音声言語情報処理　12－17
（１９９６．７．２７）

Toward more Friendly Voice-activated Telephone Operators

Zouheir Trabelsi, Toshiyuki Odaka, Akio Amano, Nobuo Hataoka

Central Research Laboratory Hitachi, Ltd.
E-mail: zouheir@crl.hitachi.co.jp

Abstract: This work discusses the development of a dialogue control component for Hitachi automatic telephone operator. The proposed dialogue control component uses a number of strategies that define clearly how the system manages the flow of information while interacting with the users. A spoken language corpora has been developed, by recording 100 calls from 10 users who are from the population that the system is intended to serve and who did not receive any training and instructions on how to use the system. This corpora allowed us to more understand the system's environment, how the users interact with the system, and the users behaviors and expectations, and to evaluate the efficiency of the adopted control dialogue strategies. Performance measures for assessing the efficiency of those strategies include mainly dial completion rate, users confusions rate, help messages access rate, and barging rate. As a whole, the dialogue control component has contributed considerably to the improvement of the efficiency of the system. Even thought some strategies should be improved, since they have lead the system to provide inefficient service in some situations.

Key words: Voice-activated automatic telephone operator, Dialogue control strategies, Spoken language corpora.

能率的な自動交換オペレータ

トラベルシ　ズハエル、小高俊之、天野明雄、畑岡信夫

（株）日立製作所　中央研究所
E-mail: zouheir@crl.hitachi.co.jp

あらまし：本稿は、音声認識を用いた自動交換オペレータにおける対話制御部の開発について述べている。提案する対話制御部はシステムとユーザの間の円滑な対話を可能とする種々の対話制御方式を用いている。本研究では、対話コーパスの構築と提案する対話制御方式の第一次評価を行うことを目的に、システムを初めて使う１０人の被験者から各々１０試行、計１００試行のデータを収集した。第一次評価の評価尺度として、完了率と、強制中止率、ヘルプの利用頻度、及びバージイン率を用いた。本報告では、提案する対話制御方式の詳細と第一次評価結果に関して報告する。

キーワード：*音声認識、自動交換オペレータ、対話制御、対話コーパス*

1. Introduction

A well-designed user interface is essential to a successful application. Such an interface can overcome many of the limitations of the current technology to produce a successful outcome from the user's point of view, even when the technology works imperfectly. A poor user interface can render a system unusable. Particularly, spoken dialogue systems can be more efficient and cooperative by including robust dialogue control components. The dialogue control component is the component in a system that manages the flow of information between the user and the system.

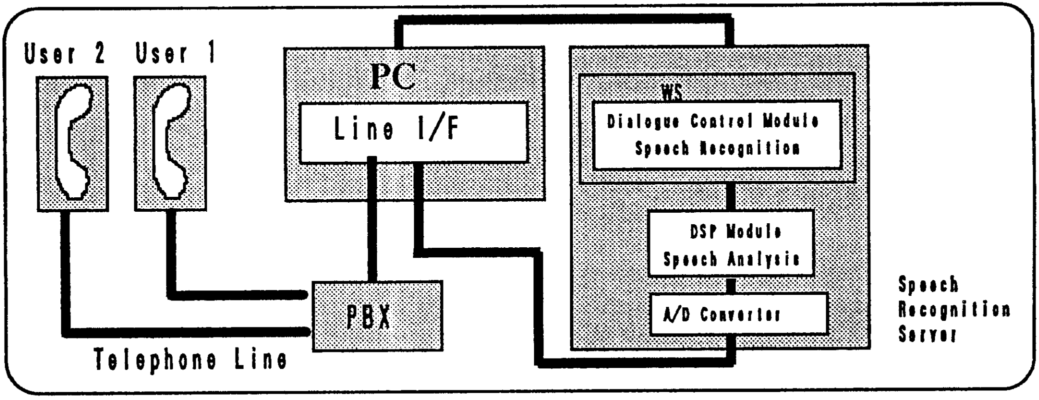
A number of voice-activated automatic telephone systems have been developed [1,2,3,4,5,6,7]. There have been two major focus in designing such systems. The first focus is to design systems that have high speech recognition performances. The second focus is to reduce the cost of using the systems. However, they use simple and/or poor strategies to manage the flow of information between them and the users. Hence, their abilities to provide cooperative and efficient services are very poor and limited. Consequently. the users are not satisfied with the way those systems interact with them.

