THE PHONETIC BASIS OF ARTIFICIAL RUSSIAN SPEECH, ITS GENERATION BY COMPUTER AND ITS APPLICATION

Kálmán Bolla and Gábor Kiss

Department of Phonetics
Linguistics Institute of the Hungarian Academy of Sciences

Abstract

The authors describe their research experiences and results in their study (analysis and synthesis) of the phonetic structure of Russian speech in recent years. Based on their research findings, they developed a Russian language text-to-speech system, called RUSSON. The paper discusses some key phonetic questions related to RUSSON (letter-sound-phoneme-microelement, word stress, segmental and suprasegmental structure, palatalization-pharingalization) and describes the computer program of RUSSON.

Introduction

Production of artificial speech does not amount to a special scientific achievement. Lately, attention is focussed rather on the application of synthetic speech and on automatic speech recognition. In Hungary, the first sound and speech synthetizer systems were developed in the late seventies, early eighties as a result of research conducted at the Department of Phonetics of the Linguistics Institute of the Hungarian Academy of Sciences. Their primary aim was to aid scientific study of the sound structure of speech.
The present paper is an account of our research experiences and results accumulated in the past few years in the phonetic analysis and synthesis of Russian speech. Preliminary work and earlier results were reported in our book titled "A Conspectus of Russian Speech Sounds" published in 1981, as well as papers in the series "Hungarian Papers of Phonetics" No. 1-16. (1978--1986).
The instruments used for the analysis and synthesis of Russian speech were those available at the Departments of Phonetics of the Linguistics Institute of the Hungarian Academy of Sciences. The most important ones are as follows: a dynamic sound spectrograph, a pitch meter, a intensity meter, a four channel mingograph, a twelve channel oscillograph. The speech synthesis was done on a PDP11/34 computer and a OVE III/c formant speech synthesizer.

The authors first showed the RUSSON system to the public at an exhibition held in Moscow in 1985 to commemorate the 40th anniversary of Hungary's liberation.

RUSSON as a phonetic research aid

RUSSON was meant as a computer model of Russian phonetic processses. It provided a means to verify our analysis and to use the analysis-by-synthesis method. The synthesizing method enables us to alter any of the individual acoustic features of speech at will, to extract and analyse its physical and phonetic elements and structures, to filter out those constituents and features which have no linguistic function; to establish the language specific rules of sound linkage, the concomitance relations and compensatory ways obtaining between various constituents of sounds, the combination and variability of elements; to analyse the structural relevance of sound elements and the sound structures made up of these.
On some phonetic issues relating to RUSSON we can only touch upon some phonetic questions which relate directly to either the development of the application of RUSSON. (A more detailed version of the present paper will appear in No. 17 of Hungarian Papers in Phonetics.)

1. Writing, phonological system, sounding speech, acoustic structure, speech perception

The Russian writing system is a syllabic and morphophonemic system using the Cyrillic alphabet. One variant of our synthetic speech system produces sounding speech taking orthographic text in Cyrillic letters (including punctuation signs). This is the well-known text-to-speech system.

2. Segmental-suprasegmental sound structure

The two structures are relatively independent of each other, which means either can be extracted from the complex acoustic signal alone, or either can be produced separately.

3. Russian wordstress and temporal structure

Word stress in Russian is quantitative stress with special features of intensity and melody. The position of word stress is free varying in cases even depending on accidence.
The synthesized samples clearly suggested that lengthening always indicate stress, although in certain positions the duration of the stressed vowel (particularly in two syllable words) may be equal to, or even less than that of unstressed vowels. The reason for this is that stress is tied to the word form and is present in actual use even if unrealized by phonetic means.

4. The consonantal nature of the sound system palatalization and pharingalization

The Russian sound system is consonantal. In harmony with the consonantal character the articulatory and perceptual basis of Russian consonants is dominated by the consonants. The sound structure of Russian speech is basically determined by two factors: its duration is determined by its stress, its vocalic structure by the palatal-pharingal articulation.

5. Intonational structures, prosodemes

The text-to-speech system RUSSON uses the following matrix to produce the actual intonation forms. If our intonation experiments so require, the values of the matrix can be adjusted.

