Building of Turn-Taking Avatars that Express Utterance Attitudes

A Social Scientific Approach to Behavioral Design of Conversational Agents

Masahide Yuasa and Naoki Mukawa

School of Information Environment, Tokyo Denki University,
2-1200 Muzai Gakuendai, Inzai,
Chiba 270-1382, Japan
{yuasa,mukawa}@sie.dendai.ac.jp

Abstract. In everyday communication, humans comprehend the attitudes of others conveyed via nonverbal behavior, such as facial expression, body posture and gaze behavior. In this paper, we describe a model for comprehending participants' desire to start to speak or to listen based on nonverbal behavior during conversation. We use a social scientific approach that is based on both an analysis of a video observation and an experiment using avatars. We explain the building of the model. We discuss detecting participant's attitudes using computer vision and the expression of their attitudes using their avatars' facial expressions and body postures.

Keywords: animated agent, avatar, turn-taking, nonverbal behavior, conversation.

1 Introduction

In everyday life, we communicate with our family, friends, co-workers, and others. We enjoy talking and eating together and sometimes have confrontational discussions. Such shared emotions are important for human beings as they contribute to the development of better relationships and create a sense of unity. In order to share emotions, we use both verbal and nonverbal behavior such as gaze direction, head orientation, facial expressions, and body posture. Nonverbal behavior plays an important role in comprehending subtle emotion. By using nonverbal communication, we express attitudes and can comprehend the degrees of others' feelings.

In our previous research about turn-taking in communication, we proposed a conversational utterance attitude model [14]. The model has two demands: "I want to start to speak" and "I want someone to start to speak." We use abstract animated agents that mimic human turn-taking in conversations, to confirm the validity of our model. However, the agents' shape was a simple sphere; which was simply not realistic enough. In order to improve upon our previous model, we needed a system that has animated avatars, agents, or humanoid robots.

Therefore, in this paper, we describe turn-taking avatars that express utterance attitudes. The avatar system detects participants' utterance attitudes using a camera, and

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 101–107, 2011.

the attitudes are then expressed by their avatars. Participants can use several utterance attitudes during conversation and can communicate smoothly without revealing their actual faces.

We used a social scientific approach in developing the model. We recorded an actual three-party conversation and observed the video. On the basis of the observation, we developed an utterance attitude model for demand such as "I want to speak"/ "I want you to speak" and built animated avatars using the model. Here, we discuss the value of the avatar system and our proposed model for conversational robots or agents.

2 Previous Research

Avatars are widely used in order to make communication more natural and fun. An avatar is the embodiment of a person. Avatars are used in online games, online communities, and 3D virtual worlds. Users can chat and interact virtually with other users. For example, BodyChat [12] uses embodied avatars to mimic face-to-face communication. There are also several virtual worlds; such as Second Life[8], Worlds.com[13] and There[9]. The avatars in these virtual worlds are capable of realistic nonverbal behavior (body posture, eye movements, head orientation, etc.) that closely resemble that in human-to-human interactions. Users can select the type of interaction for their avatar using a keyboard or a mouse. In addition, facial expressions created via facial motion-capture data are used in order to enrich communication between avatars. Yun et al. investigated the effectiveness of a local 3D facial avatar for a global audience [15].

However, the behavior of avatars is limited using these tools. Users can only choose from predefined basic behaviors. Designing ways for avatars to express their utterance attitudes using eye-gaze behavior or facial expressions during turn-taking is very important to make communication more natural and fun for users.

The psychological literature has analyzed the relationship between turn-taking and eye-gaze behavior [1, 3]. Sacks states that a listener being watched by a speaker tends to start speaking [7]. Thus, there are several existing agents and robots with a turn-taking function that detect a speaker's gaze using eye-tracking equipment [5]. Peters et al. developed animated agents in a virtual world with a mathematical model of gaze behavior. In their research, when one agent looks at another hearer, the gaze behavior of the first attracts the attention of another, who then takes the next turn to interact. Traum et al. [10] developed a conversational agent that takes turns in a multiparty conversation. The agent gives a simple rule-based response to the user; that is, when a user looks at an animated agent, the agent responds to the user. Poggi et al. talk about specific communication signals called "Mind Markers," including facial expressions and gazes that indicate a speaker's beliefs, goals, and emotions [6]. However, their research only used a limited number of basic behaviors and their implied meanings.

Duncan et al. described several turn-taking categories (e.g., turn-maintaining, turn-yielding) and the complexity of turn-taking itself. We need to take into account more complicated rules of turn-taking.