In this paper, first we introduce Hitachi Voice-activated Telephone Operator. Then, we discuss the development of the dialogue control component of this system. To manage the flow of information, the dialogue control component uses a number of strategies. In addition, we discuss the development of a spoken language corpora which has allowed us to more understand the system's environment. Finally, we evaluate the efficiency of the proposed dialogue control strategies, using the data collected in the corpora, and outline the issues that will be tackled during future works.

2. Overview of Hitachi Automatic Telephone Operator

The system [1,2] incorporates speaker-independent isolated word speech recognition technology. The vocabulary size is around 1000 words. Each person in Hitachi Central Research Laboratory (HCRL) may call another person solely by saying his/her department and/or name. Once the callee's department and/or name are identified, the system looks for his/her telephone extension number and establish automatically the appropriate telephone connection.

Fig. 1: Block-diagram of Hitachi Automatic Telephone Operator



A block-diagram of the system is shown in the Fig. 1. The speech recognition server consists of a speech analysis process, HMMs (Hidden Markov Models) recognition process and the dialogue control component. The HMMs-based recognition server has been implemented on a commercialized WS (workstation) and a DSP (Digital Signal Processing) module. The server is connected to a PBX through a telephone interface controlled by a PC (personal computer).

The early version of this system included a poor dialogue control component, since it uses simple and poor strategies to manage the user's interactions. Hence, the system's ability to provide cooperative and efficient services was very poor and limited. The dial trials of the early version of the system showed that poor dialogue control component has developed incomprehensible dialogues between the user and the system, a low dial completion rate, a high user confusions rate, and a high barging rate. The need for a robust dialogue control component to enhance the efficiency of the system is therefore unquestionable.

3. Dialogue Control Strategies

The dialogue control strategies define clearly how the system manages the flow of information while interacting with the users. The following sections discuss the adopted strategies, namely the directory structure strategy, the service-starting strategy, the help messages contents and control strategy, the feedback and confirmation strategy, the time-out strategy, and the input expressions strategy. Several other strategies will be discussed in future works, once they are finalized and implemented in the system.

3.1 The Directory Structure Strategy

The directory contains information about each person working in the HCRL, such as the family name, the first name, the department name, and the telephone extension number. In addition, for some persons, mainly some researchers, there may be special messages. That is, if a user attempts to call a researcher who went abroad for a visiting research position, the system automatically prompts a special message stating that the callee is abroad right now. Moreover, the directory contains a list of some old names of departments which do not exist right now. Such departments have been renamed, removed or split after some restructure or reforms steps. The transcript in the Table 1 shows how the system handles such old department names.

Table 1: Old Department Name Handling

User: 知能システム研究部
System: 申し訳ありません、知能システム研究部は
マルチメディアシステム研究部に変わりました。
マルチメディアシステム研究部のだれですか。

3.2 The Service-Starting Strategy

A brief welcoming prompt to the service followed by a prompt telling the user to say the word "HELP", at any point of the dialogue, after the beep sound, if he/she does not know how to use the service. The novice users are expected to access the help messages, to learn further how to use the service, at different points of the dialogues.

Then, the system prompts a message asking the user to say the department name of the callee, or the word 「わかりません」 or 「わからない」, if he/she does not know or remember the callee's department name. The transcript in Table 2 shows the messages prompted once the service starts.

