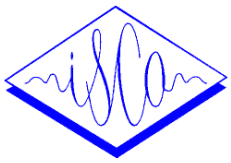

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

CHANGES IN USER'S RESPONSES WITH USE OF A SPEECH DIALOG SYSTEM

Kazuhiro ARAI

NTT Human Interface Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa, 238-03 JAPAN

ABSTRACT

In assessing a speech dialog system, changes in the user's responses to exposure to the system must be considered. In this paper, the changes in user's responses with the repeated use of a speech dialog system are described focusing on different prompt styles. Experiments examine four different styles of prompts to compare their effectiveness. In the experiments, the dialogs between a user and the system are based on the Wizard-of-Oz method. The result shows that at least two back to back trials are necessary to achieve a stable response. This paper describes the experiments and its results.

1 INTRODUCTION

Recently, many services using speech recognition over telephone networks have been proposed [1] [2] [3]. In such services, the users can only use speech for responding to the system. The users tend to utter additional words/phrases which are not essential. In order to deal with such utterances, continuous speech recognition and dialog control have been studied. It is reasonable, however, to consider that the user changes his responses with repeated use. Therefore, when assessing a speech dialog system that uses only speech, the system designer should consider changes in the user's responses with repeated use.

The goal of this study is to elicit responses suitable for speech recognition by optimizing the prompts of a speech dialog system, and estimating the changes in user's responses with repeated use of the system. The relation between prompt style and user's responses has been reported but without consideration of repeated use[4] [5] [6]. Though some research has addressed such changes[7], differences in prompt style were not examined.

This paper examines the changes in user's responses observed with repeated use of a speech dialog system focusing on the effect of prompt style. Four experiments were conducted using four different prompt styles. The experiments were based on the assumption that with experience the user tends to use only those words accepted by the speech recognizer.

The configuration of the system is shown in section 2. Section 2 also describes the vocabulary of the system and the dialog used in the experiments. The prompt styles used in the experiments are described in section 3. The results of the experiments are described in section 4.

2 The Voice Activated Call Answering System

2.1 Outline of the System

Figure 1 shows the configuration of the voice activated call answering system used in the experiments. Prompts from the system are produced in the *Speech Synthesizer* and are sent to the caller through the *public telephone network*. The experiments used the Wizard-of-Oz method to facilitate the dialog between the caller and the system: an operator listened to each caller's response and selected the text for the *speech synthesizer*. The *speech synthesizer* uses a Japanese text-to-speech synthesis board based on the *COC* synthesis method[8] for producing the prompts.

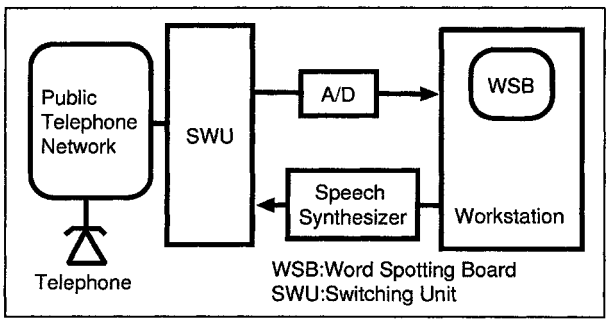

Figure 1: Configuration of the voice activated call answering system

Table 1: Japanese words in the vocabulary

Item	Number of Words
Callee's family name	22
Position's name	12
Honorific title	2
Predicate	9
Confirmation words (Positive)	12
(Negative)	29
Others	7
TOTAL	93

2.2 Words in the vocabulary

The numbers of Japanese words known by the system are shown in Table 1. The vocabulary was constructed by referring to the utterances recorded beforehand by the Wizard-of-Oz method based on the need to understand the responses.

2.3 Experiments Into Call-Answering Dialogs

160 dialogs were recorded to analyze the changes in user's responses over repeated use. Sixteen adult male subjects were used as callers and instructed to call the system. When a line between a caller and the system was established, the system asked the caller who he wished to contact so that it could make the connection. In the experiments, the callee was never at his seat so it was necessary for the system and the caller to continue their dialog. The system issued prompts to ask the caller whether he needed:

1. To have the callee return his call
2. To have the system record his phone number
3. To have the system record a message for him

Regardless of the user's responses, the system always issued a following prompt for confirmation. As follows, the total dialog consisted of 11 prompts.

- A prompt for asking the callee name.
- Three prompts for the questions described above.
- Four prompts for confirmation.
- Two prompts for instructing the user to tell the phone number and the message.
- A prompt for the concluding salutation.

The experiments focused on the first response indicating the callee's family name, and the caller's responses to the confirmation prompts. The other responses were not analyzed. Each caller was asked to make a phone call to 10 different people.

3 Prompt Styles

Several prompt styles were examined for eliciting the most suitable responses for speech recognition. In the following subsections, the styles are described in detail.

