

MAGYAR FONETIKAI FÜZETEK 17

HUNGARIAN PAPERS IN PHONETICS

**STUDIES
IN PHONETICS**

Kiadja az MTA
Nyelvtudományi Intézete
Budapest 1987

MAGYAR FONETIKAI FÜZETEK

Hungarian Papers in Phonetics

17.

STUDIES IN PHONETICS

Papers by Hungarian phoneticians submitted for the 11th
International Congress of Phonetic Sciences

Edited by

KÁLMÁN BOLLA

LINGUISTICS INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES

BUDAPEST 1987

Technical editor: ÉVA FÖLDI

Revised by PÉTER SIPTÁR

Technical assistant: PÉTER NIKLÉCZY

HU ISSN 0134--1545

ISBN 963 8461 25 X

© Az MTA Nyelvtudományi Intézete, Budapest, 1987

Felelős kiadó: HERMAN JÓZSEF, az MTA Nyelvtudományi
Intézetének igazgatója.

Készült 400 példányban, 19,06 (A/5) ív terjedelemben,
térítésmentes terjesztésre.

Hozott anyagról sokszorosítva

8717367 MTA Sokszorosító, Budapest. F.v.: DR. HÉCZEY
LÁSZLÓNÉ.

PREVIOUS CONGRESSES

First:	Amsterdam	1932
Second:	London	1935
Third:	Ghent	1938
Fourth:	Helsinki	1961
Fifth:	Munster	1964
Sixth:	Prague	1967
Seventh:	Montreal	1971
Eighth:	Leeds	1975
Ninth:	Copenhagen	1979
Tenth:	Utrecht	1983

INTERNATIONAL PERMANENT COUNCIL

President: P. Ladefoged, USA

Vice-President: E. Fischer-Jørgensen, Denmark

General Secretary: R. Gsell, France

Honorary Members:

U.A. Artemov, USSR

A. de Lacerda, Portugal

A. Martinet, France

P. Moore, USA
I. Ochiai, Japan
M. Onishi, Japan
W. Pee, Belgium
A. Rosetti, Rumania
G. Straka, France
E. Zwirner, F.R. Germany

Members:

A.S. Abramson, USA
K. Bolla, Hungary
R. Charbonneau, Canada
L.A. Chistovich, USSR
A. Cohen, the Netherlands
G. Fant, Sweden
U. Fromkin, USA
H. Fujisaki, Japan
M. Halle, USA
P. Janota, Czechoslovakia
W. Jassem, Poland
M. Kloster-Jensen, Norway
K. Kohler, F.R. Germany
J. Laver, UK
B. Lindblom, Sweden
B. Malmberg, Sweden
I.M. Nicolajeva, USSR
K.L. Pike, USA
M. Rimmel, USSR

A. Rigault, Canada

M. Rossi, France

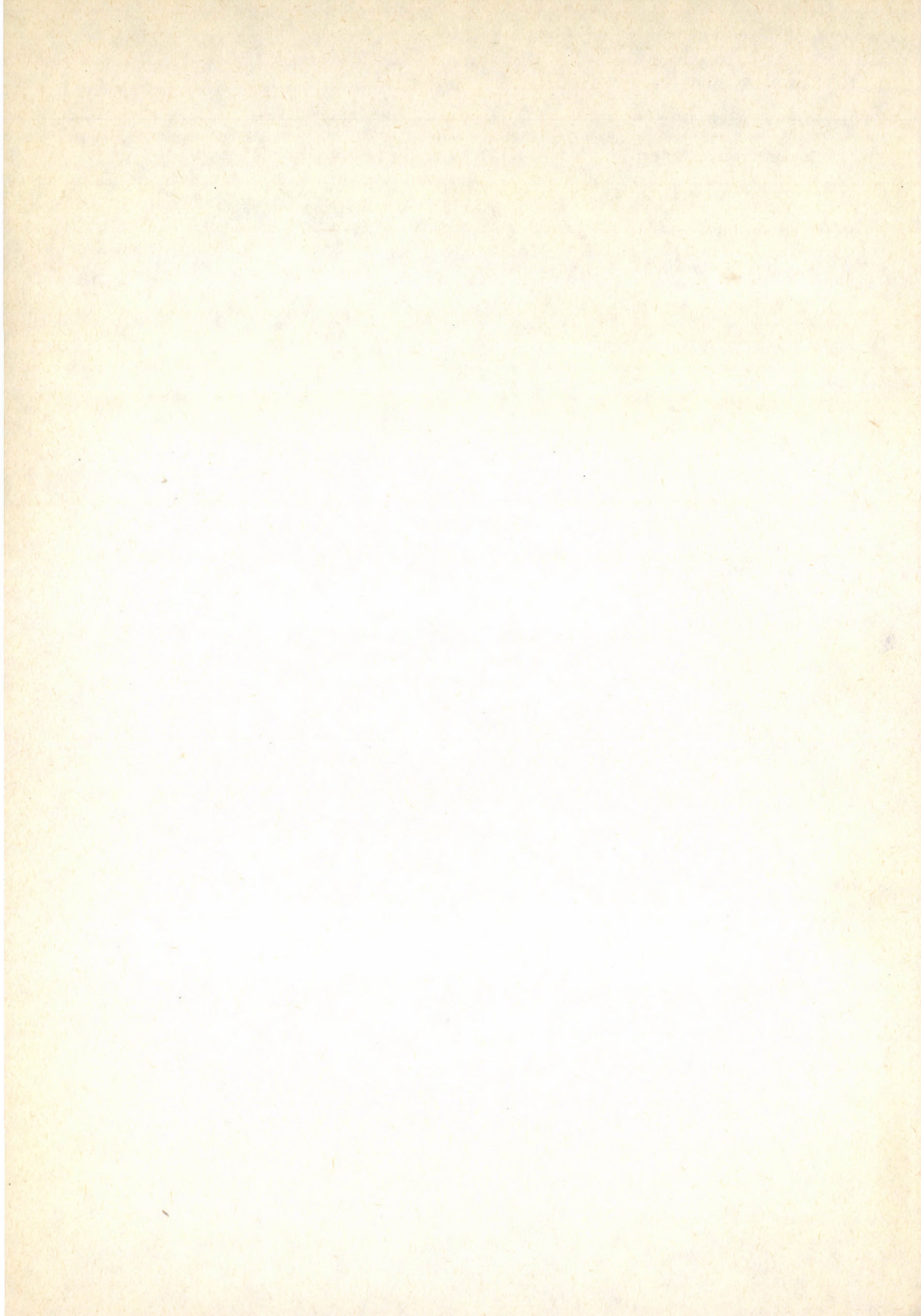
M. Sawashima, Japan

A. Sovijärvi, Finland

K.N. Stevens, USA

G. Straka, France

Wu, Zong-shi, Rep. China



STUDIES IN PHONETICS

Papers by Hungarian phoneticians submitted for the 11th
International Congress of Phonetic Sciences

CONTENTS

Kálmán BOLLA--Gábor KISS: The phonetic basis of artificial Russian speech, its generation by computer and its application	5
Éva FÖLDI: Polish palatalization and pharyngealization in an interlingual comparison	44
Mária GÓSY: High frequency speech perception: phonetic aspects and application	64
Mária GÓSY--Gabor OLASZY--Jenő HIRSCHBERG--Zsolt FARKAS: Phonetically based new method for audiometry: the G-O-H measuring system using synthetic speech	84
Ilona KASSAI: On the tonosyntax of a Hungarian child's early questions.(A preliminary report).....	102
Gábor KISS--András ARATÓ--Jozsef LUKACS--Janos SÜLYAN--Terez VASPÖRI: Braille-Lab, a full Hungarian text-to-speech microcomputer for the blind	116
Gabor KOZMA: An interlingual typological examination of vowels	132
Gabor OLASZY--Geza GORDOS: On the speaking module of an automatic reading machine	163

András SOPRONI: Errors in word stress in the Russian speech of Hungarians	192
László VALACZKAI: On the acoustic structure of German plosives and nasals	207
Domokos VÉKÁS: On nasality in French	229

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 dynamic 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 synthesizer 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 processes. 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 algorithm 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 constituents. The number of acoustic parameters playing 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 FO 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 accident. 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 pharyngealization.

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 diphthongal and triphthongal 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-pharyngeal 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.

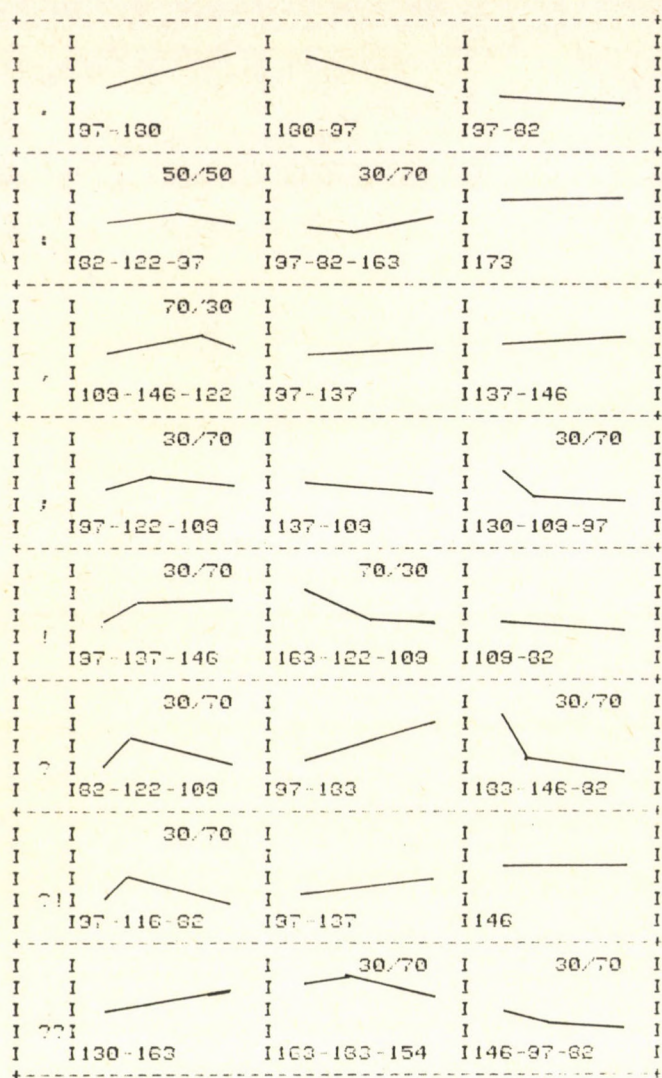


Fig. 0. The fundamental frequency (F_0) contours corresponding to the sentence final punctuation marks of the Russian text-to-speech system RUSSON

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 U 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.

