agent's additional function to solve that problem, which is developed after the experiment. The agent with this function tries to provide a topic related to the ongoing conversation, when a pause occurs. To prevent the resumed conversation from shrinking quickly, the agent provides a topic that has not appeared yet in the conversation. This function consists of the following two modules.

A->B: Have you been to Arashiyama?
B->A: No, I have not.
A->B: Have you been to Higashiyama, then?
B->A: I think I have been to ...
(A silent pause occurs, and then the agent approaches)
X->B: Have you been to Ginkakuji temple in Higashiyama?
B->X: No.
X->B: You should visit there.
X->A: Have you been to Ginkakuji temple?
A->X: Yes.
X->A: You can tell your partner about Ginkakuji temple.

Fig. 15. Example of a recorded conversation.

- Conversation monitoring
  This module has a speech recognition engine to monitor a conversation and detect keywords in it. Those keywords are typical proper nouns of a particular category, and the agent provides topics related to only that category. This constraint raises the accuracy of speech recognition, and clarifies the agent's characteristics.
- Topic-tree generation
  Topic-tree is a tree-structure of a question and answers that the agent speaks. To generate it, this module uses two keywords, one of which is detected by the conversation monitoring module, and another of which is related to the detected keyword and has not appeared in the conversation. Those two keywords are set in the template of topic-tree prepared beforehand. This method can make a topic related to an ongoing conversation and facilitating it. The module chooses the template according to the types of keywords to be set. The examples of the type of keyword are 'place,' 'person,' and 'food.' The designer of the agent has to prepare each template corresponding to each combination of those types.

By using the function described above, we implemented an agent providing topics of sightseeing in Kyoto. We selected 'region,' 'temple,' and 'shrine' as the types of keywords. Fig. 15 is an example of recorded conversations in testing this agent. A and B is a talking couple and X is the agent. 'A→B' means A talks to B. In the example, the agent's question includes 'Higashiyama' appeared in the conversation, and 'Ginkakuji' that is a temple located in Higashiyama region in Kyoto, and has not appeared in the conversation.

5. FreeWalk3: virtual city simulator

The studies on 3D virtual spaces have mainly included on-line communities (Isbell et al., 2000), military training (Macedonia et al., 1994; Rickel and Johnson, 2000), and education (Lester et al., 1997). In these applications, a virtual space is an imaginary or unfamiliar world. We are exploring the potential of virtual spaces that imitate the real world (Ishida and Isbister, 2000; Ishida, 2002a). Evacuation drill is a typical application that needs such imitated virtual spaces. In the real world, it is impossible to provoke actual disasters to conduct evacuation drills and it is very costly to carry out a large-scale drill in which hundreds of people participate. Virtual spaces can provide a virtual scene of the disaster accessible to geographically dispersed users. Since a virtual scene consists of a 3D model of the scene and agents, it is easy to prepare and modify a specification of an evacuation drill. The virtual scene enables many people to participate in an evacuation drill at their home. Furthermore, agents can increase the number of participants and play their roles in the scene.

Functionalities of conventional crisis simulation systems are visualization and model-based calculation of disasters. Visualization reproduces disasters based on collected data or simulation results of the disasters (Ahrens et al., 1997). Model-based calculation explains phenomena occurred in past disasters and predicts damage in future disasters (Noda, 2001). However, these simulation systems ignore effects of social interaction among people at the scenes of disasters, although the degree of damage is affected by social interaction (Sugiman and Misumi, 1988). Virtual scenes of disasters can bring social interaction into crisis simulations.

We developed FreeWalk3 that is a virtual space simulating group behaviour including social interaction. Evacuation drill is the primary application of it. The different characteristics of FreeWalk3 from conventional multiuser virtual environments (Greenhalgh and Benford, 1995; Waters and Barrus, 1997) are: (1) hundreds of agents participating in social interaction; (2) and the virtual space imitating the real space. FreeWalk3 has capabilities of simulating agents' low-level behaviours suitable for their prepared high-level behaviours and drawing a geometrical virtual scene where people and agents can behave. This situation and the agents' high-level behaviours are designed manually and FreeWalk3 simulates their low-level behaviours in the scene according to the design (Ishida, 2002b; Ishida and Nakanishi, 2003;). Fig. 16 shows the shopping street in Kyoto simulated by FreeWalk3. In this figure, a young female agent meets another agent while a young male agent is passing nearby. In this case, the situation is a shopping street and the agents' high-level behaviours are meeting and passing. FreeWalk3 provides the agents motor skills to walk along the predefined direction and perceptual skills to recognize the positions of other agents. In the case of an evacuation drill, a scene of disaster, roles of agents, and behaviour rules of them are provided to FreeWalk3 to simulate the virtual disaster scene in which human participants can take part. The rest of this section explains how FreeWalk3 simulates group behaviours of people and agents, describe a prototype system, and present some implementation issues found in our preliminary experience to use it.

