ORGANIZING SELF-MOTIVATED DIALOGUE WITH AUTONOMOUS CREATURES

Noriko SUZUKI, Kazuo ISHII and Michio OKADA
{noriko, ishii, okada}@mic.atr.co.jp

ATR Media Integration & Communications Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 6190288 JAPAN

ABSTRACT

This paper discusses the effectiveness of human-computer interaction in our prototype system "Talking Eye", which is based on social behaviors in self-motivated dialogue. Talking Eye system consists of autonomous creatures having the emergent computation architecture with the advantage of self-motivated dialogue. We performed an experiment using subject's impression to illustrate the effectiveness of self-motivated dialogue with the Talking Eye system. The result was obtained that they could find personal nature during interaction even if they could not accomplish their conversational purpose. The main goal of this work is to build a mechanism for autonomous creatures that promotes a more consensual feeling through its interaction with humans.

1. INTRODUCTION

Recent developments in speech-based interface technology have opened up the possibility of a new interactive style between humans and computers. Most conventional spoken dialogue systems are based on goal-oriented techniques and mostly pursue efficiency and accuracy to complete explicit tasks [1]. There has been very little research that focuses on other aspects of spoken dialogue.

Within this context, we have considered an alternative spoken dialogue system based on a theory of everyday activities [2]. In particular, we have focused on self-motivated dialogue as another communication mode of actual dialogue with the potential to enhance an informal interactive style between humans and computers [3]. An essential property of self-motivated dialogue is the emergence of the meaning of conversational behavior through participants' responses during interaction. Such an emergent property of self-motivated dialogue provides social behaviors of interaction between humans.

We have been studying the mechanism of the emergent property of self-motivated dialogue to achieve an autonomous creature for a human conversational partner on the computer [4][5]. We have constructed a prototype system called "Talking Eye" which consists of autonomous creatures to illustrate the effectiveness of self-motivated dialogue. Each creature has the architecture that implements for self-motivated dialogue by using a dynamic action selection network [6]. The creature is shaped like an eyeball and is generated by 3-D computer graphics. It can perceive human conversational behaviors by detecting simple prosody and phrases [7]. Furthermore, it can produce about 500 vocal phrases from prerecorded speech for self-motivated dialogue. It can also produce about 20 types of eye movements to indicate emotional states and attention.

We perform an experiment between subjects and the Talking Eye system to evaluate impressions about social behaviors from the MOS (mean opinion score) evaluation of a post-experimental questionnaire given to the subjects. We demonstrate that the subjects could find the personal nature of the Talking Eye system during their interaction even if their conversational purpose was not accomplished.

2. EMERGENT PROPERTY OF SELF-MOTIVATED DIALOGUE

Most conventional interactive systems have considered goal-oriented aspect of spoken dialogue. These systems are designed to be human assistants or knowledge navigators that can achieve prepared goals with high efficiency and accuracy according to a formal interaction style, ATIS being a typical example [1]. Such an interactive style involves the exchange of conversational behaviors that are assigned unique and constant meanings for information transmission.

On the other hand, actual spoken dialogue has not only goal-oriented aspect but also self-motivated aspect, which is for simple making conversation with others, for example, table talk, gossip among housewives, etc. We believe that unique and constant meaning can not always be pre-assigned to conversational behaviors from the beginning. Rather, their meanings emerge through each partner's responses such as utterances, gestures, or facial expressions during interaction. We regard it as an essential property of self-motivated dialogue.

We focus on self-motivated dialogue as a key aspect for providing an informal interactive style between humans and computers [3]-[5]. We study the mechanism of self-motivated dialogue and its implementation on a computer. Therefore, we have tried to construct architecture using the emergent property of self-motivated dialogue on a computer.

3. TALKING EYE SYSTEM

The Talking Eye system is a prototype for the creation of self-motivated dialogue between humans and artificial creatures. The system consists of autonomous creatures with the architecture of emergent computation for the property of self-motivated dialogue by using a dynamical action selection network [4][6]. Each autonomous creature is shaped like an eyeball and is generated by 3-D computer graphics (Fig. 1). Figure 2 shows an example of an implementation of the Talking Eye system. The system uses two workstations, one for prosody detection, emergent computation, motion generation and phrase generation, and the other for phrase recognition.

