A SLOVENIAN SPOKEN DIALOG SYSTEM FOR AIR FLIGHT INQUIRIES*

I. Ipšić & F. Mihelič & S. Dobrišek & J. Gros & N. Pavešić
e-mail: ivoi@fe.uni-lj.si
University of Ljubljana, Faculty of Electrical Engineering
Tržaška cesta 25, SI-1000 Ljubljana
SLOVENIA

10.21437/Eurospeech.1999-587

ABSTRACT

In this paper we give an overview of a Slovenian spoken dialog system, developed within a joint project in multilingual speech recognition and understanding. The aim of the project is the development of an information retrieval system, capable of having a dialog with a user. The paper presents the work on the development of the Slovenian spoken dialog system. Such a system has to be able to handle spontaneous speech, and to provide the user with correct information. The information system being developed for Slovenian speech is used for air flight information retrieval. The system has to answer questions about air flight connections and their time and date.

In the paper we present the developed modules of the Slovenian system and show some results with respect to word accuracy, semantic accuracy and dialog success rate.

Keywords: Spoken Dialog System, Word Recognition, Linguistic Analysis, Dialog Management, Speech Synthesis

1. INTRODUCTION

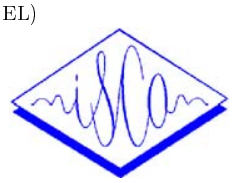
The Slovenian spoken dialog system is being developed within the joint project in multilingual speech recognition and understanding *Spoken Queries in European Languages* (SQEL-Copernicus -1634). The final aim of the project is to built a multilingual and multifunctional system, capable of having a dialog with the user in one of the four European languages (German, Slovenian, Czech and Slovak) about a task oriented topic. Such a system has to be able to handle spontaneous speech, and to provide the user with correct information. The information system being developed for Slovenian speech is used for flight inquiries.

The development of a spoken dialog system concerns solutions to speech recognition problems as well as to speech understanding and human machine interaction problems. One approach to spoken dialog systems design is to solve the problems in different communicating modules. The architecture of the Slovenian system is based on the Erlangen Train Time Table Inquiry System [2], developed within the European Esprit project Sundial [5]. The German system can perform human machine continuously spoken dialogs about train connections via telephone. Figure 1 shows the architecture of the spoken dialog information retrieval system. The major modules perform word recognition, linguistic analysis, dialog management and speech synthesis.

The information system being developed for Slovenian speech is used for air flight information retrieval. The system has to answer questions about air flight connections and their time and date. The main differences between the German system and the Slovenian are: a word recognition module for the Slovenian language, a Slovenian parser and Slovenian text-to-speech synthesis. The first step in spoken dialog system design is the collection of speech material, which is used for dialog modeling as well as for building statistical models for the word recognizer. We have collected recordings of dialogs between anonymous clients and telephone operators at the Adria Airways information service. From these dialogs we created the Slovenian speech database GOPOLIS of 5000 sentences, which was read by 50 speakers [1]. The read speech database was used for the creation of acoustic and language models, as well as for testing of the parser and the dialog manager. We have tested the separate modules with microphone and telephone quality speech. The overall system was tested with cooperative users and dialog completition and success rate was estimated.

In the following subsections we describe the main modules, the adaptation steps and show some results.

*This work was partly funded by the Commission of the European Community under COP-94 contract No 01634 (SQEL)



6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

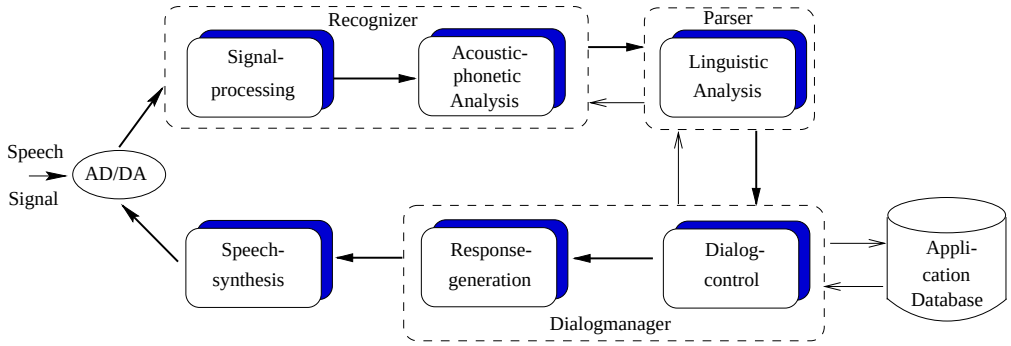


Figure 1: Architecture of the Slovenian spoken dialog information retrieval system.

