A Mechanism for Smooth Conversation

Kohji Dohsaka[†], *Norihito Yasuda, Noboru Miyazaki, Mikio Nakano, and Kiyoaki Aikawa*

Abstract

Spoken dialog systems provide an interface where computers communicate with their users in everyday speech. A weather information system “HUME” is a spoken dialog system equipped with a new mechanism for establishing smooth conversation between computers and their users. This mechanism allows a spoken dialog system to convey information to users in as short a dialog as possible depending on the system’s database. The results of dialog simulations prove the method’s effectiveness. The shortened dialogs afforded by this mechanism bring the computer closer to everyday speech.

1. Introduction

An interface where computers communicate with their users in everyday speech would indeed be one very familiar to us all. Spoken dialog systems perform tasks like retrieving information and making reservations in a collaborative way through speech communication with their human users. NTT Communication Science Laboratories studies spoken dialog systems with the aim of establishing smooth conversation between computers and their users.

This paper addresses the issue of dialog smoothness. A smooth dialog is one in which a system and its user can communicate their intentions at all times during the dialog and they can converse in an efficient manner. Dialog smoothness has been hindered by the fact that conventional systems have difficulty changing the dialog flow when the user interrupts the system while it is speaking, which prevents users from obtaining the necessary information as rapidly as they would like.

Another problem is that, due to speech recognition errors, a system must continually confirm that it understands the content of a user’s query. Otherwise, the system could misunderstand the user’s query and convey the wrong information. Although confirmations are helpful in avoiding misunderstandings, unnecessary confirmations by the system should be avoided because they interfere with the smooth flow of the dialog. Confirmation of some portion of a query may be unnecessary if the system can utilize the contents of its database. Conventional methods have problems confirming any part of the content of a query even in such cases.

NTT Communication Science Laboratories has developed a spoken dialog system called “DUG-1” that overcomes the first problem [1]. DUG-1 can respond to interjectory user utterances, like “yes” and “right”, and user interruptions in a timely manner. It does so by utilizing a novel language understanding method called “incremental understanding” [2] and a novel language generation method called “incremental production” [3]. Both these methods were developed by NTT Communication Science Laboratories. Incremental understanding is a method for analyzing user utterances incrementally, which is accomplished by means of the partial parsing technique and a multi-context model to deal with ambiguities in understanding dialog states. Incremental understanding enables a system to comprehend a user query before the user completes an utterance and to make acknowledgements in the middle of user utterances. Users can make a query while recognizing that the system understands their utterances. Incremental production

† NTT Communication Science Laboratories
Atsugi-shi, 243-0198 Japan
E-mail: dohsaka@atom.brl.ntt.co.jp

Fig. 1. Scene where a user is conversing with HUME.

allows a system to respond in a short utterance unit by means of a hierarchical planning technique while it records what information has been conveyed. This makes it possible for the system to change the flow of a dialog according to interruptions by the user in the middle of its utterances. As a result, the user can obtain the desired information rapidly.

After developing DUG-1 and thereby solving the first problem, we concentrated on solving the second problem and developed a spoken dialog system called "HUME". This inherits the mechanism for timely and adaptive responses from DUG-1 and features a novel approach to dialog control called the dual-cost method (4). The dual-cost method allows HUME to convey the desired information in as short a dialog as possible based on what is in its database. Figure 1 shows a scene where a user is conversing with HUME.

2. HUME

HUME is a weather information system that converses with users by spoken dialog. It recognizes five hundred words. The system's database, which is updated daily, stores information about the weather, temperature, and probability of rain forecast for the next week, and current warnings issued for all prefectures and major cities in Japan. Figure 2 shows the system architecture. The system consists of a speech understanding module, a dialog control module, an utterance production module, and the weather information database. The speech understanding module recognizes users' queries from their speech and records the current system's understanding of a query. The dialog control module decides the system's action based on the system's understanding at each point of a dialog. Examples of system actions are: i) confirming a user's query as in "Are you interested in Kanagawa?", ii) soliciting information from the user as in "Which city are you interested in?", iii) or conveying information to the user as in "No warnings have been issued for anywhere". According to the action decided upon by the dialog control module, the utterance production module generates a series of natural language phrases and outputs them as speech.