The emergent computation consists of three parts: the intentional context, a set of situated agents, and the

environmental context. The intentional context provides constraints on the mental state and motivations of the situated agents, while the environmental context provides constraints on the actions of the situated agents. The set of situated agents is units of conversational behaviors. Each situated agent has a simple rule of action-perception unit. The two contexts and the set of situated agents influence one another according to the activation/inhibition dynamics through interaction. As a result of such a local interaction, one action is released dynamically to the environment.

At present, the autonomous creature has two modalities: utterance and motion generation in actions applied to the real world. It can produce about 500 Japanese vocal phrases (KANSAI dialect) for self-motivated dialogue using prerecorded speech; these include “That's too bad.", “How about it?", “That sounds good!", etc. Motions are generated by the graphical animation of the eye to indicate emotional states such as surprise, disappointment, etc., as well as social signals such as attention.

Each autonomous creature can independently perceive the current conversational state after the release of an action to the surrounding environment. Actions for replies from the environment are treated as elements of the environmental context for the autonomous creature. It can also perceive human conversational behaviors by using simple prosody and phrase detection with ATR SPREC [7].

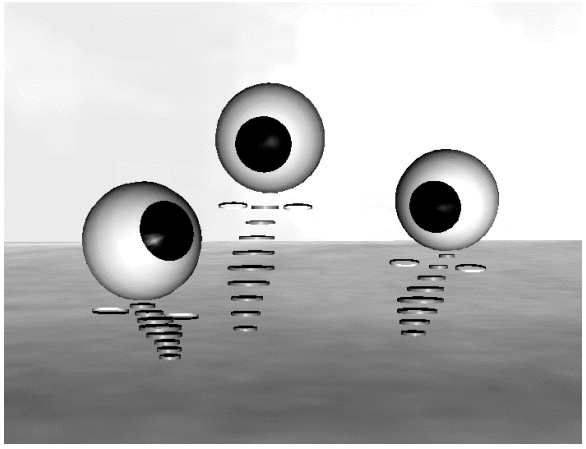

Figure 1: Appearance of Talking Eye System

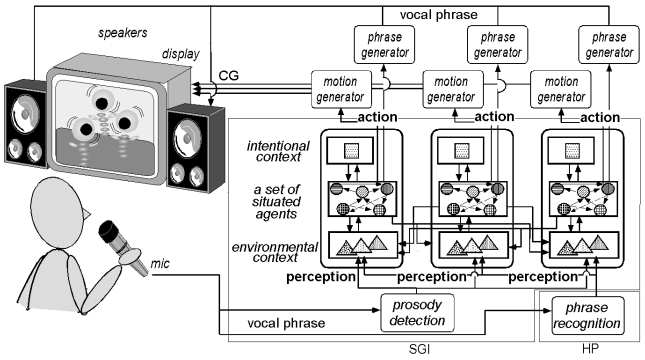

Figure 2: System Implementation

4. EXPERIMENT

We carried out an experiment between humans and the Talking Eye system to evaluate impressions about social behaviors on interaction. We will show its background, the method, the results, and the discussion.

4.1. Background

In most conventional approaches of interactive systems, it has been assumed that accurate information and cooperative responses of a system give humans good impressions about the friendliness and naturalness of the system. These systems are designed from the beginning to create a good attitude toward humans. However, the quality of interaction with the system is not always measured by the efficiency and accuracy.

On the other hand, we feel a strong personality or nature emerges from unexpected behaviors through the process of interaction rather than a good attitude during. Accordingly, we demonstrate that self-motivated dialogue between humans and autonomous creatures can be created based on the emergent property by using the following hypothesis:

[Hypothesis] A human can find the personal nature of the Talking Eye system as a result of interaction, even if his/her conversational purpose is not accomplished.

For evaluation of the hypothesis, we sat up the strong parameter of the intentional context in every autonomous creature for being selfish.

We cannot expect sequences of conversational behaviors by the autonomous creatures in the Talking Eye system because of for emergent computation. We focused on that the origin of the emergent property of self-motivated dialogue comes from spoken dialogue among humans. And we had a plan to compare the following two conditions: (i) making conversation with the Talking Eye system using the emergent computation (TEYE), and (ii) making conversation with the Talking Eye system controlled by humans based on the WOZ (wizard of oz) method (WOZ) [8].

We will get the following prediction if our hypothesis will be applied: the subjects have same trend about impression of social behaviors under the above two conditions.

4.2. Method

We used 26 university students from 18 to 28 years old as subjects. Each subject sat in front of the Talking Eye system and talked with the creatures by using a microphone. The speech voice of the creatures was output from speakers, and their motions were displayed on a computer screen. The experimenter instructed each subject to attract the attention of the creatures and to make them be quiet while the subject spoke to them (exact instructions are given below). Each subject was given the conversational purpose of the manner training by the instructions.

