THE SYNTHESIS-BY-RULE DEVELOPMENT SYSTEM WITH EXPERT CAPABILITIES

ARVO OTT

Dept. of Computer Control
Institute of Cybernetics
Tallinn,Estonia,USSR 2oo1o8

IMRE SIIL

Dept. of Software
Institute of Cybernetics
Tallinn,Estonia,USSR 2oo1o8

ABSTRACT

A flexible speech synthesis development system is described. It is a production system in which the two components - the declarative and the procedural knowledge base must be created by the domain expert. A simple rule language, interactive graphics, acoustical AB tests and explanation capabilities of the system are at the disposal of the expert. The production system runs on the I8080 microprocessor in real time.

INTRODUCTION

Futher progress in speech synthesis obviously depends on powerful and flexible development tools [1,2].
It is useful to reduce the role of the speech synthesis engineers in the process of obtaining various linguistic knowledge and give the linguists the possibility to explicitly model and immediately apply their professional knowledge. Besides, the synthesis system must be observed from two different aspects - the synthesis system, meant for real time, on-line implementation, and the system as a tool for investigations to examin the adequacy of phonetic or phonologic descriptions.
Selecting a suitable representation for the domain knowledge is one of the major problems to be encountered in building a knowledge-based system aimed at disposal of experts. Several high-level rule languages have been created for speech synthesis [1,3,4,5]. The general methodology on which these languages are based, is in principle the same. It is close to the technique of the production systems, known in the field of artificial intelligence. Indeed, in speech synthesis the terminology of production systems and expert systems has been used too [5].
The production rule system, representing the knowledge of speech synthesis control has been under development since 1982 in the Institute of Cybernetics of the Estonian Academy of Sciences. These studies were aimed at creating a flexible speech system, using terminal-analog speech synthesizer. Also, there was the task to minimize the calculation resources of microprocessor I8080, used to control the synthesizer.
We will focuse on representation of knowledge needed on different control levels of the synthesizer. All discussions have been made, taking into account the technical limitations of the realized synthesis system on the one hand, and at the same time to give the system maximum flexibility and to make minimal ad hoc solutions on the other hand.

1.VOCAL TRACT MODEL

The configuration of the vocal tract model of the formant synthesizer FS-05 was chosen and determined by a set of experiments with digital model, realized on a general purpose computer ES-1010. The resulting serial/parallel formant model is somewhat similar to the model, used in synthesizer OVE 3. There are 3 turnable (F1,F2,F3) and 2 fixed formant filters in vocal branch, a turnable resonator (FF) in fricative branch, fixed resonator in nasal branch, 5 switches for amplitude control (AV,AH,AF, AN,MO), fundamental frequency control PT and 3 transition times for pitch, formant frequencies and amplitudes (TP,TF,TA) [6]. Every control parameter is determined by one byte per 10 msek.

2.KNOWLEDGE BASE

Knowledge base is a specialized body of knowledge (facts, relationships and rules) embodied in computer memory.
Aquisition and maintenance of a domain-specific knowledge body is a critical problem for all knowledge-based systems [7]. Just putting an initial knowledge base together in a suitable representation for experts seems a formidable task. Moreover, the system must offer powerful and at the same time quite simple tools from the point of view of nonprogramming linguists for keeping the knowledge base accurate and current.
Our system works with linguistic knowledge encoded in the bases of declarations and production rules. The production rule has the form:

IF <condition> THEN <action>

The knowledge base is structured in the way which takes into consideration both logic of fonematic description and the need to process in real time the descriptions obtained. The knowledge representation and the structure of knowledge base to be filled should be sufficiently comprehensible to the domain expert. He will be familiar with the fundamental structure, organization and use of production rules, but may understand it only at the conceptual level and not in terms of performance program.
The knowledge base consists of three main parts (see Fig.1). The first part stores parametric descriptions of phonematic units and the rules determining positional variations and coarticulation of units. The second one is for knowledge about speech prosody forming. In the third part the explanations of correspondences between symbols to determine the system of spelling, allowed in the input of the synthesizer are maintained. The vocabulary of proclitics and enclitics for phonetic word forming is foreseen as well.

