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

Kálmán Bolla and Gábor Kiss

Linguistics Institute, Hungarian Academy of Sciences

INTRODUCTORY REMARKS

Production of artificial speech does not amount to a special scientific achievement. Microelectronics and computer technology has developed the technical requirements (ie large memory and storage capacity, fast processing speed, small speech synthesizer hardware). With the use of cineradiography and dinamic sound spectrographs linguistic phonetics came to acquire decades earlier the knowledge about the phonetic structure of the sound segments that synthetic speech production required. Now attention is focussed rather on the application of synthetic speech.

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 in Phonetics" No. 1--16. (1978--1986).

THE INSTRUMENTS USED FOR THE PHONETIC ANALYSIS AND SYNTHESIS

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 PDP 11/34 computer and a OVE III/c formant speech synthesizer. The operative memory of the computer is 32 kwords. The system configuration includes two floppy disk drives, a line printer type LA-36 or a VT-55 video display unit. The computer is linked via a 16 bit parallel interface to the Swedish-made OVE III speech synthesizer. This is a formant synthesizer, which can be controlled through 15 acoustic parameters (A0, AC, AH, AN, F0, F1, F2, F3, N1, AK, K1, K2, B1, B2, B3) using 12 bits. 4 bits serve to choose a particular parameter and the remaining 8 bits define the value for the parameter selected. The PDP computer runs under the operating system RT-11 V 2.0. The RUSSON program was written in the Fortran IV language. The program consists of 1 main segment, 24 subroutines and 4 BLOCK DATA SEGMENTS amounting to 15000 lines altogether. Due to the limited memory capacity of the computer the program relies on overlaying.

With a view to industrial applications, the Russian text-to-speech system has also been implemented by the authors on a SYSTER computer and a VOX-08 speech generator commissioned by the Hungarian Budapest Electroacoustic Factory (BHG). The personal computer SYSTER is operating under the CP/M system. Both the computer and the speech synthesizer are made in Hungary. The RUSSON system implemented on the SYSTER computer was first shown to the public at an exhibition held in Moscow in 1985 to commemorate the 40th anniversary of Hungary's liberation.

In recent times we have been making efforts to implement RUSSON on a Commodore 64 personal computer and a MEA 8000 speech synthesizer chip with a view to educational applications.

RUSSON AS A PHONETIC RESEARCH AID

RUSSON was meant as a computer model of Russian phonetic processses. Synthetic Russian speech not only can verify our analysis but also provides a means 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. Synthetic speech can also be used in the study of speech perception and comprehension.

ON CERTAIN PHONETIC PROBLEMS RELATING TO RUSSON

We can only touch upon some phonetic questions which relate directly to either the development of the application of RUSSON.

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

Sounding speech can be produced from various basis: a) written text, b) phonemic symbols and c) phonetic symbols. The relationship between speech and writing is rather intricate and varies from language to language despite the fact that both are realisations of the same linguistic system and that the written form is based on the spoken speech. In other words, there is no simple and direct mapping between sound elements and graphemes. A special algorhithm is needed to map speech to its written form just as converting written text to speech requires its own algorithm.

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.

In order to model phonetic processes and phonological systems RUSSON can also be made to accept phonemes or speech sound, which means that speech is produced by through phonological or phonetic transformations. The phonemic variant is based on the phonological theory of the Moscow

school, defining the vowels on the basis of five vowel phonemes. On the other hand, in the phonetic variants input consists of 35 kinds of vowels and 52 different consonants (representing the Russian speech sounds).

Finally, the ultimate constituent in the inventory of elements in the RUSSON system is the microelement. They number 288. A microelement is a homogeneous slice of the speech stream which is extracted from it on the basis of the dynamic changes of the acoustic constitutents. The number of acoustic parameters palying a role in the internal structure of microelements ranges between 1 and 23. The homogeneous nature of microelements derives from their constancy or a unidirectional change. Depending on the acoustic quality of the microelement four types can be distinguished: pauses, voiced element, noise element, and elements of a mixed structure. Pause elements allow for quantitative variations only, while the other three elements can yield countless segments of different quality and content.

RUSSON can be used to study the entire vertical range of the sound structure, from the acoustic microstructures to more abstract phonetic, phonological and graphemic relationships, from the encoding of the sound structure to decoding realized in auditive testing.

2. Segmental and suprasegmental sound structure

Our phonetic study and synthesis of Russian speech have confirmed our hypothesis that the sound body reveals two linguistically relevant structures: the segmental and the suprasegmental structure. The first is constituted by the

serial combination of discrete sound elements. Instead of conceiving them as a loose string of beads they should be viewed as a structure consisting of elements which modify each other to varying degrees and extent at the points they connect to each other. The language specific aspects of segmental structure include: the stock of phonemes and speech sounds (their number and quality), phonotactic rules of phoneme and sound combinations, the nature of word stress, the positional boundness of phonemes and sounds etc. The suprasegmental sphere includes the sound structures and tonal quality produced by changes in the temporal, melodic and intensity phenomena.

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. The following experiment was carried out to demonstrate this point.

a) with the help of the instruments listed above speech recorded on tape was produced in the following ways

-- changes in F0 over time, ie the intonation contour was played at constant intensity;

-- intensity changes in time were reproduced at constant pitch, and finally

-- both changes in pitch and intensity were produced in terms of time.

b) The suprasegmental characteristics of the played sentences could be reproduced by humming, which also proves that they have a measure of independence on the

level of perception, in the process of phonetic decoding.

c) We have conducted several synthesizing experiments to separate the segmental and the suprasegmental structures.

-- From measured data various intonation contours were produced and tested.

-- Only the segmental structure was used to generate sound sequences (ie the time, intensity and pitch values used were those of the so-called sound specific values).

-- The complete sound sequence was produced but in setting it to voice the suprasegmental structure of the utterance alone was also produced.

d) The following experiment was designed to observe the phonetic constituents of the suprasegmental structure and their linguistic function. First, the segmental structure of the sentence was produced and listened to, then by varying the temporal data, the rhythmic structure and later, the tempo was formed. Finally, by varying the pitch the sound sequence was fitted with an appropriate intonation contour. The acoustic effect of each alteration could be perceived and and evaluated straight away. This experiment led us to the conclusion that the prosodic stock of the language is constituted by the sound patterns and structure types derived from the totality of suprasegmental features (tempo, rhythm, intonation, intensity, pause, tone).

3. Word stress and temporal structure in Russian

It is well known that 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 sound body of words is basically determined by stress, which also defines the temporal structure and rhythm (the combination of long, short and reduced duration of vowels) and has a major role in conditioning the positional variants of vowels.

By means of synthesis we investigated the phonetic characteristics of word stress. The aim of our experiments was to find out which of the three factors, ie duration, intensity and pitch play the dominant role in the realization of word stress in Russian. We synthesized words of two and three syllable varying duration and intensity, and produced the acoustic features of stress first in the stressed then in the unstressed syllable. The synthesized samples were evaluated in auditive tests. The results suggested that lengthening unfailingly 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 or if its acoustic realization is not very prominent.

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

It is well established that the Russian sound system is consonantal. This question cannot be discussed in detail here. In harmony with the consonantal character the articulatory and perceptual basis of Russian consonants is dominated by the consonants. In the articulatory processes of the vocal tract the articulatory movements determining the softness and hardness of the consonants greatly affect the production of vowels as well. The generation of this consonantal feature requires the frequent movement of the tongue body up and front as well as down and back and such great shifts in tongue positions produce large transitions in the production of vowels. This fact lends Russian vowels their diphtongal and triphtongal character. The heterogeneous nature of Russian vowels derive, then, from the effect of the consonantal environment. 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. This is why 35 vowels were adopted in the database for the synthesis with each having a transition from the F1, F2 and intensity matching matrix.

