From the Proceedings of the '95 *RWC Symposium*, Tokyo, Japan

MASSIVELY–PARALLEL KNOWLEDGE PROCESSING FOR COMPLEX PATTERN UNDERSTANDING

J. Fischer, V. Fischer

Universität Erlangen-Nürnberg, Lehrstuhl für Mustererkennung (Informatik 5)
E-mail: fischerv@informatik.uni-erlangen.de

1. INTRODUCTION

Automatic understanding of complex pattern, specifically spontaneous speech and motion images, will become increasingly relevant for a great number of applications in the future. Such applications include service roboters to aid handicapped individuals, multi–modal telecooperation tasks, etc. In order to apply such systems in the real world, real–time processing capabilities are indispensable.

The work presented in this paper concentrates on the automatic understanding of spontaneous speech. A new dialog system will be presented. The any–time behaviour of the system will strongly be supported by the exploitation of parallelism on different levels and the use of linguistic and prosodic constraints.

The paper is subdivided into 4 parts: the introduction is followed by a review of past research that provides the background for our work. The structure of the dialog system is then presented, followed by a description of the parallel control algorithm. The paper will be concluded by some final remarks.

2. BACKGROUND RESEARCH

The dialog system presented here is based on the background research specified below:

- The speech understanding and dialog system EVAR[1]. This system is able to answer inquiries about the German intercity train time–table and serves as a general framework for our work.The declarative knowledge of the knowledge base of EVAR, divided into declarative and procedural knowledge, is entirely built using the semantic network formalism of ERNEST. In an integrated approach, both linguistic analysis and dialog management are handled by the use of the ERNESTcontrol algorithm.

- The ERlangen semantic NEtwork SysTem ERNEST[2] is a system shell for general pattern understanding tasks. It provides a semantic network formalism for knowledge representation and a problem–independent control algorithm, which is based on a modified A^*–algorithm in the original (sequential) version. The original control algorithm alternates between a bottom–up and a top–down search for the instantiation of concepts (i.e. the mapping of the input string into the internal represented knowledge). This system is applied also to image analysis [3].

- Recently, a new parallel control algorithm to be used with a subset of the semantic network language of the ERNEST–system was developed [4]. This algorithm employs a massively–parallel instantiation scheme for the concepts of a knowledge base, as well as parallel optimization techniques for the stepwise improvement of intermediate instantiation results. It was successfully applied to an image understanding task dealing with the interpretation of real world traffic scenes.

In our approach for a real–time dialog system, linguistic analysis (i.e. interpretation) of a spoken utterance will be controled by the above mentioned parallel control algorithm. This requires, since in its first application the algorithm was developed to be used only on a restricted subset of the ERNEST formalism, expansion of the algorithm in order to use it on the entire network formalism.

For the management of the different dialog steps, a new autonomous dialog module will be developed.

3. THE NEW DIALOG SYSTEM

Figure 1 shows the basic structure of the new parallel dialog system being developed. The two main components are the acoustic processing and the linguistic analysis. Input of the system is continuous speech. The system's *knowledge–base*, modeled in the semantic network formalism of ERNEST, consists of 5 abstraction levels: *Word–hypotheses, syntactic, semantic, pragmatic* and *dialog* level. The declarative knowledge is represented by *concepts* which are connected by *links*. The procedural knowledge is programmed in C. The word–hypotheses level represents the interface between speech recognition and speech understanding. The dialog level is the highest abstraction level and models the dialog acts.

The *acoustic processing* will analyze the speech signal and generate, by means of Hidden Markov Models and a stochastic grammar, a word hypotheses graph. This word hypotheses graph, in combination with prosodic information also extracted from the speech signal, will serve as input for the linguistic analysis component. For further details of the acoustic processing see [5].

The *linguistic analysis* will be controled by the parallel iterarive control algorithm. It will build instances of concepts for word chains of the word hypotheses graph (i.e. map word chains into internally represented knowledge), searching for the best valuated instance of a goal–concept. A goal–concept is in general a concept on dialog level, representing a dialog act. This instance of a goal–concept will then be delivered to the dialog manager, which will control the reaction of the system. It will retrieve

the train–connection from the database if requested and will put constraints on the analysis of the next utterance based on the present dialog state.

For instance, let the first statement of a user and input for the system be "Good evening". Let the word hypotheses graph generated by the acoustic processing component contain the word–chain *good evening*. The linguistic analysis will then instantiate the concept of the knowledge–base representing users greetings. According to this instantiated concept, the dialog manager will activate the system's response: "Hello, this is the automatic train time–table information system. Please go ahead with your inquiry". The next dialog step then expected by the system will be an information request. Thus it will restrict the analysis of the next utterance in order to support the instantiation of a goal–concept representing an information request.

The main idea of the parallel iterative control algorithm is briefly sketched in the following section.

