A Voice Man Machine Dialogue System Generating Written Medical Reports

P. Nugues, E. Bellegarde, H. Coci, R. Pontier, and P. Trognon

Institut des Sciences de la Matière et du Rayonnement
Caen, France

***Abstract* – In this paper we present the architecture and implementation issues of a dictation machine prototype for medical reports. This machine features a voice recognition input and a voice synthesis output. The physician dictates the results of a medical analysis to the system, controls the construction of the report as well as the system operation by an oral dialogue in natural language. At the end of the dictation, the system generates a written report.**

In addition to signal processing, voice recognition and synthesis hardware, this prototype has a knowledge-based architecture whose Knowledge Sources perform the dialogue, the syntactic analysis and the semantic analysis of sentences. The entities handled by the system are represented in an object oriented form.

INTRODUCTION

The production of medical reports is inherent in most of the medical analyses, in radiology, cardiology, neurology, etc. These analyses are generally described by the physician using a tape recorder. The report must be subsequently transcripted by a medical secretary.

Attempts are made to generate automatically the text of the report, directly from the dictation of the physician. This involves speech recognition and language processing techniques. Ideally this technique would save a lot of time because the report is immediately created, available for rereading, possible modifications or corrections.

Speech processing involves several levels of representation. The first level is acoustic. It includes data acquisition and sampling. The phonetic level corresponds to the recognition of sounds (phonemes). These phonemes are assembled into words belonging to a vocabulary at the lexical level. The location of words within sentences corresponds to the syntactic level. Semantics is the meaning of what the sentences describe.

Electronic circuit cards and systems which can perform real time speech recognition are appearing on the market. These systems, operating on a limited vocabulary, can recognize isolated words (a slight pause must separate them) or connected words. Present medical applications allow report dictations based on trigger words [1].

DESCRIPTION OF THE SYSTEM

The system we implemented uses a word recognition circuit. It parses the sentence syntax and performs certain semantic checking. In addition, the system manages a dialogue with the physician to compensate for possible failures from the recognition circuit (words not recognized or word confusions), to remove syntactic ambiguities as well as semantic inconsistencies,…[2,3]

Questions or answers from the system are both generated by a voice synthesis circuit and displayed on the screen. It allows the physician to dictate the report while visually analyzing images, for instance. Besides, through this dialogue, the physician has access to edition controls such as sentence corrections, word deletions, pauses, resumptions,…

THE KNOWLEDGE SOURCES ARCHITECTURE

The system is based on a modular architecture composed of three main Knowledge Sources (KSs): Dialogue KS, Syntax KS, and Semantics KS. The Dialogue KS serves as a user interface and manages the interactions through a word recognition circuit and a text-to-speech synthesis circuit. It controls the activation of the Syntax and Semantics KSs.

Syntax and Semantics KSs process the natural language sentences of the report. The Syntax KS uses a lexicon and grammar rules to parse the words provided by the Dialogue KS and to build the sentences. The Semantics KS uses a rule base and an agenda to check the consistency of these sentences and their arrangement (their order) in the report.

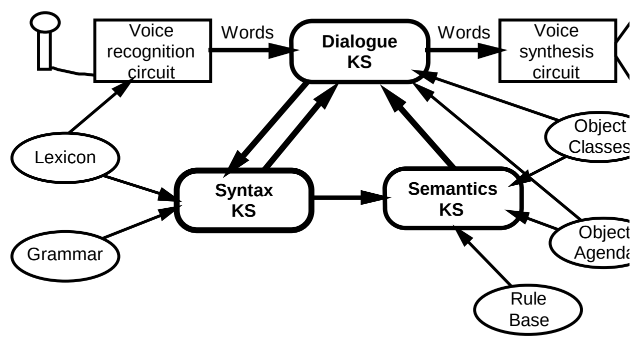

Fig. 1. The Knowledge Sources Architecture

KNOWLEDGE REPRESENTATION

The application has been limited to a specific domain: reports of Holter ambulatory monitoring. The domain knowledge of medical reports was acquired by interviewing a specialist physician. The concepts, the entities mentioned and described by the physician, are represented as object classes: for instance, the object class "heart rhythm". An object maps approximatively a short sentence. Object slots contain the characteristics of the instances of these classes.

Slots contain notably the description of the object instance through a list of values, for example: [irregular] or [irregular, slow], and the possible values of this list. Slots also concern the structure of the whole report: the slot "compulsory" signals if the object must be mentioned by the physician, the slot "next possible objects" allows to determine the dictation order (the "agenda") of the report.

13th Annual International Conference IEEE/Engineering in Medicine and Biology Society, Orlando, 1991

Different "rule" slots contain rule names that are inferred by the Semantics KS when a new instance is created, to check certain consistencies, to deduce other facts, or to modify the agenda. Slots contain messages that may be synthesized by the Dialogue KS. In case of error, an "error" slot then provides recovery points to the dialogue.