5. Intonational structures, prosodemes

Intonation is conceived here in a wider sense. The term 'suprasegmental sound structure' is considered a more unequivocal term. Suprasegmental sound structures mean

the organization of the sound body superimposed on the sound sequence, and plays a role in the creation of higher level linguistic structures (such as phonetic syntagmas and larger phonetic units). 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 in Fig. 0 can be adjusted.

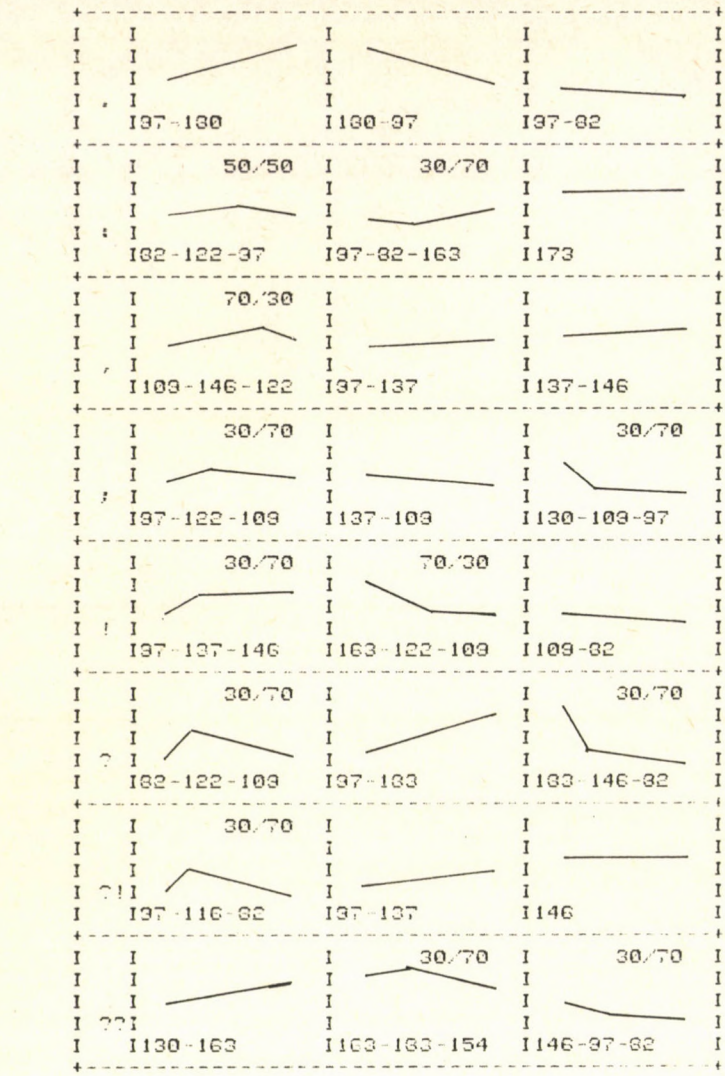


Fig. 0. The fundamental frequency (FØ) contours corresponding to the sentence final punctuation marks of the Russian text-to-speech system RUSSON

RUSSON, a Russian language text-to-speech computer system

The text-to-speech system was developed on a PDP 11/34 computer at the Phonetics Laboratory of the Linguistics Institute of the Hungarian Academy of Sciences. The operative memory of the computer is 32 kwords. The system configuration includes two floppy disk drives, a line printer type LA-36 or a VT-55 video display unit. The computer is linked via a 16 bit parallel interface to the Swedish-made OVE III/c speech synthesizer. This is a formant synthesizer, which can be controlled through 15 acoustic parameters (A0, AC, AH, AN, F0, F1, F2, F3, N1, AK, K1, K2, B1, B2, B3) using 12 bits. 4 bits serve to choose a particular parameter and the remaining 8 bits define the value for the parameter selected.

The PDP computer runs under the operating system RT-11 V 2.0. The RUSSON program was written in the Fortran IV language. The program consists of 1 main segment, 24 subroutines and 4 BLOCK DATA SEGMENTS amounting to 15000 lines altogether. Due to the limited memory capacity of the computer the program relies on overlaying.

With a view to industrial applications, the Russian text-to-speech system has also been implemented on a SYSTER computer and a VOX-08 speech generator commissioned by the Hungarian Budapest Electroacoustic Factory (BHG). This is a Hungarian made personal computer operating under the CP/M system.

Introduction of RUSSON running on the SYSTER computer

The RUSSON system implemented on the SYSTER computer was first shown to the public at an exhibition held in Moscow in

1985 to commemorate the 40th anniversary of Hungary's liberation.

Description of the operation of the program

The text-to-speech system can be started on the PDP computer with the command RUN DX1:RUSSON. The system displays a ~ (tilde) to indicate its readiness to generate sound for some Russian text or to execute some user option. When this character appears on the terminal, the user can start typing in either the text to be generated by the system or the control character of some user option (., %,+,$).

Defining the text to be generated.

The text to be generated must be entered in Russian orthographical notation sentence by sentence. However, word stress must be indicated by typing a ' after the stressed vowel. In addition every sentence may include a so-called sentence stress. This must be indicated with two '' characters placed after the stressed vowel. Sentences must be terminated with punctuation marks. In order to facilitate the generation of the correct suprasegmental structure the following double punctuation marks are also accepted: ?? ?! . Entering the sentence is terminated by pressing the RETURN key.

An error message is displayed if the user has mistyped a character (i.e. it is not a letter in the Russian alphabet or any of the above control characters). An error message is generated also if the sentence is too long for the computer's memory capacity. At present the system can generate speech

lasting approximately 5 s at one go.

Description of the user options

1. Control of speech tempo (%)

This parameter is set as a default to the normal tempo of everyday colloquial Russian speech, which corresponds to the value of 100. However, the user is free to produce slower or faster rates as well. If the value specified after the '%' mark is over 100, speech rate becomes slower and vice versa. Speech can be increased up to three times while there is practically no limit to the extent it can be slowed down.

2. Replay (.)

This option allows the user to listen to the sentence again without having to reenter it. This function is selected by typing a '.' (period).

3. Saving the sentence onto disk (+)

By entering '+' (plus) the user may save on disk the sentence just typed and produced. The sentence must be given a number which is used to identify it in subsequent recall.

4. Sound production of sentences from disk ($)

Previously stored sentences can be recalled from disk by typing the character. Then the program loads the sentence of the given number and produces it.

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 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 - to be used in the selection of the right sequence of microelements - as well as the inventory of microelements themselves. The system of rules is implemented in the form of tables and look up procedures. The stock of microelements contains the speech sounds (ie. phoneme realizations) 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 37 * 4 + 35 * 4 = 288 elements. The inventory of microelements used by the program is displayed in Fig. 2.