5.1. *Design of virtual group behaviour*

The three issues for designing the environment for group behaviour of people and agents are: (1) constructing an environment where people can behave naturally, (2)

Fig. 16. Shopping street in Kyoto simulated by FreeWalk3.

controlling many agents effectively, and (3) establishing communication among people and agents. How FreeWalk3 is designed to solve those issues is described below.

(1) Virtual space as an information resource to determine behaviour

The graphical representation of a virtual space is a major information resource for avatars that are people in the space. Accordingly, when we design the graphical representation, we should concentrate on how the space provides necessary information. Although the reality of a space is effective in making users serious and tense, the recognizability of the signs placed in a space is even more important. For example, in evacuation simulation, the validity of the simulation result is denied if the virtual space does not have the sign of an emergency exit placed in the real space.

In FreeWalk3, the 3D model of a virtual space consists of pictures taken in the real space. Those pictures are used as textures mapped on the 3D polygons of the space. The appropriate quality of each picture depends on its role. For example, the quality of bus stop pictures in the 3D model of a street is high enough to be readable (see Fig. 17a). The quality of direction signs in the 3D model of a subway station is also high (see Fig. 17b).

(2) Split control of multiple agents

Bus stop Direction sign in a subway station

Fig. 17. 3D objects as information resources.

Since the required behaviour of the agent depends on its role, e.g. a participant of an evacuation drill, we have to program each agent for each role. Although several general mechanisms to control the agent's behaviour have been studied (Laird, 2001), it is still needed to add rules and adjust parameters for each agent. This method comes at a price and makes it difficult to design the coordination of multiple agents.

FreeWalk3 has the 'split control interface' for outside software instructing agents. This interface provides simulating functionality as well as separates low-level programming from the instruction for agents. For example, when we provide the agents with the goals to walk to, such physical action as avoiding collision with other agents is automatically simulated inside the FreeWalk system. If we decide to change the agent's behaviour when the agent sees a certain object, such perceptual processing as checking the object projected on the agent's view is also simulated. As described above, the split control interface divides the control of agents into the instruction for them and the simulation of their low-level actions. This interface is helpful in developing software to edit agents' behaviour. Currently, *Q* language (Ishida, 2002b) is connected with FreeWalk3 by the interface.

(3) Individual identification of agents

When people communicate with each other through voice and video, a tone of voice and facial characteristics help them to identify each individual. This enables people to gather to talk without confusion. However, individual identification is not easy when many agents participate in communication. Agents use a speech synthesizer to speak in FreeWalk3. The variation of the tone that a general synthesizer can generate is poor. The advanced speech

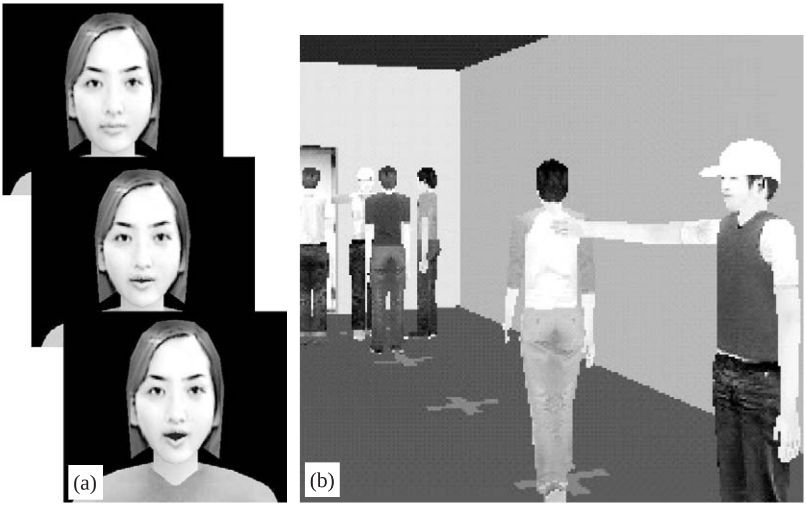

Fig. 18. Complementary cues to identify speaking agents.

synthesis technology has not been deployed for practical use yet (Black and Taylor, 1994). The current technology does not allow us to use the tone of the agents' voice as a resource to identify each agent.

