SITUATED DIALOGUE COORDINATION FOR SPOKEN DIALOGUE SYSTEMS

*Michio Okada, Noriko Suzuki and Jacques Terken**

ATR Media Integration & Communications Research Laboratories,
2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-0288 Japan

*IPO, Center for Research on User-System Interaction, Eindhoven University of Technology
P.O.Box 513, 5600 MB Eindhoven, The Netherlands

ABSTRACT

In this paper, we present a general framework and architecture for maintaining dialogue coordination in spoken dialogue systems, in which intended behaviors and goals are incrementally performed during the course of maintaining dialogue coordination. The dialogue structure emerges as a result from interaction between user and the dialogue system. The key feature of this design for the systems is to use multiple situated-agents for coordinating communicative acts that are realized as a hierarchy of autonomous behaviors by using a subsumption architecture. In this architecture it should be noted that the lower-level behaviors act autonomously for maintaining the dialogue coordination and are linked to the specifications from higher-level behaviors for dialogue management. In order to make the behavior of the system social, in general, the maintaining of dialogue coordination takes priority over the realization of intended goals of the system as a dialogue participant. We introduce an under-specification strategy for controlling the preference of the concurrent behaviors. This is in contrast to the classical, top-down approach to dialogue coordination.

1. INTRODUCTION

Interactive media are increasingly becoming populated by autonomous software agents which will be able to engage in real-time spoken dialogues with the users. In these applications, interactional rhythm of turn taking, back-channel response in natural dialogue plays an important role in maintaining both the flow of the conversation and a conversational field. Unfortunately, the majority of spoken dialogue systems developed did not adequately address the problems of dialogue coordination. One reason for this situation is that research efforts in spoken dialogue systems have been concentrated on dialogue management that is crucial both for navigating user's behaviors to achieve their goals and for reducing active vocabularies of the speech recognizer. However, the explicit dialogue management tends to force the users a rigid interaction.

In this paper, we present a general architecture for maintaining dialogue coordination in spoken dialogue systems, in which intended behaviors and goals are incrementally performed during the course of maintaining dialogue coordination. The dialogue structure emerges as a result from interaction between user and the dialogue system.

2. EVERYDAY DIALOGUE

In general, the study on spoken dialogue systems has concentrated on task-oriented or goal-oriented dialogue performing human-computer interaction in the question-answering style. Our project focuses rather on the construction of an interaction manner in our everyday dialogue that engages in social exchanges with maintaining dialogue coordination.

The basic framework of spoken dialogue systems has been based on so called "coding/decoding model" in which they are designed to interpret the meanings of other's utterance correctly, and to produce adequate responses to other. In the model, we tacitly assume that each utterance conveys complete meanings between dialogue participants in which the meanings are given by the speaker in one-sided way. As a result, the interaction style in a conversation with humans becomes imperative tone of voice lake a command-based interaction with computers that is far from that of familiar participants in our everyday conversation. As one of features of our everyday conversation and activities to distinguish from rigid human-computer interactions, we are focusing on the nature of our spontaneous behaviors that tend to entrust their meanings and values to their environment.

2.1 Entrusting Behavior

We first illustrate the basic framework for our everyday activities from a view of ecological psychology. While taking a walk, nobody would think the meanings and the values of each step prior to every step. We seem to aware the meanings and the values of our behaviors during interacting with our environment. From an ecological view of point, our behaviors are navigated by the "information" emerged from interactions between the action and its environment. In our spontaneous behaviors, since we are not able to see the emerging meanings of our behaviors, we try to entrust the meanings of the behaviors to our environment in order to find them with our prospective awareness. We call the basic strategy of our spontaneous behaviors "entrusting behavior". On the other hand, the role of environment that embodies the meanings of behaviors is called "grounding". Our spontaneous behaviors are organized from these two processes; entrusting behavior and grounding.

In general, our skilled, coordinated behaviors are not always navigated by well-prepared plans, rather they are regulated by routine activities with coordinated action-perception cycle with their environment. For instance, well-trained car drivers would not try to investigate the meanings of turning their steering wheel, rather they could aware the meanings from relations between unintentional motions of the steering and changes of appearance of the outside. The unintentional motions are a kind of entrusting behaviors, and the changes of appearance according to the motions work as grounding process to the meanings and the values of the entrusting behaviors.

