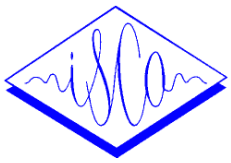

3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

THE DESIGN OF A REAL WORLD WIZARD OF OZ EXPERIMENT FOR A SPEECH DRIVEN TELEPHONE DIRECTORY INFORMATION SYSTEM

T. Andernach*, G. Deville** and L. Mortier***

* *University of Twente, Department of Computer Science, P.O. Box 217 7500 AE Enschede, the Netherlands.*
** *Facultés Universitaires de Namur, rue de Bruxelles 61, B-5000 Namur, Belgique.*
*** *Lernout & Hauspie Speech Products, Rozendaalstraat 14, B-8900 Ieper, België.*

ABSTRACT

This paper reports the design of a Wizard of Oz simulation of a spoken dialogue system in the context of a telephone directory information service. The experiment described here was carried out within the framework of a Belgian Telecom R&D project, the aim of which is to develop a prototype system for the automatic handling of directory information queries with voice input/output facilities.

The authors' task in the project consists in designing the information dialogue models which will be used in the eventual prototype. The target application is the database of the Belgian Telecom Information Centre.

The authors advocate some crucial methodological choices, which clearly distinguish this experiment from most current alternative approaches. The dialogue analysis is discussed and conclusions are drawn, the focus being on (i) the adequacy of the methodology to the design of dialogue models, (ii) the pertinence of such models in an automated environment, and (iii) the evaluation of the dialogue sessions by the subjects.

Keywords *: Spoken dialogue, Wizard of Oz experiment, human factors.*

1. INTRODUCTION

Given the current state of art of continuous speech recognition techniques, it still unfeasible to investigate human behaviour in fully integrated spoken natural language understanding systems in a real environment, as such systems are not available. Therefore, man-machine dialogue simulations seem to be a fruitful alternative in designing robust speech interfaces suitable for task-oriented applications.

The experiment discussed in this paper is a contribution to a Belgian Telecom (Belgacom) R&D project, the aim of which is to develop a prototype system for automatic handling of directory information queries with voice input/output facilities (e.g. given the name and address of a person, how to get his phone number). The final prototype will be a speaker independent continuous speech application integrating a powerful alphabet recognizer, advanced real-time keyword spotting algorithm [1] and optimized text-to-speech techniques for the pronunciation of proper names and addresses [2]. The authors' task in the project consists in designing and testing the dialogue model that will be used in the final prototype. The target application is the database of the Belgacom information service containing person's and institution's references like names, addresses and telephone numbers that are currently provided by human operators.

Prior to the Wizard of Oz simulation described here, a real scale experiment was carried out in a Belgacom Information Centre in order to get a representative sample of the target language used in this type of task-oriented discourse and to 'bootstrap' the design of a dialogue model in an automated environment. This preliminary experiment on site includes interviews of the experts-operators at the Belgacom Information Centre, the recording and transcription of actual requests for information, and the linguistic and interactional analysis of the obtained man-man informative dialogues. A detailed account of this first experimental phase is given in [3].

The motivation of the Wizard of Oz simulation is threefold: (i) to test and compare a restricted vs. free dialogue model in a real scale situation in order to optimize the modeling for the final prototype (only the directed dialogue model is discussed in this paper); (ii) to analyse the user's conversational behaviour when dealing with a machine, paying special attention to the ergonomic dimension; and (iii) to find out how the dialogue system as a whole is evaluated by the subjects in terms of robustness and user-friendliness.

The present paper gives first a functional description of the Wizard of Oz experiment. Some crucial methodological choices are then discussed, focusing on the originality of the approach. Finally, the dialogues in this automated environment are analyzed and critically discussed.