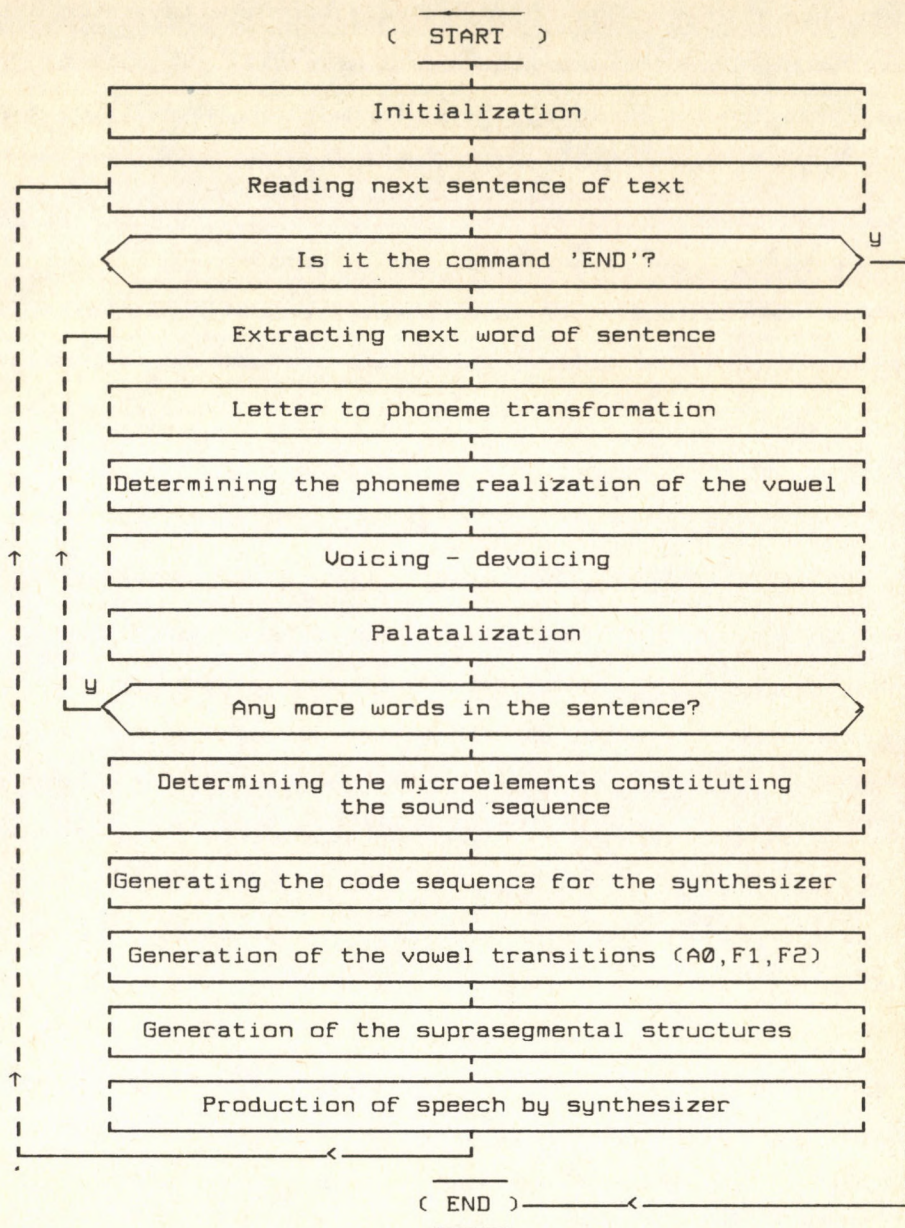


Fig. 1. The main steps of the operation of the Russian language text-to-speech system RUSSON

Consonants

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

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

Vowels

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

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

Number	Letter	Number of phoneme	Softening
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

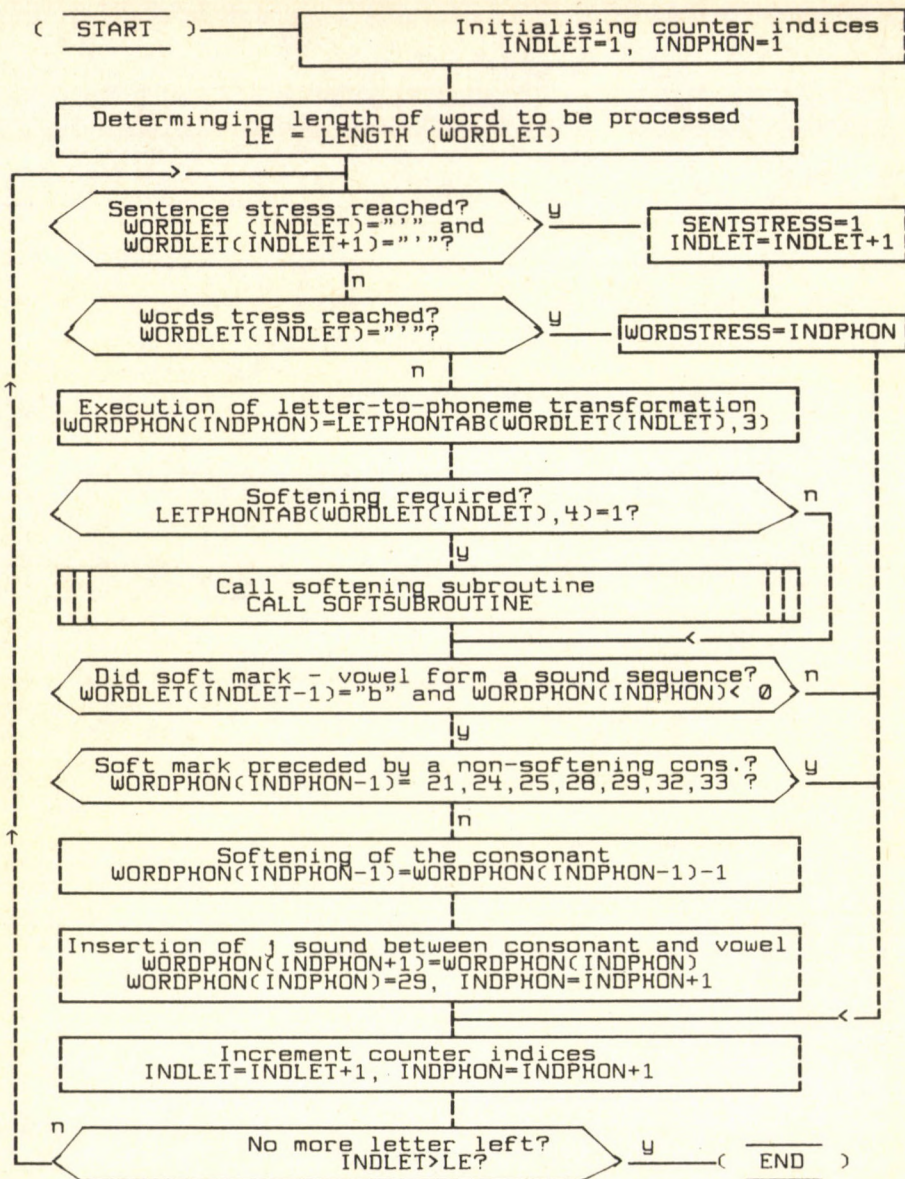


Fig. 4. Letter to phoneme transformation in the Russian language text-to-speech system RUSSON

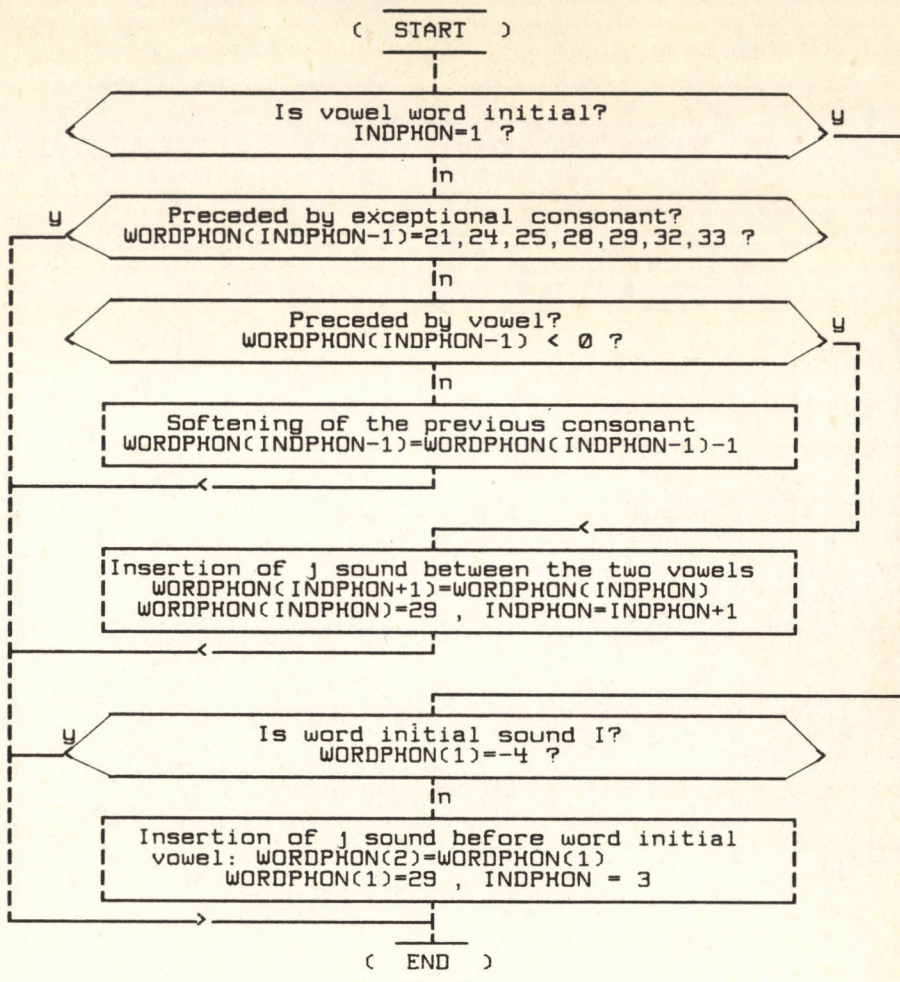


Fig. 5. Softening subroutine called by the letter-to-phoneme transformation part of the Russian language text-to-speech system RUSSON

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.