2.2 Entrusting Behaviors in Social Interaction

The perspective to our spontaneous behaviors can be extended to our social interactions like our everyday conversation. Assume that you say "Hello" to your colleague as a greeting in the morning. If your colleague went away in complete disregard of your greeting, your "Hello" would not perform as a greeting you expected. The meanings and values of your utterances are supported by the responses of your dialogue partner. When you spontaneously utter a sentence to other, you have a prospective awareness for the content of other's response and the emerging meanings of the utterance. However, you would not see the complete meanings prior to getting a response, so that your utterances are always produced speculatively. These speculative utterances are regarded as a kind of entrusting behaviors that entrust the meanings and values of the utterances to social others. Simultaneously, these behaviors are to require your dialogue partner to be a conspiracy who supports and shares the emerging meanings during exchanging the utterances.

2.3 Conversational Field

Conversational field between dialogue participants is emerged from the strained relation between the prospective awareness and the emerging meanings for the entrusting behaviors. In our everyday conversation, its implicit, primary goal would be to maintain their bound relation as a conspiracy performing a joint activity, as well as to convey messages to each other, and share them. It is a kind of conversation that we intend to enjoy doing itself. We call it "self-motivated dialogue" in contrast with a goal-oriented dialogue. According to dialogue situations and participants, we would choose communication manners such as formal/informal conversations, debate/casual chattings and mother-infant interactions. In casual chattings, the nature as a self-motivated dialogue becomes dominant, and we seems to understand other's thought by using tacit communication through a joint activity to maintain the conversational field.

It is noted that we unconsciously feel a sort of responsibility of responding other's approaches to us. It is too hard to ignore our colleague's greeting to us as an intended attitude. However, artifacts such dialogue systems would not feel the responsibility to respond our utterances for grounding the meanings. In order to generate an analogous interaction for human-human conversation, dialogue systems have to work to maintain a mutually regulated interaction. Here, the role of dialogue coordination is to maintain the conversational field between human and system.

3. SITUATED DIALOGUE COORDINATION

The focus of this research is to construct models of dialogue coordination for social agents. Here, a behavior-based approach is taken to model dialogue coordination in natural interaction with humans.

The modeling of dialogue coordination by a behavior-based approach is motivated by two considerations. Firstly, dialogue coordination is a highly skillful behavior in our everyday activities. The behavior cannot be achieved by well-prepared plans and scripts. Secondly, the communicative acts of turn-taking, interactional rhythm, and spontaneous self-repair may arise from emergent properties of concurrent interactions of primitive behaviors situated in the conversational field.

3.1 Dialogue Coordination as Skilled Behaviors

In this model, the concept of a conversational field is organized as a coordinating, self-sustained structure between a subsumption architecture for the selection of primitive behaviors and unpredictable human behaviors. A set of these primitive behavior is also called multiple situated-agent that has a primitive role or local goal. In order to pick up information about opportunities to behave, each situated-agent tries to entrust the meanings of the behavior to its environment that is composed of its surrounded situated-agents and real environment. And then, the behavior of each situated-agent is navigated by the opportunities emerged as a result of local interactions among multiple situated-agents. The mechanism, so called "emergent computation", provides a kind of the field for novel constraint satisfaction in the unpredictable environment. Skilled, coordinated behaviors can be navigated by this internal field emerged from interactions among situated-agents. The fundamental idea of situated dialogue coordination is inspired from the mechanism for an emergent computation using multiple situated-agents. The dialogue coordination can be achieved by creating and maintaining a self-sustained field with humans.

3.2 Architecture for Emergent Computation

The realization of the emergent computation is based on a spread activation network of behavior modules (= situated-agents) similar to that of Maes[2]. We are applying the architecture to represent the entrusting behavior and its grounding process in our everyday dialogue, as I have mentioned before.