Table 2: Prompts once the Service Starts

System: *こちらはマルチメディアシステム研究部の自動交換オペレータです。*
System: *使い方はわからない場合は、「ヘルプ」とおしゃってください。*
System: *相手の所属をどうぞ。所属がわからない場合は、*
「わかりません」とおしゃってください。

3.3 The Help Messages Contents and Control Strategy

An optimal system should be able to provide the option of accessing the help at any point of the dialogue, to offer automatically the help whenever the user's behavior suggests that it is needed, and to allow the user to leave the help at any time since he/she may not need to listen the entire help messages.

The strategy used in this system consists into allowing the user to access the help at any point of the dialogue, by simply saying the word "HELP". The user can leave the help at any time by simply saying the word "STOP". In addition, in order to allow the user to listen the help messages as many times as he/she wants, the system continues to re-prompt them until he/she says the word "STOP". The table 3 shows the contents of the help messages prompted to the users.

Table 3: The Help Messages

1 今からお聞き頂くのは使い方をご案内するメッセージです。
2 このメッセージは「HELP」とおっしゃると、いつでも聞くことができます。
また、終了させるには「STOP」とおっしゃって下さい。
3 相手の所属を聞いた後で、名前をお聞きします。
4 ビープ音をお聞きになってから、お答え下さい。そうでないと、聞きとれません。
5 お答えになる時は、「えー」、「あのー」、などの言葉をおっしゃらないように、御注意下さい。
また、つっかえないようにもご注意下さい。
6 確認のメッセージ、あるいは質問が正しくなければ、「違います」あるいは「いいえ」とおっしゃって下さい。
7 これでご案内を終ります。

3.4 The Feedback and Confirmation Strategy

The feedback and confirmation strategy adopted in this work consists into not waiting the confirmation of the user to execute the following step. That is, once the user says an input, the next system's prompt includes the system's interpretation of the input and what he/she should do or say. For example:

Step 1: ***System:*** 相手の所属をどうぞ。所属がわらない場合は、
「わかりません」とおしゃってください。
User: マルチメディアシステム研究部
Step 2: ***System:***マルチメディアシステム研究部のだれですか。

The system's prompt includes the speech recognition result of the user's input, and, in the same time, asks the user to say the name of the callee. If the user detects an error or he/she is not satisfied with the system's interpretation of his/her input or with the system's impending actions on his/her inputs, then he/she may say the word 「違います」 or 「いいえ」, to express his/her dissatisfaction and therefore brings back the dialogue into the previous stable stage, which is the step 1.

3.5 The Time-out Strategy

During the interaction with the user, the system often asks the user to say an input, such as the callee's name or the callee's department. However, the user may not say any input while the system is waiting.

In order to allow the system to be able to take control of the conversation during such a situation, if within four seconds the user did not say any input, the system prompts again the same message. The system gives the user three chances. If he/she did not say any input after three chances, the system prompts a message informing the user that the conversation will be stopped, and is going to connect him/her to a human operator to find out what is the problem.

3.6 The Input Expressions Strategy

A number of user's inputs expressions can be accepted and recognized by the system. For example, to say the name of a callee, the user may say the name followed by ｛さん、くん、部長、｝, and/or ｛です、お願いします、お願い｝. The actual strategy for input expressions is simple. A more efficient system should include a robust natural language processing component that can recognize a wide variety of user's input expressions.

3.7 Example of Dialogue

The transcript in the table 4 shows an example of dialogue with a novice user who did not receive any instructions on how to use the system. The example has been recorded during the dial trial.