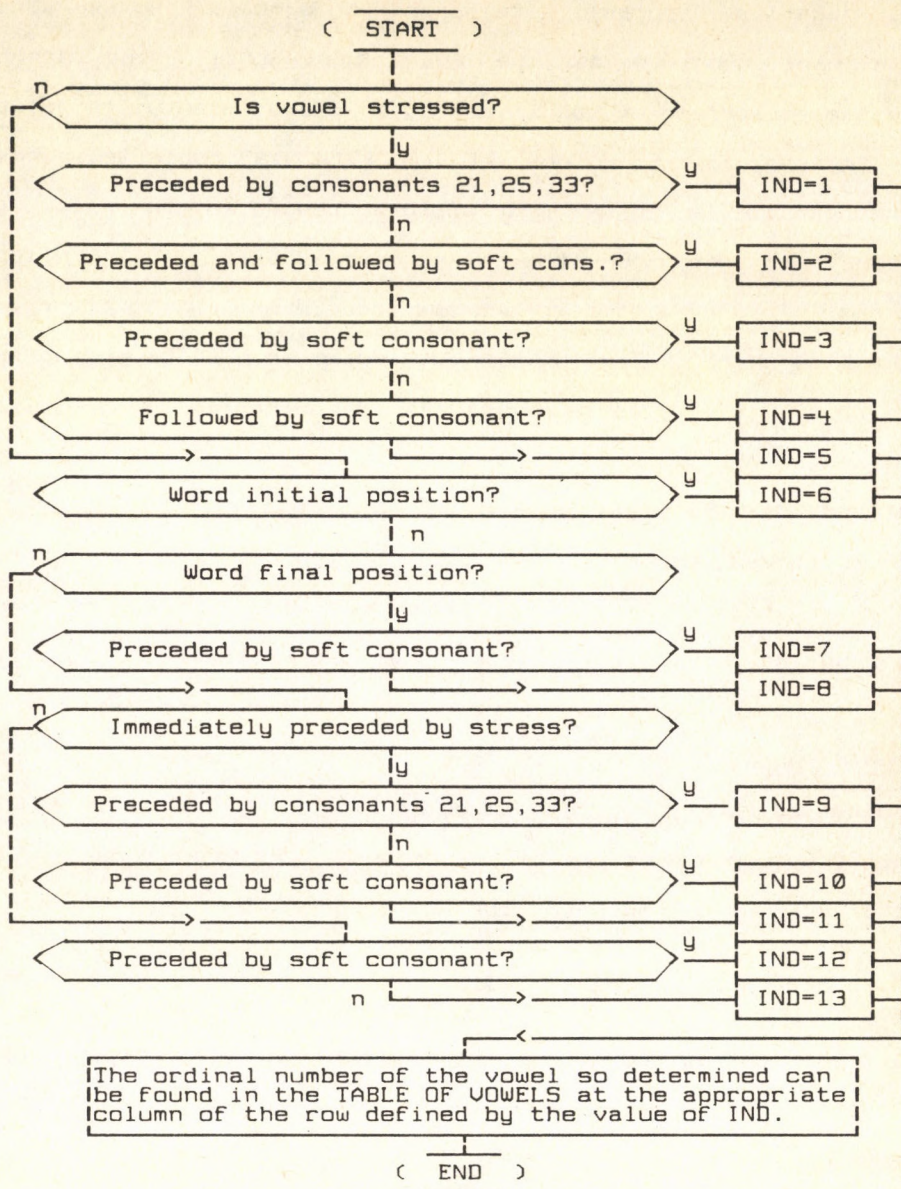


Fig. 6. Determination of vowel phoneme realizations on the basis of their phonetic position

Number	If A	If O	If U	If I	If E
1.	1 a	2 o	3 u	19 i	21 ie
2.	8 æ	12 ö	16 ü	25 ı	31 e
3.	7 a ₊	11 o ₊	15 u ₊	4 i	30 e
4.	6 a ₊	10 o ₊	14 u ₊	20 i	29 e
5.	1 a	2 o	3 u	19 i	5 e
6.	9 ʌ	9 ʌ	17 ü	26 ü	32 ẽ
7.	33 ʌ	9 ʌ	18 u	28 i	33 ʌ
8.	9 ʌ	9 ʌ	17 ü	22 ie	33 ʌ
9.	23 i	9 ʌ	17 ü	22 ie	23 i
10.	27 ie	9 ʌ	35 ü	26 i	27 ie
11.	9 ʌ	9 ʌ	17 ü	22 ie	32 ẽ
12.	33 ʌ	34 ə	18 u	28 i	33 ʌ
13.	34 ə	34 ə	18 u	24 i	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. b'	3. p'	1. b'	2.	1. b'
2. b	4. p	2. b	2.	-1. b'
3. p'	3. p	1. b'	1.	3. p'
4. p	4. p	2. b	1.	-3. p'
5. m'	5. m'	5. m'	-2.	5. m'
6. m	6. m	6. m	-2.	-5. m'
7. d'	9. t'	7. d'	2.	7. d'
8. d	10. t	8. d	2.	-7. d'
9. t'	9. t'	7. d'	1.	9. t'
10. t	10. t	8. d	1.	-9. d
11. n'	11. n'	11. n'	-2.	11. n'
12. n	12. n	12. n	-2.	-11. n'
13. g'	15. k'	13. g'	2.	13. g'
14. g	16. k	14. g	2.	-13. g'
15. k'	15. k'	13. g	1.	15. k'
16. k	16. k	14. g	1.	-15. k'
17. v'	19. f'	17. v'	2.	17. v'
18. v	20. f	18. v	2.	17. v'
19. f'	19. f'	17. v'	1.	19. f'

- *
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. ɸ	25. ɸ	21. ɸ	2	21. ɸ
22. x'	26. s'	22. x'	2	22. x'
23. x	27. s	23. x	2	22. x'
24. ʃ'	24. ʃ'	24. ʃ'	1	24. ʃ'
25. ʃ	25. ʃ	21. ɸ	1	25. ʃ
26. s'	26. s'	22. x'	1	26. s'
27. s	27. s	23. x	1	26. s'
28. i	28. i	28. i	-2	28. i
29. j	29. j	29. j	-2	29. j
30. x'	30. x'	30. x'	1	30. x'
31. x	31. x	31. x	1	30. x'
32. tʃ'	32. tʃ'	32. tʃ'	1	32. tʃ'
33. ts	33. ts	48. ts	1	32. ts
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. l	37. l	37. l	-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

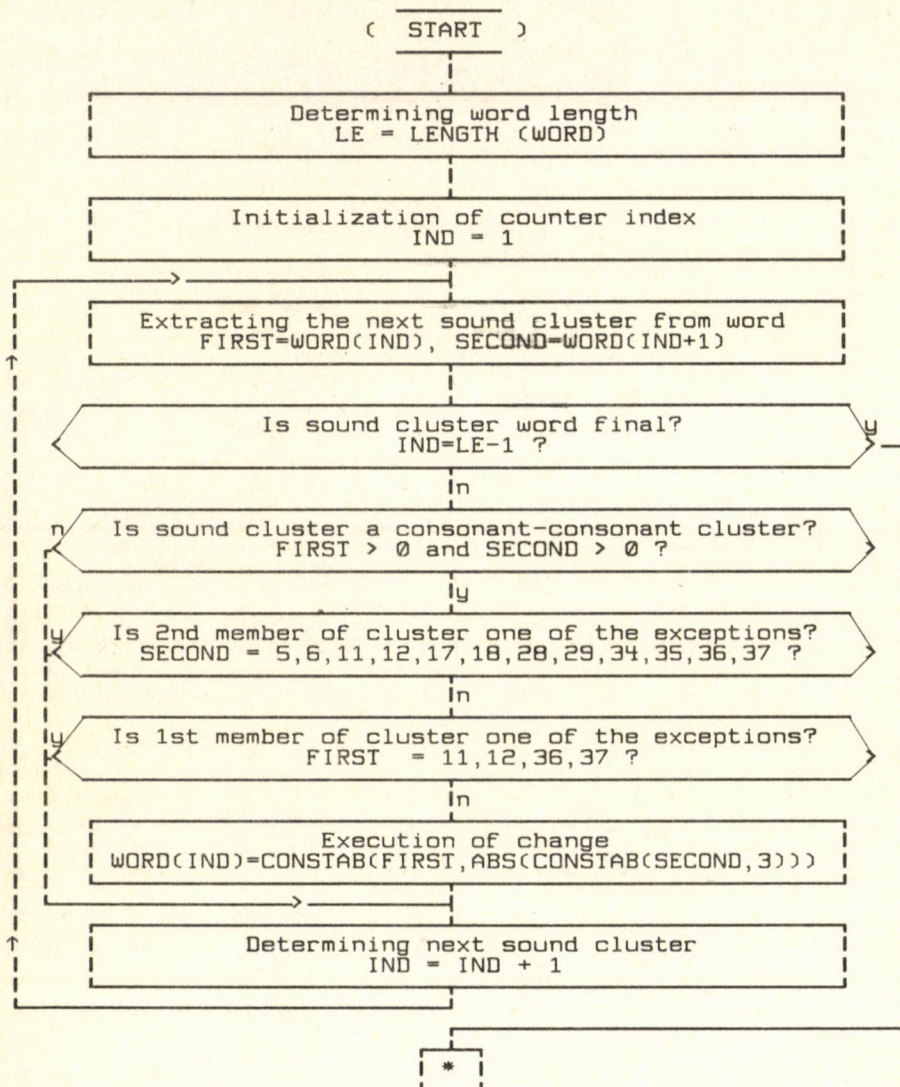


Fig. 9. The subroutine executing voicing and devoicing in the Russian language text-to-speech system RUSSON

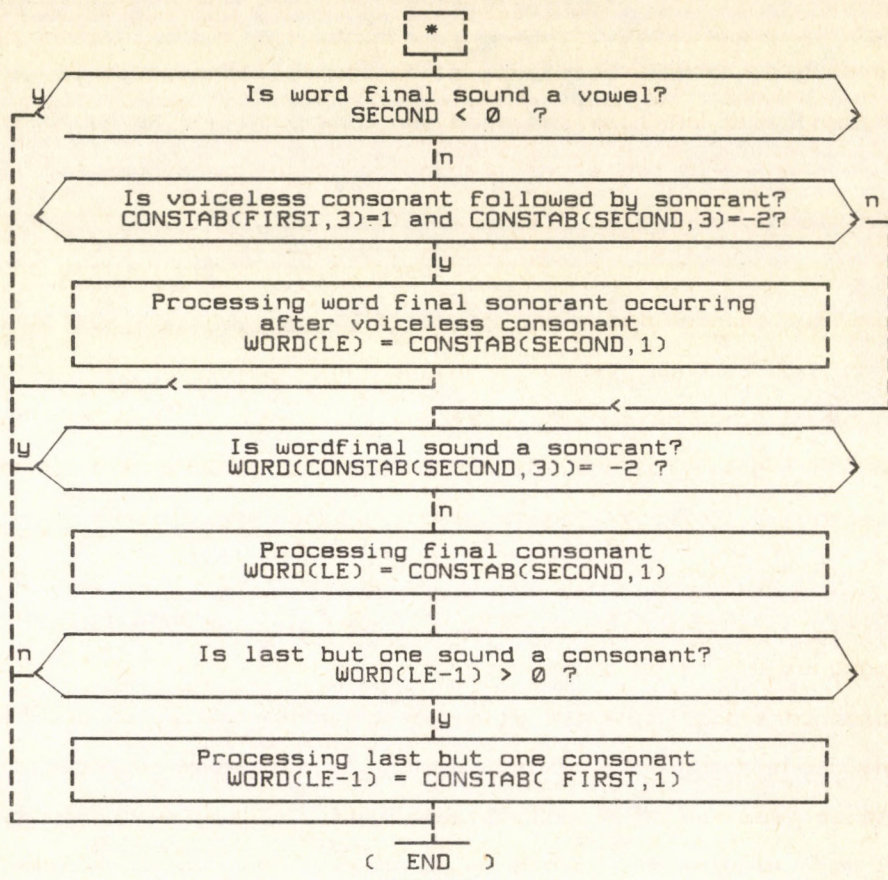


Fig. 10. The subroutine executing voicing and devoicing in the Russian language text-to-speech system RUSSON

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 palatalization. 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