Figure 1. shows a schematic view of our model for dialogue coordination. The overall architecture for a dynamic action selection consists of a set of behavior modules (= situated-agents), and two types of internal contexts: intentional context and environmental context. The intentional context

has primitive goals and motivations of the agent that is referred as a resource for dynamic action selection, and the environmental context is including dialogue histories and events from external world such as social others.

Each behavior module has a primitive motivation and local constraints as preconditions in achieving its act. Basically, these modules have a chance to be activated when all of their preconditions are satisfied and the activation level of the module exceeds an activation threshold. As well as behaving in event-driven style, a behavior module autonomously acts in order to be satisfied its precondition. By using a spreading activation, a behavior module affects its surrounding behavior modules; that is, it activates behavior modules that make up the preconditions, and inhibits them that interrupt to do so. These autonomous behaviors can be regarded as a kind of entrusting behavior that entrusts the opportunities of activity to its environment consists of other behavior modules. The grounding in the behavior module is achieved when its preconditions are all satisfied. As a result, these sustained relations organized by multiple goals and motivations of every behavior module construct a field that navigates skillful coordinated behaviors.

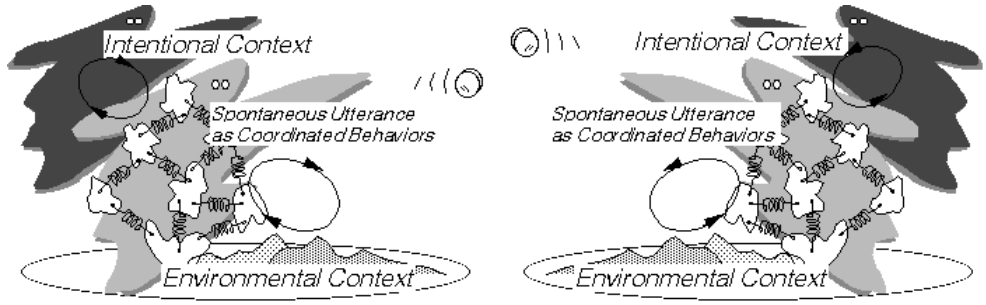

Figure 1. Schematic view of the model for dialogue coordination, and for under-specified control for dialogue flow.

3.3 Under-Specified Control for Dialogue Flow

One distinguishing feature in organizing behaviors in this architecture is that specified goals and motivations do not always regulate to organize behaviors. Rather, rational behaviors are organized in bottom-up way, and goals and motivations to contribute the organizing behaviors are revealed as the result of the local interactions and competitions. This nature of emerging motivations plays an important role in modeling self-motivated dialogues in contrast with goal-oriented dialogue with goals and motivations prepared in advance.

In order to regulate goals and motivations emerging during the dialogue, and that are prepared in advance, the same architecture mentioned above can be used as layered sets of primitive behaviors. And, to make the behavior of the system social, in general, the maintaining of dialogue coordination takes priority over the realization of intended goals of the autonomous agent as a dialogue participant. We introduce an under-specification strategy for controlling the preference of the concurrent behaviors in order to control dialogue flows.

In this architecture it should be noted that the lower-level behaviors act autonomously for maintaining the dialogue coordination with humans and are linked to the specifications from higher-level behaviors for dialogue management. That is, the behaviors of higher-level layer provide constraints for the intentional context of lower-level layer in the way of under-specification. The intended goals and motivations are incrementally performed during the course of maintaining dialogue coordination. This is in contrast to the classical, top-down approach to dialogue coordination.

4. IMPLEMENTATION

There are two ways to investigate the mechanism and its nature of the dialogue coordination proposed here. Firstly, it is called “constructive approach to dialogue phenomena” in which we reconstruct the conversational field emerged from two communicating autonomous agents. Secondly, it is a way to understand its properties by constructing the conversational field in interaction between the autonomous agent and human.