3. Efficient dialog depending on the system's database

Speech recognition errors force a system to confirm a user's query. The system can determine the content of a query only by means of a user's acknowledgement, like "yes", to the system's confirmation. After determining the content of a query, the system makes a response to convey information to the user. A dialog thus consists of the sub-dialog for confirmation and the system's response afterwards. Although the sys-

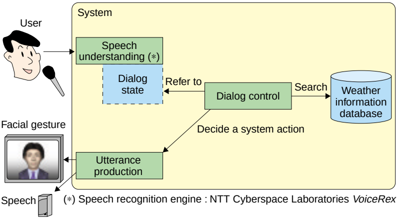

Fig. 2. HUME system architecture.

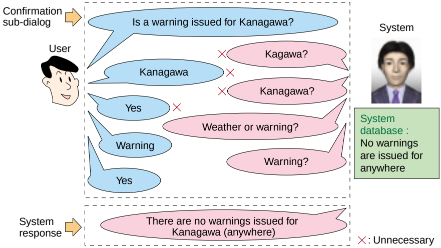
Confirmation sub-dialog
User
System
Is a warning issued for Kanagawa?
Kagawa?
Kanagawa
Kanagawa?
Yes
Weather or warning?
Warning
Warning?
Yes
System database : No warnings are issued for anywhere
System response
There are no warnings issued for Kanagawa (anywhere)
×: Unnecessary

Fig. 3. Inefficient dialog.

tem's confirmations are helpful in dealing with speech recognition errors, it would be better to avoid unnecessary confirmations because they hinder dialog smoothness.

Conventional dialog control methods concentrate on reducing the length of a confirmation sub-dialog and do not allow for the length of the system's response after a confirmation. They make unnecessary confirmations since the length of a confirmation sub-dialog runs counter to that of a system's response. The lengths of both a confirmation sub-dialog and the system's response must be taken into consideration in order to reduce the total dialog length. The length of a system response depends on the database content. Therefore, the system must control the dialog so as to reduce the total dialog length as much as possible according to its database.

Figure 3 shows an example of an inefficient dialog that conventional methods might carry out. Assume no warnings have been issued for anywhere and this data is stored in the database. The user asks whether a heavy rain warning has been issued for Kanagawa prefecture. The system first confirms the place in question and then confirms whether the information the user wants is about weather in general or about warnings. However, the system actually does not need to confirm the place because once it has confirmed that the user is asking about warnings, it could answer regardless of the place because no warnings have been issued for anywhere. If several warnings have been issued for different places, then the system must, of course, confirm the place. But doing so when it is irrelevant makes the system's response unnecessarily long, which is frustrating for the user. Instead, the system should control the dialog depending on the content of the system's database.

4. Efficient dialog control: dual-cost method

We have developed a dialog control method called the dual-cost method, which enables the system to avoid unnecessary confirmations, depending on the content of its database. The dual-cost method controls a dialog so as to minimize the sum of the length of a confirmation sub-dialog (confirmation cost) and the length of the system response (information transfer cost). In HUME, the confirmation cost is estimated as the number of items the system must confirm, and the information transfer cost is estimated as the number of content words in the system's response. We have a more refined method of estimating confirmation cost based on the speech recognition rate. The method implemented in HUME is a simplified version of the refined one.

The dual-cost method considers all dialog plans that are possible within the system's understanding at each point of a dialog. A dialog plan is a sequence of system actions, and it represents the way the system confirms the content of a currently recognized user query and then responds to the user. The dual-cost method computes the confirmation and information transfer costs for each dialog plan. These costs have roughly a trade-off relationship. The dual-cost method chooses a dialog plan that yields the minimum sum of the two costs. According to the chosen plan, the system takes the next action.