2. A WIZARD OF OZ EXPERIMENT

2.1. Experimental set up

In a Wizard of Oz experiment, subjects interact with a machine without knowing that the turns of the machine are simulated by a so-called wizard (comparse). The experiment was carried out at the Belgacom Information Center of Aalst (Flanders regio), directly involving four persons :

(i) the subject was any customer calling the Belgacom Information Service requesting for a telephone number by means of an over-the phone dialogue with the information system. Note that the customers is initially unaware of the fact that he is the potential subject of the experiment.

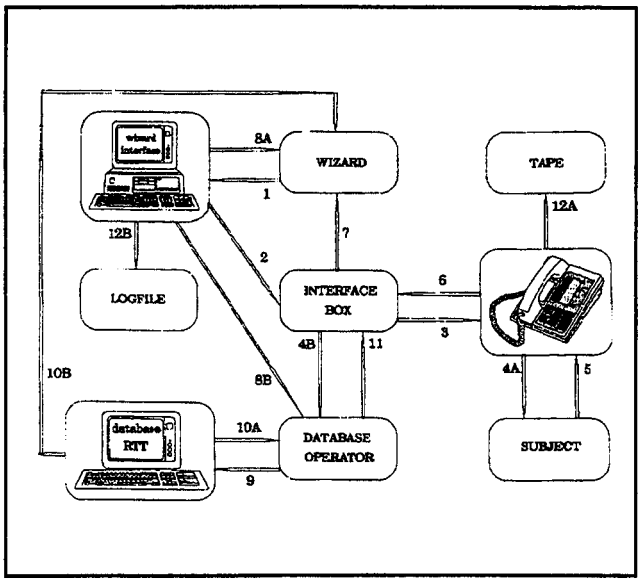

Figure 1. Experimental setup of the Wizard of Oz experiment

(ii) The wizard is a computational linguist whose task was to simulate an automatic spoken information system with the help of the Wizard Interface. This task consists in the elicitation, recognition and understanding of a user's utterances (audible via a headphone) and, subsequently, the choice of the next path in the dialogue model corresponding to a step in the course of the information dialogue. More specifically, through a dialogue with the subject, the wizard must acquire the data needed to query the Belgacom database. The system's utterances were provided via the Wizard Interface.

(iii) The database operator was one employee of the Information Service. Her task was to get the input (query) for the database system only via a window of the Wizard Interface from the wizard. So she has read access to both the database system and the Wizard Interface. The database operator is allowed to enter the query in the database system only if the wizard demands her to do so.

(iv) A connection operator was an employee of the Information service, and her task was to filter people who made a request for a telephone number and who were willing to participate in the experiment. After confirmation of the subject's participation, the connection operator shortly briefed the subject and connected him with the system. Precautions were taken at that stage for not informing the subject about the course of the experiment.

2.2. Experimental phases

In a pre-experimental phase the experimental setup was determined and the Wizard Interface as well as the dialogue models were designed and implemented. Then the hardware (PC containing the Text-To-Speech board) and the telephone-connection were installed on the site. Next, the database operator and the wizard were instructed about how to act in the course of a dialogue session with a subject. Finally a one-day pilot experiment was carried out in order to test the setup and adapt it, if need be. After the pilot study, the actual experiment started.

2.3. The dialogue sessions

In this two-day experiment, about one hundred dialogue sessions with different customers were held and recorded on site. A session consists of three phases : a connection phase, a proper dialogue phase and an evaluation phase :

(i) In the first phase the caller was connected with the connection operator checking if the caller was asking for a telephone number. In that case, the operator requested the caller to participate to an experiment to test an automatic information system. In case of affirmative answer, the caller became subject of the experiment and was told that after the session with the system, he has the opportunity to ask questions and verify the obtained information if necessary, and that he would be asked some questions. In case the caller did not wish a phone number or refused to take part to the experiment, a standard information dialogue took place between the connection operator and the caller.