In our research, we developed a conversational avatar system with effective and efficient turn-taking based on the utterance attitude model. Using the model, we developed conversational avatars that can communicate with humans in a lively and emotional manner. In our previous research, we proposed an utterance attitude model

that presents a participant's utterance attitudes that "he/she wants to start to speak" or "he/she wants someone to start to speak" based on an analysis of the observed conversations. On the basis of this model, we built an animated avatar system that can express participants' attitudes detected by a camera. The participants can converse on the basis of the attitudes of each other's animated characters; they do not need to reveal their real faces and they can express their own attitudes. This system allows smooth communication and promotes good relationships among participants.

3 A Social Scientific Approach to Behavioral Design of Conversational Agents

In this section, we describe our social scientific approach to building our conversational avatars. Our approach consists of four processes that involve analyzing actual human behavior, developing a model, building avatars as a prototype system, and evaluating the system.

3.1 Analysis of Actual Chatting

We observed and analyzed a video-recorded 20-min conversational scene of three female university students using ethnographic conversation analysis. We observed and carefully transcribed every action of these participants, including words (what the participants were saying) and nonverbal actions, such as gaze, head orientations, and facial expressions. After the analysis, we proposed an utterance attitude expression model.

Fig. 1. Conversation scene with three female university students

3.2 Utterance Attitude Model [14]

Figure 2 shows our proposed utterance attitude model [14]. In this model, utterance attitudes can be categorized into nine classes on a two-dimensional plane: the horizontal axis represents the expressions of a person who "wants to speak/not to speak," and the vertical axis represents a person who "wants someone to speak/not to speak." The plane also shows that expressions can be classified into two types: (1) subtle, implicit, expressive behaviors that are expected to be noticed by others and (2) direct and explicit behaviors that intentionally control the utterance behaviors of others. The inner ring of the model shows implicit attitudes displayed by participants, for example, (1) I want to speak. The outer ring indicates explicit attitudes that can control the other participants, for example, the (7) I want him/her to speak.

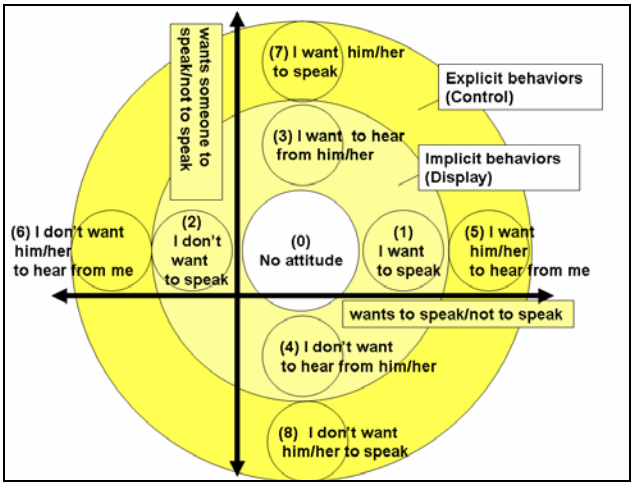

Fig. 2. Utterance Attitude Model

3.3 Building Turn-Taking Avatars

Utterance Attitudes for Avatars. We developed a system that enables three participants to converse using avatars. We implemented four expressions of utterance attitudes in our avatar system. Figure 3 shows avatar expressions that represents the utterance attitudes of participants.

As shown in Fig. 3 (a), the avatar's body expanding in height represents a positive utterance attitude. The avatar's contracting body shown in Fig. 3 (b) represents a

Fig. 3. Avatar's expressions and the representation of a participant's utterance attitudes

negative utterance attitude. The effects of expanding/contracting behaviors are well known in the literature about animation movies. In should be noted that the movements of the body reveal the avatar's implicit attitudes that "he/she wants to start to speak" or "he/she does not want to start to speak."

In addition, the avatar has hands for purposively controlling partners' utterances as shown in Fig.3, (c) promoting and (d) blocking. These movements represent explicit attitudes in order to control the other's turn. The avatar's promoting hand means it "wants you to speak." and the avatar's blocking hand means "please stop speaking."

The system can easily detect these four participants' attitudes using computer vision technology and can express the avatars' attitudes.

Fig. 4. User Interface

User Interface. This system has six cameras to detect a participant's behaviors, which are transmitted to the other participants. An audio link was established and the participants wore headsets with a microphone to converse.

One participant can see the behaviors of the other two participants on the display. The system was arranged so that when a participant looks at the avatar to the left, the participant represented by that avatar can see the front face of the avatar. The other participant represented by the avatar on the right can see the profile of that avatar's face (Fig. 4).

Detecting a Participant's Face and Hands. Our system detects head orientation (right/left), the height of the head (high/middle/low), and hand gestures (promoting/blocking) using cameras. We use six cameras set around two displays in front of a participant. The system determines whether a participant gazes at the right or the left display and estimates the height of the body by detecting the height of the participant's head. The avatar's body expanding or contracting in height is based on the participant's head movement in term of height. In addition, the system can detect hand gestures using a colored marker, which is attached to the participant's finger.