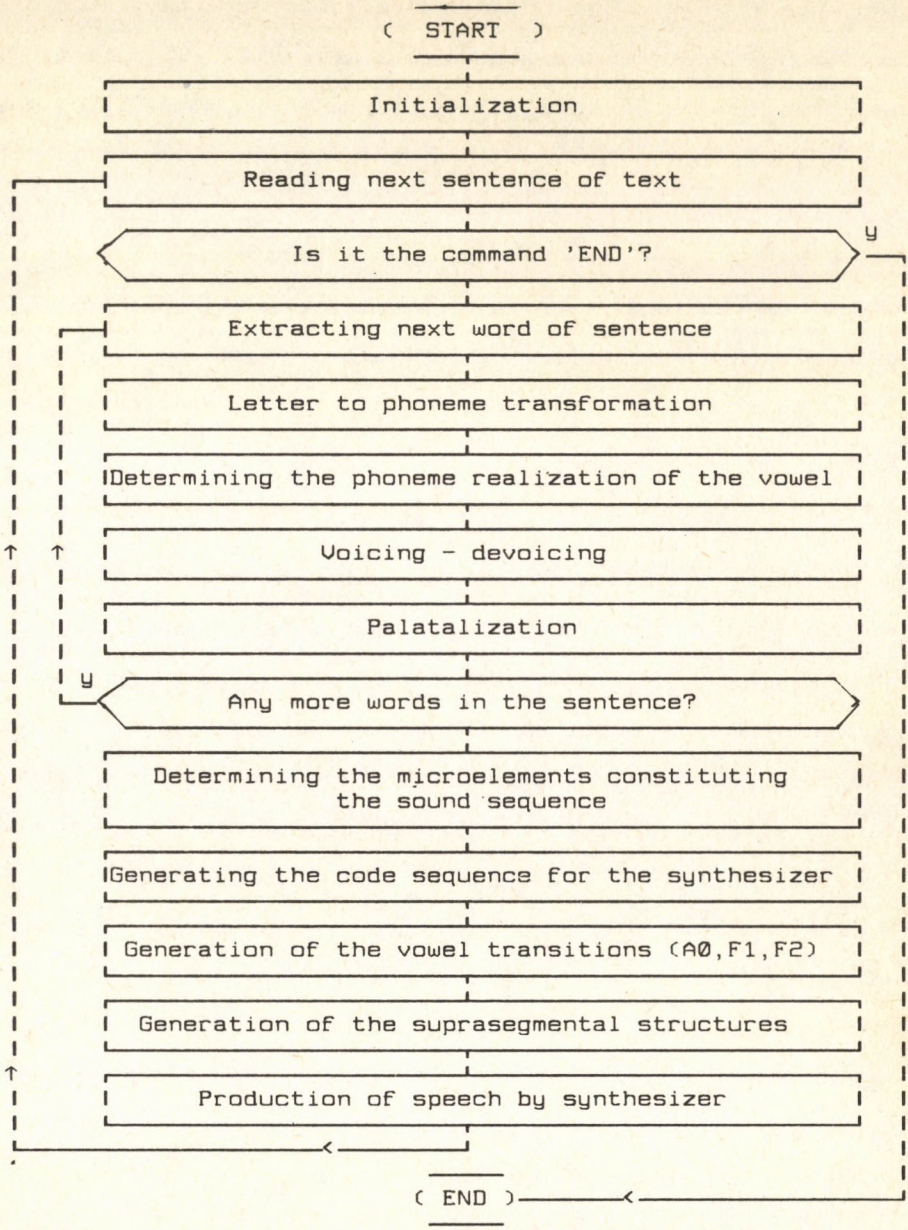


Fig. 1. The main steps of the operation of the Russian language text-to-speech system RUSSON

Consonants

Number of sound	Phonetic symbol	Number of microelement realising the sound
1.	ɓ'	1, 2, 3, 4
2.	ɓ	5, 6, 7, 8
3.	p'	9, 10, 11, 12
4.	p	13, 14, 15, 16
5.	m̓	17, 18, 19, 20
6.	m	21, 22, 23, 24
7.	d'	25, 26, 27, 28
8.	d	29, 30, 31, 32
9.	t'	33, 34, 35, 36
10.	t	37, 38, 39, 40
11.	n'	41, 42, 43, 44
12.	n	45, 46, 47, 48
13.	g'	49, 50, 51, 52
14.	g	53, 54, 55, 56
15.	k'	57, 58, 59, 60
16.	k	61, 62, 63, 64
17.	ʋ'	65, 66, 67, 68
18.	ʋ	69, 70, 71, 72
19.	f'	73, 74, 75, 76
20.	f	77, 78, 79, 80
21.	ʒ	81, 82, 83, 84
22.	ʒ'	85, 86, 87, 88
23.	z	89, 90, 91, 92
24.	ʃ:	93, 94, 95, 96
25.	ʃ	97, 98, 99,100
26.	s'	101,102,103,104
27.	s	105,106,107,108
28.	i̯	109,110,111,112
29.	j	113,114,115,116
30.	x'	117,118,119,120
31.	x	121,122,123,124
32.	t͡ʃ'	125,126,127,128
33.	t͡s	129,130,131,132
34.	r'	133,134,135,136
35.	r	137,138,139,140
36.	l'	141,142,143,144
37.	ɬ	145,146,147,148

Fig. 2/a. The structure of the stock of microelements

Vowels

Number of sound	Phonetic symbol	Number of microelement realising the sound
1.	a	149,150,151,152
2.	o	153,154,155,156
3.	u	157,158,159,160
4.	i	161,162,163,164
5.	ɛ	165,166,167,168
6.	a+	169,170,171,172
7.	ą	173,174,175,176
8.	æ	177,178,179,180
9.	ʌ	181,182,183,184
10.	o+	185,186,187,188
11.	o̜	189,190,191,192
12.	ö	193,194,195,196
13.	ŏ	197,198,199,200
14.	u+	201,202,203,204
15.	u̜	205,206,207,208
16.	ü	209,210,211,212
17.	ŭ	213,214,215,216
18.	u̯	217,218,219,220
19.	i̜	221,222,223,224
20.	i̯	225,226,227,228
21.	iɛ	229,230,231,232
22.	i͡ɛ	233,234,235,236
23.	ĭ	237,238,239,240
24.	ĭ̯	241,242,243,244
25.	ı	245,246,247,248
26.	ı̆	249,250,251,252
27.	ie	253,254,255,256
28.	ĭ̯	257,258,259,260
29.	ɛ	261,262,263,264
30.	e	265,266,267,268
31.	e	269,270,271,272
32.	ɛ̆	273,274,275,276
33.	ɩ	277,278,279,280
34.	ə	281,282,283,284
35.	ŭ̜	285,286,287,288

Fig. 2/b. The structure of the stock of microelements

Pauses

Number of sound	Phonetic symbol	Number of microelement realising the sound
1.	P1	288
2.	P2	289
3.	P3	290
4.	P4	291
5.	P5	292

Fig. 2/c. The structure of the stock of microelements

The letter-to-phoneme transformation

As shown by Fig. 1, the processing of a given sentence up to the stage where the sequence of microelements is determined is carried out word by word. The letter to phoneme transformation is also based on words. The program makes recourse to the LETTER-PHONEME TABLE (LETPHONTAB). The first column of this table includes the ordinal number, the second contains the letters in the Russian alphabet. The third column contains the ordinal number of the phonemes which directly correspond to the letters. There are 21 consonant

