

Speech-Based Conversation Environment for Dynamic Knowledge Interaction

Tomohiro FUKUHARA
Synsophy Project*/
Nara Institute of Science and Technology

*Communication Research Laboratory
Ministry of Posts & Telecommunications
Kobe, Hyogo 651-2401, Japan
tomohi-f@crl.go.jp

Toyoaki NISHIDA
The University of Tokyo**/
Synsophy Project

**School of Engineering
The University of Tokyo
Bunkyo-ku, Tokyo 113-8656 Japan
nishida@kc.t.u-tokyo.ac.jp

Abstract

We propose a speech-based conversation environment called Voice Cafe for facilitating dynamic knowledge interaction. Dynamic knowledge interaction is a key concept for creative interaction. Conversation is one of important interactions that facilitates dynamic knowledge interaction. In Voice Cafe, users can talk with conversational agents being equipped with automatic speech recognition (ASR) and text-to-speech (TTS) systems. Users can listen to a conversation between agents, and join in their conversation by speech interaction. Design concepts and an overview of Voice Cafe are described.

1 Introduction

Interaction plays an important role in our creative work. We often hit on new ideas or solutions to problems in casual conversation such as talking with friends or colleagues at cafe or bar. This kind of interaction is important because it often includes clues to solve problems. Dynamic knowledge interaction is a key concept for creative interaction. Nishida described that co-evolution of human and knowledge networks is important for knowledge-creating community[1]. We derive creativity from dynamic knowledge interaction that blends dynamic knowledge of humans and static knowledge of documents.

Conversation is one of important interactions for facilitating dynamic knowledge interaction. We can exchange not only questions and answers, but also intuitive ideas and thoughts on problems. This kind of knowledge helps our divergent thinking that is a way of thinking seeking variety of

answers to problems¹.

Talking and listening are important interactions in a conversation. By talking something to others, we can convert our abstract thoughts into concrete ones because we are forced to explain our thoughts. By listening to the others' talks, we can convert concrete thoughts such as stereotypes or prejudices into elastic thoughts because we can imagine variety of things from the others' speaking. With talking and listening, we can broaden or deepen own viewpoints and thoughts, and finally we can get new ideas.

Previous works has been done to solve a problem on how to facilitate a creative conversation. Nishimoto described a conversational agent that facilitates creative a conversation between humans[3]. Advantage of this agent is that the agent monitors a conversation between humans, and suggests new information to them. However, this agent has limitations on the situation that the agent is used. The situation of this agent is limited to at on-line. Furthermore, interaction method is limited to character-based interaction. These limitations are not suitable for facilitating a daily informal conversation. Casual and natural interaction method for human is needed.

We propose a speech-based conversational environment called Voice Cafe for facilitating dynamic knowledge interaction. In Voice Cafe, users can listen to a conversation between agents, and join in their conversation. Each agent is equipped with automatic speech recognition (ASR) and text-to-speech (TTS), and talk with users and the other agent by using synthetic speech. We discuss design concepts and an overview of Voice Cafe in the following sections.

¹ The definition of divergent thinking is described in [2].

2 Design concepts for creative conversation

We describe merits of speech-based interaction for a creative conversation, and design concepts of Voice Cafe.

2.1 Merits of speech in a creative conversation

The reasons why speech-based interaction is important for a creative conversation are described from a viewpoint of differences of interaction method such as speech and keyboard.

1. Speech is one of natural interactions for humans.
2. Input speed of speech is faster than that of keyboard.
3. Speech does not limit the situation of use.

First, speech is one of natural interaction ways for humans. It needs little training or knowledge to use speech. In contrast, keyboard is not innate interaction method, and requires much training to make full use of it. Cognitive and physical burdens for keyboard are much higher than speech.