For example, in Fig. 3, after the system has recognized that the user's query is about warnings, it com-

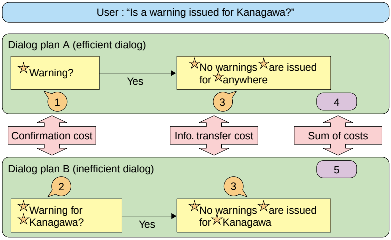
User : "Is a warning issued for Kanagawa?"
Dialog plan A (efficient dialog)
☆Warning?
Yes
☆No warnings ☆are issued for ☆anywhere
1
3
4
Confirmation cost
Info. transfer cost
Sum of costs
Dialog plan B (inefficient dialog)
5
2
3
☆Warning for ☆Kanagawa?
Yes
☆No warnings ☆are issued for ☆Kanagawa

Fig. 4. Example showing how the dual-cost method works.

pares two dialog plans. In plan A, the system confirms only the type of information that the user is interested in and then makes a response. In plan B, it confirms both the type of information and the place that the user is interested in and then makes a response. Plans A and B are shown in Fig. 4.

Since the fact that no warnings have been issued for anywhere is stored in the database, the information transfer costs of plans A and B are equal. The confirmation cost of plan B, however, is higher since the system must confirm the place. The dual-cost method therefore chooses plan A, which yields the minimum sum of the costs, and the system avoids unnecessary confirmation about the place.

When several warnings have been issued for different places and this information is stored in the database, plan B is chosen because plan A would increase the information transfer cost. As described above, the dual-cost method allows the system to choose an appropriate dialog action according to its database and to avoid unnecessary confirmations that conventional methods cannot get around.

5. Evaluating the dual-cost method

We evaluated the performance of the dual-cost method in dialog simulation experiments where a system and an agent simulating a user converse to perform a given task. The communication was done in such a way that they exchanged internal representations of the contents of their utterances. The internal representations were written as a set of attribute-value pairs. We controlled the recognition rate of attributes included in the user utterances and intentionally caused system misrecognition of the user utterances according to the given recognition rate.

The task was a weather information query. The simulated user asked about the weather, temperature, rain probability, or current warnings by specifying four parameters: place, date, type of warning, and type of information. The database contained the names of fifty cities, two dates, such as today and tomorrow, ten types of warnings, and four types of information. It stored the information about the weather category, predicted highest and lowest temperatures, and rain probability for six-hour periods on each date for each place. It also stored the data that no warnings had been issued anywhere.

We compared two conventional methods, the lump-sum method and the piecemeal method, with the dual-cost method. These conventional methods attempt to confirm a user query independently of the system's database. The lump-sum method tries to confirm as many items as possible at once. The piecemeal method tries to confirm items one by one. The confirmation strategies exploited in other conventional methods are generally some combination of these methods.

Figure 5 shows a result when the simulated user enquired about a warning. For each recognition rate, 2000 simulation dialogs were carried out. The length of the dialog was measured as the sum of the content words that the system and the user exchanged. The results indicate that the dual-cost method could perform the task using a shorter dialog than the conventional methods.

When a user enquires about temperature or rain probability, even the dual-cost method cannot dispense with unnecessary confirmations, because the system must confirm all attributes specified by the

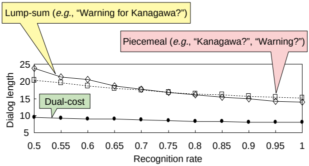

Fig. 5. Evaluation result: cases where a user enquires about a warning.

user in this case. However, other simulations have indicated that, even in such cases, the dual-cost method is not any worse than the conventional methods.