Number	Letter	Number of phoneme	Softening
1.	а	-1	0
2.	б	2	0
3.	в	18	0
4.	г	14	0
5.	д	8	0
6.	е	-5	1
7.	ё	-2	1
8.	ж	21	0
9.	з	23	0
10.	и	-4	1
11.	й	29	0
12.	к	16	0
13.	л	37	0
14.	м	6	0
15.	н	12	0
16.	о	-2	0
17.	п	4	0
18.	р	35	0
19.	с	27	0
20.	т	10	0
21.	у	-3	0
22.	ф	20	0
23.	х	31	0
24.	ц	33	0
25.	ч	32	0
26.	ш	25	0
27.	щ	24	0
28.	ъ	-	1
29.	ы	-4	0
30.	ь	-	0
31.	э	-5	0
32.	ю	-3	1
33.	я	-1	1

Fig. 5. LETTER - PHONEME TABLE (LETPHONTAB) used in the letter to phoneme transformation of the Russian language text-to-speech system RUSSON

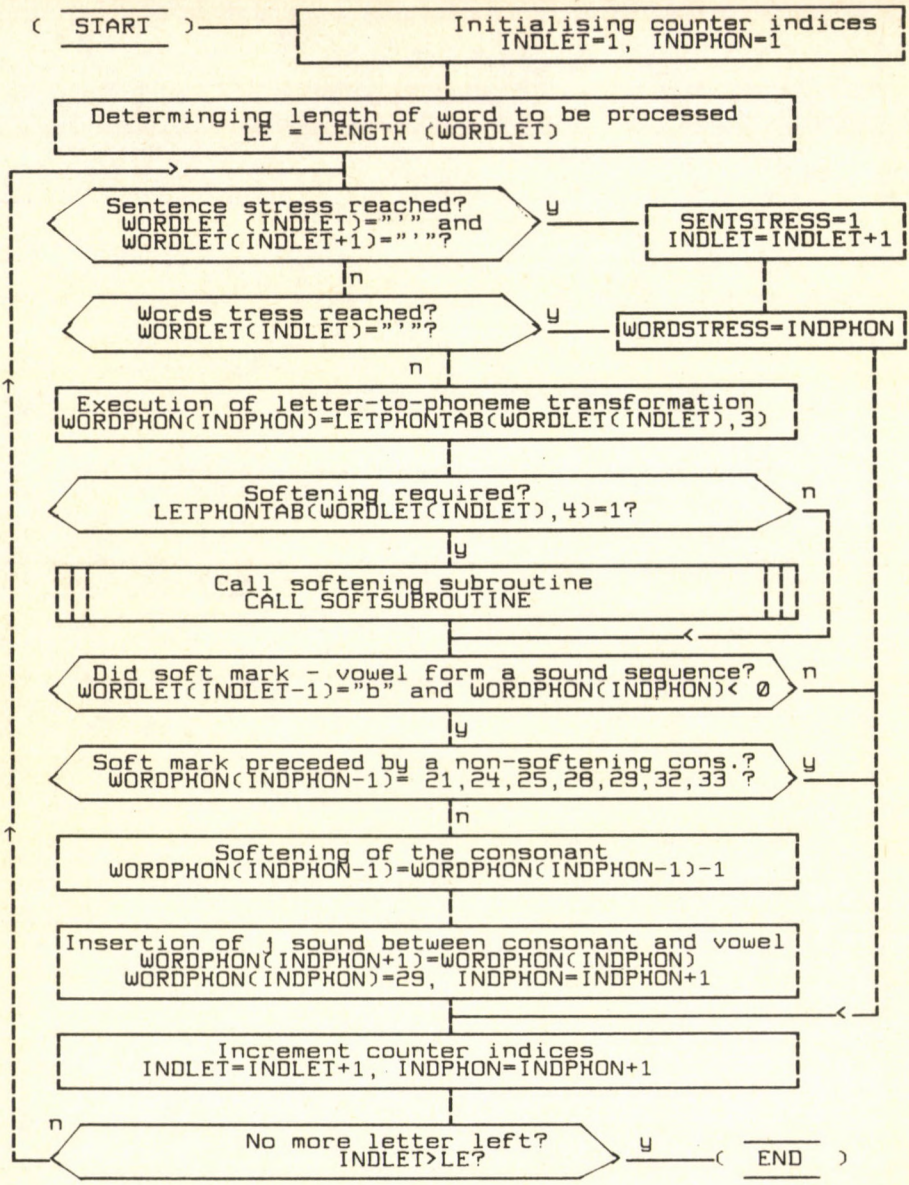


Fig. 4. Letter to phoneme transformation in the Russian language text-to-speech system RUSSON

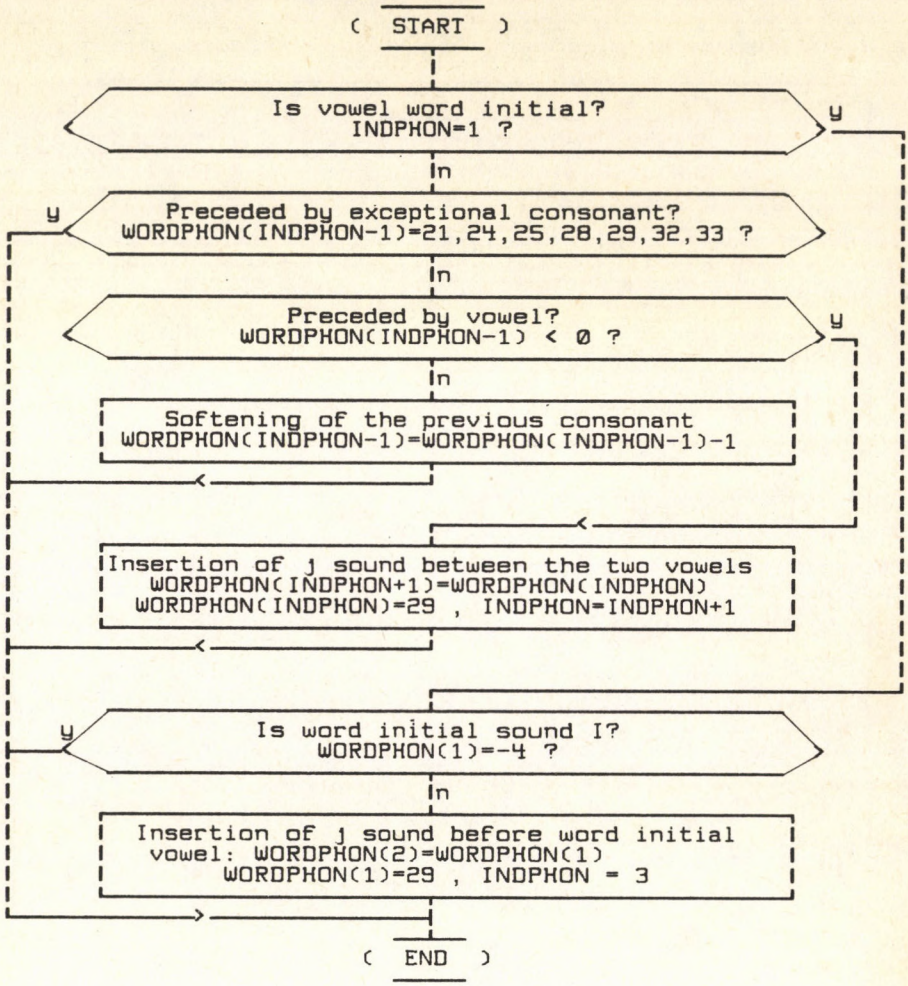


Fig. 5. Softening subroutine called by the letter-to-phoneme transformation part of the Russian language text-to-speech system RUSSON

and 5 vowel phonemes of this type (represented in the table by negative figures). The phonetic symbols representing consonant phonemes can be found under this number in the consonant section of the inventory of microelements.The value 1 in the fourth column of the LETPHONTAB table indicates whether it is necessary to apply softening in the following sound. If not, then the column carries a '0'. The detailed steps of the letter to phoneme transformation are given in the flow chart of Fig. 4. The particular word to be processed is held in the variable WORDLET in its orthographic form. The sequence of phonemes corresponding to this word will be stored in the variable WORDPHON. The program also registers word stress as well as possible sentence stress by storing the ordinal number of the stressed vowel. If the softening subroutine is also called, it makes use of the fact that the soft phoneme realizations precede their hard counterpart, so their ordinal number is one less. The operation of the softening function is displayed in Fig. 5. The transformation of the whole word yields the sequence of phonemes making up that word, and this forms the basis for further operations.