3.4 Preliminary Test of the Avatar

We tested the effectiveness of our avatar system using three participants. We recorded the participants' conversation using the avatars and observed the recorded video. We found that participants used expressions of utterance attitudes efficiently and an avatar's expressions were recognized by the others as utterance attitudes; therefore, our

expressions of turn-taking were useful in avatar communication. For example, when a participant had an opinion and wanted a turn to speak, his/her head or body leaned forward, and his/her avatar's head and body were moving up. The participants recognized the movement as meaning "he/she wants to start to speak." When a participant did not want to start to speak, he/she stretched out his/her hand to promote conversation and his/her avatar did the same. This movement was interpreted as "he/she wants me to start to speak." Thus, the utterance attitudes we proposed were used effectively to communicate in the avatar system. In an interview with the participants, they stated that they felt there was less collision (when more than one person starts to speak at once) in turn-taking than during the case of voice chatting. One participant commented that he could prepare to start to speak and it was easy to speak at the turn-taking point of the conversation.

4 Discussion

We used only limited expressions (four attitudes) for avatars in the preliminary test. Even though participants could use only four expressions of utterance attitudes, they could still converse. We should use other expressions of attitudes, including subtle nonverbal behaviors during conversation. However, it is difficult to extract these behaviors because nonverbal behaviors can be very subtle. Even though we selected only four attitudes based on psychological findings, it was very difficult to select appropriate expressions for the other attitudes. In the preliminary test, our expressions of turn-taking were useful in avatar communication. Therefore, four expressions are enough, because it is the first step using both implicit and explicit attitudes in an utterance attitude model. Additional experiments are required to design more understandable expressions based on research about nonverbal behavior.

We did not take into account combinations of attitudes; such as "expansion" (the wants-to-speak attitude) and "blocking" (the does-not-want-a-partner-to-speak attitude). Further experiments on the combinatorial effects of attitude are required.

On the basis of these findings and future research, we will improve the avatar system to make communication and turn-taking smoother.

5 Conclusions

We proposed a new avatar system that detects participants' utterance attitudes and generates their avatars' utterance attitudes. We selected appropriate expressions of utterance attitudes for the avatars on the basis of psychological findings. In our test, participants could take turns in conversing using the avatar system and the expressions were used effectively.

References

1. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press, Cambridge (1976)
2. Duncan Jr., S.: Some Signals and Rules for Taking Speaking Turns in Conversations. Journal of Personality and Social Psychology 23, 283–292 (1972)
3. Kendon, A.: Some Functions of Gaze Direction in Social Interaction. Acta Psychologica 32, 1–25 (1967)

4. Ma, C., Osherenko, A., Prendinger, H., Ishiizuka, M.: Chat system based on emotion estimation from text and embodied conversational messengers. In: Proc. of the 2005 International Conference on Active Media Technology (ATM 2005), pp. 546–548 (2005)
5. Matsusaka, Y., Fujie, S., Kobayashi, T.: Modeling of Conversational Strategy for the Robot Participating in the Group Conversation. In: Proc. of Eurospeech 2001, pp. 2173–2176 (2001)
6. Poggi, I., Pelachaud, C., Caldognetto, E.M.: Gestural Mind Markers in ECAs. In: Proc. AAMAS 2003, pp. 1098–1099 (2003)
7. Sacks, H., Schegloff, E., Jefferson, G.A.: A Simplest Systematics for the Organization of Turn-taking for Conversation. Language 50(4), 696–735 (1974)
8. Second Life, http://www.secondlife.com/
9. There.com., http://www.there.com/
10. Traum, D., Rickel, J.: Embodied agents for multiparty dialogue in immersive virtual worlds. In: Proc. of AAMAS 2002, pp. 766–773 (2002)
11. TVML, http://www.nhk.or.jp/strl/tvml/
12. Vilhjalmsson, H., Cassell, J.: Bodychat: autonomous communicative behaviors in avatars. In: Proc. of the Second International Conference on Autonomous Agents, pp. 269–276 (1998)
13. World.com., http://www.world.com/
14. Yuasa, M., Mukawa, N., Kimura, K., Tokunaga, H., Terai, H.: An Utterance Attitude Model in Human-Agent Communication: From Good Turn-taking to Better Human-Agent Understanding. In: CHI Extended Abstracts 2010, pp. 3919–3924 (2010)
15. Yun, C., Deng, Z., Hiscock, M.: Can local avatars satisfy a global audience? A case study of high-fidelity 3D facial avatar animation in subject identification and emotion perception by US and international groups. Computers in Entertainment 7(2) (2009)