(ii) In the second phase the proper dialogue with the system takes place. The designed setup is shown in figure 1. The digits between brackets refer to this figure. After the subject's confirmation of his participation in the experiment the wizard starts the dialogue by activating the Wizard of Oz system (1) which sends a message by Text-To-Speech via the interface box (2) and the phone (3) to the subject (4A). The subject then responds via the phone (5) and the interface box (6), his reply being audible by the wizard (7). In order to take the appropriate decisions and to interpret the subject's answer adequately, the wizard is guided by messages and questions on the screen of the Wizard Interface (8A).

The cycle 1-8A is iterated as long as the wizard needs additional data for elaborating a valid request in the database. When sufficient information is available, the system produces a message indicating that the database search process is activated. At that moment the database operator is allowed to execute the query (9). The result of the query is visible both by the wizard (10A) and the database operator (10B). The process 1-8 starts again to inform the subject about the result of his query. The process ends when the subject has no more requests for a phone number.

(iii) In the third phase after the dialogue session with the system, the subject is connected with the database operator (11). First the subject has the opportunity to verify the answer to his request provided by the system, and to have it corrected in case of wrong information. Then a short oral evaluation dialogue takes place between the database operator and the subject, where the following questions are asked :

- *Did you have the impression that your request was treated by a computer? If yes, why? If no, why not?*
- *Are you satisfied about the way your inquiry has been treated? If yes why? If no why not?*
- *Are there certain aspects of the dialogue you are very satisfied about ? If yes why?*
- *Are there certain aspects of the dialogue you are very unsatisfied about ? If yes why?*
- *By whom do you prefer your request to be treated : the computer or a human operator? Why?*

After the answering to the questions, the dialogue session is closed by thanking the subject for his participation.

Both dialogues in phases (ii) and (iii) were recorded on tape (12A) and the history of the dialogues in phase 2 were stored in a special logfile for further analysis (12B).

2.4. The Wizard Interface

The Wizard Interface consists of a module (called the dialogue manager) that helps the wizard to control the dialogue simulation session and a component (called text-to-speech or TTS for short) that converts the system's utterances into synthesized speech. The functions of the Wizard Interface are :

1. guiding the wizard through the dialogue path by displaying decision points in a window and allowing the wizard to enter the decisions. Each time a decision point is reached in the dialogue model a question will be asked to the wizard. The answer to the question corresponds to the taking of a new path invoking another decision point or an action of the system (depending on the point in the dialogue);
2. keeping track of the course of the dialogue by storing the path that is followed throughout the full dialogue. Every question related to a decision point and every action taken by the system is stored in a file the moment they are invoked;
3 allowing the wizard to enter the phone number yielded by the database query, number that will be sent to the subject by TTS. Special intonation models (a.o. for telephone number announcements) have been incorporated [4];
4 selecting a stored system utterance and displaying it in a window, each time the system invokes an action to send a message to the subject;
5. generating the system's utterances by text-to-speech, with a speech quality higher that the vocoder-distorted speech and similar to the quality of the productions of the eventual system. The TTS technique frees the wizard from having to speak the messages himself.

The Wizard Interface is a program written in both Pascal and C and runs on a PC [4]. It contains a Dialogue Manager which reads the dialogue model from file and stores it in RAM as a transition network describing all possible paths that can be taken in the dialogue. A state in this dialogue model can be a decision point that has to be activated by the wizard, or can trigger an action with the following function : uttering a system message, letting the wizard enter data (a required phone number) and giving the database operator the opportunity to enter the selected parameters.

2.5. The database

The database used in the experiment is the database used by the professional operators of the Belgacom Information Service. This database can be queried by entering a phone number or by entering a locality and a name or address of a client. The query is entered in a command line in which the relevant data provided by the caller, called the *selection parameters*, can be specified in random order. This enables the database operator to enter each selection parameter at the moment it is given by the user. Every selection parameter is preceded by a one character code to identify that parameter, as illustrated in the following example of query :

for, njansen, skerkstr, lantwerpen

As reaction to this query the database system will look in the city *(l)* Antwerpen for the name *(n)* Jansen and will select all clients with this name in the kerkstraat *(s)*.