2. Word Recognition

The speech signal is sampled at 16 kHz with 16 bit and mel–cepstrum features and their derivatives are computed every 10 ms. The speech signal vectors are transformed into symbols using a soft vector quantization technique. The recognizer is based on hidden Markov models of context dependent phone units - polyphones. A polyphone consists of a phone with arbitrary length of left and right context of phones, and it is determined from the training database. The only criterion to form a polyphone is the minimal number of its occurrences in the words of the training sentences. To define the context dependent units we have proposed a set of 33 Slovenian phones and their HMM models, which give the highest recognition accuracy. To train the polyphone models we use the ISADORA system [7]. The 827 words are modeled with 2086 polyphone models of different length.

The recognizer is built by parallel concatenation of word models, which are constructed from context dependent phone models. Best word sequences are generated using Viterbi beam search and a categorical bigram language model with perplexity 10. We have defined 127 word categories to classify the words from the Slovenian recognition vocabulary. The categories group words with same grammatical and semantical characteristics. The bigrams were trained on 10000 sentences comprising 827 different vocabulary words. The hypothesized word sequences are then verified using a trigram language model. The result is the most probable word sequence [4]. The elementary hidden Markov models of polyphones are trained with read speech using the Slovenian speech database GOPOLIS.

The recognizer has been evaluated on microphone signals as well as on telephone speech. The speaker independent word recognition accuracy is over 90%, while the word accuracy for telephone quality speech signals is in average 76%.

3. Linguistic Analysis

Input to the linguistic module is the best matching word sequence, its output is the semantic interpretation of the utterance in the semantic interpretation language (SIL) expression [2]. Besides the known problems which occur during the design of a linguistic analysis system additional problems arise due to the nature of the Slovenian language: a large number of inflected word forms and a rather free word order. Therefore we have proposed a robust semantic-driven analyser to select the most important information from the recognized sentences [6]. The semantic-driven analyser performs three different steps:

- parsing of temporal expressions,
- parsing of simple noun words. Their meaning is often deduced from a simple semantic category, which they belong to, and from their position in the sentence. This method is used for departure/arrival city determination.
- The rest of the sentence is parsed using a very simple parser that tries to locate as many keywords as possible. These keywords are used in further processing to determine the sentence type.

The semantic-driven analyser is based on the DCG formalism, and is implemented in Prolog. It can han-

dle also ungrammatical and colloquial expressions. Phrases, which the analyser cannot parse are:

- sentences where the time phrase is divided throughout the sentence,
- determination of arrival or departure town can fail in those rare examples where there is no difference between the first and fourth case of a noun and
- some complex time expressions.

We have tested the linguistic analyser on different sets of recognized read sentences and spontaneous dialog utterances. Most of the relevant information needed for an air flight inquiry was parsed correctly. Since the parser is semantic-driven also a set of wrongly recognized sentences can be parsed correctly.

4. Dialog Management

The dialog manager takes the semantic representation of the user utterance and performs the interpretation within the current dialog context. The dialog manager from the German demonstrator is language and application independent, so it can be used for the Slovenian dialog system. The dialog manager consists of a number of communicating sub–components [3]:

- the linguistic interface, which enables the input from different parsers,
- the belief module, responsible for interpreting user utterances in the current dialog context,
- the task module, responsible for handling the task and performing a database query,
- the dialog module, responsible for the pragmatic interpretation and planning of system utterances and
- the message planer, responsible for message generation.

The adaptations for a new language and a new application have to be done within the task module and the message planer. For the Slovenian dialog system the flight timetable, provided by the Slovenian airline company Adria Airways is used. The timetable, which is available over the WWW, is transformed into Prolog facts representing the timetable data.

First we have tested the dialog manager with typed input. The dialog manager performs a database query when having all the relevant parameters needed for a successful database access. In the nesxt step the dialog manager has been connected with the recognizer using microphone input. The recorded dialogs show that in about 50% of dialogs the users receive the desired information.

5. Message Generation and Text–to–Speech

The SIL representation of the system answer is converted to text via sentence tabloids. The text is transformed to speech with a simple concatenation technique of prerecorded time signals for isolated words. In the German system this technique proved to provide intelligible system answers and can easily and quickly be implemented in the other languages. Another possibility is the use of synthesized speech. A system for Slovenian synthesis was developed.