Operation of the Russian language text-to-speech computer system RUSSON

The program produces sentences of any content entered in correct Russian orthography in the following three main steps.
a) First, using a set of rules the program maps the letter sequence into a series of so-called microelements, which will ultimately form the segmental basis of artificial speech.
b) Next, on the basis of the sentence final punctuation mark the suprasegmental

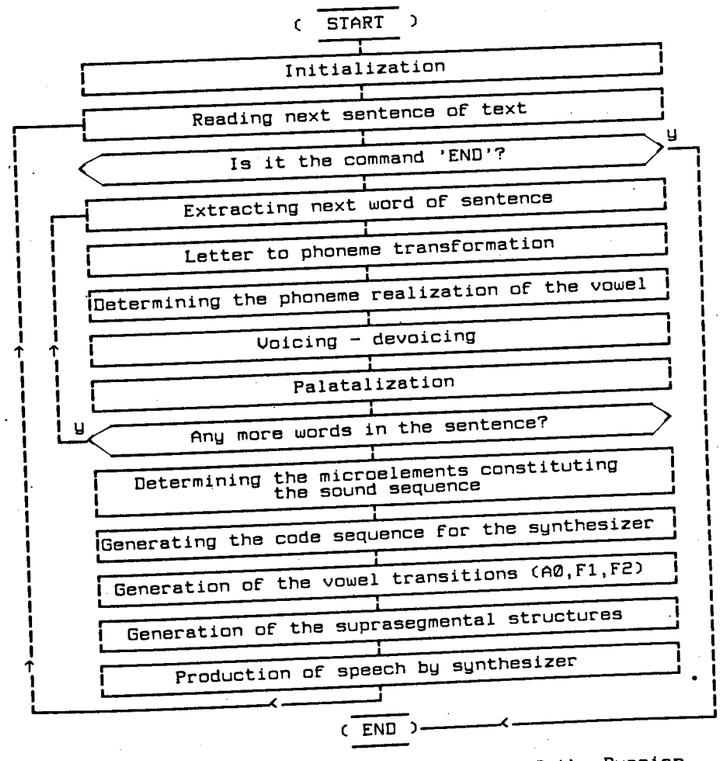

Fig. 1. The main steps of the operation of the Russian language text-to-speech system RUSSON

structure is generated and then integrated with the segmental structure.
c) Finally, the code sequence resulting from the above two steps, which now maps the complex acoustic phenomena, is passed to the synthesizer, which will produce the sentence.
The operation of the program in more detailed steps is illustrated in the flowchart in Fig. 1.

The stock of micro elements

The control program produces the given sentence with the help of a system of rules and the inventory of microelements. The system of rules is implemented in tables and look up procedures. The stock of microelements contains the speech sounds and the pauses. Each sound is built up of 4 microelements. The RUSSON program produces the sound structure out of a possible set of 37 consonant and 35 vowel phoneme realizations. The pauses between words and sentences are generated out of 5 microelements of different length. Thus, the inventory of microelements must contain 34 * 4 + 35 * 4 = 292 elements.

The letter-to-phoneme transformation

The Russian text may consist of 31 letters as well as a soft and a hard mark. Going through the string of letters in the sentence the program selects out of the 21 consonants and 5 vowels those which correspond to the letters, simultenously carrying out any softening where requied. At this stage the program also registers word stress as well as possible sentence stress by storing the ordinal number of the stressed vowel.

Selection of vowel phoneme realizations

The program segment designed to establish the correct vowel phoneme realizations takes as input data the word to be processed and the vowel phonemes making up the word as yielded by the letter-to-phoneme transformation. They can be of the following five types: A, O, U, I, E. Taking these five vowels and their phonetic positions inside the given word the program selects one of the 35 possible vowel realizations.In defining the phonetic positions the program considers stress, pre- stress, word initial and word final positions as well as the quality of the preceding and following sound (whether it is soft or hard).

Selection of the consonant phoneme realization

The consonant phoneme number yielded by the letter-to-phoneme transformation is identical to the phoneme realization number. However, in the course of later processing the sequence of consonants may undergo change as a result of the program segments which check for voicing or palatalization.

Voicing and devoicing

The program extracts the two-member sound clusters from the words of the sentence one by one. If the cluster is made up of two consonants, both members will be checked to see if either of them belong to the exceptions. If the first member is listed as one undergoing no modification or the second member belongs to the set of consonants that do not change the preceding consonant, then the program passes on to the next cluster. When a modification is called for, it is carried out with the help of a table. Word-final consonant-consonant clusters require special treatment. First, the word-final sonorant is devoiced (if necessary) and then the preceding consonant is processed.