2.6. The dialogue model

In designing the dialogue model two types of constraints were kept in mind : the current state of the art in automatic speech recognition, and the fact that a prototype has to be developed in a relative short time lapse. Due to the first constraint the so-called *word spotting* technique will be used in the final prototype to 'interpret' the utterances of the user. The presence of a (set of) key word(s) in a given utterance triggers the choice for another path in the dialogue.

Due to the second constraint, the dialogue model implemented in the Wizard Interface is called a 'directed' model, where little initiative is taken by the subject, whereas the dialogue control is mostly taken by the system. Every step in the dialogue is deterministic and the wizard's task is to interpret subjects' utterances strictly in terms of the dialogue model.

The directed dialogue model is characterized by two types of decision points (one where the wizard has to interpret the subject's reply, and one where the result of the database query has to be interpreted) and three kinds of actions that can be taken (message to the subject, message to the database operator and the spotting of the subject's reply).

The chart in figure 2. illustrates the global flow of the dialogue model. After the opening phase (where the system introduces itself) starts the data acquisition phase. When the data acqusition phase makes a database query possible (i.e. when there is an optimal number and combination of selection parameters necessary to yield a valid database request), the query result handling phase takes place. The process starts again until no other phone number is requested. In that case the dialogue will be closed. A detailed account of the directed dialogue model is given in [3].

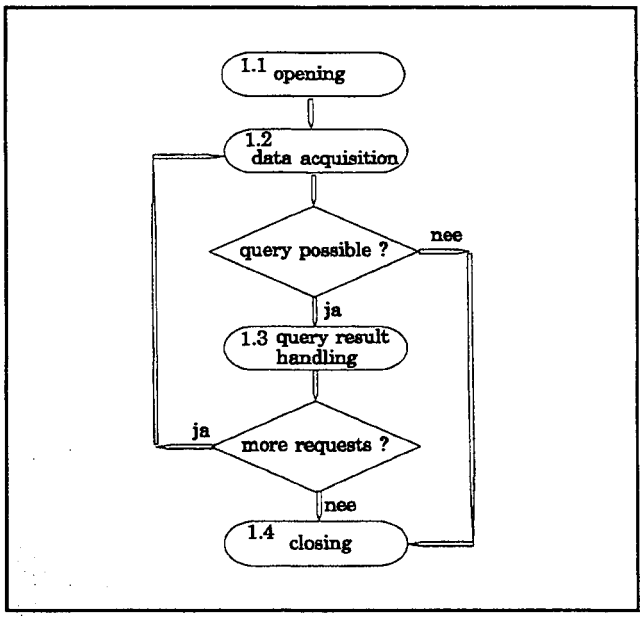

Figure 2. Chart of the directed dialogue model

3. METHODOLOGICAL APPROACH

The underlying methodology is articulated along a series of assumptions, which clearly distinguish this experiment from most current alternative approaches :

- the initial choice is to support our research on dialogue modeling with informative dialogues between two human beings : a Telecom Information Service operator and a client calling for an inquiry. Such conversations have proved useful for gaining insight into the global dialogue structure and exchange mechanisms occurring in a finalized environment such as checking, repairing and spelling;
- the second key aspect is to keep the experimental setting as authentic as possible. This means that the subjects, the dialogue task and the database used in the simulation are similar to the potential users, task and database of the eventual system. The same applies to the text-to-speech technique used under Wizard of Oz conditions;
- a third interesting dimension is the semi-automatic execution of the experiment, in the form of a window-based Wizard Interface, a stored dialogue transition network, a text-to-speech system output and the logging of the history of the dialogue. The simulated tasks are the subject's request understanding and interpretation, the decision to progress in the dialogue transition network, and the consultation of the database;
- an additional instructive facet of the experiment is the evaluation of the dialogue session by the subjects, in terms of (i) the subject's *impression* of the identity of the dialogue partner, (ii) the subject's *attitude* towards the dialogue partner and (iii) the subject's *preference* regarding the dialogue partner.