3.1 Callee's Name

If the response begins with a word in the system's vocabulary, it is easy for the speech recognition system to achieve high accuracy. In order to make the caller utter the callee's family name at the beginning of the reply, the prompts were based on the following styles [4].

Style A1 Emphasizing the word to be entered by using pause insertion and the expression "~ *dake*" which obligates the user to response with only the keyword.

Style A2 Using incomplete sentences. This style is effective in making the caller complete the sentence with the desired information.

Style A3 Using sentences which do not have the above.

3.2 Positive or Negative Responses For Confirmations

In order to evaluate the diminishment of superfluous utterances in the confirmation responses, four different styles of prompts for eliciting positive or negative responses were prepared. Table 2 shows the prompts that were prepared. The styles are as follows:

Style B1 This prompt style explicitly indicates the words with which the user should respond. A typical example of this style is *"Say A or B, please"*.

Style B2 This style repeats the first part of the previously issued prompt. Utterances of this style can often be observed in Japanese dialogs.

Style B3 This style uses "tag" questions, either negative or positive depending on which is *safer*. For instance, if the system records the user's message even if the user does not want to have it recorded, no problems are likely to occur. On the other hand, however, if the user wants to have his message recorded and the system fails to do so, problems are very likely. Therefore, according to the situation, the system will respond with a tag question that is less likely to result in a harmful misunderstanding.

Style B4 Prompts of this style are sentences which can be used for any confirmation.

Table 2: Prompts for Confirmation

Style	Return Call	Record Phone Number	Record Message
B1	*Say yes or no, please.*	*Say yes or no, please*	*Say yes or no, please.*
B2	*Shall I ?*	*Does he ?*	*Do you ?*
B3	*I will, won't I ?*	*He doesn't, does he ?*	*You do, don't you ?*
B4	*please say again.*	*please say again.*	*please say again.*

Table 3: The combinations of the prompt styles

Combination Pattern	1	2	3	4
Callee's Name	A1	A2	A2	A3
Confirmation	B1	B2	B3	B4

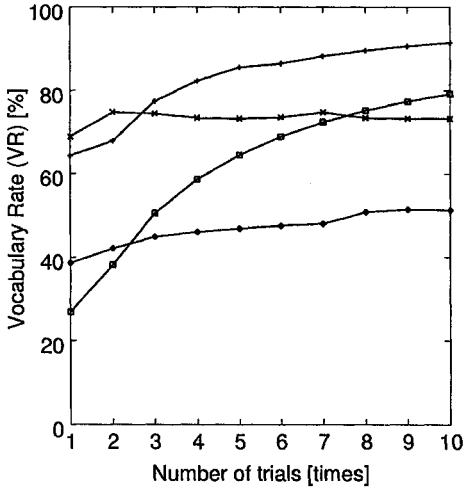

Figure 2: VR changes with combination pattern 1

Table 3 shows the combinations of prompt styles that were presented to the callers. In the experiments, each pattern was given to four different callers. Each caller was asked to make a phone call to 10 different people.

4 Results

Except for the responses to instructions and salutations, all responses were divided into *"bunsetsu"*, the linguistic unit in Japanese, to determine the vocabulary rate (*VR*): the percentage of the full and partial *"bunsetsu"* in the response held in the system's vocabulary. *VR* was calculated for each response by dividing the accumulated number of words in the vocabulary by the total number of *"bunsetsu"*. High values of *VR* indicated that the user used proportionally more of the words held in the vocabulary of the system.

Figure 2, 3, 4 and 5 show the *VR* changes in each combination pattern. Each line shows the *VR* change of one user. Each *VR* value is the average of the eight