Table 4: Transcript of an Example of Dialogue

System: こちらはマルチメディアシステム研究部の自動交換オペレータです。
System: 使い方はわからない場合は、「ヘルプ」とおしゃってください。

System: *相手の所属をどうぞ。所属がわからない場合は、*
「わかりません」とおしゃってください。
User: *マルチメディアシステム研究部。*
System: *メディカルシステム研究部のだれですか。*
User: *ヘルプ。*
System: (* The system prompts the help messages *)
User: *ストップ。* (* After listening twice the help messages, the user said the word
"STOP" to stop the help messages *)
System: *メディカルシステム研究部のだれですか。*
User: *違います。*
System: *もう一度、相手の所属をどうぞ。所属がわからない場合は、*
「わかりません」とおしゃってください。
User: *マルチメディアシステム研究部。*
System: *マルチメディアシステム研究部のだれですか。*
User: *井沢さんお願いします。*
System: *井沢ですね。*
User: *はい。*
System: *かしこまりました。少々お待ち下さい。。。*

4. Data Collection: Spoken Language Corpora

The collection of spoken language corpora is an important research area and represents a significant portion of the work in developing a spoken language system. Progress in spoken language understanding is closely linked to the availability of spoken language corpora.

To build such a corpora, we asked 10 users to use naturally and freely the system, without providing them any training and instructions on how to use it. We have recorded 100 calls -10 calls per user-, which constitute the contents of the spoken language corpora.

4.1 The Speech Recognition Rate

Table 5 shows the speech recognition rate of the first version of the system used to build the spoken language corpora. The speech recognition rate is more than 70% and the baseline recognition rate is around 91%. The speech recognition rate is relatively low because the number of the inputs out of vocabulary is high and the environment, such as the telephone line, is noisy. But we believe that this rate is acceptable for building the first spoken language corpora and evaluating the first version of the system.

It is important to notice that the number of user's inputs out of the vocabulary is relatively high. This is because the lack of spoken language corpora while designing the system. We believe that with the availability of spoken language corpora, the next version of the system will be able to identify further inputs expressions. And, consequently, the speech recognition rate of the system would improve considerably.

Table 5: The Speech Recognition Rate

Vocabulary size	*: 132 Words*	*Correct recognition rate*	*: 70%*
Total number of inputs	*: 122*	*Miss-recognition rate*	*: 30%*
Miss-recognition		*Out of the vocabulary rate*	*: 20%*
- Pure miss-recognition	*: 13*		
- Out of the vocabulary	*: 24*		
Correct recognition	*: 85*	*Baseline recognition rate*	*: 91%*

4.2 The Recorded Calls Status

This corpora allowed us to identify new user's inputs expressions which we did not identify them during the development of the system. In addition it allowed us to evaluate the efficiencies and inefficiencies of the system while interacting with the users. Hence, based on this study, the future versions of the system will tackle all the problem that have lead the system to provide in some occasions inefficient services. The status of the 100 recorded calls of the 10 users is shown in the Table 6. We expect that as more calls are recorded, our understanding of the system environment and how the users interact with the system, will improve; and consequently the system's efficiency will improve considerably.

Table 6: Recorded Calls Status

Number of Users: **10**
Number of succeeded calls (10 calls per user) : **65 calls succeeded, 35 calls failed**
Number of inputs expressions not supported : **15 expressions**
Number of times the users were confused : **19**
Number of times the users attempted to access the help : **16**
When the first attempt to access the help took place : **During the 3rd call (Average)**
Number of times the users spook inputs during outgoing prompts : **8**

5. Evaluation and Future Works

5.1 New User's Inputs Expressions

While analyzing the above spoken language corpora, Table 6, fifteen new user's inputs expressions have been identified. Table 7 contains some new identified inputs expressions. The current version of the system does not support those expressions and therefore cannot recognize them. The number of those expressions is relatively high for this application, automatic telephone operator. This is due to the lack of spoken language corpora while designing the system, to the limited number of inputs expressions recognized by the system, and to the limited vocabulary size. The next version of the system will attempts to reduce the numbers of the user's inputs out of the vocabulary. This can be done by recognizing new inputs expressions, by including new words in the vocabulary, and by carefully designing the prompts and the help messages so that the user becomes more instructed on how to say his/her inputs and what inputs expressions can be recognized. In addition, by using advanced techniques, such as word spotting, to detect key words in the user's inputs the number of inputs out of the vocabulary can be significantly reduced.