Determining the phoneme realizations.

The next step concerns the selection of the right phoneme realizations constituting the word in question. First, the vowel phonemes are processed. The consonants are processed in two steps, first voicing then palatalization must be established.

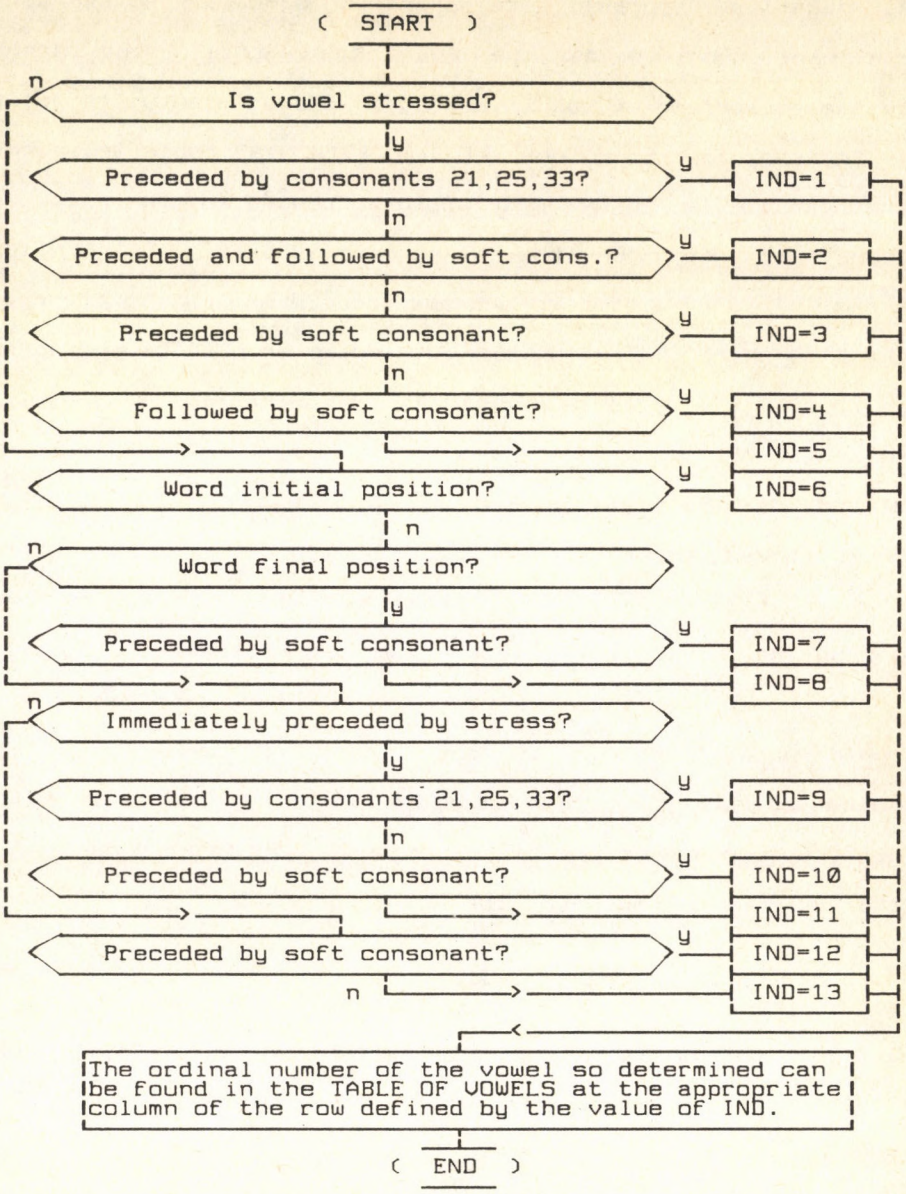
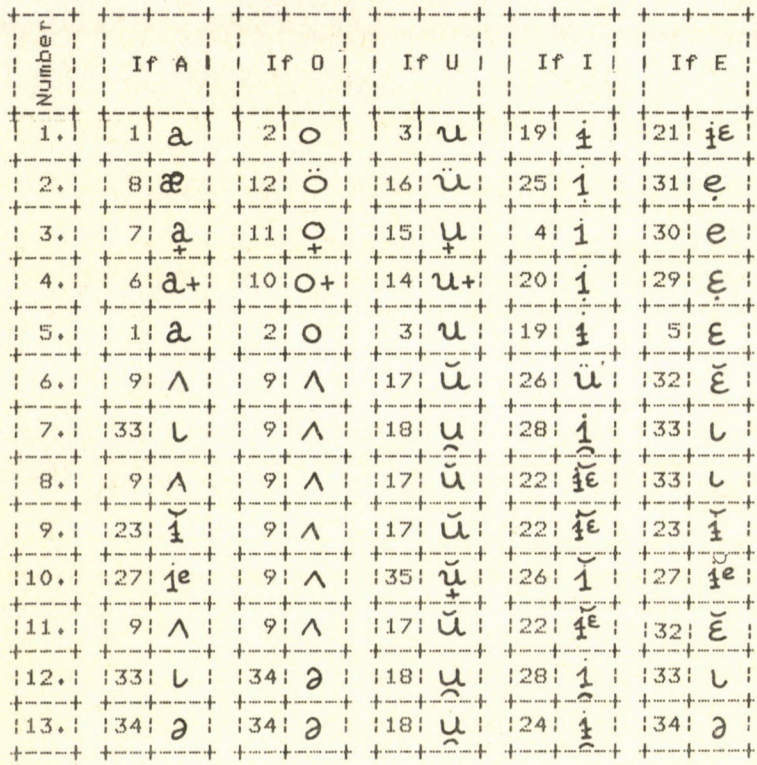


Fig. 6. Determination of vowel phoneme realizations on the basis of their phonetic positon



Number	If A	If O	If U	If I	If E
1.	1 a	2 o	3 ʊ	19 ɨ	21 ɨɛ
2.	8 æ	12 ö	16 ʊ̈	25 ɪ	31 e̞
3.	7 a̟	11 o̟	15 ʊ̟	4 i	30 e
4.	6 a+	10 o+	14 ʊ+	20 i̞	29 ɛ̞
5.	1 a	2 o	3 ʊ	19 ɨ	5 ɛ
6.	9 ʌ	9 ʌ	17 ʊ̆	26 ʉ̞	32 ɛ̆
7.	33 ɤ	9 ʌ	18 ʊ͡	28 i͡	33 ɤ
8.	9 ʌ	9 ʌ	17 ʊ̆	22 ɨ̆ɛ	33 ɤ
9.	23 ɤ̆	9 ʌ	17 ʊ̆	22 ɨ̆ɛ	23 ɨ̆
10.	27 ɪe	9 ʌ	35 ʊ̟̆	26 ɪ̆	27 ɨe
11.	9 ʌ	9 ʌ	17 ʊ̆	22 ɨ̆ɛ	32 ɛ̆
12.	33 ɤ	34 ə	18 ʊ͡	28 ɨ͡	33 ɤ
13.	34 ə	34 ə	18 ʊ͡	24 ɨ͡	34 ə