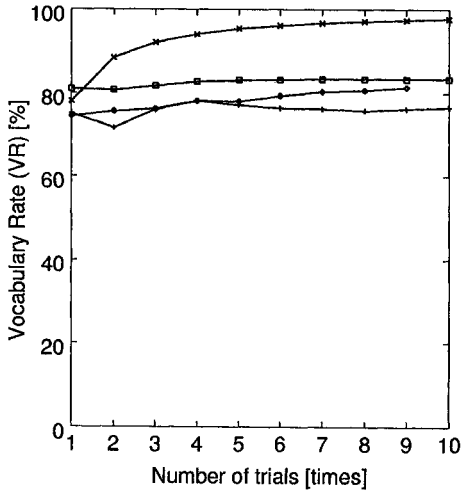

Figure 3: VR changes with combination pattern 2

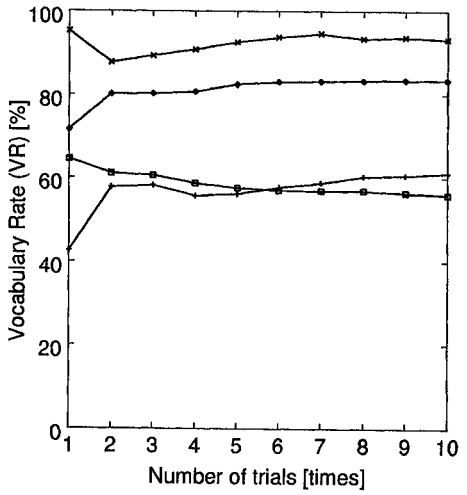

Figure 4: VR changes with combination pattern 3

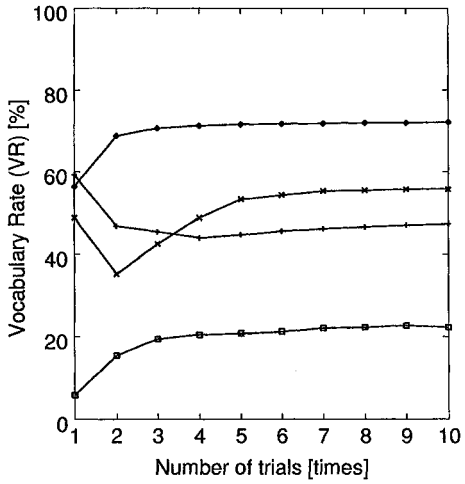

Figure 5: VR changes with combination pattern 4

responses recorded in one trial. This allows to compare the changes in *VR* with the number of trials. The following conclusions can be reached.

Pattern 1 yielded the most remarkable increase in *VR*. This increase shows that the user initially responded inappropriately but improved with experience. This result indicates that restrictive prompts are effective for eliciting the suitable responses for speech recognition, even though they are not natural.

Pattern 2 yielded high *VR* values from the first trial. The small dispersion among the users indicates that the users in this pattern accepted this prompt style easily from the initial stage. This result shows that it was easy for the user to accepted the prompt styles observed in dialogs between humans.

With pattern 3, the *VR* values of two users were very low. The low values must be caused by the inappropriate confirmation prompt style; this pattern used the same prompt style as pattern 2 to obtain the callee's name. This result shows that it is difficult for some users to respond to prompts that uses tag questions that change according to the situation.

The maximum *VR* value with pattern 4 was only 72%. The large dispersion among the users indicates that this prompt style is not effective for eliciting responses suitable for speech recognition. The user's response styles vary widely in the first and the second trials. Therefore, about two trials must be performed by each user before the speech dialog system can be accurately evaluated.

5 Discussion

When assessing a speech dialog system that can use nothing but speech, the system designer should consider changes in the user's responses with repeated use of the system. This paper has described the changes in user's responses focusing on different prompt styles. The experiments revealed the following findings.

1. Emphasizing the keyword and indicating the words with which the user should respond are the most effective style for eliciting responses suitable for speech recognition.
2. Incomplete sentences and repeating the first part of the previously issued prompt are well accepted by the user.
3. It is difficult for some users to respond to prompts that, depending on the situation, use tag questions.
4. About two trials must be performed by each user before the speech dialog system can be accurately evaluated.

The user's aptitude for interacting with speech dialog systems is one of the most important factors in experiments of this kind. Aptitude consists of experience in using the speech dialog system, the level of understanding the task and so on. Because experimental results are strongly influenced by the user's aptitude, future research should comment on the user's aptitude. Changes in the user's responses with an actual speech recognizer will be analyzed while focusing on the interaction that includes recognition errors and mis-understanding.

Acknowledgment

The author wish to thank Dr. Nobuhiko Kitawaki, director of Speech and Acoustics Laboratory and Dr. Noboru Sugamura, group leader, for their encouragement during this study.

References

[1] K. Kinder, S. Fox and G. Batche, "Accessing Telephone Service Using Speech Recognition: Results From Two Field Trials," AVIOS'93, pp.83-89, (1993).

[2] G. W. Smith, M. Bates, S. Hamilton, P. Jeanrenaud and M. Vandewegh, "Voice Activated Automated Telephone Call Routing," AVIOS'93, pp.99-103, (1993).

[3] M. Conway and T. B. Schalk, "Voice Recognition at the Cellular Switch: Automating 411 Call Completion," AVIOS'93, pp.155-161, (1993).

[4] K. Arai, M. Kitai, S. Nakajima and H. Nishi, "Effects of Question Style on Speech Dialog," International Workshop on Speech Processing, pp.149-154, (1993).

[5] S. Basson, "Prompting the User in ASR Applications," Workshop on Speech Recognition over the Telephone Line, (1992).

[6] D. Lauretta and G. Deffner, "Spoken Commands: Format and Frequency of Use," AVIOS'92, pp.201-206, (1992).

[7] W. S. Meisel, "User Acceptance of Restrictions on Speaking Style in Speech Recognition Applications," AVIOS'93, pp.163-168, (1993).

[8] K. Hakoda, S. Nakajima, T. Hirokawa and H. Mizuno, "A New Japanese Text-to-Speech Synthesizer based on COC Synthesis Method," ICSLP'90, pp.809-812, (1990).