Dialogue Model on Corpora for Several Tasks and Interface in a Dual Task Situation

Katsuhiko SHIRAI
(Research Assistants : Shu NAKAZATO and Shigeki OKAWA)
School of Science and Engineering, Waseda University

1. Introduction

It is important for designing a spoken dialogue system to consider the situation of user. The dialogue mode strongly depends on the task or the environment. Even in the same environment, pragmatic and syntactic forms of dialogues may be different of the goals of the tasks. And also the same is true for the variation of the user's environment. For example, in the human interface of car-navigation system, the requirement is different depending on whether the driver is driving or parking the car. User interface must be adaptable for user's situation.

In section 2, we propose a non-task-oriented general dialogue model based on corpora for several tasks. In section 3, we examine the functions of the spoken dialogue interface in a dual task situation where two different tasks are being carried on at the same time.

2. Comparison of dialogue models on corpora for several tasks

To realize a natural dialogue like dialogues between humans on the computer, it is necessary to make use of not only acoustic information but also the linguistic knowledge and more high-level knowledge of dialogue management. The collection and analysis of natural dialogue corpora is the main method of clarifying how humans actually use such knowledge.

From this point of view, many analyses have been done on dialogues, but such analyses are made independently on dialogues based on one single task. Therefore, no assertion can be made on how general the information acquired from these analyses are. For this reason, it is difficult to make a judgment whether models obtained from these analyses are task-oriented or not. Therefore, we conceive a non-task-oriented dialogue model by analyzing corpora based on several tasks and extracting the utterances which are not task-oriented.

2.1 Dialogue corpora

We analyzed following four dialogues in Simulated Spoken Dialogue Corpus Vol.1.

- Task 1: Crossword puzzle task

  Two subjects have across or down hints of same crossword puzzle, and exchange each knowledge in order to accomplish the puzzle.

Table 1: Ratio of Task-oriented Utterance

	task 1	task 2	task 3	task4	all
Non-task-oriented	836	132	122	195	1285
Task-oriented	694	53	61	110	918
Total	1530	185	183	305	2203
Task-oriented/Total	45%	29%	33%	36%	42%

- Task 2: Scheduling task

  Two subjects set the date for a meeting with each schedule tables.

- Task 3: Telephone shopping task

  Two subjects play a customer or a clerk. The clerk asks the customer some information for the shopping.

- Task 4: Reservation guide task

  Two subjects play a travel agent or a customer. The customer applies a tour.

Functional labels which show functional roles in the dialogue were attached to all utterances in the corpus. It might be happened that one utterance had several labels.

We could classify these utterances to two types. One was task-oriented utterances which conveyed indispensable information for the task. For example, in crossword puzzle task hints or answer were task-oriented utterances. The other was non-task-oriented utterances which did not relate to the task. For example, greeting or acknowledgment were non-task-oriented utterances.

Table 1 shows ratio of task-oriented utterance. We found out that from 60% to 70% of all utterances were not task-oriented, by analyzing them in terms of their roles in the dialogues based on several tasks. Furthermore, the non-task-oriented utterances were classified into labels related to Turn-taking.

Table 2 shows list of label of non-task-oriented utterance.

Table 2:Labels of Non-task-oriented Utterance

	Label	Turn-Taking
1	Topic-Start	unsettled
2	Topic-Pause	utterer
3	Topic-End	unsettled
4	Information-Addition	utterer
5	Information-Renewal	utterer
6	Information-Lack	utterer
7	Recognition-Success	companion
8	Recognition-Repeat	companion
9	Recognition-Failure	companion
10	Contents-Affirmation	companion
11	Contents-Confirmation	companion
12	Contents-Denial	companion

2.2 Dialogue Model

By making a Markov model of those labels that are common to several tasks, we made a non-task-oriented model. Then we examine the propriety of the labels by comparing the model with the Markov model of characteristic labels which are obtained from the words in utterances. By using this model, it is possible to classify tasks not by the contents of them but by the tasks character. Therefore, with this classification we don't need to collect a large number of dialogue corpora for every different tasks and we can use the general dialogue models.

3. Spoken Dialogue Interface in a Dual Task Situation

In this section, we examined the functions of the spoken dialogue interface in a dual task situation where two different tasks are being carried on at the same time. Spoken dialogue interface is expected to be used in various situations in the future. However, it is important to know what kind of situation is suitable to use spoken dialogue interface. One of the situations in which speech interface is expected, is a dual task situation. In order to develop a spoken dialogue system, it is important to consider such a situation and the effects of the spoken dialogue interface in it.

It has been said that the merit of spoken dialogue is that it doesn't make the user's manual and visual modalities busy. Furthermore, it is indicated in the field of human factors that the efficiency of dual task performance was maximized when operators could use different modality in one task from the modality used in the other. Therefore, it is expected that spoken dialogue interface is effective in a dual task situation. We made an experiment to prove this hypothesis.

3.1 Experiment 1

Twelve subjects performed a dual task experiment. Four kinds of modes with different input and output modality were selected for investigation: (1) a task with manual input and visual output (MV); (2) a task with vocal input and visual output (VV); (3) a task with manual input and auditory output (MA); (4) a task with vocal input and auditory output (VA).

Table 3: Experimental condition

Combination	Input	Output
MV	manual	visual
VV	vocal	visual
MA	manual	auditory
VA	vocal	auditory

In this task the system presented stimuli to subjects by display or synthetic voice. The subjects were required inputting to the system by voice or mouse. A total of 24 trials were made. We examined which dual task situation is effective for the use of spoken dialogue interface by comparing the efficiency of dual task performance with that of single task performance.

As the result of this experiment, it became clear that the spoken dialogue interface was effective not only when the modalities of one task was MV, but also in the tasks using other types of modalities. However, the VV combination was not effective when the auditory output modality was used in the other task.