In FreeWalk3, 3D sound, lip animation, and gesture help people to know which agent is speaking. First, the 3D sound enables people to guess the location where the voice is produced. Next, the lip animation enables people to detect which agents are speaking (see Fig. 18a). Finally, the consistency between the utterances and the gestures (see Fig. 18b) complements information necessary to judge the agent speaking.

5.2. Implementation of virtual group behaviour

Since the server administrates only registration and deletion of the avatars and the agents in its virtual space, client–server communication scarcely occurs. Almost all the data including voice, posture, and gesture cues are carried in peer-to-peer transmission between clients. A single client can process multiple agents. The maximum number of the agents running on a client depends on the limitation of CPU and memory resource. Only one avatar can run on a single client because of the limitation of input and output devices as a screen, keyboard, and mouse. To implement FreeWalk3, we added the following components to FreeWalk2.

- Split control interface
  FreeWalk3 provides the functions, which can be called by outside software instructing agents. Currently, the function call mechanism is implemented as

shared-memory communication. The FreeWalk client repeats a cycle consisting of the calculation phase for processing the behaviour of agents and avatars, and the rendering phase to draw a frame based on its result. In that repeated cycle, it is difficult to start a function immediately when it is called. In FreeWalk3, the function calls are stored once in a shared memory, and those calls are picked up together to run when the calculation phase begins.

The split control interface has two function call modes: blocking mode and nonblocking mode. Outside software can choose either mode according to its convenience. In blocking mode, if you instruct the agent to walk to a certain destination, the walking function returns when the agent arrives at the destination. It is not possible to give the agent another command while it is walking. In nonblocking mode, the walking function returns immediately, and it is possible to ask the agent another behaviour anytime without waiting for the agent to finish walking. Currently, concurrent function calls is not allowed to prevent the behaviour control of agents from being too complex.

- Human walking animation library

  Walking animation is vital to the representation of group behaviour of people in daily spaces as a street and a station. FreeWalk3 uses the human walking animation library developed by Tsutsuguchi et al. (2000). When a walking animation generator works to make normal computer graphics animation, it can generate the animation based on the static walking path prepared beforehand. However, the generator can have only information about the next step in a virtual world as FreeWalk, since the walking path is determined dynamically. The library we use can generate the walking animation based on only the 3D position information of the next step.

- Magnetic model module

  Magnetic model module developed by Okazaki and Matsushita (1993) simulates the physical actions as avoiding collisions. In this module, each agent and each avatar is modeled as a particle with the force of repulsion. The degree and the range of the force of each agent and each avatar can be set through the function parameters of the split control interface. Based on the values of those parameters, the magnetic model module automatically modifies the walking direction. This module enables FreeWalk3 to draw walking crowds in a street and simulate fan-shaped crowds rushing into an emergency exit.

- Speech synthesis and recognition engines

  FreeWalk3 cooperates with speech synthesis and recognition engines to have the capability of voice communication between people and agents. The speech recognition engine converts the utterance of a person controlling his/her avatar into text. The engine is running on the client computer the avatar is running. Then, the client sends the text to agents standing near to the avatar. Finally, the agents can use keyword matching technique to analyse the text. Since voice is converted into text at not the hearing side but the speaking side, the amount of transmitted data is small. Similarly, the text of an agent's utterance is transmitted to avatars, and then the speech synthesis engine running on the avatar's client play it.

- VRML library

  We use VRML as the format of the 3D model data of virtual spaces. FreeWalk3 uses a graphics library to interpret and draw VRML data. Since the level-of-detail function of the library roughly draws human models placed far from the avatar's viewpoint, many human models can be drawn on the screen. Consequently, many people and agents can enter into the same virtual space. Currently, we simply use the VRML 'switch' node changing the avatar's head model to implement lip animation.

5.3. Experience in designing virtual group behaviour

In the preliminary experience to use FreeWalk3, the problems about the split control interface and the model data of virtual spaces are found.