Fig. 7. TABLE OF VOWELS of the Russian language text-to-speech system RUSSON

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. The phonetic symbols of the 35 vowel phoneme realizations which capture the quality of the sound can be found in the vowel sections of the inventory of microelements. 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). The program segment displayed in the form of flow chart in Fig. 5 shows the determination of the value of an index. The value of the index defines one of the rows of the TABLE OF VOWELS (VOWTAB) (Fig. 6, 7). The column of the VOWTAB table is defined by the phoneme being processed. This is set to 1,2,3,4 and 5 corresponding to the vowels A, O, U, I, E respectively. At the interjunction of the two indices in the VOWTAB table defines the serial number of the vowel phoneme realization. This value is assigned a negative sign from this point on.

	CONSONANT		VOICELESS PAIR		VOICED PAIR	*	PALATALIZED **
1.	б'	3	п'	1	б'	2	1 б'
2.	б	4	п	2	б	2	-1 б'
3.	п'	3	п	1	б'	1	3 п'
4.	п	4	п	2	б	1	-3 п'
5.	м'	5	м'	5	м'	-2	5 м'
6.	м	6	м	6	м	-2	-5 м'
7.	д'	9	т'	7	д'	2	7 д'
8.	д	10	т	8	д	2	-7 д'
9.	т'	9	т'	7	д'	1	9 т'
10.	т	10	т	8	д	1	-9 д
11.	н'	11	н'	11	н'	-2	11 н'
12.	н	12	н	12	н	-2	-11 н'
13.	г'	15	к'	13	г'	2	13 г'
14.	г	16	к	14	г	2	-13 г'
15.	к'	15	к'	13	г	1	15 к'
16.	к	16	к	14	г	1	-15 к'
17.	в'	19	ф'	17	в'	2	17 в'
18.	в	20	ф	18	в	2	17 в'
19.	ф'	19	ф'	17	в'	1	19 ф'

*
1 INDICATES VOICELESS SOUNDS

2 INDICATES VOICED SOUNDS

MINUS VALUE INDICATES SONORANTS

**
MINUS VALUE INDICATES PLOSIVES

Fig. 8/a.

TABLE OF CONSONANTS (CONSTAB) of the Russian language text-to-speech system RUSSON

	CONSONANT	VOICELESS PAIR	VOICED PAIR	*	PALATALIZED **
20.	f	20 f	18 v	1	19 f'
21.	ʒ	25 ʃ	21 ʒ	2	21 ʒ
22.	z'	26 s'	22 z'	2	22 z'
23.	z	27 s	23 z	2	22 z'
24.	ʃ'	24 ʃ'	24 ʃ'	1	24 ʃ'
25.	ʃ	25 ʃ	21 ʒ	1	25 ʃ
26.	s'	26 s'	22 z'	1	26 s'
27.	s	27 s	23 z	1	26 s'
28.	i͡	28 i͡	28 i͡	-2	28 i͡
29.	ɟ	29 ɟ	29 ɟ	-2	29 ɟ
30.	ɕ'	30 ɕ'	30 ɕ'	1	30 ɕ'
31.	ɕ	31 ɕ	31 ɕ	1	30 ɕ'
32.	t͡ʃ'	32 t͡ʃ'	32 t͡ʃ'	1	32 t͡ʃ'
33.	t͡s	33 t͡s	48 t͡s	1	32 t͡s
34.	r'	34 r'	34 r'	-2	34 r'
35.	r	35 r	35 r	-2	34 r'
36.	l'	36 l'	36 l'	-2	36 l'
37.	ɫ	37 ɫ	37 ɫ	-2	36 l'

*
1 INDICATES VOICELESS SOUNDS

2 INDICATES VOICED SOUNDS

MINUS VALUE INDICATES SONORANTS

**
MINUS VALUE INDICATES PLOSIVES

Fig. 8/b.

TABLE OF CONSONANTS (CONSTAB) of the Russian language text-to-speech system RUSSON

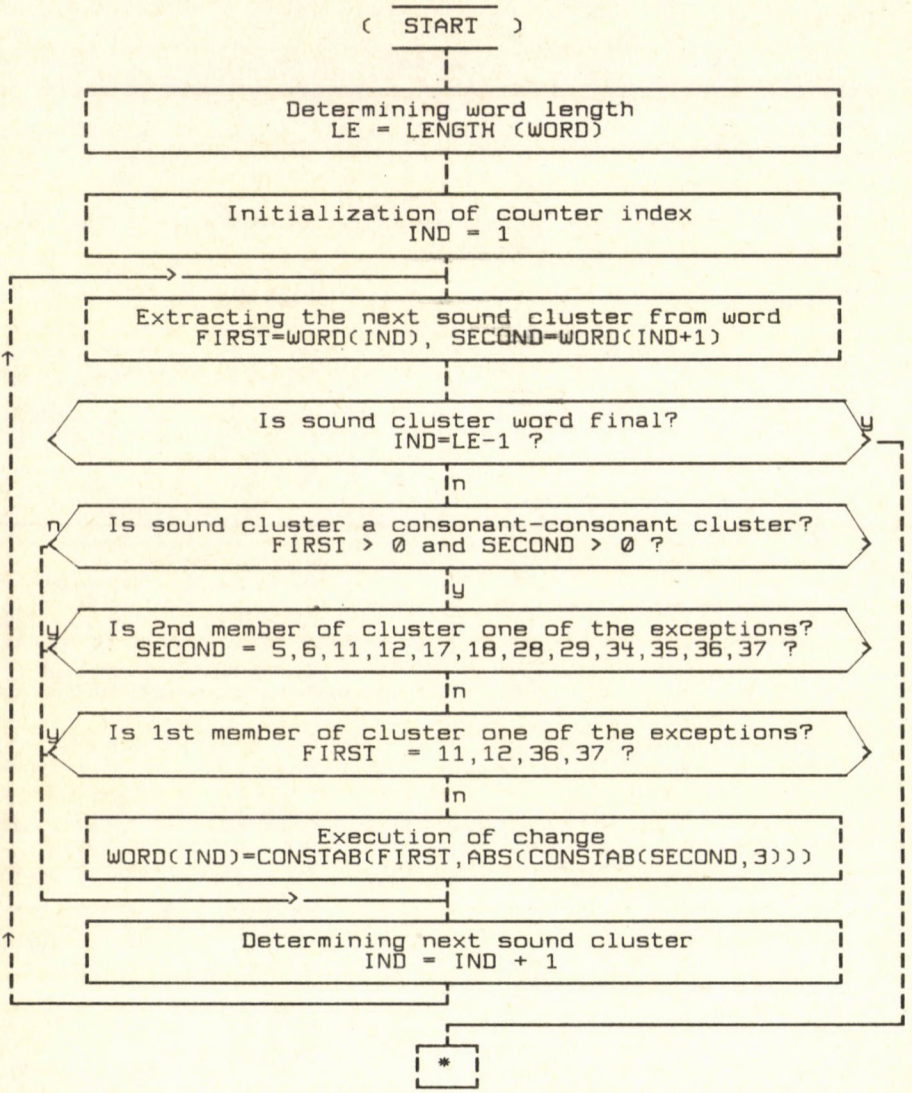


Fig. 9. The subroutine executing voicing and devoicing in the Russian language text-to-speech system RUSSON