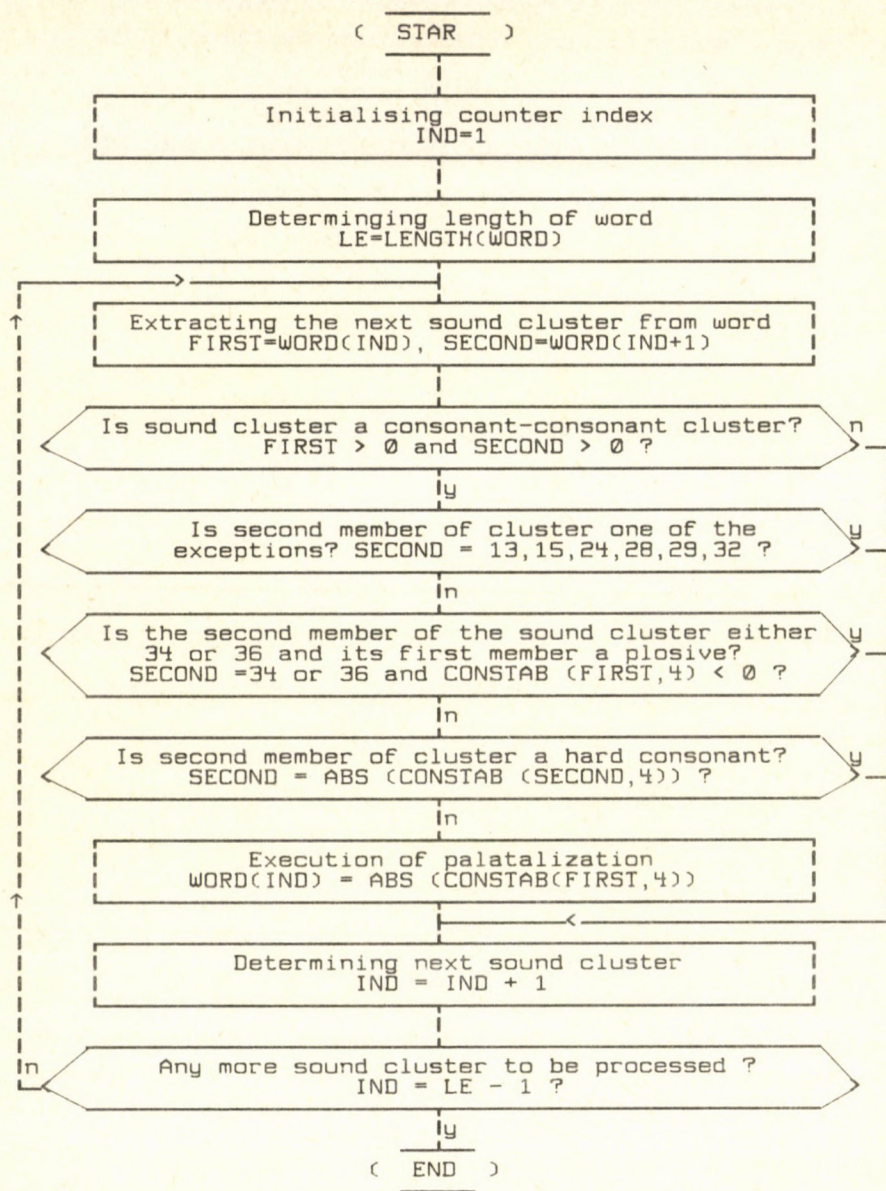


Fig. 11. The subroutine executing palatalization in the Russian language text-to-speech system RUSSON

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 20 and is 150 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 (A_0) and the first two formants (F_1 , F_2) 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 A_0 , F_1 , F_2 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 A_0 , F_1 , F_2 values for the transition of all realizations of a given vowel phoneme. Fig. 12 shows the word

Examples for inputting Russian texts.

Те'тя пѣ'т ру''сский ча'й.

Те'тя пѣ'т ру''сский ча'й?

Сады' цвету'т весно''й.

Сады' цвету'т весно''й?

Ната'ша пое'хала нада''чу.

Ната'ша пое'хала нада''чу?

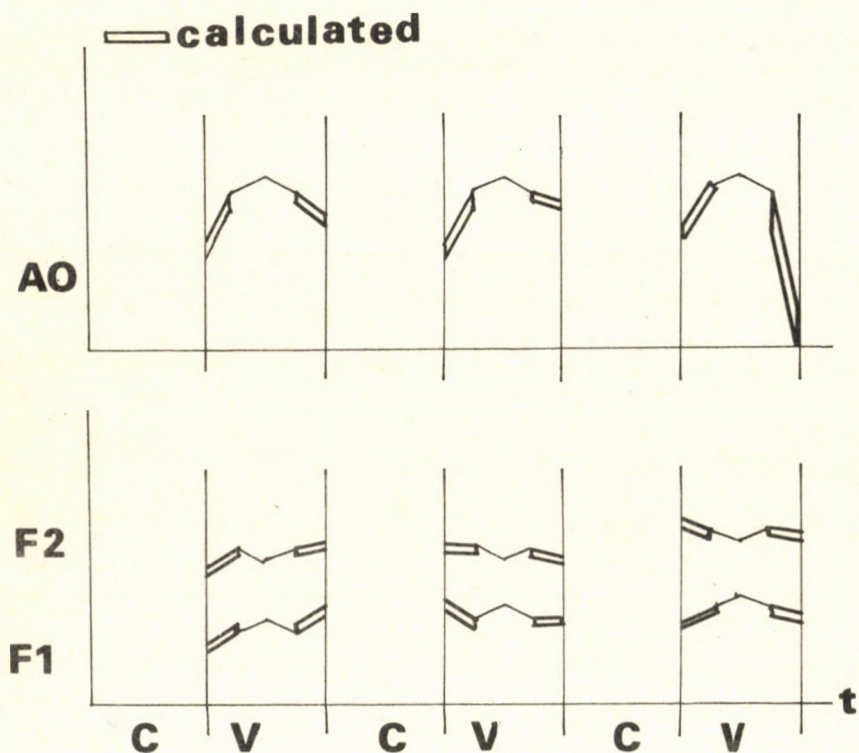


Fig. 12. The vowel transitions constructed

"Hara'ma" 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.

POLISH PALATALIZATION AND PHARYNGEALIZATION IN AN
INTERLINGUAL COMPARISON

Éva Földi

Phonetics Department, Eötvös Loránd University

ABSTRACT

This paper examines the articulatory and acoustic realization of palatalization and pharyngealization on Polish material and in an interlingual comparison (with Hungarian and Russian). The analysis is primarily based on cineradiographic and spectrographic data and has been carried out with the aid of a computer.

INTRODUCTION

Various views have been put forward on the opposition of 'soft' and 'hard' consonants, a phenomenon existing in several Slavonic languages and having considerable effects on the sound and phoneme systems concerned. Opinions differ especially with regard to the phonetic and phonological status of palatalized sounds and phonemes. The articulatory processes of palatalization and pharyngealization and their acoustic correlates have not yet been clearly defined in Polish studies. The most controversial issues include the synchronic vs. asynchronic character of palatalization and

the degree of palatality, whereas pharyngealization is generally not analysed in detail at all. An especially heated debate surrounds the problem of the soft bilabial consonants [b', p', m']: several scholars deny the existence of these sounds, respectively phonemes, and consider their articulation physiologically impossible (Jassem 1974; Rocławski 1976 and 1984; Koneczna 1965), while others take the opposite view (Baudouin de Courtenay 1922; Szober 1931; Wierzchowska 1971 and 1980).

In this paper I shall analyse the articulatory and acoustic realization of palatalization and pharyngealization in the total articulatory process of the sounds concerned. The material of the examination consisted of Polish words containing the consonants in question in word initial or word internal position. I compiled the corpus so that all manners of articulation (plosives, fricatives, affricates, nasals, liquids, and rolled sounds) are represented in it (e.g. <pasek -- piasek -- pili, wara -- wiara -- widmo, dżonka -- dzionka, mała -- miała -- miś, lato -- list, rym -- ring> etc.).

The articulatory processes were analysed by cineradiographic methods, the acoustic features by spectrographic ones, and both procedures were supplemented by examinations by computer. The cineradiographic recordings were made by a Siemens Sirescope 2 radiographic apparatus and a Siemens Sirecord S video recorder. The spectrograms were made by a 700 type spectrograph, and the computerized examinations by a Commodore 64 personal computer. The results obtained from

these analyses were compared with other -- typologically similar (Russian) or different (Hungarian) -- languages which were examined in the same way on the basis of Kalman Bolla's conception within the framework of an interlingual phonetic research project (Bolla 1982; Bolla--Földi 1981; Bolla--Földi--Kincses 1986). Nineteen native Polish speakers whose pronunciation met the norms of educated Standard Polish took part in the experiment. The cineradiograms and spectrograms reproduced in this paper are based on the pronunciation of a male speaker identified as "St".

DISCUSSION

The Polish language, similarly to Russian, is of a consonantal character. This means -- among other things -- that the quality of the surrounding consonants does not depend on the vowels, but -- on the contrary -- the quality of the consonant determines the type of the vowel that can co-occur with it. For example, a palatal vowel cannot follow a pharyngeal consonant (cf. 1. [bɛwɪ] -- 2. [b'il'i]). This condition, as can be seen from the above example, has linguistic relevance as well (1. '/they/ were', 2. '/they/ struck'). It is to be noted further that 1. there are only monophthongal vowels in Standard Polish; 2. it is more precise to speak of pharyngealization than of velarization since the movement of the tongue in the course of articulation is more significant in the direction of the pharynx than in that of the velum.

Taking the above-mentioned facts into consideration, I

tried to answer the following questions:

1. Is there a synchronic or asynchronic palatalization in Polish, with special regard to the bilabial soft plosives and nasals. How is the articulatory process reflected in the acoustic result?

1. The analysis of the video recordings consisted of the following steps:

- a) the delimitation of the sound under examination within the word, and the measuring of its duration;
- b) the division of the sound into five frames on the basis of its duration -- whereby the whole process of sound formation can be well observed;
- c) the representation of the individual frames on sketches taken from the screen;
- d) the indication of the measurement and reference points on the sketches;
- e) the measurement of the relevant data, and their processing according to the purposes of the examination.

(For details about the cineradiographic methods applied in the interlingual phonetic research project referred to above, and the way the resulting data are processed by computer, see Bolla--Földi--Kincses MFF 15. 1986. 155--65.)

The examination of the Polish 'hard' consonants revealed the following:

-- The process of pharyngealization can be well observed on the cineradiographic recordings throughout the time of articulation: the tongue gradually moves in the direction of

the pharynx and, consequently, the surface of the pharyngeal cavity reduces and that of the palatal cavity increases (see Fig. 1, 4, 7). This observation is also supported by the data obtained from the examinations of the surface or rather surfaces of the supraglottal cavities by computer (see Fig. 1/b, 4/b, 7/b). The supraglottal cavities were divided into five regions -- labial, palatal, velar, pharyngeal and nasal -- and the relative surface data, expressed in percentage values, were examined in comparison with the total surface of the supraglottal cavities and with each other.

