THE SYNTHESIS-BY-RULE DEVELOPMENT SYSTEM

WITH EXPERT CAPABILITIES

ARVO OTT
Dept. of Computer Control
Institute of Cybernetics
Tallinn, Estonia, USSR 200108

INRI SIIL
Dept. of Software
Institute of Cybernetics
Tallinn, Estonia, USSR 200108

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 are clearly separated from the domain knowledge. A simple rule language, interactive graphics, acoustical AR tools and explanations capabilities of the system are at the disposal of the expert. The production system runs on the 18800 microprocessor in real time.

INTRODUCTION
Further 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 engineer in the process of obtaining various linguistic knowledge and give the linguist the possibility to exploit various resources of microprocessor computer systems - to control the synthesizer.

We know that the absence of representation of knowledge is 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 it an intelligent solution in 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 a digital model on a general-purpose computer ES-1010. The resulting model of the vocal tract in a synthesizer AV 3. There are 3 formants (F1,F2,F3) and 2 fixed formant filters in vocal tract. The formant filter (FF) in the fricative branch, fixed resonator in nasal branch, switches for vowel side control (Y,AA,AR,NO), fundamental frequency control P and 3 transition times for glottis frequency and amplitudes (TF,TF,TG) [8].

2. KNOWLEDGE BASE

Knowledge base is a specialized set of knowledge on facts, relationships and rules. It describes the domain in which the synthesis is done. We know the hierarchy of the production systems and expert systems has been used too [3].

At the rule-base level, 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 computer to control the synthesizer. We know that the knowledge is 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 it an intelligent solution in the other hand.

By 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 phonetic description and the parts to process in real time. The declarative information obtained by the knowledge representation and the structure of knowledge base to be filled should be sufficiently compatible 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 - the set of parameters: - the set of parameters of the phonetic units and the control parameters, the rules determining position variations and coarticulation of units.

The second module is for knowledge about speech prosody formulating. 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 formulating is foreseen as well.

The knowledge base is used by 3 functional modules of the system which perform transformations defined in production rules: A - graphemes to phonetic units (phonemes) transformations (process P1 in the Fig.3); B - phonetic unit to terminal unit (aliphonme) transformations (process P3); C - prosody transformations (process P2).

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

terminated to enable the unit stream in the control process.

The linguist gets the possibility to augment initial knowledge base according to his phonological conception using some convenient formalism. The system works with specific knowledge about every subbase in the knowledge base structure and it can be used as an assistant of expert to fill out subbases with specific knowledge.

The special rule language was described by us in [8]. The translation of the rules into the ON-LINE SYSTEM

DEVELOPMENT SYSTEM

ON-LINE SYSTEM

Fig.1 Knowledge base structure

Fig.2 Structure of the development and on-line systems

Every module uses corresponding knowledge subbases and has the interpreters of rules and interpreters of declarations. The sub-
bases of declarations to module A are for:
- default correspondences between graphemes and phonemic units, grapheme grouping
- abbreviations (base ABBREV)
- symbols, conditionals and comments (base CLT) to form phonemic word

Module B uses declarative base of default control parameters of phonemic units (DEL). The condition part of the record of the rule base ORFULE for module A describes the situation in 16 byte analysis window. The 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 ORFULE makes structural changes in the string of phonemic units, derived, using the base of declarations DEF. The action part can contain only one rule, the one that produces an analysis window - it can determine some additional flags to describe the units in grapheme grouping. For instance, 14 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 phonemic units. The form of rule is in module B is displayed in Fig. 2, used in parametric rules of the system SRS. The following parametric production rules for B & D representations of phonemic units which are the pointers to the nodes of the unit (phoneme). To use the tree structure describing the properties of the phonemic unit we may minimize the size and time work of the B module rules.

In the third module C the production rules are used to present only to determine the time model of the speech - the pitch and loudness rule system is in development. The left hand side of the timed rule is similar to the condition part of the rules for module A.

To determine the sequential durations we use the formula:

\[ D_f = f_1 \cdot D_1 + \ldots + f_n \cdot D_n \]

where: \( D_f \) - segmental duration; \( f_1, \ldots, f_n \) - factors, fixed by the condition part of the timing rule; \( D_1, \ldots, D_n \) - value of time, determined by the action part of the rule.

\( D_f \) is determined by one production rule and \( f_1, \ldots, f_n \) for each factor determining the speech tempo or the interval 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 and syntactic processes (see Fig.2). The last modules trace the rule 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 is the process of development the internal representation of these units in 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 phonemic units of speech synthesis. We distinguish three groups of units:

- Fig. 3 External and internal representation of the units

The graphical representation of the parameter PI of the control word D - duration of the segment. TP - onset transition time.

It is the task of an expert to decide - if he is the indicator of grapheme grouping - how many control words, phonemic 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 specified in a rule base ORFULE for the external representation of the units. Using interactive computer graphics the domain expert can work with concrete acoustic descriptions. The graphical representation of the speech fragment is in essence the explicit input specification for the vocal tract model. It is possible to modify control parameters immediately and we can in the course of perceptual experiments.

5. IMPLEMENTATION AND USE

Most of the production systems are realized using LISP or PREGO language, which easily allows to describe the conditions and action rules. Nevertheless, we support the viewpoint that in that the expert systems (and also production systems) will find more real if they are programmed in some common language. Indeed, for example, the user needs a lot of look around resources and is usually slow. Especially for the tasks described in this work, we find it important to program some parts of the production system in microprocessor ASSEMBLER.

The interpreters of productions (inference engines) were programmed in microprocessor ASSEMBLER and are exactly the same for the development system, based on the microcomputer and the system for real applications (see Fig.1.2). The kernel program was used to drive different vocal tract models: FS-05, FOM model on the signal processor 1280 etc.

The development system runs on the personal computer LAYART under MS-DOS operating system and is linked with 40 megabytes of memory. Created rule language has been used in several experts. For example the set of Russian text-to-phonemic units rules were fixed by the phonetician of Moscow State University. This rule system with interpreter, declarative tables, for 20 abbreviations and 61 clitics takes only 6 Kbytes of ROM and 512 bytes of RAM in 16860 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 on 16860 system.

The system has proved to be a powerful and flexible rule development tool which reduces the load of resources of any microcomputer.

REFERENCES