Simulation in Automated Speech Processing Systems

L.J.M. Rothkrantz R.J. van Vark
J.P.M. de Vreught E.J.H. Kerckhoffs

Knowledge Based Systems
Computer Science
Delft University of Technology
alparon@kgs.twi.tudelft.nl

Keywords: Automated Speech Processing, Simulation, Speech Recognition.

Abstract

Design and implementation of Automated Speech Processing (ASP) systems takes a lot of time and effort. To assess the impact of certain decisions, it is suitable to simulate those ASP systems in a so–called Wizard of Oz experiment. In this paper, some simulation experiments for ASP systems will be presented. The domain of these ASP systems is a telephone inquiry system for public transport information.

1 Introduction

In recent years many advances have been made in the field of Automated Speech Processing (ASP). However, the recognition of continuous speech is still far from perfect. Recently, the first telephone inquiry systems using ASP have been introduced. Traditionally, the human operator of such systems combines near–perfect speech recognition, an enormous vocabulary, and flexible dialogue management. Current ASP systems are not powerful enough to combine all these aspects of human operators. Therefore, ASP designers are faced with the challenge to find an optimal compromise between the restrictions of current speech technology and the complexity of human-human dialogues.

Wizard of Oz experiments are often conducted to study the impact of simplifying the human–human dialogue model. In such a Wizard of Oz experiment, an ASP system is secretly simulated by a human operator and respondents, unaware of the simulated design, are requested to call the system. The appreciation of the simulated system can be assessed by conducting interviews or by administering questionnaires. Improving an ASP prototype calls for a lot of additional testing and (re–)training.

A variation of the Wizard of Oz experiment is an experiment in which the customer is simulated. Such a system can be used as a testbed to assess the performance of successive releases of ASP systems.

In this study, the design and implementation of both types of simulation are discussed. The domain in which these simulations take place, is the provision of public transport information, which is currently provided by the Dutch company OVR. Over 400 human operators handle more than 18 million calls for information on traveling by train, busses and boat, a year. At this moment, customers of OVR can call the human operated service as well as an automated system based on ASP.

2 OVR dialogue model

Periodic survey studies have shown that the existing human operated service is a very effective and efficient system resulting in a high degree of user satisfaction. Therefore, it is reasonable to develop the ASP system based on the human–human dialogue model. Modelling the operator–customer interaction was done in a corpus–based approach using over 5000 recorded human–human dialogues.

It was proven that a 'standard' OVR conversation has an underlying scenario. A request for information on travelling by public transport centers around 5 topics: place of departure, place of arrival, time of departure, time