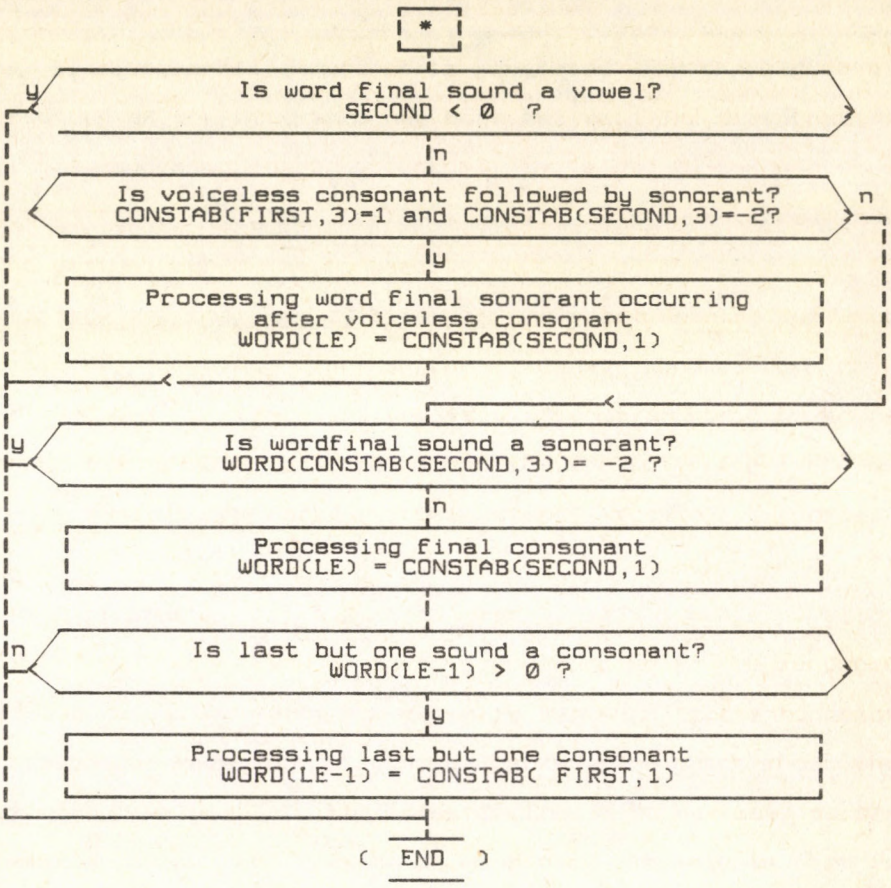


Fig. 1o. The subroutine executing voicing and devoicing in the Russian language text-to-speech system RUSSON

Selecting the consonant phoneme realization.

The selection of the consonant phoneme realization follows that of the vowels. The consonant phoneme number yielded by the letter-to-phoneme transformation is identical to the phoneme realization number. The phonetic symbol of the realizations can be found in the consonant section of the stock of microelements (see Fig. 8 a,b). 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 segment controlling voicing and devoicing makes use of the TABLE OF CONSONANTS (CONSTAB) see Fig. 8 a,b, in particular, columns 1, 2, and 3. Column 0 contains the serial number and the phonetic symbol of the consonant phoneme realizations. The corresponding voiceless phoneme realization is found in column 1 with its voiced pair in column 2. (Naturally, the voiceless pair of a voiceless sound is itself just as the voiced pair of a voiced sound amounts to the same sound.) Column 3 of the CONSTAB table has the following meaning: value 1 indicates a voiceless, value 2 a voiced sound, a negative value stands for a sonorant. The same program segment executes voicing and devoicing. However, the absolute value of column 3 shows whether the transformation affects column 1 or 2 of the CONSTAB table. The step-by-step operation of the program segment executing voicing and

devoicing is shown in Fig. 9, 10. As shown by the chart, the word to be processed is contained in the variable WORD from which two member sound clusters are extracted one by one. If the sound cluster consists of a vowel and a consonant, no change is made and the next cluster is extracted. 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. Consonants 11, 12, 36 and 37 are exceptions in initial position in the cluster, while consonants 5, 6, 12, 17, 18, 28, 29, 34, 35, 36, 37 are exceptions as the second member of the cluster. When a modification is called for, it is carried out with the help of CONSTAB in the way described above. Word final consonant--consonant clusters require special treatment. First, the word final sonorant is devoiced (if necessary) and then the preceding consonant is processed.

Execution of palatalization

The final step in the determination of the consonant phoneme realization is to determine the modifications owing to pálatalization. This program segment takes as input the word contained in the variable WORD which has been used and possibly modified in earlier steps. Here again, the program first extracts two element sound clusters which are checked