Table 7: New inputs expressions

1- えと、 {「部名」、「ユニット名」、「人名」}
2- 「部名」の「ユニット名」
3- 「ユニット名」の「人名」
4- 「部名」の「人名」
5- 「部名」+{じゃない, じゃありません}
6- いいえ + {違います、違う}
7- 「部名」+「ユニット名」+ {です、お願いします、お願い}
8- 「ユニット名」+ {です、お願いします、お願い}
9- いいえ + {「部名」、「人名」}
10- {「部名」、「人名」} + {じゃない, じゃありません}
11- {違います、違う} + {「部名」、「ユニット名」、「人名」} + {です、お願い、お願いします}
12- 「姓」+「名」+ {です、お願い、お願いします}
13- 「人名」+ 先生
14- はい、そうです + {お願い、お願いします}
15- 何も言ってません

5.2 Prompt Talk-through

The technology used to develop the system does not allow prompt talk-through. That is, it does not give the users the ability for the speech recognizer to understand spoken inputs during outgoing prompts. Therefore, the users must say their inputs after the prompts are over.

While designing the system, the contents and locations of the prompts were carefully designed so that the users do not attempt to say their inputs during outgoing prompts. In addition, each prompt is followed by a beep sound indicating to the users that they may start saying their inputs.

However, the above spoken language corpora shows that eight times the users attempted to talk during outgoing prompts. We consider that this number, eight times, is low for one hundred recorded calls. Hence, this indicates that the prompts are almost well designed.

The analysis of the eight calls where the prompt talk-through took place shows that during the outgoing of two prompts some users attempts to speak. Those prompts are:

Prompt 1: *こちらはマルチメディアシステム研究部の自動交換オペレータです。*
使い方はわからない場合は、「ヘルプ」とおしゃってください。
(Once this prompt is over, some users attempt to say the word 「ヘルプ」, even though they did not hear the beep sound)
Prompt 2: *相手の所属をどうぞ。所属がわからない場合は、*
(Once this prompt is over, some users attempt to say the department name, even though they did not hear the beep sound)
「わかりません」とおしゃってください.

Prompt 1 misled some users since it is telling them to say the word help whenever they need help, without indicating that they must speak after the beep sound. Also, *Prompt 2* misled some users since it is telling them to say the department name, without indicating that they must speak after the beep sound. Hence, the next version of the system will attempt to find out how those prompts should be improved in order to not mislead the users and therefore reduce the possibilities of prompt talk-through. For example, by including, in the service- starting prompts, a new prompt which indicates clearly that the users should speak after the beep sound.

5.3 The Confusion of the Users

The spoken language corpora, Table 6, shows that nineteen times the users were confused during the one hundred recorded calls. We believe that this number is acceptable and does not affect considerably the efficiency of the whole system.

Mainly, three reasons have allowed to develop situations during which the users were confused. First, some users could not know how to express their dissatisfaction with the system's interpretation of their inputs or impeding actions on that inputs. Even though, the help messages indicate clearly how the users can express their dissatisfactions. Second, some users do not know how to get out from situations where the speech recognition process fails several times to recognize correctly their inputs. Third, some users could not understand why the system is attempting to call a person whom they do not want to call; and they could not know how to stop the outgoing attempt dial.

Consequently, as a future work, the system will attempt to tackle those problems. For example, the users will be allowed to stop an incorrect attempt dial, by the system, by pushing any button on the phone. On the other hand, the system will attempt to automatically prompt particular help messages in particular situation to help the user to accomplish successfully his interactions.

5.4 The Prompts

Once the recording of the calls of a user is done, we had a conversation with him/her about the system. A number of users had informed us that the prompts used by the system are relatively lengthy. That is, they have to wait until those lengthy prompts are over to start saying their inputs. For example, once the user pick up the phone to use the operator, they have to wait until the service-starting prompts are over to start saying the department names. Hence, it would be more efficient to allow those users, who do not want to hear again those prompts, to stop them and therefore bypass them and progress directly to the next step. To provide more efficient service, the next version of the system will attempt to allow the users to stop any prompt and bypass it , by solely pushing any button.

