Helper Agent: A Chat Assistant for Cross-cultural Conversations

Katherine Isbister and Hideyuki Nakanishi

This article introduces a collaborative project between NTT's Open Lab, Kyoto University's Department of Social Informatics and Stanford University's Communication Department. We created an agent prototype that was designed to support human-human communication in virtual environments. The prototype mimics a party host, trying to find a safe common topic for guests whose conversation has lagged. We performed an experimental evaluation of the prototype's ability to assist in cross-cultural conversations, using the Global Enhanced Multifunctional Network (GEMnet) broadband line between Stanford University, Kyoto University and NTT's Open Lab. In this overview, we will discuss the design of the prototype, as well as the evaluation.

Introduction

Communication contexts are becoming an ever more prominent part of the computer interface. It is increasingly common for people to meet for the first time through a computer interface. Because it easy to arrive at a virtual meeting place from many entry points, it is often hard for visitors to assume much about one anothers' cultural backgrounds, group memberships, and other aspects of social identity. Some commercial chat rooms use human moderators to help fulfill this need. However, human moderators are a scarce resource. Social interface agents could provide ongoing, in-context help in forming social connections and building common ground between visitors to virtual environments. Our project is a first step toward exploring this new application space for interface agents.

Target Application—Cross-cultural Conversations

For our first prototype, we focused on 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. What is a safe topic in one culture, may be very awkward 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[1][2]. We developed an agent prototype that could provide safe-topic suggestions, if the conversation was faltering. We focused on conversations between Japanese and Americans. These two national groups are known to have very different interaction styles and cultural norms[2].

Design Overview

The agent basically acts in the same way a busy human party host does, looking for clues that the guests' conversations are going badly. It tracks audio from a two-person conversation, looking for longer silences that will trigger its conversation aid. Then it directs a series of text-based, yes/no questions to both conversation partners in turn, and uses their answers to guide its suggestion for a new topic to talk about. Finally, it retreats until it is needed again.

Communication Environment

Our prototype works within an existing 3-D virtual meeting space called FreeWalk (**Figure 1**), which was developed by H. Nakanishi[3]. Users are represented as three-dimensional pyramid objects, with their video image mapped onto one face of the pyramid (Figure 1). In the

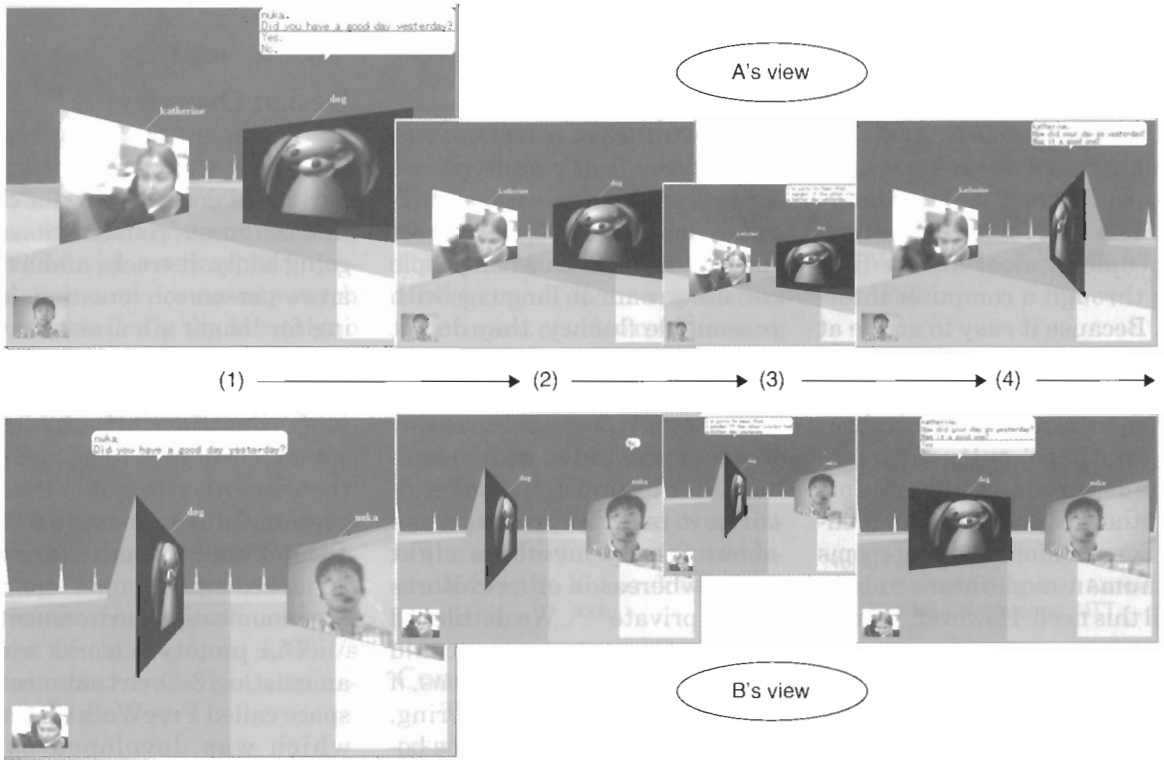
Figure 1. FreeWalk: Virtual Meeting Space Environment