The different phases of the synthesis task are performed by separate independent modules, operating sequentially. Input text is translated into a *series of phonemes* in two consecutive steps. First, abbreviations are expanded to form equivalent full words using a special list of lexical entries. A text preprocessor converts further special formats, like numbers or dates, into standard grapheme strings. The rest of the text is segmented into individual words and basic punctuation marks. Next, word pronunciation is derived, based on an user extensible pronunciation dictionary and letter-to-sound rules. The dictionary covers 16.000 most frequent Slovenian words along with 350 most frequent proper names. A prosodic generator assigns pitch and duration values to individual phonemes.

Special attention is payed to segmental duration determination, where the effect of speaking rate on phone duration were widely studied. The fundamental frequency is first predicted on a word basis – modeling the words' tonemic accent, later a global intonation countour is superponed. Once appropriate phonetic symbols and prosody markers are determined, the final step within a text-to-speech system is to produce audible speech by assembling elemental speech units. This is achieved by taking into account computed pitch and duration contours, and synthesising a speech waveform. A concatenative diphone synthesis technique was used. The TD-PSOLA (Time Domain Pitch Synchronous Overlap and Add) [8] scheme enables pitch and duration transformations directly on the waveform, at least for moderate ranges of prosodic modifications without considerably affecting the quality of synthesised speech.

First tests were performed, assessing the intelligibility and naturalness of the synthesized speech [9].

6. Conclusion

In the paper we have presented the architecture of the Slovenian spoken dialog system. The system is based on the German SUNDIAL demonstrator, which can perform dialogs about intercity timetables. The main adaptation steps when implementing the comparable system are: definition of subword units and the language models for Slovenian word recognition, a Slovenian parser and a Slovenian text-to-speech module. So far experiments with speaker independent Slovenian speech recognition and parsing of recognized utterances have been performed. The dialog manager was modified for the Slovenian language and for the air flight information retrieval task. Experiments with typed and spoken input have shown that database queries can provide the users with the desired information. These results are appropriate for the Slovenian information system, which enables a dialog with a user over the telephone line. So far we have collected a number of spoken dialogs over the telephone. The results are encouraging and show that in about 50% of dialogs the users get the desired information.

REFERENCES

[1] S. Dobrišek, J. Gros, F. Mihelič, N. Pavešić: *Recording and Labeling of the GOPOLIS Slovenian Speech Database*, First International Conference on Language Resources and Evaluation, edts.: A. Rubio, N. Gallardo, R. Castro, A. Tejada, May 1998, Granada, Spain, Vol. II, pp. 1089 – 1096.

[2] W. Eckert, T. Kuhn, H. Niemann, S. Rieck, A. Scheuer, and E.G. Schukat-Talamazzini. A Spoken Dialogue System for German Intercity Train Timetable Inquiries. In *Proc. European Conf. on Speech Technology*, pages 1871–1874, Berlin, 1993.

[3] Wieland Eckert. Customizing the Erlangen Dialog Manager. Technical report, Lehrstuhl für Mustererkennung, FAU Erlangen–Nürnberg, 1996.

[4] I. Ipšić, F. Mihelič, N. Pavešić, and E. Nöth. Slovenian Word Recognition. In *Proceedings of the 3^{rd} Slovenian-German and 2^{nd} SDRV Workshop*, Ljubljana, 1996.

[5] J. Peckham. Speech Understanding and Dialogue over the Telephone: an Overview of Progress in the SUNDIAL Project. In *Proc. European Conf. on Speech Technology*, volume 3, pages 1469–1472, 1991.

[6] K. Pepelnjak, F. Mihelič, N. Pavešić: *Semantic Decomposition of Sentences in the System Supporting Flight Services,* Journal of Computing and Information Technology, Vol. 4, No. 1, 1996, pp. 17 – 24.

[7] E.G. Schukat-Talamazzini. *Automatische Spracherkennung - Grundlagen, statistische Modelle und effiziente Algorithmen.* Künstliche Intelligenz. Vieweg, Braunschweig, 1995.

[8] E. Moulines, F. Charpentier: *Pitch-Synchronous Waveform Processing Techniques for Text–to–Speech Synthesis Using Diphones*, Speech Communications **9**, pp. 453 – 467, 1990.

[9] J. Gros, N. Pavešić, F. Mihelič: *Text–to–Speech Synthesis: A complete system for the Slovenian Language*, Journal of Computing and Information Technology, Vol. 5, No. 1, 1997, pp. 11 – 19.