[Instructions] This is a system for a “leadership test". You are a teacher in primary school and the artificial creatures on the screen are your students. Please attract their attention and to have them listen quietly while being spoken to. If they show

resistance or are abusive, please do not get angry and continue to talk to them calmly.

Although the relationship between a teacher and students generally determines the particular social roles from the beginning, it is possible to shift to a personal relationship among them. Therefore, we carried out a post-questionnaire about social behaviors using the following evaluation items:

Social attitude: this evaluates impressions of the social attitudes of the creatures' utterances and behaviors as a result of interaction.

Social act: this evaluates impressions of the social acts of the creatures' utterances and behaviors as a result of interaction.

Personal nature: this evaluates impressions on the personal natures of the creatures' utterances and behaviors as a result of interaction.

We used two kinds of evaluation items for the social items, because social attitudes and social acts are not always consistent through interaction based on a previous work [9]. The meanings of behaviors are defined as a result of the interaction. The consistency of social attitude and act determines the bias to the general or specific area according to the current context [10]. On the other hand, we used only one evaluation item for the personal item because there is little inconsistency of attitude and act for a personal aspect. Table 1 shows the details of each evaluation item.

We performed an experiment to study the subjects' impressions of their interaction with the Talking Eye system through a post-experiment questionnaire involving the MOS (mean opinion score) evaluation. The questionnaire asked them to rank their impressions from –5 (negative) to 5 (positive).

Table 1: Items for Evaluation

Personal Nature	gloomy	<->	bright
	cheerless	<->	cheerful
	shy	<->	sociable
	pessimistic	<->	optimistic
Social Act	gentle	<->	lively
	inactive	<->	active
	passive	<->	aggressive
	internal	<->	diplomatic
Social Attitude	emotional	<->	intellectual
	hasty	<->	steady
	childish	<->	strong
	insincere	<->	serious

4.3. Results

Results are presented in Table 2. Figure 3 shows three kinds of histograms of the ten grades for every evaluation item. No significant difference with the average MOS rates was found between the two conditions TEYE and WOZ for three evaluation items.

Table 2: Results of Experiment

average(SD)	TEYE	WOZ
Personal Nature	2.50(1.06)	1.93(1.27)
Social Act	2.39(1.04)	2.13(2.43)
Social Attitude	-1.59(1.86)	-1.41(2.30)

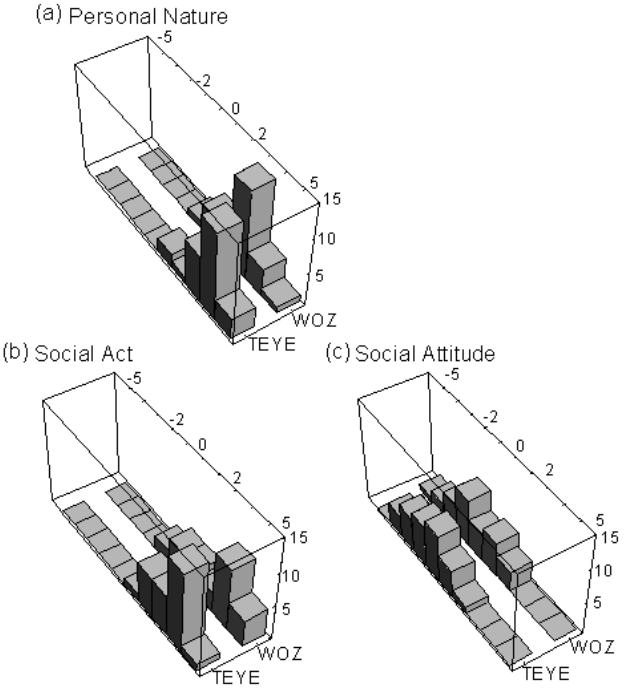

Figure 3: Histogram for Each Evaluation Item: (a) Personal Nature (upper), (b) Social Act (below, left), (c) Social Attitude (below, right)

4.4. Discussion

One result of the experiment was that the subjects obtained almost the same rates under both conditions. This result suggested that the subjects followed the same trend for the two conditions. We considered that the result supported our prediction and the validity of our hypothesis was illustrated. As the proof for the sufficiency of our hypothesis, we focus on the trend of the subjects' reactions to the utterances and behaviors of the autonomous creatures. The parameter of intentional context on every autonomous creature was set to a strong value to have each of them behave selfishly, resistively and excitedly. The wizard also tried to select the same behaviors as the autonomous creatures based on emergent computation. The subjects evaluated the personal natures and the social acts as positive although they evaluated the social attitudes as negative. This suggested that a contradiction was produced between social act and attitude. The subjects could find the personal nature of the Talking Eye system as a result of interaction in spite of its bad attitude.