In a script of Fig. 19, physical and perceptual functions currently provided by the split control interface are written in a bold font. The script is an agent's behaviour to lead the agent A to the exit after meeting with it at the appointed place. The split control interface is necessary for abstract instructions such as this example. The remaining problems of the interface are described below.

(1) Occurrence of deadlock

Deadlock may occur inside the simulation of the split control interface. There is the case that the agent cannot complete the instruction to walk, since the density of agents is too much to walk through. Even if the walking path is chosen safely, deadlock may occur because of the avatars' movement that cannot be predicted. It is hard for outside instructing software to avoid this problem. Thus, the split control interface should have the mechanism to detect the occurrence of deadlock to solve it. If the distance to the destination is not shortening for a certain period of time, the walking function can be judged in the deadlock.

(2) Inflexibility inside the functions

The split control interface's simulating ability provides the designers of agents' behaviour an abstract controlling functionality to instruct agents. To construct the agents' behaviour, the designers have only to combine the functions provided by the interface, while they can only use a few parameters of the functions to control the inside simulations. This causes inconvenience when the designer exactly adopts the simulation to the targeted application, since he/she cannot directly modify the algorithm of simulation. For example, the

Walk to the coordination (12.0 2.5), which is the meeting place.
Turn to the agent *A* if the distance to the **Position** of it is less than 2.0.
Approach to the agent *A* until the distance to the **Position** of it becomes 0.5.
Face (-5.5 18.0), which is the exit.
Point to (-5.5 18.0).

Fig. 19. Example of a script.

current algorithm of the magnetic model is not appropriate for simulating walking movement in a crowded shopping street. Since the function parameters are sometimes not useful enough to control agents' behaviour, the inside algorithm should be exchangeable.

FreeWalk3 dose not have the scheme to use spatial structural data complementary to the VRML data. The problems caused by this are described below.

(1) Lack of symbols
To call such functions as walking, turning, and pointing, you mainly use coordinates and angles to indicate the destination and the targeted object. Since it is not efficient to set immediate values to the parameters of coordinates and angles every time, you may define symbols representing points and objects scattered around the 3D space model. If the space is very large and complex, the number of symbols increases too much to define them. To reduce this work, the basic set of symbols and the VRML data of a space should be constructed together.
(2) Lack of attributes
The agent controlled by the walking function goes up slopes and steps of stairs automatically. So, it is not critical to care about the geographical shape of a virtual space to move agents. However, the agent cannot judge whether the bump on the ground is a step of a stair or not, since it can know only the difference of level from the VRML data. Consequently, the agent may climb over the signboard standing on the ground. To eliminate this abnormal behaviour, each object should have an attribute to prohibit agents from riding on it.

6. Conclusions

We developed a virtual space where people and agents can communicate with each other, and engage in group behaviour together. First, we developed a virtual meeting space FreeWalk1 that makes social interaction more casual and relaxed than that in telephone-like environments. Next, we developed FreeWalk2, which is equipped with a social agent that is a character acting as an go-between for people to reduce the problem of the low social context in virtual meeting spaces. Currently, we are developing FreeWalk3 to conduct an evacuation drill on the Internet.

In FreeWalk1, the positions of participants make spontaneous simultaneous conversations possible. The spatial communication model is tightly combined with the video-mediated communication. FreeWalk1 facilitates many familiar behaviours seen in casual meetings. All of these behaviours are almost impossible in normal videoconferencing environments. The freedom of three-dimensional virtual spaces seems to stimulate participants into talking easily. They formed concurrent multiple groups to greet and chat with others when they did not have any common topics to discuss together.

In FreeWalk2, when the agent finds the awkward pause in a conversation, it approaches those in conversation to direct a series of yes/no questions to both conversation partners in turn, and uses their answers to guide its suggestion for a new topic to talk about. The statistical analysis of a questionnaire collected in the cross-cultural study shows that our agent strongly influenced people's impressions of the agent, their partners, and stereotypes about their partner's nationality. We found an agent's presence affects participants' style of behaviour. It may be possible to mold user behaviour with the choices one makes about how the agent behaves and what it talks.

In FreeWalk3, the split control interface divides the control of multiple agents into the instruction for them and the simulation of their low-level actions. This simulating ability provides the designers of agents with an abstract controlling functionality to instruct many agents efficiently. To construct the agents' behaviour, the designers have only to combine the functions provided by the interface. The remaining problems of the interface are the following: (1) deadlock may occur inside the simulation of the interface. (2) It is not possible to modify directly the algorithm of simulation.