The cineradiographic recordings prove that the Polish 'hard' consonants -- similarly to the Russian ones -- are of a pharyngeal articulation.

-- To examine the synchronic vs. asynchronic character of palatalization, I selected a sample of words in which the consonant in question occurred not only before [i], but also before a different vowel (e.g. <pili -- piasek, miś -- miara, kicz -- kiedy> etc.).

This phonetic position -- soft consonant + a vowel different from [i] -- is the most debated, especially in the case of bilabial soft plosives and nasals. The researchers who claim that there are no [b', p', m'] sounds or phonemes in Polish usually refer to this position, among other things, and hold the opinion that, for example, the word <piasek> contains a [pja] sound combination where a variant of the phoneme /j/ occurs after the hard, rather than soft, consonant. On the other hand, those who accept the existence of [b', p', m'],

still differ in the question of the synchronic vs. asynchronic character of palatalization. For example, according to the authors of the Polish pronouncing dictionary, synchronic palatalization can occur only before [i] and other front vowels but asynchronic palatalization takes place before back vowels.

A thorough observation and analysis of the cineradiographic recordings, following the process of sound formation shot by shot (1 shot = 20 ms), proves the existence of the Polish soft consonants, including [b', p', m']. It has also been revealed by the data that the articulation of the Polish soft consonants involves synchronic palatalization, even in the case of the much-debated bilabial soft plosives and nasals (see Fig. 2, 3, 5, 6), independently of the quality of the vowel that follows the consonant in question. It can be well observed on the recordings that the articulatory configuration characteristic of palatalized articulation takes shape already in the initial phase of sound formation. This holds true even for voiceless bilabial soft plosives -- the tongue takes up a palatalized articulatory position at the beginning of the articulation, i.e. in the silent phase. The results of the cineradiographic analyses were confirmed by the data obtained from the examinations of the articulatory processes of the supraglottal cavities by computer (see Fig. 2/b, 3/b, 5/b, 6/b, 8/b, 9/b): a significant decrease of the palatal surface and a considerable increase of the pharyngeal surface can be observed for every manner of articulation (plosives,

fricatives, etc.).

The data obtained from the cineradiographic analysis of the Polish material examined have proved the existence of synchronic palatalization in every case.

2. The cineradiographic analysis was supplemented by a spectrographic examination.

The following problem arose in the course of the analysis of the spectrograms, during the segmentation of the sound sequences: after bilabial soft plosives -- and other plosives of a different place of articulation -- if the following vowel is different from [i], a 40--80 ms [i̯]-like segment, whose proper status is debated, can be seen, between the consonant in question and the following vowel. This [i̯] segment is considered by some scholars to be a variant of the phoneme /j/ which occurs after a hard (pharyngeal) consonant. This opinion is disputable, among other things, since no trace of this [i̯]-type sound can be seen on the cineradiographic recordings.

I have also analysed the temporal structure of sounds on the spectrograms and obtained the following data concerning e.g. [b, b', p, p'] in the pronunciation of a male informant identified as "St":

- the total length of [b] was 135 ms, of which plosion took 15 ms;
- the total length of [b'] before [a], a vowel different from [i], was 170 ms, of which plosion took 20--25 ms, and the [i̯] segment 65 ms;
- the total length of [b'] before [i] was 140 ms, of

which plosion took 25 ms;

- the total length of [p] was 120 ms, of which plosion took about 10--15 ms;
- the total length of [p'] -- also before [a] -- was 160 ms, of which the silent phase took 80--100 ms, the plosion 20--25 ms, and the [i] segment 60 ms; --
- the total length of [p'] before [i] was 130 ms, of which plosion took 20 ms.

The above duration data show that the post-consonantal [i] can hardly be considered to be an independent sound, i.e. it cannot be one of the realizations of the phoneme /j/. It cannot be an element of a vowel diphthong either, since there are no vowel diphthongs in Standard Polish.

The conclusion that can be drawn from a comparison of the data of the cineradiographic and spectrographic analyses is that the [i] segment appearing on the spectrogram is the acoustic reflex of the palatalized articulation of certain consonants -- in particular, plosives or nasals -- since no trace of a corresponding segment can be observed on the cineradiographic recordings.

The above conclusion is supported by certain data of the historical development of Polish (see e.g. Baudouin de Courtenay 1922; Koneczna 1965), and it is also backed up by the rules of Polish orthography: the grapheme *i* is used to denote the soft consonants if they are followed by a vowel different from [i] (e.g. <piasek, kiedy, miód> etc.) and it has to be in the same written syllable with the preceding consonant, i.e. they cannot be divided from each other (e.g.

<ma-la-ria> rather than <ma-la-ri-a> or <ma-lar-ia>).

The analyses made with cineradiographic and spectrographic methods and completed with examinations by computer prove that the sounds [b', p', m'] are the realizations of the phonemes /b', p', m'/ and not the variants of /b, p, m/ before the phoneme /j/ in the Polish language.

The diagrams of Hungarian material reflect the absence of the opposition of palatalization/pharyngealization in that language. On the other hand, Polish palatalization/pharyngealization has not proved to be different in comparison with Russian, as far as synchronicity is concerned. This has to be pointed out since, according to some scholars' opinion, in Polish -- as opposed to Russian -- a so-called 'delayed' palatalized articulation takes place, especially in the case of bilabial consonants (cf. Awdiejew 1982, 47).

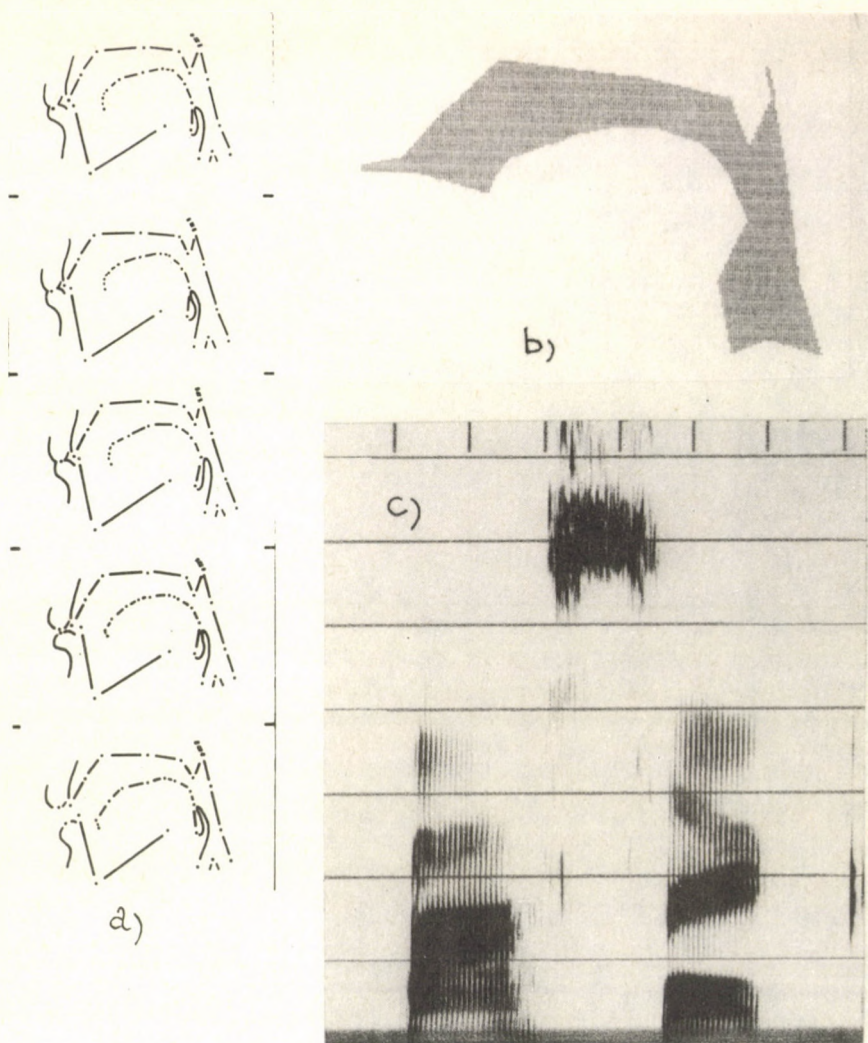


Fig. 1. a) The cineradiogram of [p] as in <pasek>
 b) The surface of the supraglottal cavities
 in the pure phase of the articulation of
 [p] as in <pasek>
 c) The spectrogram of the word <pasek>

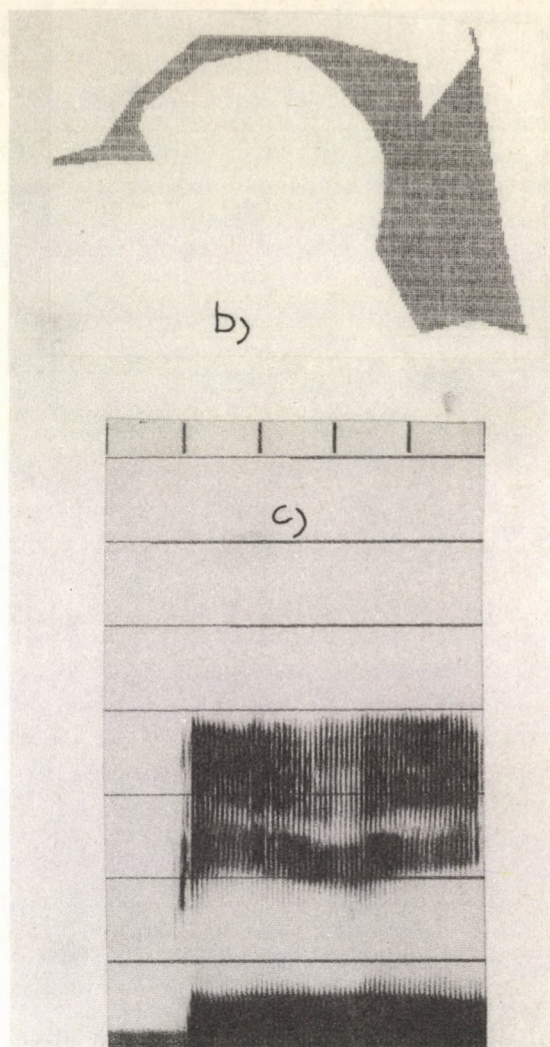
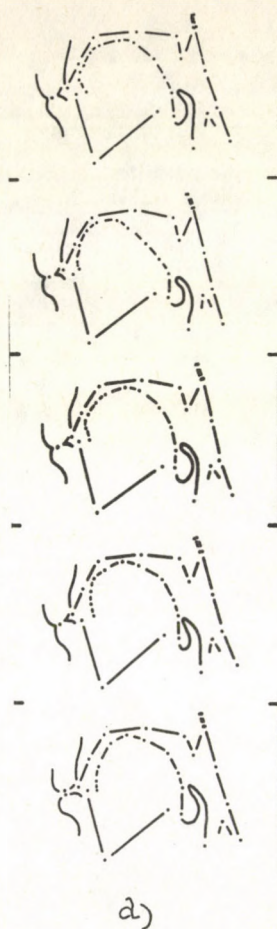


