

AN EXPERIMENTAL DIALOGUE SYSTEM: WAXHOLM

*Mats Blomberg, Rolf Carlson, Kjell Elenius, Björn Granström,
Sheri Hunnicutt, Roger Lindell and Lennart Neovius**

*Department of Speech Communication and Music Acoustics,
Box 700 14, KTH, Stockholm, S-10044 Sweden*

**Names in alphabetic order.*

Recently we have begun to build basic tools for a generic speech dialogue system. The main modules, their function and internal communication have been specified and tested. The different components are connected in a computer network. A preliminary version of the system has been tested using simplified versions of the modules.

THE DEMONSTRATOR APPLICATION, WAXHOLM

We intend to build a generic system where speech synthesis and speech recognition can be studied in a man-machine dialogue framework. In addition, the system should facilitate the collection of speech and text data that are required for the development of the system. The demonstrator application, that we call WAXHOLM, gives information on boat traffic in the Stockholm archipelago (see Figure 1). Besides the speech recognition and synthesis components, the system will handle graphic information such as pictures, maps, charts and time-tables. The standardised query language (SQL) is used to access the database, which contains boat schedules, port locations and other information about the Stockholm archipelago. The application has great similarities to the ATIS domain within the DARPA community and other similar tasks in Europe. The possibility to expand the task in many directions is an advantage for future research on interactive dialogue systems. An initial version of the system based on text input has been running since September 1992. The speech recognition component will handle continuous speech with a vocabulary of about 1000 words.

SYSTEM OVERVIEW

The dialogue system consists of a number of independent and specialised modules that run as servers on our computer system. A notation has been defined to control the information flow between them. The structure makes it possible to run the system in parallel on different machines and facilitates the implementation and testing of alternate models within the same framework. The communication software is based on UNIX de facto standards, which will facilitate the reuse and portability of the components. Currently part of our efforts are spent on making the communication

software robust with efficient error handling. Simultaneously the recognition module needs to be integrated into the system by an improved notation, making data transfer fast and efficient.

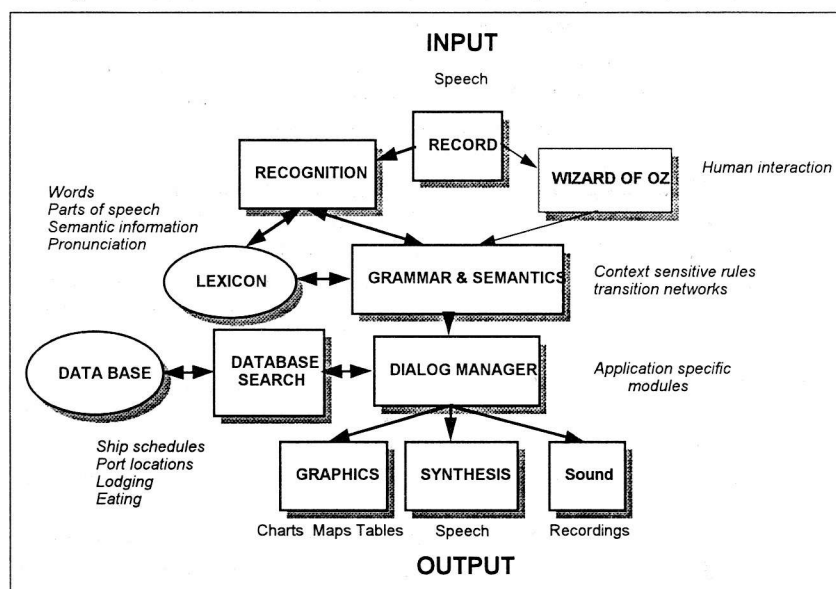


Figure 1. Block diagram of the demonstrator application, Waxholm.

NATURAL LANGUAGE COMPONENT

Our initial work on a natural language component is focused on a sublanguage grammar, a grammar limited to a particular subject domain: that of requesting information from tables about transportation. This component provides syntactic and semantic knowledge to the recogniser.

We are aiming to develop a parser that is technically robust -- a parser that is efficient and fast, that is statistically sound, and that fails gracefully. We are also stressing the interactive development in order to have control over the system's progress as more components are added.

The fundamental concepts are inspired by TINA, a parser developed at MIT (Seneff, 1989). At the last Swedish Phonetics Conference, we presented our parser, STINA, i.e., Swedish TINA, (Carlson and Hunnicutt 1992.) STINA is knowledge-based and is designed as a probabilistic language model. It contains a context-free grammar which is compiled into an augmented transition network (ATN). Probabilities are assigned to each arc after training. Features of STINA are a stack decoding search strategy and a feature-passing mechanism to implement unification where both syntactic

and semantic features are defined. The continued work on this component will be guided by the results from our pilot experiments that are now in progress.