Generally, a contradiction was not produced in goal-oriented dialogue, e.g., a contradiction between social act and attitude, because unique and constant meanings were assigned to conversational behaviors from the beginning. However, we consider that self-motivated dialogue interaction has the possibility to produce the contradiction between social act and attitude during interaction.

Figure 4 shows an example of appearance for self-motivated dialogue between a subject and Talking Eye system and Table 3 shows a sample dialogue between a subject (H) and autonomous creatures (T1-T3) during the experiment.

Figure 4: Appearance of Self-Motivated Dialogue between Subject and Talking Eye System

Table 3: Sample Dialogue between Subject (18 years old, female) and Talking Eye System

H:	Haai, kotti muitee.	Hi, attention please!
T1:	Nan'de kounarunen.	Why are we doing this?
T2:	Hottoiteena.	Leave me alone.
H:	Shii, shizukanii.	Hush!
T3:	Urusainaa.	Don't bother us.
T2:	Sou sou.	That's right.
H:	Oshaberiwa damedesuyoo.	Hold your tongue.
T1:	Son'nakoto iwanto.	Calm yourself.
T2:	An'taniwa kankei arahen.	It's none of your business!

5. CONCLUSTION

In this paper, we focused on self-motivated aspect of spoken dialogue and proposed an informal interactive style between humans and computers based on a mechanism of an emergent property of self-motivated dialogue. We implemented the mechanism on a prototype system called Talking Eye by using emergent computation based on a dynamic action selection network. We carried out an experimental test with subjects to evaluate impressions about social behaviors to the Talking Eye system under two conditions: 1) chatting with autonomous creatures using emergent computation, and 2) chatting with autonomous creatures controlled by humans based on the WOZ method. As a result, we obtained the same trend in the evaluation under the two conditions and demonstrated that the emergent property of self-motivated dialogue produced a contradiction in the evaluation of social act and attitude. The result suggested that humans could have positive impression about interaction with the interactive system, even if the system was not behave just as humans want to.

In the future work, we will add a learning mechanism for the acquisition of communication skills through interaction to simulate human development.

ACKNOWLEDGEMENT

We would like to thank President Ryohei Nakatsu, Dr. Yasuhiro Katagiri and members of Department 4 for their continuous support in this work. Especially, we thank Mr. Yugo Takeuchi for his advice of section 4.

REFERENCES

1. Zue, V., Seneff, S., Polifroni, J., Phillips, M., Pao, C., Goddeau, D., Glass, J. and Brill, E. "Pegasus: a spoken language interface for on-line air travel planning", *ARPA Workshop on Human Language Technology*: 196 – 201, 1994.
2. Suchman, L. A.: *Plans and situated actions*, Cambridge University Press, 1997.
3. Okada, M.: *Hesitating Computer* (in Japanese), Kyoritsu Syuppan, 1995.
4. Suzuki, N., Inokuchi, S., Ishii, K. and Okada, M. "Chatting with interactive agent", *EuroSpeech'97*, Vol. 4: 2243 – 2246, 1997.
5. Suzuki, N., Ishii, K. and Okada, M. "Talking Eye: Autonomous creature as accomplice for human", APCHI'98, 409 – 414, 1998.
6. Maes, P. "How to do the right thing", Connection Science, 1(3): 291 – 323, 1989.
7. Shimizu, T., Yamamoto, H., Masataki, H., Matsunaga, S. and Sagisaka, Y. "Spontaneous dialogue speech recognition using crossword context constrained word graphs", ICASSP'96, 145 – 148, 1996.
8. Maulsby, D., Greenberg, S. and Mander, G. "Prototyping an intelligent agent through wizard of oz", *INTERCHI'93*, 277 – 284, 1993
9. La Piere, R. T. "Attitudes vs. actions", Social Forces, 13: 230 – 237, 1934.
10. Ajze, A. and Fishbein, M. "Attitude-behavior relations: A theoretical analysis and review of empirical research", Psychological Bulletin, 84: 888 – 918, 1977.