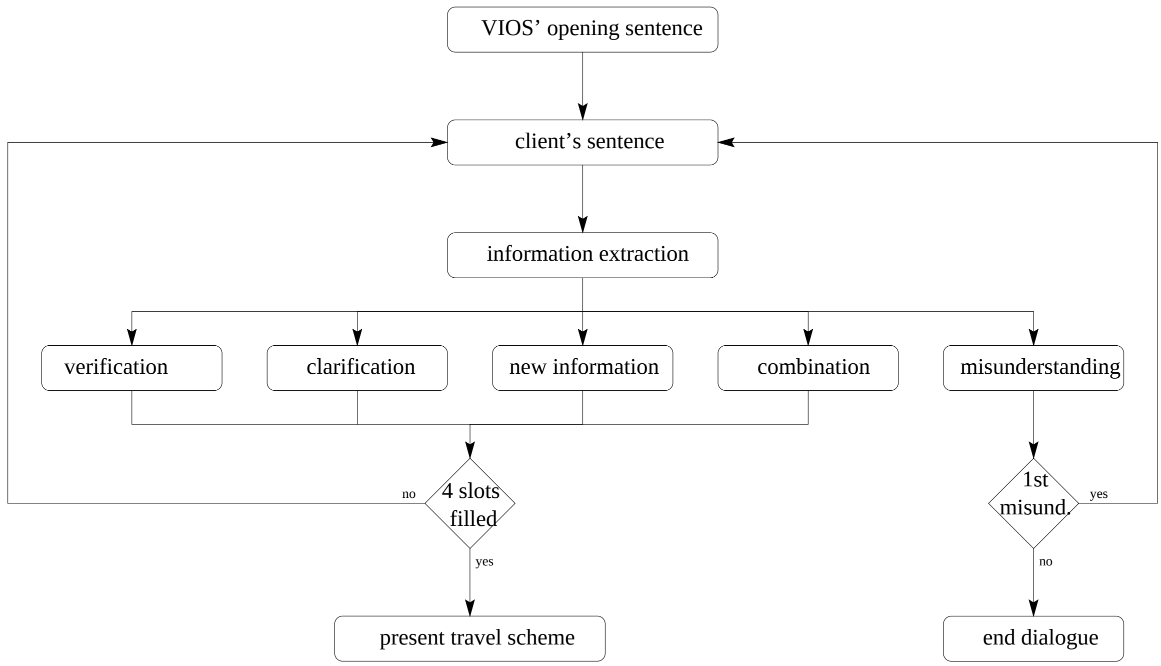

Figure 1: Data flow in a VIOS dialogue

of arrival, and the date of travelling. Both the human operator and the ASP system should be focussed on filling these five slots.

VIOS [5], the ASP prototype of OVR, uses a so–called 'blackboard' approach, in which the customer is requested to provide information to fill each slot. The ASP system extracts the relevant information from the customer's utterances which can occur in any order. In the current application, only exact arrival and departure times can be requested by the customer. If all appropriate slots are filled, the ASP system queries a database to retrieve a travel schedule and provides the information to the customer.

Although it is possible for the customer to provide all information in one utterance, the information is usually provided step-by-step. After every utterance of the customer, the ASP system asks for verification, clarification or new information. Because the available information system is restricted in place and time, it acts as a finite state machine. The dialogue manager of the ASP system can generate a finite number of prompts, which can be ordered as follows (see figure 1):

Additional information
The ASP system requests further information if two or more slots are open. The last open slot will be the arrival or departure time, which is the information requested by the customer. These prompts are typically given as wh-questions: "Which railway station do you want to travel to?" and "At what time do you want to travel?"

Verification
The system always asks for verification of the information extracted from the customer's utterances. For example, "You want to travel on [date] from [departure place] to [arrival place]?"

Clarification
The ASP system reduces ambiguity by requesting additional information or verifying the systems interpretation. For example, "Do you want to depart at 8 o'clock a.m. or 8 o'clock p.m.?"

Correction
The system requests correction of wrongly extracted information. For example: "You don't want to travel from Delft, please would you be so kind to make a correction."

Combination
The ASP system combines prompts from

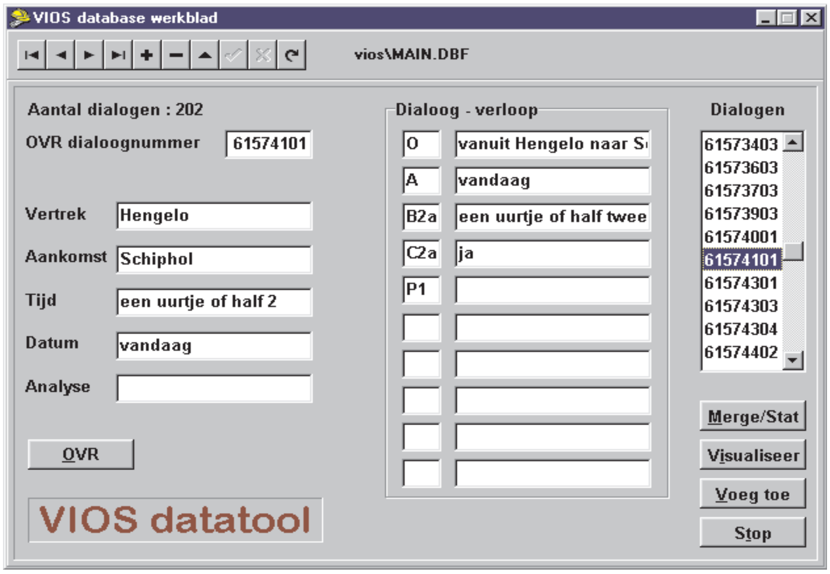

Figure 2: Graphical user interface of Simcall

different categories. The system can ask for verification followed by a request for additional information. Direct request for verification of every piece of extracted information is time consuming and can irritate the customer. Consider for example the combination of “You want to depart from Delft. At what time do you want to depart?” into “At what time do you want to depart from Delft?”