Second, input speed of speech is faster than that of keyboard. Diehl and Stroebe pointed out that there are factors causing process loss in a creative conversation such as brainstorming[4][5]. One of factors that causes process loss is *blocking effect*. Blocking effect is a phenomenon that a user forgets her ideas while she is waiting for her turn to speak in a group conversation. Similarly, input speed may become a factor of process loss because a user forgets his or her ideas if s/he is unable to input his or her ideas speedy. Thus, speedy input of speech is suitable for a creative conversation.

Finally, speech-based interaction does not limit situations of use. Speech-based interaction system does not need special devices to input. By using speech, we can tell new ideas, which flashed into his or her mind, to the system at anytime and anywhere. Although there are situations that using the systems is forbidden², speech-based interaction systems basically allow us to use them in casual situations such as while walking down

² Such as in concert hall or in buses or trains. Using cellular phone is forbidden or refrained in public transportations in Japan.

the street. This is important feature because we can utter our ideas on the spot when we hit on new ideas.

2.2 Design concepts of Voice Cafe

Following are the design concepts of Voice Cafe.

1. Listening to conversations between agents is allowed.
2. Joining in conversations between agents is allowed.
3. Various kinds of topics are provided for users.

Listening to someone's talk is important for a creative conversation. We can find differences and similarities of thoughts by listening to someone's talk. In Voice Cafe, a function that users can listen to a conversation between agents is implemented. Each agent talks to the other by associating with a keyword that the other suggested. Our aim here is that users can get various viewpoints from unexpected keywords suggested by agents.

Joining in a conversation between agents is important for creative conversation. The purpose of this concept is to prevent a user to forget his or her ideas. The ideas that a user hits on should be externalized immediately, otherwise s/he will forget his or her ideas. We plan to implement a function that users can interrupt and join in the conversation between agents. We aim to solve the blocking effect this concept.

Finally, various topics are provided for users. We implemented topics on city information on Akashi³, and travel information around the world. We aim that these topics cause lively conversations between users and agents.

3 Voice Cafe

We describe a prototype system called Voice Cafe. Voice Cafe enables users to listen to a conversation between agents, and join in their conversation. An architecture of Voice Cafe and examples of conversations are described.

3.1 Architecture of Voice Cafe

Figure 1 shows an overview of Voice Cafe. Voice Cafe comprises two note PCs that each is

³ Information on Akashi city is described at http://web.pref.hyogo.jp/kankou/frame/ef_tra.html



Figure 1: Overview of Voice Cafe.

equipped with ASR, TTS, and a set of microphone and speaker. Each note PC plays a role of conversational agent. A set of microphone and speaker is used for capturing and sounding their speech. All conversations between agents are talked by synthetic speech.

Two types of conversations are prepared. One is *associative conversation*, and the other is *travel chat*. In associative conversation, users can listen to a conversation between agents. Each agent associates with a keyword that the other suggested. In travel chat, users and agents talk with each other on travel information such as world heritages⁴, foods and drinks in the country, and so on. In this conversation, users can join in a conversation between agents.

3.2 Associative conversation

An Example of this conversation is shown in Table 1. There are two types of agents in the table, i.e., *association agent* who associates new keyword with original one, and *explanation agent* who explains the keyword. In the table, Kenji plays the association agent, and Naoko plays the another. Conversation between association and explanation agents is performed as following steps.

1. User suggests a keyword at the beginning.
2. Association agent recognizes the user's keyword, and associates new keyword with the utterance.

⁴ <http://www.unesco.org/whc/>

Table 1: Example of associative conversation.

| Talker | Utterance |
|--------|--|
| User | <i>What do you associate with Akashi?</i> |
| Kenji | Talking of <i>Akashi</i> , I associate <i>the Akashi strait bridge</i> with it. |
| Naoko | The Akashi strait bridge is the longest suspension bridge in the world. |
| Kenji | Talking of <i>the Akashi strait bridge</i> , I associate <i>Akashi strait</i> |

3. Description agent recognizes the utterance of the other, and associates a description for the keyword.
4. Repeat the process between agents.