Dialogue handling is also implemented in STINA making use of the lexical semantic features. Topic selection is done by a probabilistic approach that needs application-specific training. Thus, the data collection is of great importance for the progress of the project. The dialogue will be naturally restricted by application-specific machine capabilities and the limited grammar. We also assume that the human subjects will be co-operative in pursuing the task. Recovery in case of human-machine "misunderstandings" will be aided by informative error messages generated upon the occurrence of lexical, parsing or retrieval errors. This technique has been shown to be useful in helping subjects to recover from an error through rephrasing of their next input (Hunnicut, Hirschman, Polifroni and Seneff, 1992).

APPLICATION-SPECIFIC DATA COLLECTION

We are now in the position of starting data collection of speech and text data using the WAXHOLM system. Initially, a "Wizard of Oz" (a human simulating part of a system) will be replacing the speech recognition module, Figure 2 and Table 1. Utterance size speech files will be stored together with the text entered by the Wizard. This corpus is being used for grammar development, for training of probabilities in the language model in STINA, and also for generation of an application-dependent bigram model to be used by the recogniser. The collected text is also being used to train word collocation probabilities. Our plan is to replace explicit formulations of semantic coupling by a collocation probability matrix.

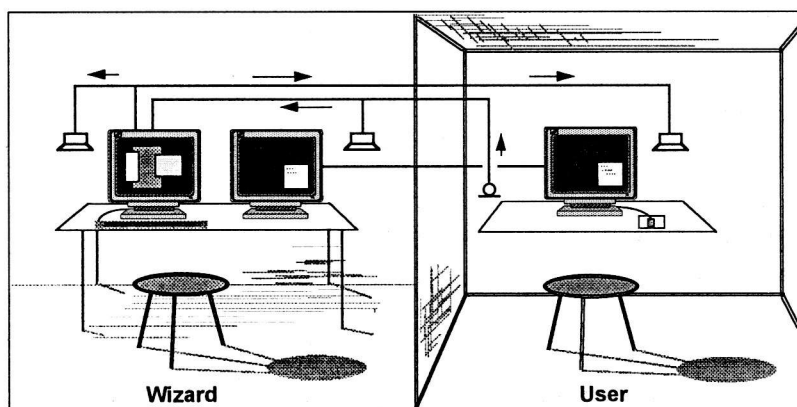


Figure 2. Hardware setup for data collection, with the help of a Wizard.

<synthesis>	Du vill åka ut i skärgården under helgen. För att ha det bekvämt vill du bo på hotell. Du slutar klockan 3 på fredagen. Ta reda på var man kan bo och när du kan åka.
<speech input>	var finns det hotell i skärgården . SEMANTIC_PARSE: (TOP-LEVEL (QUESTION (QUEST-WHERE "var"/PLACE/) (Q-VERBAL (SUBJECT) (OBJECT "hotell"/HOTEL/)) (AT-PLACE "skärgården"/REGION/))) proposed topic EXIST
<synthesis>	Programmet söker nu i data-basen .
<SQL>	EXIST PORT = * HOTEL = END
<synthesis>	Informationen om hotell visas i tabellen .
<VDU>	 SANDHAMN
<speech input>	när går första båten till Sandhamn på fredag efter klockan 3 . SEMANTIC_PARSE: (TOP-LEVEL (QUESTION (QUEST-WHEN "när"/TIME/) (Q-VERBAL "går"/MOVE/ (SUBJECT "första"/QUANT/ "båten"/BOAT/)) (TO-PLACE "sandhamn"/PORT/) (AT-DAY "fredag"/DAY/) (FROM-TIME (AFTER-TIME "3"/QUANT/)))) proposed topic TIME_TABLE
<synthesis>	Var ifrån vill du åka.
<speech input>	från Lidingö .
<synthesis>	Jag förstår inte ordet, Lidingö.
<speech input>	jag vill åka från Klippudden . SEMANTIC_PARSE: (TOP-LEVEL (STATEMENT (SUBJECT) (VERBAL "åka"/MOVE/) (FROM-PLACE "klippudden"/PORT/))) proposed topic TIME_TABLE
<synthesis>	Programmet söker nu i data-basen .
<SQL>	TIME_TABLE FROM PORT = KLIPPUDDEN DAY = FREDAG TIME > 15.00 TO PORT = SANDHAMN DAY = FREDAG END
<synthesis>	detta är en tabell över de båtar som går från Klippudden till Sandhamn på en fredag .
<VDU>	BÅT AVGÅNG ANKOMST WAXHOLM II 17.25 20.00

Table 1. Example of a dialog gathered during the first experiments.

REFERENCES

- Carlson, R., & Hunnicutt, S. (1992): "STINA: A probabilistic parser for speech recognition," FONETIK'92, Sixth Swedish Phonetics Conference, May 20-22, 1992, Technical Report No. 10, Dept. of Information Theory, Chalmers University of Technology, Göteborg.
- Hunnicutt, S., Hirschman, L., Polifroni, J., & Seneff, S. (1992): "Analysis of the effectiveness of system error messages in a human-machine travel planning task," ICSLP 92 Proceedings, Vol. 1, University of Alberta, Canada.
- Seneff, S. (1989): "TINA: A probabilistic Syntax Parser for Speech Understanding Systems," Proc. ICASSP-89.