Misunderstanding
If the ASP system is unable to extract any information from the last utterance, it will ask for repetition. If misunderstanding occurs repeatedly, it will stop the dialogue and ask the customer whether he/she wants to be connected to a human operator.

In the VIOS prototype about 100 different templates are used. Everyone of these templates has one or more open fields, which can be filled with names of railway stations or time expressions.

3 Simulation

In a telephone inquiry system covering public transport information, both the operator and the customer can be simulated by an ASP system. Both simulations serve different purposes.

3.1 Simulation of the operator

In a Wizard of Oz design, an ASP system is simulated by a human operator. Customers are instructed to request information concerning a journey from one railway station to another. The human operator, i.e. the wizard, of the system was supplied with a standard set of prerecorded sentences. These sentences belong to one of the states as depicted in figure 1. The set of sentences is the result of an analysis of 500 human–human dialogues from the corpus of OVR dialogues. With this carefully selected set of sentences, the wizard has to manage the dialogues. For effective dialogue management, a scenario was designed, which is also based on the analysis of the OVR corpus. The scenario describes the guidelines for the wizard, i.e. the appropriate reactions to the customer’s utterances.

When all information slots are filled, a database query is defined. The relevant information from the journey planner was presented to the customer by using a speech generator based on concatenation.

3.2 Simulation of the customer

Training human operators or testing an ASP system takes a lot of time and effort. In case of testing an ASP system, it is difficult to find respondents willing to call at the appropriate time and place. There is a need for an automated testbed, i.e. a set of prerecorded customer calls. Again, a corpus–based approach is chosen to generate such a testbed. From 200 dialogues, specific sentences are selected, directly related to the five information slots (see figure 2).

This set is completed with specific sentences out of interactive dialogues. Based on the data flow diagram in figure 1, every customer utterance can evoke an operator reaction out of any of the five indicated categories. Also, an appropriate customer reaction has to be prepared to every system utterance.

Analysis of dialogues in the corpus shows most dialogues to have an underlying scenario. For every dialogue, the selected sentences are arranged according to this scenario. The basic scenario is depicted in figure 3. With a very limited set of command, symbolised by icons in a graphical user interface, a database can be queried for information.

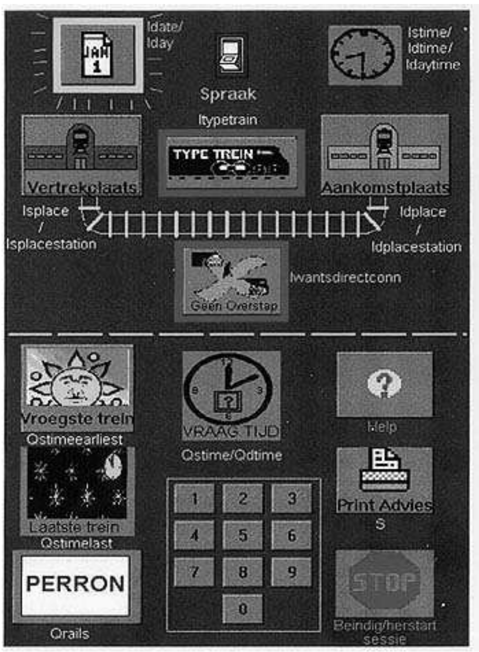

Figure 3: User interface using non–verbal scenarios

4 Experiments

Several experiments have been conducted to test the simulations described in the previous section.