lower right corner of the screen there is an overhead view of all avatars (**Figure 2**); however the user does not see his/her own avatar on the main screen. The user does see a small video window of themselves in the lower left corner, to help them adjust their camera. Each user's voice is transmitted to others around them in the virtual space. The volume of other peoples' voices is proportional to how close they are to you in the space (farther away is fainter). Users can move around in the space, and rotate the orientation of the 'face' of their own pyramid, using the mouse or the arrow keys.

Helper Agent Features

Embodiment of the Agent in the Virtual Space

Helper Agent is presented on-screen the same way users are (Figure 2). This allowed us to take advantage of nonverbal cues in designing the agent's behavior, such as turning to face users as it poses a question to them, and approaching and departing the conversation physically. The agent is an animated dog, done in a style somewhere between typical Japanese and American



(1) person A is asked the first question (2) and responds, (3) then the agent comments. (4) Next person B is asked a question. Note that the agent faces the person it is addressing.

Figure 2. Conversation from Both Participant's Point-of-view

cartoon dogs. We chose a dog because we wanted users to think of the agent as subservient, friendly, and reasonably socially intelligent. We chose stylized animation instead of more realistic, because we did not want the agent to be interpreted as a specific individual, but as more iconic and minor[4].

Nonverbal Communication Abilities

The dog has a set of animations of the proper nonverbal conversational moves for asking questions, reacting to affirmative or negative responses, and making suggestions. We crafted these animations as a supplement to the agent's speech[5], and focused on making them friendly and submissive in style[6]. The dog orients its face toward the user that it is addressing, and displays the proper animation for each phase: approach, first question, reaction, follow-up question and finally topic suggestion. After concluding a suggestion cycle, the agent physically departs from the conversation zone, and meanders at a distance from the interaction, until it detects another awkward silence. This makes it clear to the conversation pair that the agent need not be included in their discussion[7].

Topic Knowledge

We gathered topics using an internet survey, that university students from Japan and America 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 during interaction, drawing users into conversation. Safe topics included: movies, music, the weather, sports and what you've been up to. Unsafe topics included: money, politics and religion. A sample safe question: "Is the weather nice where you are right now?" A sample unsafe question: "So, do you think it is alright for a country to fish for and eat whales?"

Conversation Model and Interface

The user interface for communicating with the agent is very easy to learn. The agent does not use voice—it presents questions to the participants in a text-balloon above its head (we thought text was far less intrusive than audio). The user indicates 'yes' or 'no' using the mouse to click on their answer. Both participants see all questions, but only the user addressed sees the Yes/No options. When the person answers the question, their answer is displayed in a text-balloon above their own avatar (Figure 2).

Each topic has a tree structure, with nodes that are: first 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 person. Topics were designed to draw participants into a dialogue, so each turn is tailored for this purpose. 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 safe (or unsafe) 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's call this person A. When A answers, the agent replies to A's answer (Figure 2). 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 replies to B. Finally, 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, so that it makes sense given their replies. 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.

Evaluation of the Agent

Goals

We wanted to test the benefits of our prototype in a controlled setting. Our initial expectations were:

1. 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 partner, and would form more positive impressions of one another's nationality.
2. The unsafe topic agent would make people uncomfortable, but might lead to a more meaningful and interesting conversation than the safe topic agent.

Design

We designed an 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 Helper Agent. The study was a collaboration between the NTT Open Lab, Kyoto University's Depart-

ment of Social Informatics, and Stanford University's Communication Department. We used GEMnet's high-bandwidth (1.5 Mbit/s) dedicated line between the universities. We set up a PC with a small camera and microphone/headset at each location (**Figure 3**), and installed FreeWalk and Helper Agent at both sites.