Fig. 2. a) The cineradiogram series of [p'] as <pili>
 b) The surface of the supraglottal cavities in
 the pure phase of the articulation of [p'] as
 in <pili>
 c) The spectrogram of the word <pili>

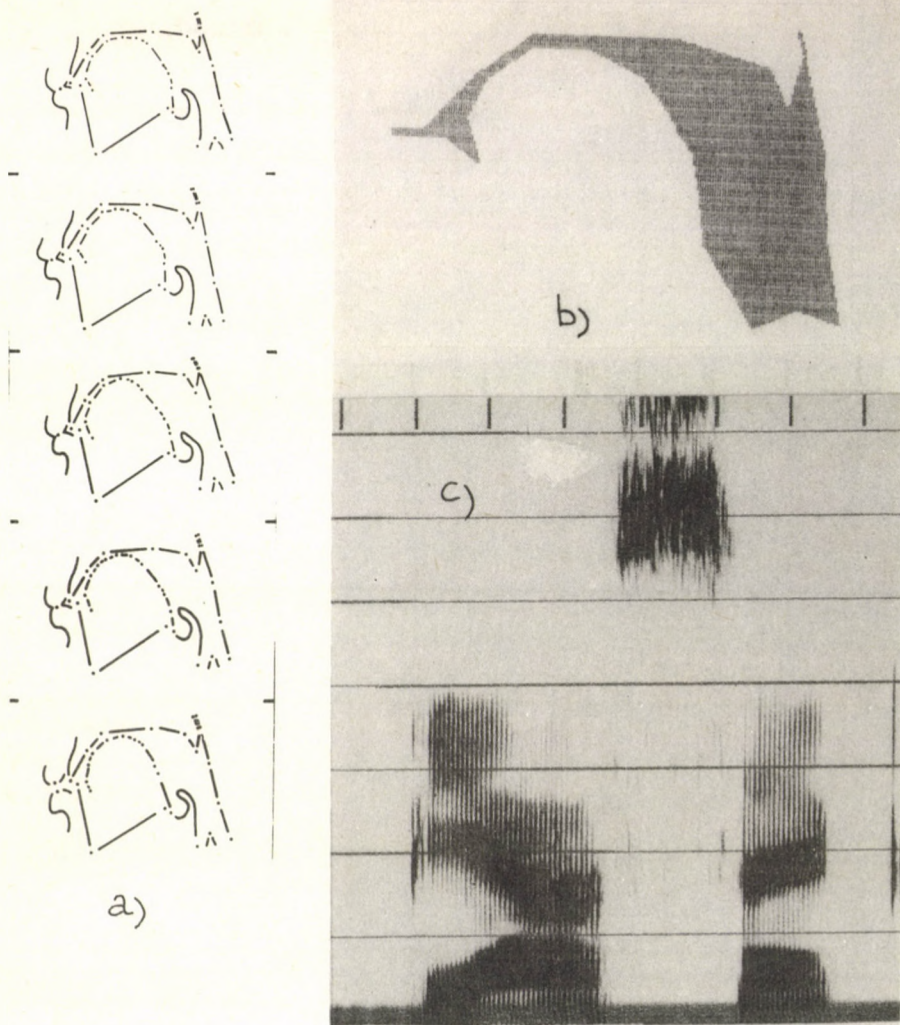


Fig. 3. a) The cineradiogram series of [p'] as in <piasek>
 b) The surface of the supraglottal cavities in the pure phase of the articulation of [p'] as in <piasek>
 c) The spectrogram of the word <piasek>

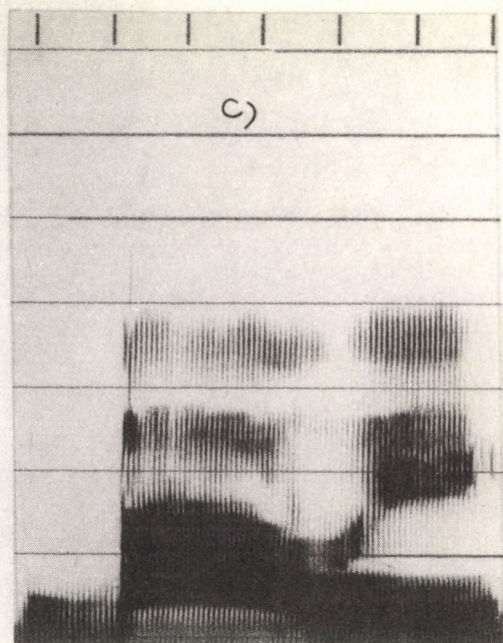
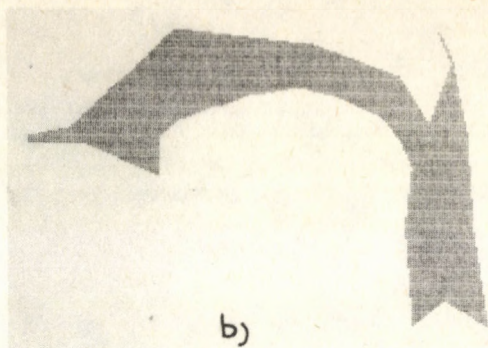
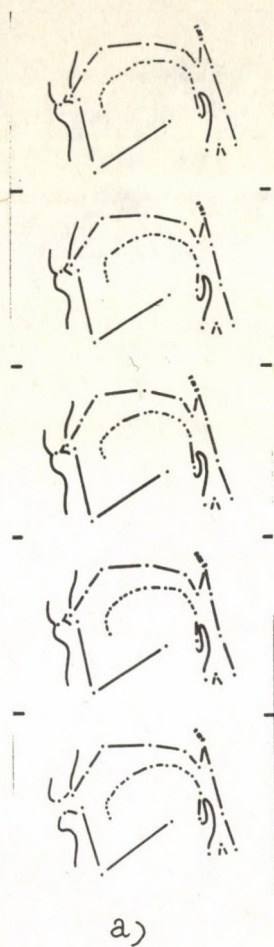


Fig. 4. a) The cineradiogram series of [b] as in <bały>

b) The surface of the supraglottal cavities in the pure phase of the articulation of [b] as in <bały>

c) The spectrogram of the word <bały>

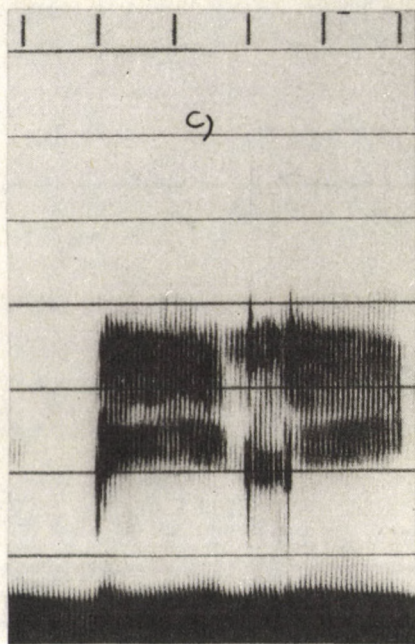
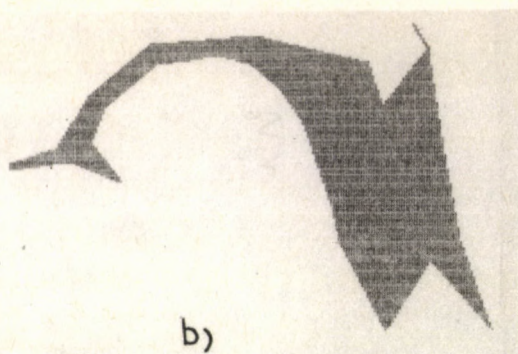
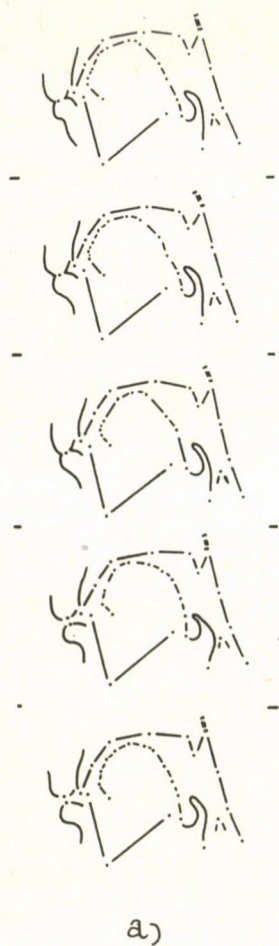


Fig. 5. a) The cineradiogram series of [b'] as in <bili>
 b) The surface of the supraglottal cavities in the
 pure phase of the articulation of [b'] as in <pili>
 c) The spectrogram of the word <bili>

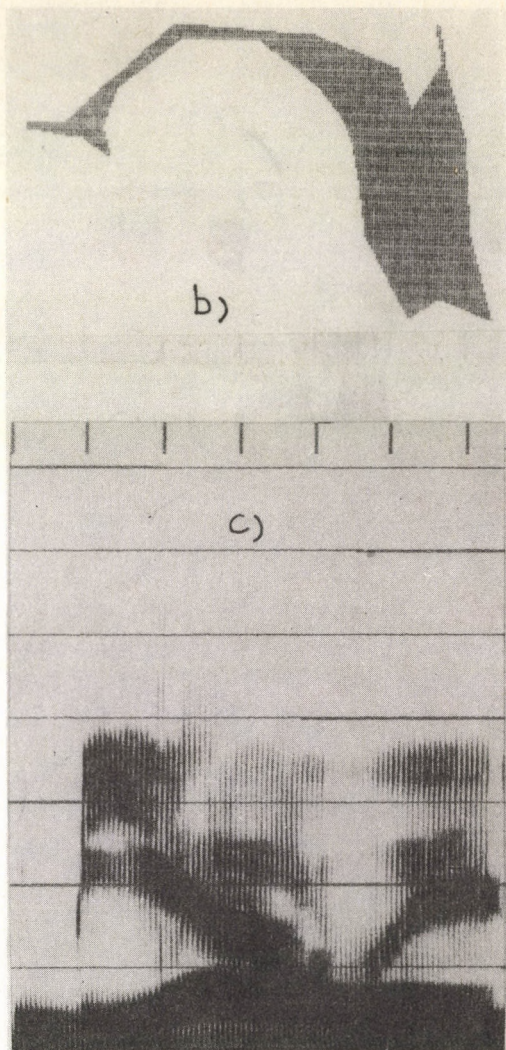
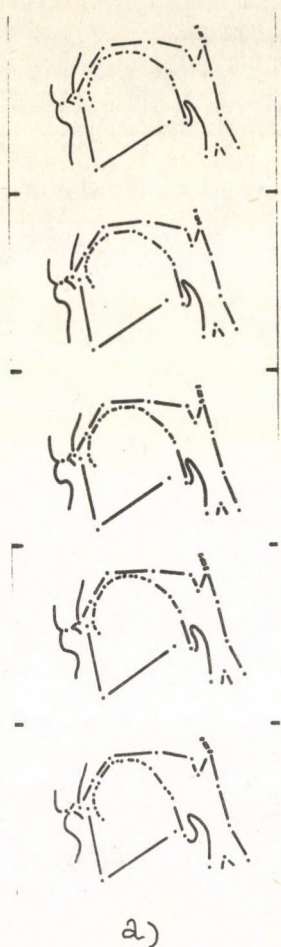