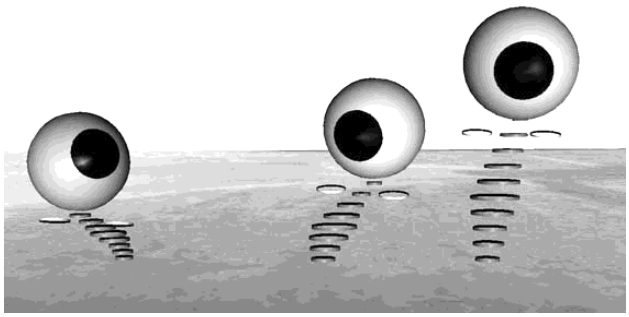

Figure 2. Overview of autonomous communicating creatures, Talking Eye.

4.1 Talking Eye

A demonstration system for the autonomous software agent called "Talking Eye" has been created for these purposes. Figure 2. shows an overview of the autonomous agents, Talking Eye. These agents are able to communicate each other, and to communicate with a human in an external environment through speech and vision. The behavior engine of Talking Eye consists of three submodules: perceptual system, behavior system and motivational system. The perceptional system extracts vocal and visual events from the behaviors of social others. The behavior system realizes these behaviors as motions of eyeball and utterances.

Motions generated by the graphical animation of a disembodied eyeball are used to convey information about attention, motivation, and to provide social signals such as head turns and nods of human. For spoken utterance generation, about 400 listed phrases are used for real-time responses that include various communicative acts such as requesting, informing, warning, suggesting, confirming, back-channel response, agree/disagree. In current implementation, the behavior system of the agent defines around 120 primitive behaviors both for generating motions and utterances, and for regulating goals and motivations in the motivational module.

These motions and utterances are generated from the behavior system of Talking Eye. And then, these events are conveyed to the environmental context of other's autonomous agents, that are cues to organize a sequence of behaviors. The conversational field, and interactional rhythm emerged in these agents has self-sustained, self-regulated properties.

Figure 3. Talking Eye as an interactive system.

4.2 Talking Eye as Interactive System

These features are investigated as an interface agent or an interactive system (Figure 3), which is sufficiently general to be used as a basis for application in different domains [5].

Speech recognition for the generation of inputs for the multiple situated-agents (behaviors) is performed using the continuous speech recognizer. The current version has a vocabulary size of 300 words. By using syntactical templates, partial phrases are interpreted into content word and modality part including sentence final particles and adjectives. These are fed into perceptual system of the behavior engine.

The vision system is also implemented based on the behavior-based vision methodology using active cameras. The system is capable of detecting visual events such as nodding and expression of agree or disagree from the movement and direction of human face. Situated agents take these visual cues as inputs for the cooperative management of dialogue coordination in interactive dialogue.

The approach is demonstrated that is computationally simple, and provides robust performance in human-computer interaction. Using the architecture, natural interaction with the human can be achieved without any explicit modeling of dialogue coordination. The interactions among these primitive behaviors within the conversational field results in the emergence of socially adaptive behaviors, such as hesitations with self-repairing behaviors for maintaining interactional rhythm, and dialogue coordination[4].

5. CONCLUSION

In this paper, we have discussed on dialogue coordination from the following two aspects: (a) the coordination between a grounding process for other's utterances and a regulating process for utterances as entrusting behaviors to other, (b) the coordination of goals and motivations that are prepared in advance, and that are emerged during the interactions with social others. Although the details of the behaviors engaged in the current system are beyond the scope of this paper, this work represents an important step toward realizing autonomous agent that can maintain dialogue coordination in social interaction with humans. This approach provides a basis for implementing a successful coordination among the modalities for maintaining dialogue coordination in spoken dialogue systems.

6. REFERENCES

1. Cassell, J. et al "Modeling the Interaction between Speech and Gesture", *Proc. of the 16th Annual Conference of the Cognitive Science Society*, 1994.
2. Maes, P. "Situated Agents Can Have Goals", *Robotics and Autonomous Systems*, Vol.6 pp.49-70 1990.
3. Reed, E. *Encountering the World, Toward an Ecological Psychology*, Oxford University Press, 1996.
4. Okada, M. *Hesitating Computer* (in Japanese), Kyoritsu Syuppan, 1995.
5. Suzuki, N., K. Ishii and M. Okada "Organizing Self-Motivated Dialogue with Autonomous Creatures", *Proc. of ICSLP-98*, 1998.