5.5 The Help Messages

The analyze of the 100 recorded calls allowed us to more understand how to design and control the help messages. That is, the designer of the system should define clearly what is the contents of the help messages, how the user can access the help messages, how to stop the help messages, how to make the user knows that there are help messages, whether or not there are particular help messages for each step in the dialogue, how to prompt the help messages, for example once a help message is prompted, allow the user few seconds to assimilate its contents before prompting the next message. In addition, it is important to define carefully the contents of the help messages in such a way that the user can identify easily what are the capabilities and limits of the system.

The help messages of the current system do not show clearly what are the capabilities and the limits of the system. For example, they do not tell the user what type of inputs expressions can be identified. On the other hand, the user have to say several times the word "STOP" to stop the help messages. This is because of the failure of the system to recognize correctly the word "STOP". It would be more effective, for example, to allow the users to stop the help messages by saying the word "STOP" or pushing any button on the phone.

In addition, some users had forgotten the existence of the help messages. Because the system tells them that they may say help whenever they need help, only when the service starts. Therefore, a number of users forget such information as the dialogues go on. As a remedy, it would be better to automatically prompt the help messages once some errors and/or problems occur. Moreover, it would be more effective to prompt only the help messages that concern the current step of the dialogue.

5.6 Future Works

The following is the summary of the issues that will be tackled during our future works.

1. The coverage of more inputs expressions and words out of the vocabulary.
2. The reduction of the situations where the users may be confused and/or misled.
3. The development of a technique to allow the users to bypass or stop the prompts.
4. The improvement of the help messages contents and control strategy.
5. The improvement of the speech recognition performance.
6. The building of a large spoken language corpora using the improved version of the system.
7. The evaluation of the efficiency of each proposed strategy and the whole system.

6. Conclusion

In this work, we have proposed a number of dialogue control strategies for Hitachi voice-activated automatic telephone operator. The strategies intend to control efficiently the flow of information between the users and the system. To evaluate the efficiency of the strategies, we have build a spoken language corpora, which, in addition, has allowed us to more understand the system's environment. As a whole, even though, the speech recognition rate of the system is relatively poor, the strategies have contributed considerably to the improvement of the efficiency of the services, and consequently to the dial completion rate. However, some strategies need to be improved, during future works, since they have lead the system to provide inefficient services in some particular situations.

References:

[1] Hataoka, N., Odaka, T., Amano, A., "Speech Recognition System for Automatic Telephone Operator based on CSS Architecture." Proceedings of the 2nd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA94), Kyoto, Japan, pp. 77-80, 1994.

[2] Odaka, T., Amano, A., Hataoka, N., "Study on Speech Recognition Application Using Telephone Line and Local Area Network.", Technical Report of IEICE, SP94-55 (1994-11).

[3] Karis, D., and K. M. Dobroth. "Automating services with speech recognition over the public switched telephone network: Human factors considerations." IEEE Journal on Selected Areas in Communications, 9:574-585, 1991.

[4] Leiser, R. G. "Improving natural language and speech interfaces by the use of metalinguistic phenomena." Applied Ergonomics, 20:168-173, 1989.

[5] Kamm Candace, "User Interfaces for Voice Applications." Voice Communication Between Humans and Machines, National Academy of Sciences, pp. 422-442, 1994.

[6] Shriberg, E., E. Wade, and P. Price. "Human-machine problem solving using Spoken Language Systems (SLS): Factors affecting performance and user satisfaction." Proceedings of the DARPA Speech and Natural Language Workshop, M. Marcus (ed.), Morgan Kaufmann, San Mateo, 1992.

[7] Vysotsky J. George. "VoiceDialing - The First Speech Recognition Based Telephone Service Delivered to Customer's Home." Proceedings of the 2nd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA94), Kyoto, Japan, pp. 149-152, 1994.