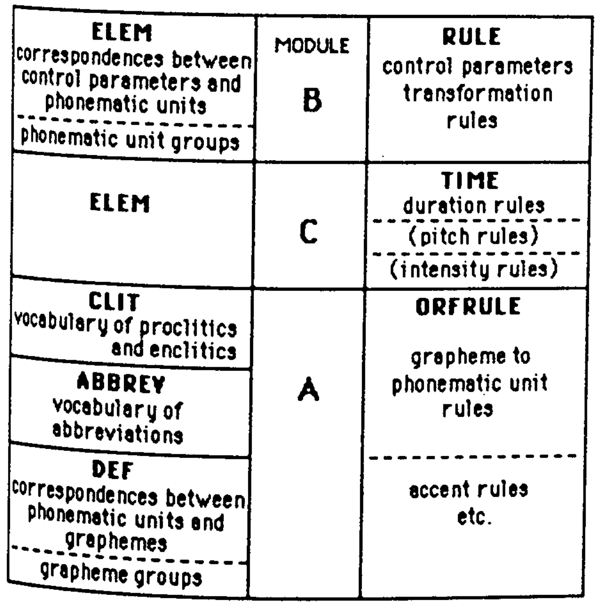

Fig.1 Knowledge base structure

The knowledge base contains initial pieces of knowledge - a set of graphemes, a set of internal and external representations of phonematic units and the control parameters which define the acoustics of these units. In other words the graphemes and initial sound representations must be determined to enable the unit stream in the control process.
The linguist gets the possibility to augment initial knowledge base according to his phonologic conception using some convenient formalism. The system has metalevel knowledge about every subbase in the knowledge base structure and it can help as an assistant of expert to fill out subbases with specific knowledge.
The special rule language was described by us in [8]. The translators of the rules forms from the rule text the compact intermediate rule tables for interpreters in kernel modules (see Fig.2).

3.KERNEL MODULES

The knowledge base is used by 3 functional modules of the system which perform transformations defined in production rules:

A - grapheme to phonematic unit (phonéme) transformations (process P1 on the Fig.3);
B - phonematic unit to terminal unit (allophone) transformations (process P3);
C - prosody transformations (process P2).

These modules constitute the kernel of the system. The modules are maximally independent - they are connected by unit strings which are observable by editing and explanation modules. The modules can be used separately, for example the module A was used as a phonetic transcriptor in tools of linguistic studies.

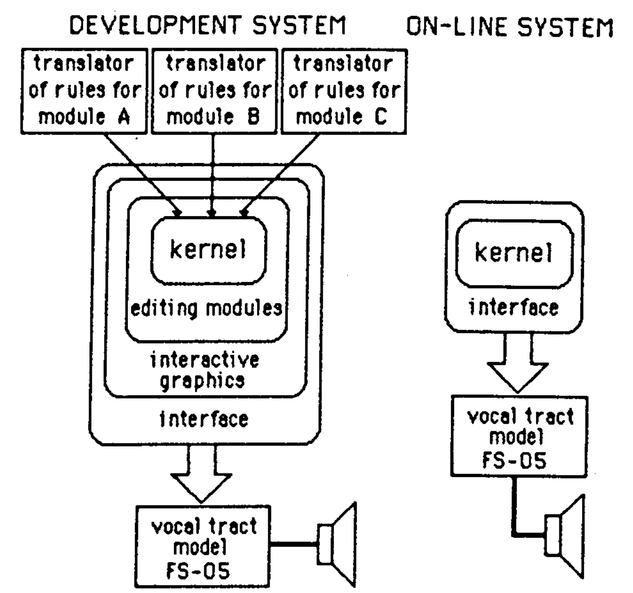

Fig.2 Structure of the devclopment and online systems

Every module uses corresponding knowledge subbases and has the interpreters of rules and interpreters of declarations. The subbases of declarations to module A are for:

-default correspondences between graphemes and phonematic units, grapheme grouping (base DEF);
-abbreviations (base ABBREV);
-proclitics and enclitics (base CLIT) to form phonetical word