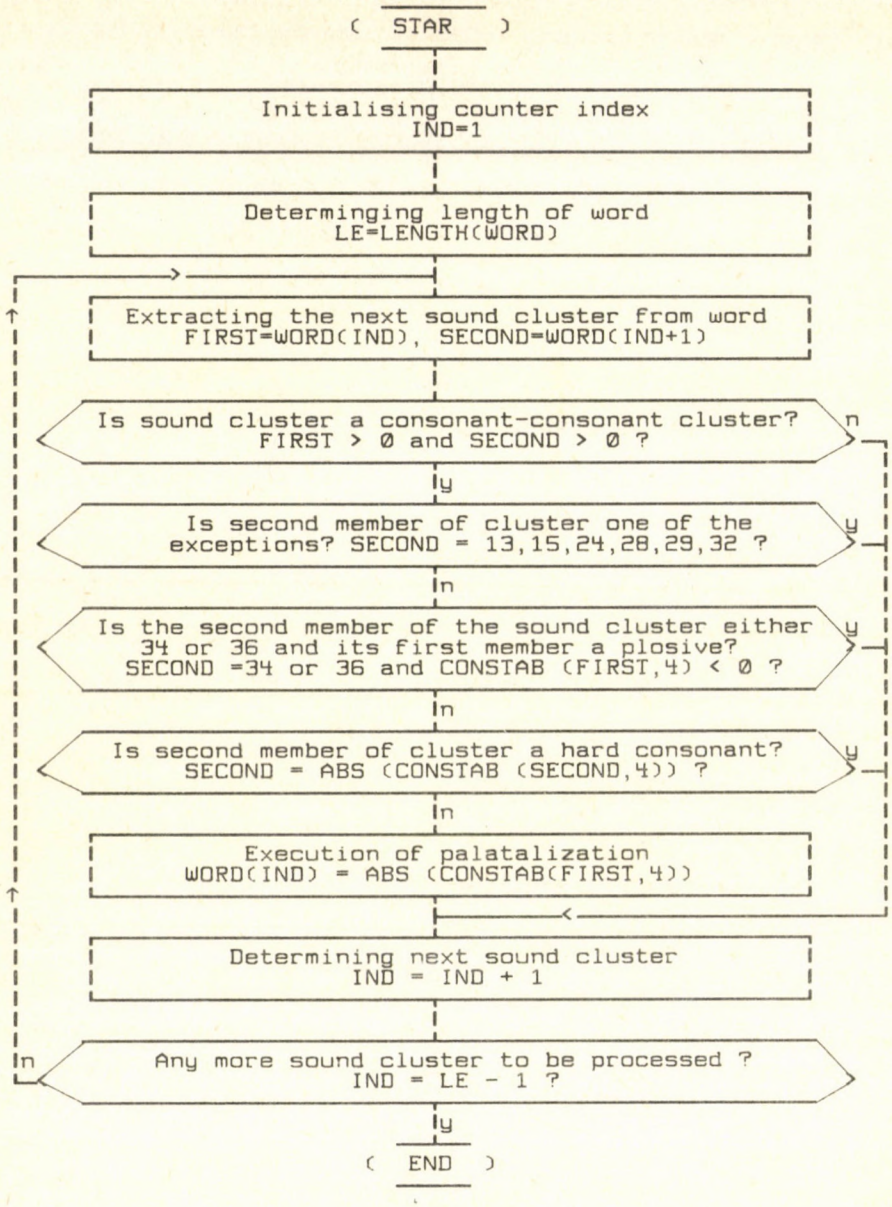


Fig. 11. The subroutine executing palatalization in the Russian language text-to-speech system RUSSON

to see if they are both consonants. If not, then the next cluster is accessed. If they are both consonants then the combinations not undergoing palatalization are filtered out. This amounts to carrying out the following three tests:

1) Is the second consonant one of the exceptions? Here the following consonants are the exceptions: 13, 15, 24, 28, 29, 32.

2) Is the second member of the cluster either numbered 34 or 36 and the preceding consonant a stop? Column 4 of CONSTAB helps to identify stops, as it contains a negative value for stop consonants.

3) Is the second member of the cluster a hard consonant? Again, column 4 of CONSTAB is checked to see if the absolute value of the cell there is identical with the number of the consonant currently processed. If the numbers are not identical, this means palatalization is not called for.

Where required, palatalization is executed by changing the number of the initial member of the cluster in column 4 of CONSTAB to its absolute value.

Having examined all the clusters in the word, the palatalization routine terminates its operation and this means at the same time that the number of both the vowel and consonant phoneme realizations making up the word have been identified.

If the sentence includes some more words to be processed, the program continues with the letter-to-phoneme transformation of that word, otherwise it proceeds to determine the sequence of microelements on the basis of the phoneme realization

numbers.

Defining the microelements of the sound sequence.

The suprasegmental structure corresponding to the sounds defined earlier is based on microelements. As can be seen in Fig. 11, 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. Therefore, for vowels it is used only in initial positions or when preceded by a voiceless consonant. Consonants 24, 31, or 32 always have only three microelements. Voiced consonants have four microelements in initial position or when preceded by a voiceless consonant.

The ordinal number of microelements are calculated on the following basis:

for consonants:

no. of microelement = (no. of consonant - 1) * 4 + INDEX

where INDEX = (1), 2, 3, 4

for vowels:

no. of microelement = ABS((no. of consonant - 1) * 4) + 148 + INDEX

where INDEX = (1), 2, 3, 4

The program inserts a pause microelement between words. The number of this microelement is 2o and is 15o ms long. Depending on the sentence final punctuation mark the program

selects an appropriate pause microelement.

The last step in the construction of the segmental structure of the utterance is the formation of the vowel transitions.

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. They are the second and the fourth microelements of a vowel (except in word initial position or when preceded by a voiceless consonant, in which case it is the first and not the second element). 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. When modifying the first microelement the initial values of A0, F1, F2 are accessed from a table while the target values of the transition are the values of the second microelement of the vowel. In adjusting the last microelement of the vowel, the initial values are supplied by the final values of the last but one microelement while the final values of the modified vowel are accessed from a table again. This table is called VOWEL TRANSITIONS (VOWTRANS) and it has 37 rows corresponding to the 37 consonants with five rows each for the five vowel phonemes. It appears then that the program specifies the same initial and final A0, F1, F2 values for the transition of all realizations of a given vowel phoneme. Fig. 12 shows the word

Examples for inputting Russian texts.

Тё'тя пъё't ру''сский ча'й. Тё'тя пъё't ру''сский ча'й?

Сады' цвету't весно''й. Сады' цвету't весно''й?

Ната'ша пое'хала нада''чу. Ната'ша пое'хала нада''чу?

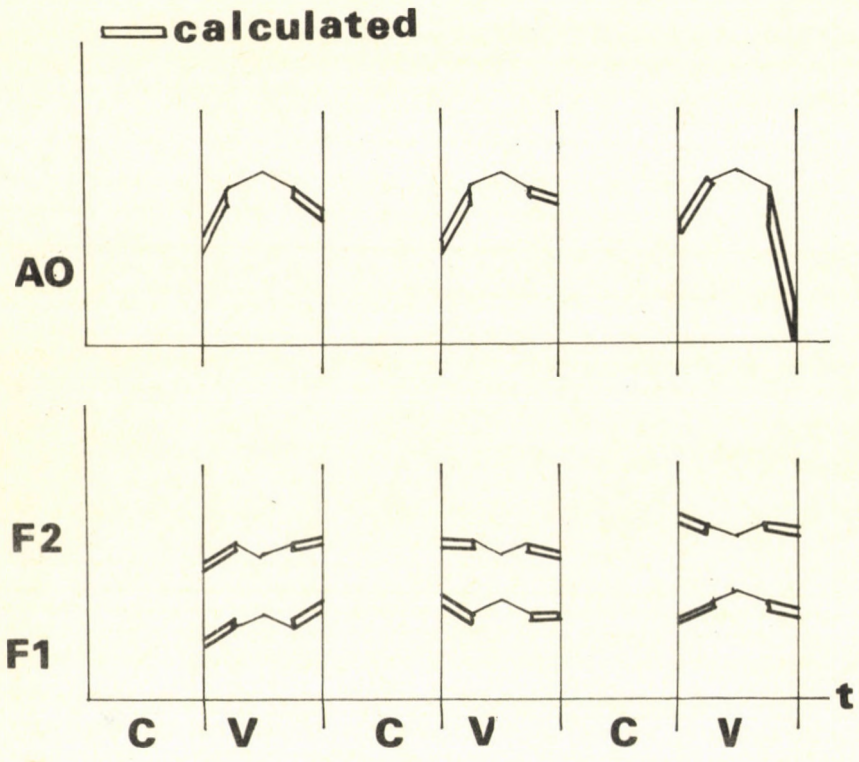


Fig. 12. The vowel transitions constructed

" Hата'ма " with the vowel transitions constructed.

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 sentence final punctuation marks: . (full stop), :(colon), , (comma), ; (semi colon), ! (exclamation mark), ? (question mark), ?! (question mark - exclamation mark), ?? (double question mark). The intonation contours corresponding to the sentence final punctuation marks are shown in Fig. 0. Each intonation contour is made up of three parts. The chart shows the frequency values of the start, the end as well as the possible break off point of each part. The middle part is fitted onto the lengthened vowel carrying the sentence stress. If the intonation contour has a break, it is positioned at the given percent value of the duration of the vowel. The initial part of the intonation contour is fitted onto the stretch preceding the sentence stress. (If the sentence begins with a vowel bearing the sentence stress, then the first part is omitted.) The third part of the intonation contour is fitted onto the stretch which comes

after the sentence stress. (If the sentence ends on a vowel carrying the sentence stress, then this part is omitted.)

With this operation completed, the complex sound structure is ready to be produced.