In total we had 90 participating students. They were trained in how to use the system, then left alone to talk for 20 minutes. Afterward, they filled out a web-based survey in their native language, with questions about the interaction, their conversation partner, and the agent. We also asked them to rate themselves, their partner, and the typical person of both participants' cultures on some commonly used stereotypic adjectives.

Results

A social Interface Agent Can Help Support Human-human Communication

Our evaluation demonstrated that a human-human communication assistant can have positive effects on perception of the experience, one's own qualities, one's partner and even one's partner's cultural group. Americans had a higher opinion of their own behavior, their partner's behavior, and the typical Japanese person's behavior, in the safe agent condition (versus the unsafe condition). However, Japanese participants did not have this positive reaction. We cannot be sure why the two groups had such different responses. One reason may be that the agent's questions were implemented in English. It's possible that Japanese subjects felt it was a two-against-one situation. This might explain why they disliked the interaction, even though it seemed to make them rate their partner as more similar to themselves (our main goal!). We would need to test the system again, using a bilingual agent that address all questions to users with both languages displayed, to be sure. In any case, the positive American reaction was a strong support of our research concept.

Provocative Help Can Be Good

Our evaluation also suggested that a communication assistant can be helpful both when it offers safe topics to talk about, and when it steers the conversation in less safe directions. In fact, the Japanese participants seemed to prefer the unsafe topic agent, and both groups found it more interesting than the safe topic agent. For overall conversational support purposes, both kinds of help may be desirable. We suspect that an agent with a model for offering both kinds of topics, depending upon the conversation flow, would be the most desirable.

User-adaptation Would Make the Agent More Effective

The two cultural groups had very different impressions of the same agent behaviors, and reacted in different ways. For example, behavior that was perceived as blunt and unfriendly by Americans was seen as nice and competent by Japanese. An effective agent for different types of people will probably need to adapt its behaviors to user subgroups, or perhaps to individuals' own interaction styles and preferences. We believe we created a more American identity for our agent by delivering its topic help in English. In future iterations, we plan to create an agent whose presentation is adapted to different user styles and preferences.

Agent Behavior May Shift User Behavior

Both the Japanese and American participants noted that Japanese seemed to act more American in the unsafe agent condition. This result indicates that it may be possible to mold user behavior with the choices one makes about how the agent will behave, creating a different conversational environment by bringing different traits to the fore. This could have very interesting implications for those interested in setting a specific group conversational tone or style

Figure 3. Set-up for the Experiment (Stanford Side)

in a virtual meeting space.

Conclusion

We built a social agent prototype, designed to facilitate human-human interaction. A cross-cultural evaluation of the prototype demonstrated its effectiveness, and raised interesting considerations for further development of this class of interface agent. We feel the support of human-human interaction in virtual meeting places is an exciting and useful new domain for interface agents. Given the proliferation of online spaces, and the interest in community formation that far exceeds the industry's ability to staff communities with human hosts, this kind of agent may become a familiar part of the virtual landscape. We hope people who are called upon to think about and design these kinds of interface agents will find our prototype design and evaluation results useful in guiding their work.

Acknowledgements

Thanks to GEMnet for providing the broadband line, to Professors Toru Ishida of Kyoto University and Cliff Nass of Stanford University for their support and guidance of this project, and to Eva Jettmar for her assistance in the evaluation.

References

(1) H. H. Clark, *Using Language,* Cambridge University Press, 1996.

(2) E. T. Hall and M. R. Hall, *Hidden Differences: Doing Business with the Japanese,* Anchor Books, 1990 reprint.

(3) H. Nakanishi, C. Yoshida, T. Nishimura and T. Ishida, FreeWalk: A 3D Virtual Space for Casual Meetings, *IEEE MultiMedia, 6* (2), pp. 20-28, 1999.

(4) S. McCloud, *Understanding Comics: The Invisible Art,* Harper Perennial, 1993.

(5) J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjalmsson and H. Yan, Embodiment in Conversational Interfaces: Rea, *International Conference on Human Factors in Computing Systems (CHI-99),* pp. 520-527, 1999.

(6) B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places,* Cambridge University Press, 1996.

(7) E. T. Hall, *The Hidden Dimension,* Anchor Books/Doubleday, 1982 (1996).

The Authors

Katherine Isbister
(formerly) Communication Science Laboratories, NTT.

Hideyuki Nakanishi
Department of Social Informatics, Kyoto University.