Acknowledgements

I express my sincere gratitude to Toru Ishida for invaluable advices. I express my special thanks to the cooperation of Clifford Nass, Katherine Isbister and Eva Jettmar. I am deeply indebted to my colleagues of the Community Computing Project, Toshikazu Nishimura, Chikara Yoshida, Hiroki Nakano, Masayuki Okamoto, Minoru Nakata and Satoshi Nakazawa. I received a lot of support in the preparation of the facilities used in the cross-cultural experiment from NTT GEMNet project, Fumio Hattori, Tomoaki Tanaka, Takeshi Oguro, Kaoru Hiramatsu, and Jun-ichi Akahani. I also received a lot of support in the development of FreeWalk3 from Ken Tsutsuguchi, Shigeyuki Okazaki, and Satoshi Matsushita. FreeWalk3 has been developed by Department of Social Informatics, Kyoto University and JST CREST Digital City Project. Finally, I would like to thank the editor Kumiyo Nakakoji and the anonymous reviewers for their helpful and thoughtful comments.

References

Ahrens, J., McCormick, P., Bossert, J., Reisner, J., Winterkamp, J., 1997. Case study: wildfire visualization. IEEE Conference on Visualization, Phoenix, AZ, pp. 451–453.

Black, A.W., Taylor, P.A., 1994. Chatr: a generic speech synthesis system. *International* Conference on Computational Linguistics II, Kyoto, Japan, 983–986.

Borning, A., Travers, M., 1991. Two approaches to casual interaction over computer and video networks. International Conference on Human Factors in Computing Systems (CHI91), New Orleans, LA, pp. 13–19.

Bowers, J., Pycock, J., O'Brien, J., 1996. Talk and embodiment in collaborative virtual environments. International Conference on Human Factors in Computing Systems (CHI96), Vancouver, Canada, pp. 58–65.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H., Yan, H., 1999. Embodiment in conversational interfaces: rea. International Conference on Human Factors in Computing Systems (CHI99), Pittsburgh, PA, pp. 520–527.

Clark, H.H., 1996. Using Language. Cambridge University Press, Cambridge.

Cranach, M.V., 1971. The role of orienting behavior in human interaction. In: Esser, A.H. (Ed.), Behavior and Environment: the Use of Space by Animals and Men. Plenum Press, New York, pp. 217–237.

Duck, S., 1998. Human Relationships 3rd Edition. SAGE Publications, London.

Foner, L., 1997. Entertaining agents: a sociological case study. International Conference on Autonomous Agents (AGENTS97), Marina del Rey, CA, pp. 122–129.

Greenhalgh, C., Benford, S., 1995. Massive: a collaborative virtual environment for teleconferencing. ACM Transactions on Computer–Human Interaction 2 (3), 239–261.

Hagsand, O., 1996. Interactive Multiuser VEs in the DIVE System. IEEE MultiMedia 3(1), Boston, MA, 30–39.

Hall, E.T., 1966. The Hidden Dimension. Doubleday, New York.

Hall, E.T., Hall, M.R., 1990. Hidden differences: doing business with the Japanese, Reprint Edition. Anchor Books, New York.

Han, J., Smith, B., 1996. CU-SeeMe VR immersive desktop teleconferencing. International Conference on Multimedia, pp. 199–207.

Isbell, C.L., Kearns, M., Kormann, D., Singh, S., Stone, P., 2000. Cobot in LambdaMoo: a social statistics agent. National Conference on Artificial Intelligence (AAAI2000), Austin, TX, pp. 36–41.

Isbister, K., Doyle, P., 1999. Touring machines: guide agents for sharing stories about digital places. Workshop on Narrative and Artificial Intelligence in AAAI99, Orlando, FL.

Isbister, K., Hayes-Roth, B., 1997. Social implications of using synthetic characters. Workshop on Animated Interface Agents in IJCAI97, Nagoya, Japan, pp. 19–20.

Isbister, K., Nakanishi, H., Ishida, T., Nass, C., 2000. Helper agent: designing an assistant for human–human interaction in a virtual meeting space. International Conference on Human Factors in Computing Systems (CHI2000), The Hague, The Netherlands, pp. 57–64.

Ishida, T., 2002a. Digital City Kyoto: social information infrastructure for everyday life. Communications of the ACM (CACM) 45 (7), 76–81.