3.2 Experiment 2

We adopted Car-Navigation task as realistic dual task. In order to evaluate our interface of car-navigation system we made experiment. Main task was a previous MV task as real driving which needs to see and touch. Sub task was operating car-navigation system which could be controlled by either voice or touch-panel. Twelve subjects classified three groups as follows.

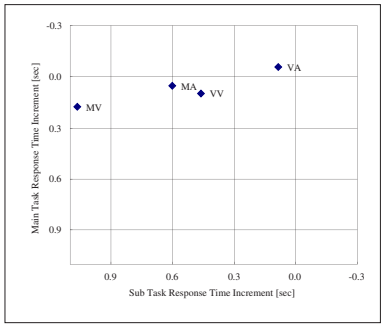

Fig. 1: Response time (Main Task MV)

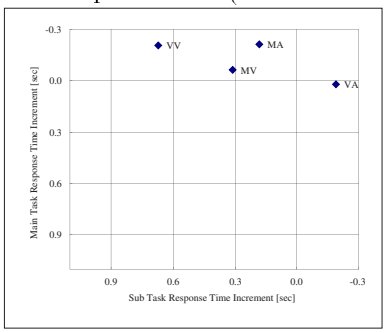

Fig. 2: Response time (Main Task VV)

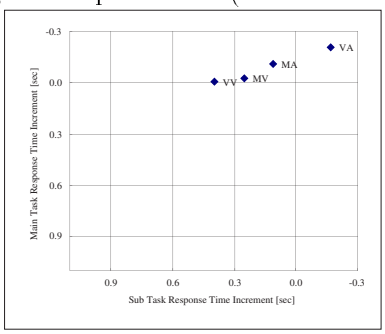

Fig. 3: Response time (Main Task MA)

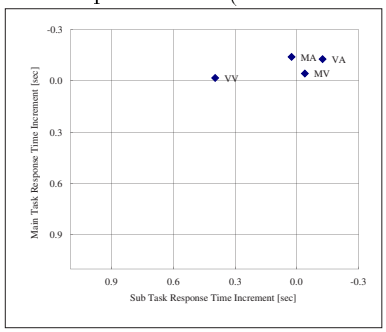

Fig. 4: Response time (Main Task VA)

Table 4: Experimental condition

Group	N	Single Task		Dual Task	
		1st	2nd	1st	2nd
A	6	multi	multi	multi	multi
B	3	multi	multi	vocal	manual
C	3	multi	multi	manual	vocal

In the single task condition which means parking situation, subjects operated only a car-navigation system. In the dual task condition which means driving situation, subjects operated the system and the MV task.

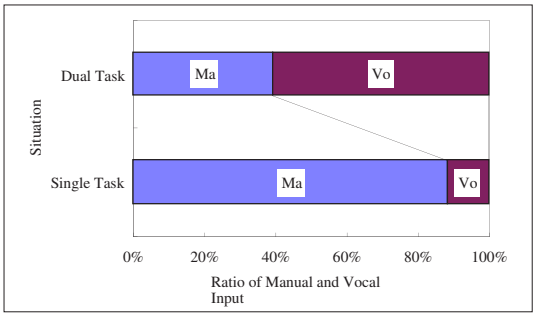

Fig. 5: Ratio of Manual and Vocal Input

Fig. 5 shows difference of input modality between single task and dual task conditions in the group A. There is obvious difference between two conditions. Under the condition of dual task ratio of vocal input is larger than manual input.

Fig. 6 shows the rate of error, Fig. 7 shows the achievement time and Fig. 8 shows the number of turn of glance in the main task for the group B and C.

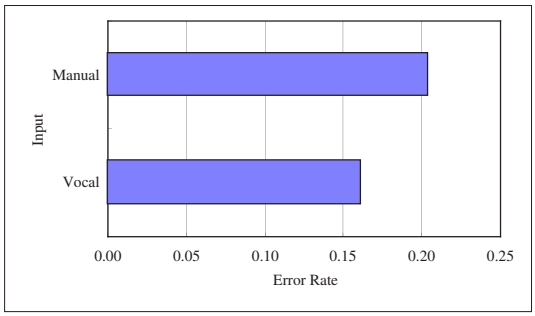

Fig. 6: Error Rate

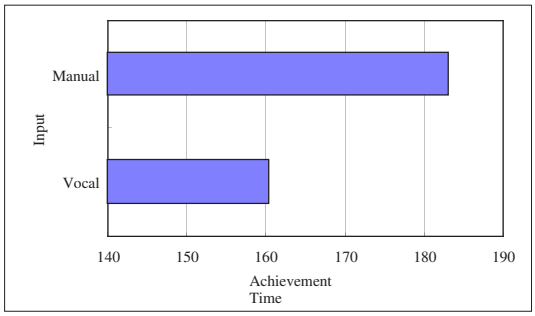

Fig. 7: Achievement Time

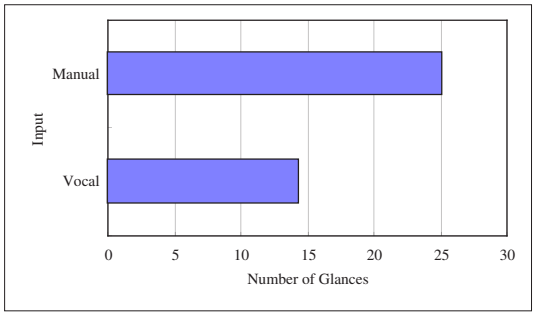

Fig. 8: Number of Turns of Glances

The vocal condition was superior at both error rate and achievement time. In number of turns of glances there was significant difference ($p<0.05$). These results shows a case of speech interface is suitable.