3.3 Travel chat

In this mode, users can join in a conversation between agents. Figure 2 shows a scene in which users and agents are talking with each other. Conversation in this mode is controlled by *script*, which is an independent scenario describing which agent (including users) to talk, and what utterance to be talked (contents of the speech). An example script is shown in Table 2. Italicized words are recognized by ASR, and the line including ‘#’ represents a comment line. In the table, there is a break point where users can join in the conversation. Users can join in the conversation by calling a name of either agents at the break point.

4 Discussion

We discuss on the following points in this section.

1. Is Voice Cafe just a Q-A system?
2. Are the implemented conversations effective for facilitating creative conversation?

First, Voice Cafe is not just a Q-A system because there's a big difference between Voice Cafe and Q-A systems. The difference is that agents can talk with each other such as a conversation shown in Table 2. In case of the Q-A system, users only query databases and get an answer. There is little interactions between users and databases in Q-A system. On the other hand, there are much more interactions in Voice Cafe.



Figure 2: Conversation with agents.

Users can not only acquire answers to queries, but also join in the conversation, and tell their questions or thoughts to agents. Although, users do not always acquire a response from agents, casual situation in which users can tell their thoughts and opinions is important for creative work.

Second, the effectiveness about implemented conversations is further research issue. The aim of the implemented conversations is to arouse users' interests, and encourage the users to join in a conversation between agents. We suppose that users who are familiar with the contents of the conversation can find new viewpoints by finding similarities and differences of thoughts between agents and users. This causes further conversation between users and agents. On the other hand, we suppose that users who are not familiar with the contents can find a lot from the conversation, and they will have more interests to the conversation. We will evaluate the effectiveness of the implemented conversations.

5 Conclusion

We proposed Voice Cafe, which is an environment for facilitating dynamic knowledge interaction. Casual conversation such as speech interaction is important for facilitating creative work. We proposed design concepts and an architecture of Voice Cafe. In Voice Cafe, users can listen to a conversation between agents, and join in their conversation. We will evaluate the effectiveness of the implemented conversations, and investigate

Table 2: Example script in travel chat.

| Talker | Utterance |
|--------|--|
| Naoko | <i>Kenji</i> , I've been to <i>Mexico</i> last year. |
| Kenji | Really? Did you visit to <i>Teotihuacan</i> ? |
| Naoko | No. <i>What is it?</i> |
| Kenji | Why didn't you visit that sights? |
| | # Break point (User can join in the conversation.) |
| User | Hey, <i>Kenji</i> . |
| Kenji | Yes? |
| User | <i>Will you tell me about it?</i> |
| Kenji | OK. <i>Teotihuacan</i> is one of pre-Hispanic cities . . . |

the situation for facilitating creative work.

Acknowledgement

The authors thank Dr. Asako Miura and Dr. Nobuhiko Fujihara for valuable discussions.

References

- [1] Nishida, T., ed., "Dynamic Knowledge Interaction", CRC Press, Boca Raton FL, 2000.
- [2] Murakami, Y., "Comparative Study of Various Creation Techniques: A New Synthesis", *Journal of Japan Creativity Society: Theories and Methods of Creativity*, Vol. 1, pp.196–206, Tokyo, 1983 (in Japanese).
- [3] Nishimoto, K., Sumi, Y., and Mase, K., "Enhancement of Creative Aspects of A Daily Conversation with A Topic Development Agent", In *Coordination Technology for Collaborative Applications: Organizations, Process, and Agents*, Vol. 1364, Lecture Notes on Computer Science, Springer-Verlag, pp.63–76, 1998.
- [4] Diehl, M. and Stroebe, W., "Productivity loss in brainstorming groups: Toward the solution of riddle", In *Journal of Personality and Social Psychology* 53, pp.497–509, 1987.
- [5] Diehl, M. and Stroebe, W., "Productivity loss in idea-generating groups: Tracking down the blocking effect", In *Journal of Personality and Social Psychology* 61, pp.392–434, 1991.