4. DISCUSSION OF THE RESULTS

In total 90 dialogue sessions took place during the experiment on site. Due to technical problems (3 dialogues) or to the fact that the subjects hanged up during the dialogue (5 cases), only 82 sessions were considered as valid (i.e. dialogues closed by the system's closing message).

4.1 Subject's conversational behavior

The first important observation was that subjects did not always provide answers we would like them to produce; either they gave too much (selection parameters in a sentence and not in isolation), too little (which prompted a correction loop) or wrong information. The spelling of proper names and street names was a recurrent difficulty, especially at the first try.

Another observation was that the same 'yes-no' question triggered a wide range of different reactions from the subjects.

We also observed that subjects have to get used to talking to a machine; 16% of the answers to the first system question contained pauses and hesitations at the beginning of the turn.

The subjects, however, rapidly adapt to machine behavior in the course of a dialogue session (less hesitations and no empty turns occurred after the second question).

4.2 Evaluation of the dialogue sessions

Almost all of the subjects (93%) replied affirmatively when they were asked whether they had the impression that they were communicating with a machine. The reason they most often gave was that the speech produced by the TTS-board did not sound fully humanly. This is due to the fact that actual speech synthesis technology cannot simulate the human voice in its complexity. But it also appears that near-human voice quality is not desirable in this type of experiment. Machine-like voice is a requirement if we want to analyse subjects' linguistic behavior when dealing with a machine. Actually, producing machine-like voice is a sufficient condition to carry out a valid Wizard of Oz experiment : non-expert subjects (general public) do not have a realistic image of the capabilities of current speech recognition systems in that they consider machine-like output to be the discriminating factor that characterizes a fully automatic dialogue system.

Furthermore, 89% of the interviewed subjects were satisfied about the way their request was treated. This figure has to be weighted down as some — probably unsatisfied — subjects could not be interviewed since they abruptly interrupted the dialogue session. Others disconnected immediately after the information dialogue (skipping the interview) and thus did not inform us about their satisfaction in using the system.

Finally, 71% of the subjects preferred to be informed by a human operator rather than by a computer. Their reasons were : *(i) speed* : the interpretation of the subjects' utterances by the wizard was not experienced to be real-time; *(ii) quality of speech* : some subjects complained about the audibility of the speech produced by the machine; *(iii) reliability* : some subjects consider human operators more reliable; *(iv) flexibility* : human operators are better for handling complex and unexpected situations; *(v) social reasons* : some subjects appreciate social contact with human operators. 9% preferred to be informed by a computer, and 20% had no preference. Surprisingly, most of the subjects were satisfied about the way their request was treated but preferred to be informed by human operators. Note that almost half of these subjects (43%) motivated their choice on the basis of subjective, non-functional criteria (social reasons).

5. CONCLUSION

We presented the design and evaluation of a real world Wizard of Oz experiment for a speech driven telephone directory information system that is applied to the database of the Belgian Telecom Information Service. Special attention has been devoted to (i) the adequacy of the proposed methodology to the design and evaluation of informative dialogue models, (ii) the pertinence of such models in an automated environment, and (iii) the evaluation of the dialogue sessions by the subjects.

6. REFERENCES

[1] Bourlard, H. (ed.) : Development of a Speaker Independent Speech Recognition Prototype for Automatic Directory Information. Progress report RTT-Belgacom R&D project, LHS, Brussels, 1990.

[2] Van Coile, B., Leys, S. & Mortier, L. : On the Development of a Name Pronunciation System. Proc. ICSLP 92, pp. 487-490, 1992.

[3] Deville, G. & Andernach, T. : A Wizard of Oz experiment in the RTT-Belgacom R&D project. BIKIT report, June 1992.

[4] Mortier, L. : Intonation in a Simulated Man-Machine Dialogue. M.A. thesis in Knowledge Technology. University of Ghent, 1992.