Тё'тя пьё'т ру"сский ча'й.

Тё'тя пьё'т ру"сский ча'й?

Сашы' цвету'т весно"й.

Сашы' цвету'т весно"й?

Ната'ша пое'хала нада"чу.

Ната'ша пое'хала нада"чу?

Execution of palatalization

Here again, the program first extracts two element sound clusters.If they are both consonants then the combinations not undergoing palatalization are filtered out. Where required, palatalization is executed by changing the number of the initial member of the cluster.

Defining the microelements

The suprasegmental structure corresponding to the sounds defined earlier is based on microelements.Four microelements are assigned to every phoneme realization. However, the program does not make use of all the four microelements in every instance. There are cases when only the second, third and fourth element is used. The function of the first microelement is to ensure a smooth, even onset of a sonorant sound.

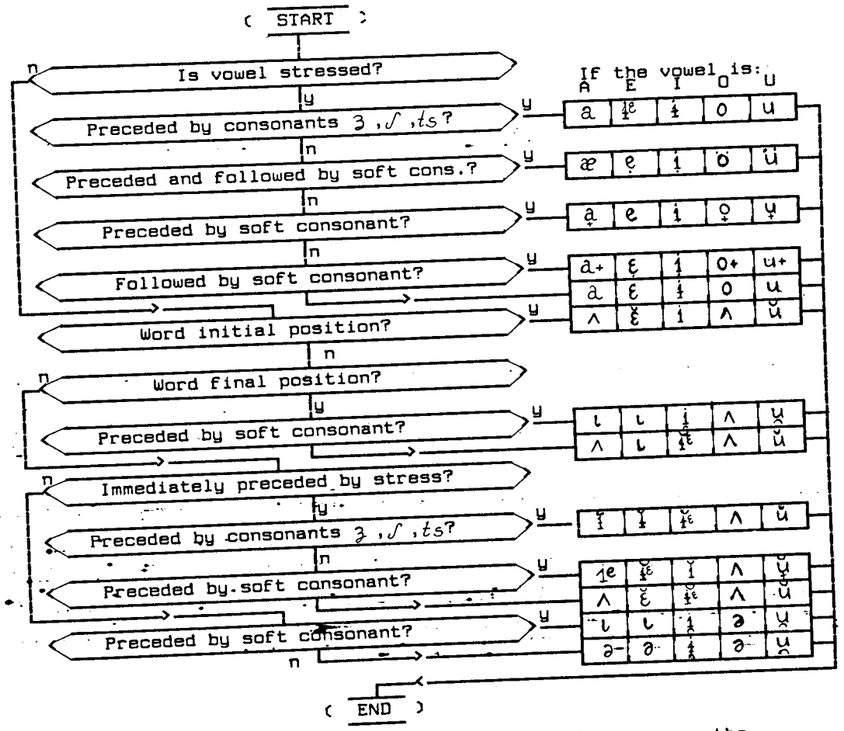

Fig. 2. Determination of vowel phoneme realization on the basis of their phonetic position

Defining the transitions between vowel realizations

The vowel transitions are composed whenever a vowel occurs next to a consonant. In order to enhance faithful reproduction the vowel realizations have to be adjusted to the actual phonetic environment. This adjustment affects the first and the last microelement of the vowel realization. The modification concerns the adjustment of intensity (A0) and the first two formants (F1, F2) in such a way that they should conform to the corresponding values of the preceding or following consonant.

Generation of the suprasegmental structure

The suprasegmental structure is generated when the segmental structure of the utterance has been defined. The construction of the suprasegmental structure is aided by the sentence stress typed in the text as well as the sentence final punctuation mark. The temporal structure of the utterance is modified so that the duration of the vowel bearing sentence stress is doubled. The sentence final punctuation mark defines one of the eight possible intonation contours to be used. The RUSSON program recognizes the following sentenre final punctuation marks: . (full stop), :(colon), , (comma), ; (semi colon), ! (exclamation mark), ? (question mark), ?! (question mark - exclamation mark), ?? (double question mark).
With this operation completed, the complex sound structure is ready to be produced.

Control of the speech synthesizer

The sequence of code thus generated is passed on to the speech synthesizer to control its operation when it sets sound to the text.