Ishida, T., 2002b. *Q*: a scenario description language for interactive agents. IEEE Computer 35 (11), 54–59.

Ishida, T., Isbister, K., 2000. Digital Cities: Experiences, Technologies and Future Perspectives. Lecture Notes in Computer Science, Vol. 1765, State-of-the-Art Survey. Springer, Berlin.

Ishida, T., Nakanishi, H., 2003. Designing scenarios for social agents. In: Zhong, N., Liu, J., Yao, Y. (Eds.), Web Intelligence. Springer, Berlin, pp. 59–76.

Kim, A.J., 2000. Community Building on the Web. Peachpit Press Berkeley, CA.

Laird, J.E., 2001. It knows what you're going to do: adding anticipation to a Quakebot. International Conference on Autonomous Agents (AGENTS2001), Montreal, Canada, pp. 385–392.

Lea, R., Honda, Y., Matsuda K., Matsuda, S., 1997. Community place: architecture and performance. Symposium on Virtual Reality Modeling Language (VRML97), Monterey, CA, pp. 41–50.

Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., Bhogal, R.S., 1997. The persona effect: affective impact of animated pedagogical agents. International Conference on Human Factors in Computing Systems (CHI97), Atlanta, GA, pp. 359–366.

Macedonia, M.R., Zyda, M.J., Pratt, D.R., Barham, P.T., Zeswitz, S., 1994. NPSNET: a network software architecture for large-scale virtual environments. Presence 3 (4), 265–287.

Nakanishi, H., Yoshida, C., Nishimura, T., Ishida, T., 1996. FreeWalk: supporting casual meetings in a network. International Conference on Computer Supported Cooperative Work (CSCW96), Boston, MA, pp. 308–314.

Nakanishi, H., Yoshida, C., Nishimura, T., Ishida, T., 1999. FreeWalk: a 3d virtual space for casual meetings. IEEE MultiMedia 6 (2), 20–28.

Nakanishi, H., Nakazawa, S., Ishida, T., Takanashi, K., Isbister, K., 2003. Can software agents influence human relations? Balance theory in agent-mediated communities. International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS2003), Melbourne, Australia, pp. 717–724.

Noda, I., 2001. Rescue simulation and location-based communication model. World Multiconference on Systemics, Cybernetics and Informatics (SCI2001), IV. Orlando, FL.

O'Conaill, B., Whittaker, S., Wilbur, S., 1993. Conversations over video conferences: an evaluation of the spoken aspects of video-mediated communication. Human-Computer Interaction 8 (4), 389–428.

Okazaki, S., Matsushita, S., 1993. A study of simulation model for pedestrian movement with evacuation and queuing. International Conference on Engineering for Crowd Safety, pp. 271–280.

Reeves, B., Nass, C., 1996. The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Cambridge University Press, Cambridge.

Rickel, J., Johnson, W.L., 2000. Task-oriented collaboration with embodied agents in virtual worlds. In: Cassell, J., Sullivan, J., Prevost, S. (Eds.), Embodied Conversational Agents. MIT Press, Boston.

Root, R.W., 1988. Design of a multimedia vehicle for social browsing. International Conference on Computer Supported Cooperative Work (CSCW88), Portland, OR, pp. 25–38.

Sellen, A.J., 1995. Remote conversations: the effects of mediating talk with technology. Human-Computer Interaction 10 (4), 401–444.

Sugawara, S., Suzuki, G., Nagashima, Y., Matsuura, M., Tanigawa, H., Moriuchi, M., 1994. InterSpace: networked virtual world for visual communication. IEICE Transactions on Information and Systems E77-D (12), 1344–1349.

Sugiman, T., Misumi, J., 1988. Development of a new evacuation method for emergencies: control of collective behavior by emergent small groups. Journal of Applied Psychology 73 (1), 3–10.

Takeuchi, A., Naito, T., 1995. Situated facial displays: towards social interaction. International Conference on Human Factors in Computing Systems (CHI95), Denver, CO, pp. 450–455.

Tsutsuguchi, K., Shimada, S., Suenaga, Y., Sonehara, N., Ohtsuka, S., 2000. Human walking animation based on foot reaction force in the three-dimensional virtual world. Journal of Visualization and Computer Animation 11 (1), 3–16.

Waters, R.C., Barrus, J.W., 1997. The rise of shared virtual environments. IEEE Spectrum 34 (3), 20–25.