Module B uses declarative base of default control parameters of phonematic units (ELEM).
The condition part of the record of the rule base ORFRULE for module A describes the situation in 16 bytes analysis window. This window acts as a shift register, where in addition to the grapheme codes are the indicators of the grapheme groups. The action part of production in ORFRULE makes structural changes in the string of phonematic units, derived, using the base of declarations DEF. The action part can change also the contents of analysis window - it can determine some additional flags in the indicator of grapheme grouping. For instance, 11 types of actions are needed to carry out all structural changes, specified in the rule system for Russian.
The production rule for module B (RULE base) determines the changes in control parameter domain, depending on the adjacent phonematic units. The formalism of rules in module B is simpler than that, used in parametric rules of the system SRS [4]. The analysis window of production rules for B is 3 reperesentations of phonematic units which are the pointers to the nodes of the unit (phoneme) tree. To use the tree structure describing the properties of the phonematic unit we can minimize the size and time of work of the B module rules.
In the third module C the production rules are used at present only to determine the time model of the speech - the pitch and loudness rule system is under development. Left hand side of the timing rule is similar to the condition part of the rules for module A.
To determine the segmental durations we use the formula:

$$D=D_1 f_1+\ldots+D_n f_n$$

where: D - segmental duration; $f_1..f_n$ - factor, fixed by the condition part of the timing rule; $D_1..D_n$ - value of factor, determined by the action part of the rule.

$D_i f_i$ is determined by one production rule and may be interpreted for example as a factor determining the speech tempo or the inherent duration of the segments etc.
The intention was to use the same speech synthesis kernel program for both - the speech synthesis development system and the applied system. The development system has in addition a set of editing and explanation modules to examine and change the units in different stages of synthesis (see Fig.2). The last modules trace the rules evoked in transformation processes and the intermediate results of every process stage.

4.UNITS REPRESENTATION

The speech synthesis control algorithm treats different representations of units in synthesis. It is quite clear that in the process of development the internal representation of these units is hardly observed - an expert who develops rules of speech synthesis must use various means to produce the external representation of units.
The expert must have a possibility to define his own abstractions - for instance how to mark the phonematic units of speech synthesis or in which terms to fix the groups of units.
Figure 3 describes the internal and external structure of the units in speech synthesizer FS-05.

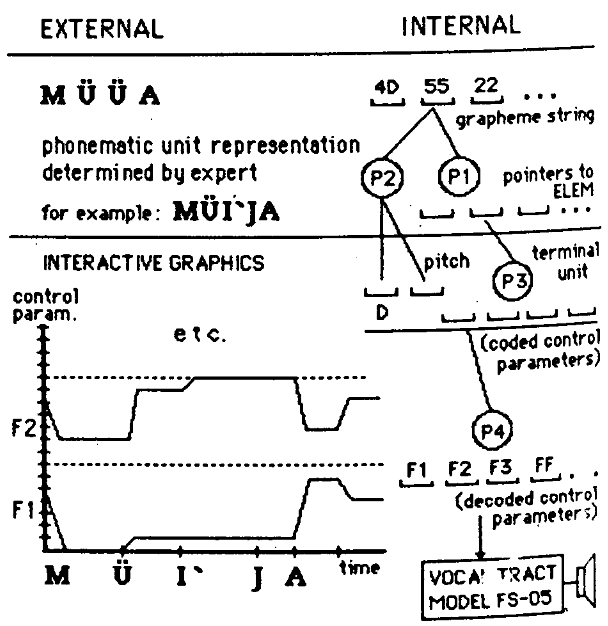

Fig.3 External and internal representation of the units

P1-P3 indicates the processes A,B,C in the kernel which was characterized above.
In synthesis process the phonematic units are the addresses to the ELEM base. As the kernel program of the synthesizer has to work in real time and must use minimal memory, the phonematic unit is internally represented by 1 byte and the control parameter base of these units (ELEM base) is situated in 256 byte ROM. For this purpose the control parameters, describing different sounds were coded and packed into one or several 4 byte control words and decoded only at the end of the control process. The maximal lenght of the ELEM base is 64 records of units (if every unit is determined by one control word). The graphical description of the parameter segment, determined by one control word is given in Fig.4.