Fig. 6. a) The cineradiogram series of [b'] as in <bialy>
 b) The surface of the supraglottal cavities in the pure phase of the articulation of [b'] as in <bialy>
 c) The spectrogram of the word <bialy>

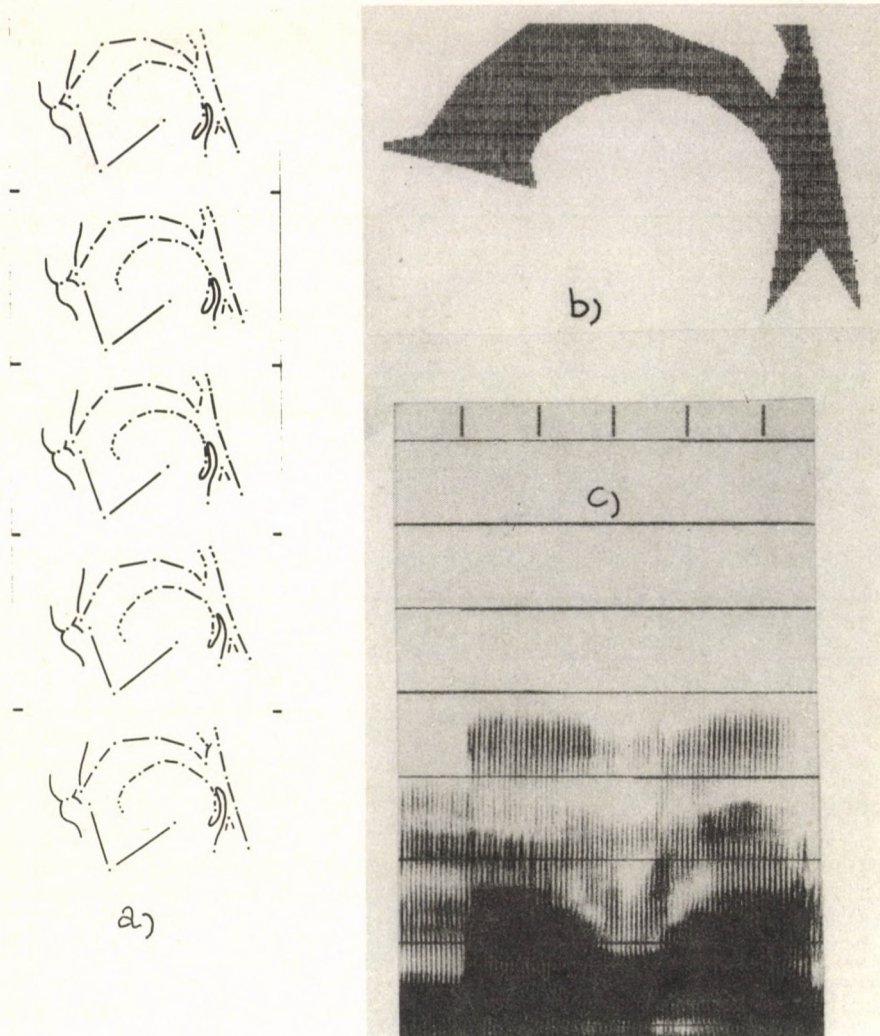


Fig. 7. a) The cineradiogram series of [m] as in <mała>
 b) The surface of the supraglottal cavities in the pure phase of the articulation of [m] as in <mała>
 c) The spectrogram of the word <mała>

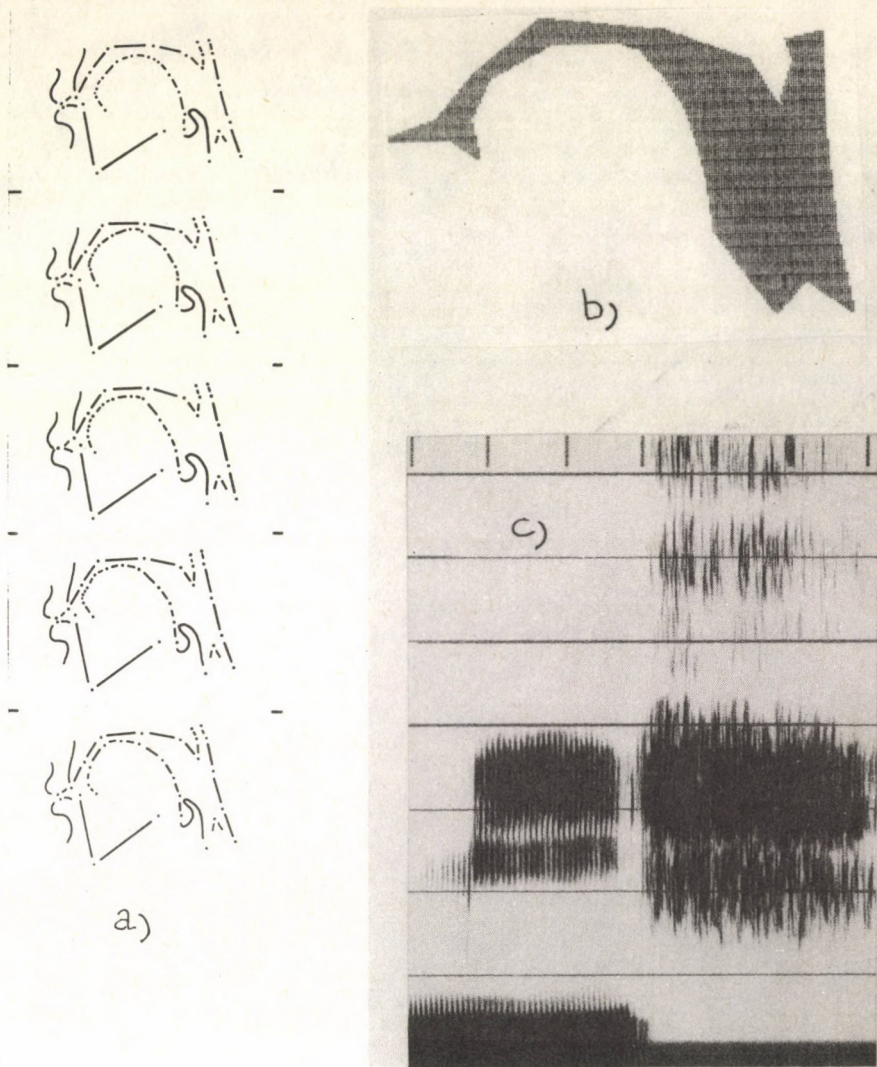


Fig. 8. a) The cineradiogram series of [m'] as in <miś>

b) The surface of the supraglottal cavities in the pure phase of the articulation of [m'] as in <miś>

c) The spectrogram of the word <miś>

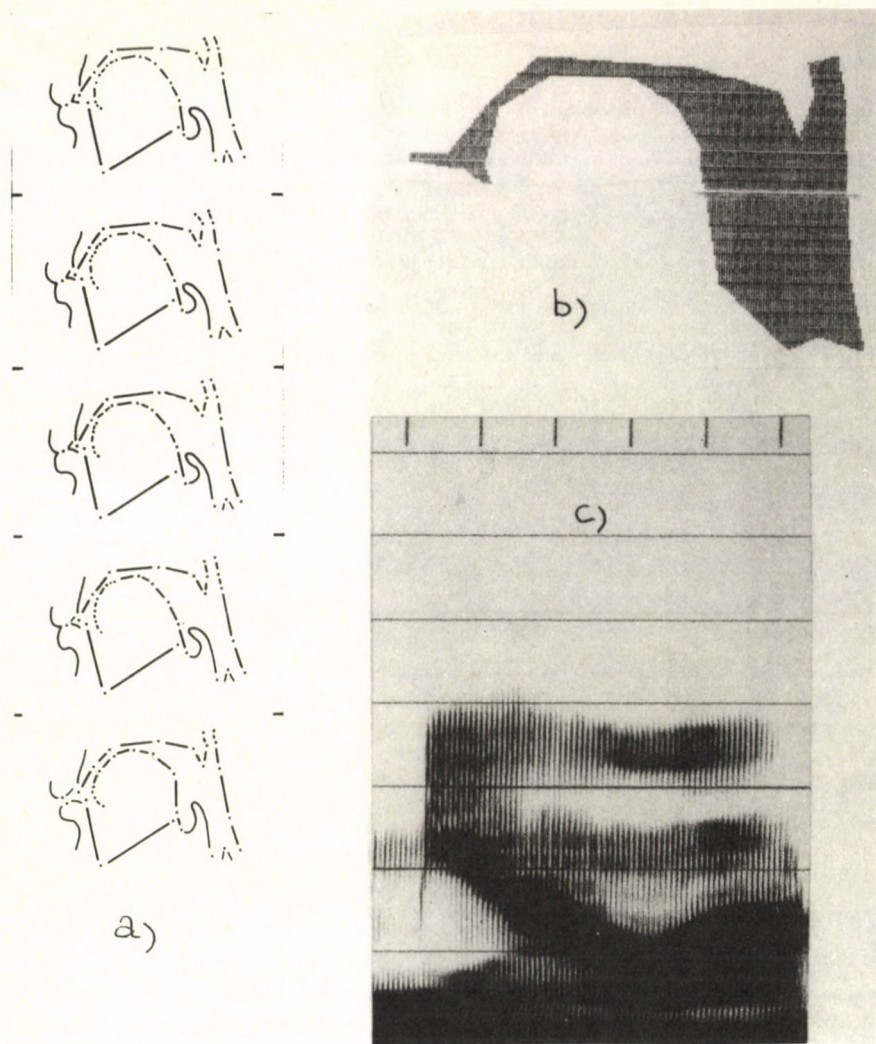


Fig. 9. a) The cineradiogram series of [m'] as in <miała>

b) The surface of the supraglottal cavities in the pure phase of the articulation of [m'] as in <miała>

c) The spectrogram of the word <miała>

CONCLUSION

The joint analyses of the articulatory process and the acoustic result unanimously prove that palatalization covers the total time of articulation even in the case of the much-debated bilabial plosives and nasals. Though the acoustic pattern as reflected on the spectrograms does not equally sharply express the presence of palatalization during the whole length of the sound, the cineradiographic recordings, and the computer-processed data taken from those recordings, prove that beyond any reasonable doubt.

REFERENCES

- AWDIEJEW, A.: Niektóre problemy fonetyki języka polskiego w świetle badań kontrastywnych. In: 2 zagadnień fonetyki i fonologii współczesnego języka polskiego. Red. GOSIENIECKA, J. Toruń, 1982, 44--8.
- BAUDOUIN DE COURTENAY, J.: Zarys historii języka polskiego. Warszawa 1922.
- BOLLA Kálmán: Drosz hangalbum. MFF 11. 1982.
- BOLLA Kálmán--FÖLDI Éva: A lengyel beszédhangok képzési és akusztikus sajátosságairól. MFF 7. 1981. 91--139.
- BOLLA Kálmán--FÖLDI Éva--KINCSES Gyula: A toldalékcsoő artikulációs folyamatainak számítógépes vizsgálata. MFF 15.1986, 155--65.
- JASSEM, W.: Mowa a nauka a łączności. Warszawa 1974.
- KONECZNA, H.: Charakterystyka fonetyczna języka polskiego na tle innych języków słowiańskich. Warszawa 1965.

