TEXT-TO-SPEECH RUSSIAN SYNTHESIS BY RULE

ALEŠ BUČEK

Tesla Electronics Research Inst.
Prague, Czechoslovakia

JEVGENIJ TIMOFEJEV

Faculty of Pedagogics
Hradec Králové, Czechoslovakia

ABSTRACT

The paper concerns the present-day state of research of automatic conversion of Russian written text into a corresponding acoustic signal.

INTRODUCTION

Our information is limited to the results of research work carried out in the framework of integrated efforts of Tesla Electronics Research Institute in Prague and Faculty of Pedagogics in Hradec Králové. As a result of the result of the research work the first version of the program, which assigns a sequence of short sounds of appropriate amplitude and spectral composition to any Russian text written in a usual form, has been developed. The sounds transmitted by microcomputer rapidly one after another are percepted by users as spoken Russian.

SPEECH SYNTHESIS

Our solution is based on approximation of speech signal on the basis of the basis of two-formant sounds, which are tabulated for one-period length in the computer memory and transmitted into loud-speaker respecting the sound combination of input text. We have used an arsenal of 12 vocal-like sounds and of 1 noise-like sound. By changing the rate of transmision of various digital patterns, by various number of patterns in one period, by various number of periods and various loudness we have obtained much more extensive set of various sounds. By means of these 12+1 sounds and by way of their transmission individual elements of spoken speech are described in the computer memory. We have defined these elements as follows:

↗h - sound initiation
h - sound body
h↘ - sound ending
cv - consonant-vowel combination /each with other/
vc - vowel-consonant combination /each with other/.

Russian word CKOPO /sko:ra - phonetic transcription/ is decomposed in comformity with this definition into the following speech elements:

↗s,s,s↘,↗k,k,ko,o:,or,r,ra,a,a↘

The above decomposition is not entered by the user - the computer carries out the operation without any intervention from the user's part.

The tables of parametric description of individual sounds, which approximate the sounds of speech, have been composed on the basis of prof. M. Romportl's and prof. L. Bondarko's works. We have also used spectrogams of natural speech. Perception tests were the decisive argument of spectrograms interpretation.

In the first version of our synthesis the high degree identification of Russian word accent /when basic prosodic parameters are absent/ is provided by means of:

- greater quantity of stressed vowels vs. unstressed vowels, final stressed a-vowels are double elongated
- stressed vowels are 6 dB louder than unstressed vowels
- stressed vowels are 1/2-tone higher than unstressed vowels
- quality alternation of unstressed o/a-vowels /a-norm pronunciation/, i-norm pronunciation can be also introduced, but the perception of word accent is not improved.

TEXT-TO-SPEECH ALGORITHM

The described method of synthesis of segmental features of speech and approximation of stressed/unstressed vowels phonetic contrast have enabled to produce a synthetic signal of spoken Russian, which has no prosodic feature and sounds somewhat monotonously, but is characterized by high degree adaptability of the users of Russian to this signal with good understanding.

Besides the input text need not be entered in a form of phonetic transcription. The automatic conversion of a usually written Russian text into input phonetic transcription is also provided.

The first version of algorithm of written text-to-phonetic transcription includes four basic stages:

1. Receipt of input text

In the input buffer the system selects only alphabetic letters of written Russian /capital letters of Russian alphabet/, character ":" for word accent, character "," for pause, character ␣ /spacing/ and CR , which ends the receipt of a text.

2. Text transmission into working memory

The text is transmitted character by character from left to right till CR . During transmission some characters or some combination of characters are processed:

- realization of some consonant combination is modified / ПРА:ЗДНИК - prazn'ik, ЧЕ:СТНЫЙ - česnyj, КАТА:ЕТСЯ - katajetca, ЛЁГКИЙ - l'oxk'ij etc./
- realization of adjective inflextion in genetive case is altered / ДОРОГО:ГО - dorogavo/
- a-norm pronunciation is introduced / ХОРОШО: - xarašo/
- realization of pronoun ЧТО and conjunction ЧТОБЫ is altered /što, štoby/
- i-norm pronunciation is introduced in a limited size /in the first unstressed syllable before stressed syllable/
- consonant combination СЧ is substituted by realization of šč /in a limited size/.

3. Text processing in working memory from left to right:

- the letter Ъ before vowels is substituted by its spoken equivalent
- doubled consonant is substituted by single one
- the orthographical Ь is conversed into its phonetic realization
- pronunciation of a preposition with unstressed vowels is realized / ОБО, ПЕРЕДО etc./
- conversion of multiciphered letters /Е, Ё, Ю, Я / is ended
- realization of final stressed a-vowels is modified
- the so-called coarticulation in vowel combination / НАУ:КА, СООБЩЕ:НИЕ etc./ is respected.

4. Text processing from right to left:

In this stage the text is processed according to deaf-sonorous assimilation laws of Russian. The text processing is finished as soon as the beginning of the text is reached.

In the present stage of development, our algorithm of automatic transcription contains more than 30 rules and occupies 1,5 KB of ROM-memory. It is universal and every Russian text can be processed. Algorithm development has been based on two methodical principles: approximation and ignoration. For example, the algorithm approximates the pronunciation of all unstressed a/o-vowels as a single realization of a weak /a/ in opposition to a strong stressed /a/. The pronunciation of some strange-origin Russian words with weak unstressed o-vowels is ignored. Nevertheless the user has an opportunity to produce realization with unstressed o-vowels: in this case accent need not be input /the qualitative alternations of unstressed vowels are conditioned by accent input/. The basic criterion for algorithmic rules extension is communicative effect of an acoustic signal and its aesthetic realization. For example, from communicative point of view it is not necessary to modify the consonant combination ЧТ, ЧН into št, šn. The user of Russian will understand text with čt,čn-realization in the same way as with št,šn-realization. But from aesthetic point of view the above modification of text should be desirable. That is why the čt,čn - št,šn conversion has a limited effect and is valid for words ЧТО and ЧТОБЫ only. The rest of words are ignored / Решение конечно. - Конечно, он прав./.

The first version of our text-to-speech algorithm contains the greatest part of Russian written text/phonetic realization differences and is being constantly improved. The practical ideal version of the algorithm is connected with further progress in miniaturization of hardware as well.