6. Conclusion

We have explained the mechanism for a spoken dialog system and its user to converse in a smooth manner. The dual-cost method enables an efficient dialog depending on the system's database. This method is incorporated into a weather information system HUME. Dialog simulation experiments verified the method's effectiveness. The technologies explained in this paper can be applied to a variety of other tasks, like reservations and information guidance, performed using everyday speech. In the future, we will broaden the application range of the spoken dialog interface.

References

[1] M. Nakano, K. Dohsaka, N. Miyazaki, J. Hirasawa, M. Tamoto, M. Kawamori, A. Sugiyama, and T. Kawabata, "Handling rich turn-taking in spoken dialog systems," Eurospeech99, Budapest, Hungary, pp. 1167-1170, 1999.

[2] M. Nakano, M. Miyazaki, J. Hirasawa, K. Dohsaka, and K. Kawabata, "Understanding unsegmented user utterances in real-time spoken dialog systems," ACL-99, College Park, MD, USA, pp. 200-207, 1999.

[3] K. Dohsaka and A. Shimazu, "A computational model of incremental utterance production in task-oriented dialogs," COLING-96, Copenhagen, Denmark, pp. 304-309, 1996.

[4] K. Dohsaka, Y. Yasuda, N. Miyazaki, M. Nakano, and K. Aikawa, "An efficient dialog control method under system's limited knowledge," ICSLP2000, Beijing, China, Vol. 2, pp. 739-742, 2000.

[5] http://www.brl.ntt.co.jp/cs/dug/dug1/index.html

[6] http://www.brl.ntt.co.jp/cs/dug/hyumu/index.html

Kohji Dohsaka

Senior Research Scientist, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.S. and M.S. degrees in information and computer science from Osaka University, in 1984 and 1986, respectively. In 1986, he joined NTT Corporation where he has worked on natural language processing and spoken dialog systems. He is a member of the Association for Computational Linguistics, the Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers, and other academic societies. He received the 1997 Best Paper Award of the Information Processing Society of Japan.

Norihito Yasuda

Researcher, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.S. degree in integrated human studies and the M.S. degree in human and environmental studies from Kyoto University, in 1997 and 1999, respectively. In 1999, he joined NTT Communication Science Laboratories, NTT Corporation, Tokyo. He is a member of the Acoustical Society of Japan.

Noboru Miyazaki

Researcher, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.S. degree in information engineering and the M.S. degree in intelligence science from Tokyo Institute of Technology, in 1995 and 1997, respectively. In 1997, he joined the Basic Research Laboratories (now NTT Communication Science Laboratories), NTT Corporation, Tokyo. He is a co-recipient of both the Best Paper Award and the Inose Award from the Institute of Electronics, Information and Communication Engineers in 2001. He is a member of the Acoustical Society of Japan and the Japanese Society for Artificial Intelligence.

Mikio Nakano

Associate Manager, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.A.S degree in pure and applied sciences, the M.S. degree in coordinated sciences, and the Sc.D. degree in information science from the University of Tokyo, in 1988, 1990, and 1998, respectively. In 1990, he joined NTT Corporation, where he has been working on natural language processing and spoken dialog systems. He is a member of the Association for Computational Linguistics, the Association for Computational Machinery, International Speech Communication Association, and other academic societies.

Kiyoaki Aikawa

Senior Research Scientist, Supervisor, Group Leader, Media Information Laboratory, NTT Communication Science Laboratories.

He received his Ph. D. from the University of Tokyo in 1980. He joined NTT Basic Research Laboratories in 1980. He was a visiting scientist at Carnegie Mellon University, PA, USA, in 1990. From 1992 to 1995 he was a senior researcher in ATR Laboratories. After staying in NTT Human Interface Laboratories from 1996 to 1998, he joined NTT Communication Science Laboratories in 1999. He is a member of the Institute of Electrical and Electronics Engineers, Acoustical Society of America, and Acoustical Society of Japan. He received the Sato Award from the Acoustical Society of Japan. He received the Telecom-System Technology Award form the Electrical Communication Foundation.