4. PARALLEL CONTROL ALGORITHM

Linguistic ambiguities and competing word hypotheses lead to a search task which is to find the best interpretation of a word chain included in a word graph. The function of the control algorithm is to guide this search. In our case the control algorithm will search for the instance of a goal–concept with the highest score. The score of an instance of a concept will combine acoustic and linguistic scores.

The operations performed by the parallel control algorithm can be divided into a *network level* and a *control level*. The network level comprises operations for the parallel instantiation of concepts; the control level provides strategies for the computation of an optimal interpretation and has to handle competing hypotheses and competing linguistic interpretations.

Parallelism on the network level is made possible by the encoding of the concept–centered representation of the knowledge–base into a dataflow network (see Fig. 1) by splitting up concepts into independently computable components neccessary for their instantiation, e.g. attributes, structural relations and concept scores. These components are represented by nodes. Parallel instantiation proceeds from initial attributes (nodes without predecessors) up to goal–concept scores (nodes without successors).

The word hypotheses linked to the instance of the goal–concept are used to compute the valuations for the goal–concepts which in turn are used for the computation of an error criterion. The latter is minimized by means of combinatorial optimization techniques (e.g. *simulated annealing*, *threshold acceptance*, *great deluge algorithm* and *genetic algorithms*), leading to an instance of a goal–concept with minimal error after some iteration steps. The optimization is parallelized in a local network using the interprocessor communication facilities of the parallel virtual machine PVM.

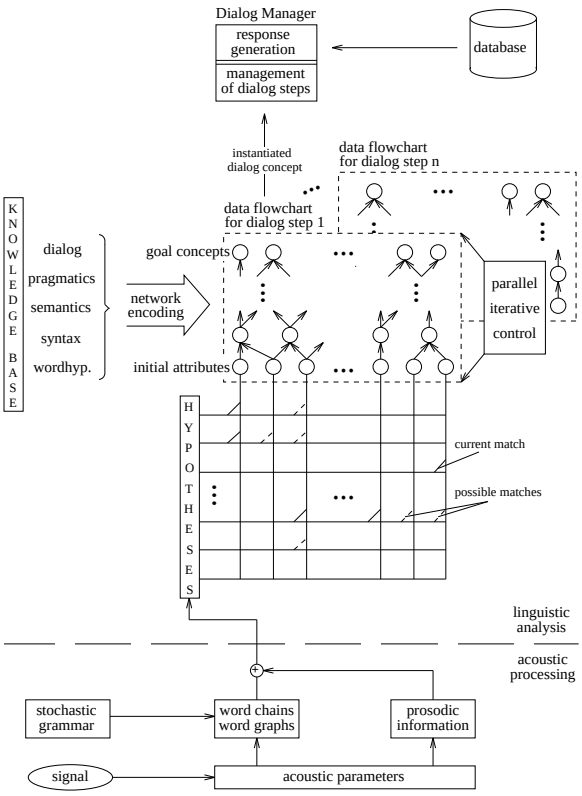

Figure 1: Architecture of the parallel speech understanding and dialog system.

5. CONCLUSION

The current state of the implementation shows the feasability of the approach described here. The fact that this parallel control approach was also successfully applied to image analysis (see section 2) confirms that it is well suited for pattern understanding tasks requireing the interpretation of multi–sensor (e.g. speech and image) input. Besides further work of adaptation of the parallel control algorithm on the entire ERNEST network formalism, future work will focus on the reduction of search space for improving the any–time behaviour of the system by the incorporation of restrictions from several sources of information (e.g. linguistic, prosodic information).

References

[1] M. Mast, F. Kummert, U. Ehrlich, G. Fink, T. Kuhn, H. Niemann, and G. Sagerer. A speech understanding and dialog system with a homogeneous linguistic knowledge base. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16(2):179–194, 1994.

[2] H. Niemann, G. Sagerer, S. Schröder, and F. Kummert. ERNEST: A semantic network system for pattern understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(9):883–905, 1990.

[3] H. Niemann, H. Brünig, R. Salzbrunn, and S. Schröder. Interpretation of industrial scenes by semantic networks. In *Proc. of the IAPR Workshop on Machine Vision Applications*, pages 39–42, Tokyo, 1990.

[4] V. Fischer. *Parallelverarbeitung in einem semantischen Netzwerk für die wissensbasierte Musteranalyse.* PhD thesis, Technische Fakultät der Universität Erlangen-Nürnberg, 1995.

[5] H. Niemann, E. Nöth, E.G. Schukat-Talamazzini, A. Kießling, R. Kompe, T. Kuhn, K. Ott, and S. Rieck. Statistical Modeling of Segmental and Suprasegmental Information. NATO ASI Series F. Springer–Verlag, Berlin, Heidelberg, New York, 1995. (to appear).