The Dialogue Knowledge source

The Dialogue KS receives all the words from the recognition circuit and runs concurrently. It processes the control words and passes on the other words to the Syntax KS. Control words correspond to dictation controls such as "begin", "quit", "pause",… answers, "yes", "no",… and to text edition controls such as "erase last word", "erase a sentence", "correct a sentence", "write report",… The Dialogue KS also processes errors detected in the Syntax KS and in the Semantics KS.

Text edition controls are interpreted and linked to operations on object instances. "Erase" corresponds to the removal of an item of the description slot (erase last word) or to the deletion of an object (erase a sentence). "Correct a sentence" corresponds to the modification of the description of the related object. The current agenda is modified to take into account the object to process.

If an error is detected by another KS, the Dialogue KS will synthesize a message corresponding to the type of error and to the current object. The Dialogue KS will also modify the agenda using heuristics and considering the recovery points.

The Dialogue KS keeps a record of the discourse history. This history is represented by the time-ordered list of the processed objects. At the end of the dictation, it is used to generate the written report.

THe Syntax Knowledge Source

The Syntax Knowledge Source uses a parser based on Definite Clause Grammars [4] and implements phrase structure rules. Five types of phrases are defined: noun, verb, adjective, adverb, and preposition.

A lexicon contains all the words of the application, coded as a base of facts. These facts describe the word categories such as, for instance: verb(exist). Morphological aspects were not taken into account but could be easily implemented.

The Syntax Knowledge Source receives a continuous sequence of new words from the Dialogue KS and parses them incrementally. It adds each word to the phrase in progress, parses it and repeats the process as long as the phrase is correct and not complete. When a sentence is complete, significant words are passed on to the semantics KS and a new sentence is started. If the phrase cannot be parsed, control is given back to the Dialogue KS.

The Semantics Knowledge Source

The Semantics Knowledge Source uses the object database and a rule base to insert the new object in the report, to check its consistency, together with the consistency of the report in progress.

The Semantics KS receives, from the Syntax KS, a list of words corresponding to the last complete sentence. This list includes notably the name of the object being described and its attributes. The Semantics KS verifies that this object is a member of the current agenda. Then it creates a new instance of the object class and fills the description slot with the attributes.

Object consistency rules are triggered considering the values of description slot. The rules check that these values are correct and do not conflict with those of previously described objects. In the same way, deduction rules are also triggered to automatically fill other object slots.

Finally, the new agenda is dynamically updated, using the "next possible objects" slot and considering the values of the description slot of the current object.

In case of failure, the type of inconsistency is notified and control is given back to the Dialogue KS.

Implementation and results

A prototype was implemented on a 386SX compatible computer. The voice recognition and text-to-speech synthesis is performed by an Introvoice IV circuit board [5]. The board is driven by a dedicated resident software which signifies that it can operate concurrently with the dialogue and language processing program. Recognition is operated on isolated words which are stored in the keyboard buffer. Each speaker achieves the training process by repeating each word of the vocabulary about 5 times.

The software was developed using Prolog language. The total number of words of the application is 180 and the average number of sentences is approximatively 10. All the processes are fast enough to give the user the impression of real time. The prototype was tested on Holter analyses for which it could successfully generate the corresponding written reports.

Conclusion

In this paper we have presented a prototype of a dictation machine, mixing voice processing capabilities with natural language understanding and dedicated to medical reports. Positive tests were made on a specific subject (Holter recording) where it proved to be feasible and promising.

Efforts are under progress to extend the vocabulary and the grammar as well as to improve the edition controls, the management of the dialogue history, and the flexibility of the report agenda. Other improvements will also probably come from new recognition hardware.

We hope that this system, as well as the techniques we described, will aid the physicians in an friendly and efficient way in their routine reporting tasks.

References

[1] VoiceMED, Kurweil AI, Waltham, 1990.

[2] J.M. Pierrel, Dialogue oral homme-machine, Hermès, 1987. (In French).

[3] L. Brunessaux et al., Satic, un Système de dialogue oral pour l'Assistance Technique à l'Intervention Chirurgicale, Proc. of Expert Systems Conference, EC2, Avignon, France, 1990. (In French).
[4] F.C. Pereira and D.H. Warren, Definite Clause Grammars for Natural Language Analysis, Artificial Intelligence, Vol. 13, 1980.
[5] Introvoice IV user's guide, The Voice Connection, Irvine, 1989.

Pierre NUGUES
Institut des Sciences de la Matière et du Rayonnement
6, boulevard du Maréchal Juin,
F-14050 CAEN CEDEX, FRANCE
Phone: 33-31-45-27-22
e-mail: nugues@ucaen.univ-caen.fr