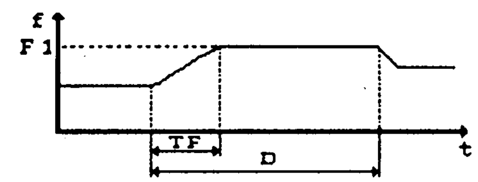

Fig.4 Graphical representation of the parameter F1 of the control word. D - duration of the segment, TF - onset transition time;

It is the task of an expert to decide - if he chose more complex units (units, described by more than one control word), phonematic units will be accordingly less than 64.
To simulate the needed variations of sounds on the acoustical level the control parameters must be changed. For testing and determining these changes in the development system the control parameters can be represented in a graphical form.
Using interactive computer graphics the domain expert can work with concrete phonetic descriptions. The graphical representation of the speech fragment is in essence the explicated input specification for the vocal tract model. It is possible to modify control parameters immediately in the course of perceptual experiments.

5.IMPLEMENTATION AND USE

Most of the production systems are realized using LISP or PROLOG language, which easily allows to describe the condition-action rules. Nevertheless, we support the viewpoint of [9] that the expert systems (and also production systems) will find more real use if they are programmed in some common language. Indeed, for example LISP needs a lot of programming resources and is usually slow. Especially for the task, described in this work, we find it important to program some parts of the production system in microprocessor ASSEMBLER.
The interpreters of productions (inference engine) were programmed in I8080 ASSEMBLER and are exactly the same for the development system, based on the microcomputer and the system for real applications (see Fig.2).
The kernel program was used to drive different vocal tract models: FS-05, formant model on the signal processor I2920 etc.
The development system runs on the personal computer LABTAM under CP/M-80 operating system and needs about 40 Kbyte of memory. Created rule language have been used by domain experts. For example the set of Russian text-to-phonematical unit rules were fixed by the phonetician of Moscow State University. This rule system with interpreter, declarative tables for 20 abbreviations and 41 clitics takes only 2 Kbytes of ROM and 512 bytes of RAM in I8080 microprocessor system.
Also the parametric rules were selected both for Russian and Estonian speech synthesis, using the aid of interactive computer graphics. The speech synthesis algorithm for Russian - the kernel program with its knowledge base and interpreters takes 6 Kbyte of ROM and 1 Kbyte of RAM of I8080 system.
The system has proved to be a powerful and flexible rule development tool which requires little resources of an ordinary microprocessor.

REFERENCES

[1] S.R.Hertz, J.Kadin, K.J.Karplus "The DELTA rule development system for speech synthesis from text" Proc. of the IEEE, 1985, vol.73, No.11, p.1589-1601
[2] J.Allen "A perspective on man-machine communication by speech" Proc. of the IEEE, 1985, vol.73, No.11, p.1541-1550
[3] R.Carlson, B.Granström "A text-to-speech system based on a phonetically oriented programming language" STL-QRSR 1/1975.
[4] S.R.Hertz "From text to speech with SRS" J.Acoust.Soc.Am., 1982, vol.72, No.4, p.1155-1170
[5] A.Aggoun et al. "Prosodic knowledge in the rule-based SYNTHEX expert system for speech synthesis" NATO ASI series, 1985, vol.F16, Springer Verlag, p.495-516
[6] О.Кюннап, А.Отт "Управляемый микропроцессором синтезатор речи" В кн.: Автоматическое распознавание слуховых образов - 12, Киев, 1982, стр.410-411
[7] R.Davis, D.B.Lenat "Knowledge-based systems in artificial intelligence", 1982, McGraw-Hill Int. Book Co.
[8] A.Ott, I.Siil "Real time speech synthesis - development and employment" Computers and artificial intelligence, Bratislava, 1987, Vol.6, No.2
[9] T.Mannel "What's holding back expert systems?" Electronics, 1986, No.28, p.59-65