4.1 Simulation of operator and caller

In this experiment, an ASP system for both the customer and operator are simulated. Two computers are connected of which one computer is equipped with a simulated operator system and the other with a simulated customer system. Both systems are composed of the following modules:

- a database containing all prerecorded utterances. These utterances are grouped according to the scenario.
- A graphical user interface. By clicking appropriate buttons a corresponding message can be sent between both computers.

In a first experiment, 50 utterances were installed on both computers. Some of these utterances have open slots for railway stations and time expressions. From the population of students from the Delft University of Technology, five student pairs were selected to act as wizard for customer or operator. The simulated customer was asked to inquire about a scheme for a certain train journey based on 25 non-verbal scenario (see figure 3. The simulated human operator had to manage the dialogues and to fill the information slots.

It was possible to fulfill 80 % of the dialogues using these 50 utterances. One can conclude that an appropriate scenario had been designed as well as a representative set of utterances. The messages are written text and we have to realise that the generated dialogues are no completely natural dialogues. The successive utterances of customer and operator have a correct semantic relation, but they are probably ill–defined from a grammatical or linguistic viewpoint.

4.2 Simulation of the operator

In a second experiment, the operator was simulated in a Wizard of Oz design. About 400 students were requested to call the simulated ASP system for the OVR inquiry system to assess the impact of some parameters on the appreciation of the ASP system. In a completely randomised factorial design, we manipulated six variables and assessed the degree of appreciation of the simulated system. The following six experimental modes were simulated:

- mistakes: the ASP system deliberately misinterpreted similar sounding station names (i.e. Weesp instead of Wezep).
- verification: the user was always asked for verification of filled slots.
- dialogue style: in the non–directive style the initiative was mostly left to the user (mixed–initiative), while in the directive style the system asked for each slot separately.
- break/tempo: before every operator utterance a 3 second delay was introduced, in order to simulate non–real–time operation of the system.
- voice: the sex of the simulated voice could be set to male or female.
- explicitation: in this mode the system informed the user about its current status.

A non–directive dialogue style was highly appreciated and the number of mistakes is negatively correlated with appreciation of the system. The other variables had no significant effect on the appreciation of the system.

4.3 Simulation of the caller

In a third experiment, the caller was simulated in a Wizard of Oz design. In total, 25 students were requested to call the OVR–system run by human operators. Almost all the calls were successful, i.e. they resulted in a correct travel schedule. The human operator handled all the simulated calls as regular calls. As the Wizard of Oz design was kept secret from the human operator, they could not be asked about their opinion on the 'simulated' calls.

5 Conclusions

From our experiments it can be concluded that it is possible to simulate human operators and customer in a telephone inquiry system environment for public transport information. A limited number of 50 utterances is sufficient to simulate the dialogues between customer and human operator.

A Wizard of Oz design for operator and customer proved to be a valuable method to assess the impact of variables on the performance and appreciation of automated speech recognition systems.

References

[1] M. Puype, *O.V.R. Dialogue Structuring —a corpus–based approach—*, M.Sc. thesis, Delft University of Technology, 1996.

[2] L.J.M. Rothkrantz, W.A.Th. Manintveld, M.M.M. Rats, R.J. van Vark, J.P.M. de Vreught, H. Koppelaar, *An Appreciation Study of an ASR Inquiry System*, To appear in the Proceedings of the EuroSpeech 1997 Conference, 1977.

[3] L.J.M. Rothkrantz, R.J. van Vark, H. Koppelaar, *Corpus–Based Test System for an Automated Speech Processing System*, Proceedings of the SALT-97 Workshop on Evaluation in Speech and Language Technology, pp. 164–171, 1997.

[4] R.J. van Vark, J.P.M. de Vreught, L.J.M. Rothkrantz, *Classification of Public Transport Information Dialogues using an Information Based Coding Scheme*, Workshop Dialogue Processing in Spoken Language Systems, European Conference on Artificial Intelligence, 92–99, 1996.

[5] G. Veldhuijzen van Zanten, *Pragmatic Interpretation and Dialogue Management in Spoken–Language Systems*, Proceedings of the 11th Twente Worshop on Language Technology, 81–88, University of Twente.