RODŁAWSKI, B.: Istota miękkości głosek. JP LVI/1. 1976,
26--36.

RODŁAWSKI, B.: Palatalność. Teoria i praktyka. Gdańsk 1984.

SZOBER, S.: Gramatyka języka polskiego. Warszawa 1931.

WIERZCHOWSKA, B.: Wymowa polska. Warszawa 1971.

WIERZCHOWSKA, B.: Fonetyka i fonologia języka polskiego.
Warszawa 1980.

HIGH FREQUENCY SPEECH PERCEPTION: PHONETIC ASPECTS AND APPLICATION

Mária Gósy

Linguistics Institute, Hungarian Academy of Sciences

INTRODUCTION

The so-called speech frequencies (100--3000 Hz) seem to be both necessary and sufficient for the perception of the vowels and consonants. The acoustic information in this frequency range is generally suitable for understanding running speech. (An English sentence can be understood about 90% correctly when appearing acoustically in a 1000 Hz wide frequency band with 1500 Hz middle frequency, see Denes--Pinson 1973, 185.)

However, a lot of comprehension problems arise if only these frequencies can be used. This can be demonstrated with the telephone where general conversation can easily be carried out without any problems in understanding. However, identification of names or comprehension of suddenly changed topic of dialogue can cause difficulty. It is known that people with hearing loss at high frequencies (above 3000 Hz) suffer from perceptual and understading difficulty. There is no doubt that the first two energy maximums, the formants, contain the main information for the identification of vowels

and certain consonants. Moreover, components of some other consonants -- like [s] or [ts] -- occurring below 3000 Hz are sufficient for their identification. The role of high frequencies (above 3000 Hz) in perception, however, has been little investigated (Fant 1973). The acoustic information contained in the high frequencies in question may be purely supplementary; alternatively it may play an independent and special role in perception. To bring this problem a little closer to a solution, experiments were carried out with Hungarian-speaking native listeners.

METHOD AND MATERIAL

The material used consisted of (i) 25 sound-sequences without meaning and (ii) 102 monosyllabic, phonetically balanced Hungarian words. The bisyllabic sound-sequences contain almost all Hungarian speech sounds. The acoustic structure of part of them corresponds to Hungarian phonotactic rules while that of another part of them contradicts them. All the words consist of three sounds: a vowel between two consonants. The words range from well-known ones, in everyday use, to ones very rarely used. They belong to different grammatical categories. Attempts were made to choose both the sound-sequences and words containing consonants and vowels in different phonetic positions and in different environments. The speech material was recorded by a male announcer who pronounced it as isolated statements in random order (see Tables 1 and 2). The recording was made with a professional tape recorder and microphone under

laboratory conditions. An 8 ms pause was left between the sound-sequences/words. The intensity level of sound-sequences and the words varied within ± 6 dB. Two types of filtration method were used for testing: pass-band and high-pass filtering by an Audio Filter. The filter slope was always 36 dB/octave. The cut-off-frequencies were

Table 1.

Words

Hungarian English Hungarian English Hungarian English

gáz	gas	híd	bridge	fáj	ache
menny	sky	gyár	factory	zsúr	tea-party
zab	oat	síp	whistle	cél	aim
szín	colour	nem	no	gőz	vapour
rög	clod	táj	scenery	csepp	drop
lom	odds	vér	blood	rét	meadow
cikk	article	busz	bus	ház	house
néz	looks	pác	pickle	függ	hangs
kád	bath	kör	circle	kacs	tendrill
gúzs	withy	gém	heron	lép	steps
bír	has	gyűr	crumples	sár	mud
tök	squash	dög	carrion	pötty	dot
szán	sledge	mér	measures	jár	walks
zseb	pocket	csík	stripe	gyík	lizard
jód	iodine	nyáj	flock	bök	pokes
sün	hedgehog	rom	ruin	sín	rail

rúg	kicks	kéz	kez	zár	lock
kép	picture	fej	head	sör	beer
les	watches	nagy	big	száj	mouth
dob	drum	bőr	skin	rügy	bud
máz	glaze	rím	rhyme	zöm	bulk
tér	square	csekk	check	zsír	fat
von	pulls	szőr	hair	sor	line
seb	hurt	fog	tooth	bal	left
vár	waits	zúz	crushes	rák	crab
sakk	chess	láb	leg	fal	wall
szem	eye	csók	kiss	zsák	sack
víz	water	dér	hoar-frost	csak	only
pék	baker	kín	pain	hűt	cools
nád	reed	mos	washes	nyír	birch
sokk	shock	bál	ball	hegy	hill
tej	milk	tór	dagger	díj	prize
cár	tsar	cím	address	pók	spider
jég	ice	géz	gauze	vád	charge

Table 2.

Sound-sequences without meaning

<u>writing</u>	<u>API-form</u>	<u>writing</u>	<u>API-form</u>
1. sikág	ʃika:g	14. natup	notup
2. vokus	vokuj	15. zsegém	ʒege:m
3. kete	kete	16. macáf	motʃa:f
4. türáp	tyra:p	17. tipa	tipɔ
5. szajuk	sɔjuk	18. bodon	bodon
6. celnü	tʃelnɥ	19. fücesz	fytʃes
7. bató	bɔto:	20. cátany	tʃa:toɲ
8. lüse	lyʃe	21. sumpös	ʃumpɔʃ
9. rivem	rivem	22. deféz	defe:z
10. furáz	fura:z	23. zipod	zipod
11. siszud	ʃisud	24. kűvecs	kyvetʃ
12. tádo	ta:do:	25. pársza	pa:rsɔ
13. csinza	tʃinzo		

Table 3.

Cut-off-frequencies of filtration (Hz)	Correct identification in %	
	words	sound-sequences
2200 h.p.*	67	49
2200--2700 p.b.	98	78
2700 h.p.	72,5	31
2700--3300 p.b.	95	75
3300 h.p.	74	
3300--3900 p.b.	95	
3900 h.p.	46	
3900--4700 p.b..	95	

*The abbreviations mean high-pass and pass-band filtration.

for words 2200 Hz, 2700 Hz, 3900 Hz and 2200--2700 Hz, 2700--3300 Hz, 3900--4700 Hz; for sound-sequences 2200 Hz, 2700 Hz, and 2200--2700 Hz, 2700--3300 Hz. These values were chosen in view of the fact that the highest acoustic cue for Hungarian vowels appears in general to be about 2200 Hz; it is the second formant for the [i:] sounds. There are 8 different materials for the words and 4 for sound-sequences. In order to examine the role of the upper frequencies, those below 2200 Hz were removed. The frequency response of the Audio Filter used is shown in Figure 1.

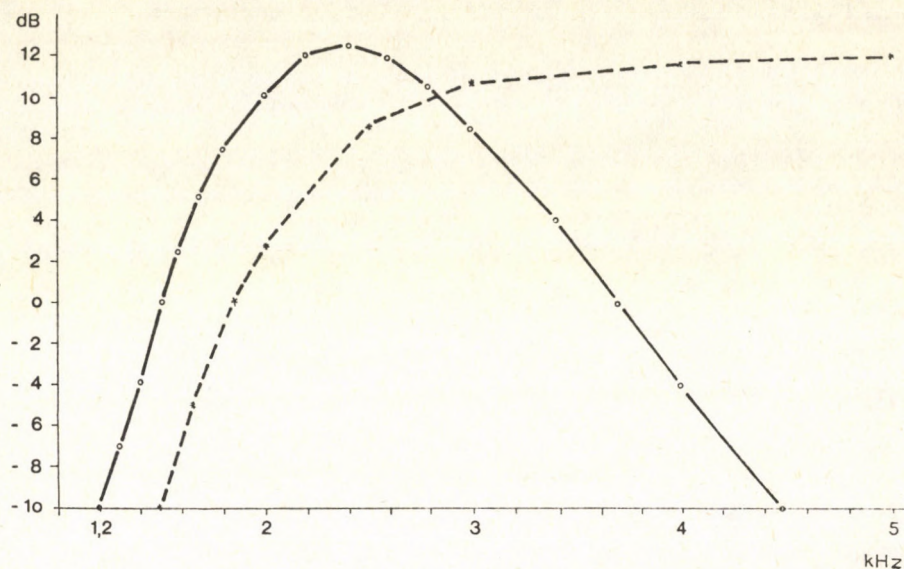


Fig. 1. Frequency responses of the Audio Filter used (with cut-off-frequencies 2200 for high-pass and 2200-2700 Hz for pass-band filtering)

Spectrographic analyses were made of filtered sound-sequences and words by the Sound Spectrograph (Type 700 of Voice Identification). Each of the 12 test materials was administered to 10 adult normal-hearing subjects, totally 120 subjects, half of them females and half males. The experiments were conducted in a silent room through a professional loudspeaker. The listeners' task was to write down the sound-sequences or words they could perceive/understand. In order to obtain statistically

significant results, we used our own Psychotest program.

RESULTS AND DISCUSSION

The experimental data for sounds-sequences and words are summarized in Table 3. These show that (i) the perception/understanding of sound-sequences/words was better under pass-band filtering than under high-pass filtering; (ii) perception/understanding decreased under high-pass filtering according to the change of the cut-off-frequency; (iii) a frequency band seems to occur with the highest perception/understanding ratio: 2200--2700 Hz. The differences between the filtered groups proved to be significant at the .01 level.

These results led us to the conclusion that there are frequency bands in which more acoustic information about the same word/sound-sequence seems to be disturbing to the decoding processes (see Rosen and Fourcin, 1983). The supposed idea is that the upper part of the acoustic structure of certain speech sounds does not remain characteristic for them when the lower part is lost. In other words: these high frequencies do not contain unambiguous information about the sounds or cannot be acoustic cues used for identification. The components appearing at these frequencies have been thought to play a supplementary role in recognition. Results obtained from examinations using the low-pass filtration method confirm this (see Gosy 1986). If this were the case, the high elements would have been

redundant (Pisoni 1981, 255). Our new results have not confirmed this assumption, and, indeed, they seem to contradict it. The data have supported an alternative hypothesis, namely that certain speech sounds and sound combinations have special 'cue-like' components above 2000 Hz. This 'secondary-cue' hypothesis was further investigated by means of spectrographic analyses. These showed that, as expected, the main difference in acoustic structure between pass-band and high-pass filtered sound-sequences (with and without meaning) lies in the presence or absence of the higher frequencies.

By way of illustration let us look at the bilabial nasal consonant [m]. The original acoustic structure of [m] contains acoustic cues at about 500 and 1500 Hz (F1 and F2). In the absence of these frequencies it is not possible to identify without elements above 2000 Hz. Spectrographic analysis of [m] shows further components at about 2800 and 3700 Hz (Fn, Fn+1). The word <mos> ('washes') was understood accurately when frequencies below 2000 Hz were removed by filtration. When the component at 2800 Hz was reduced in intensity by further filtration, however, identification of the consonant became impossible (see Figure 2).

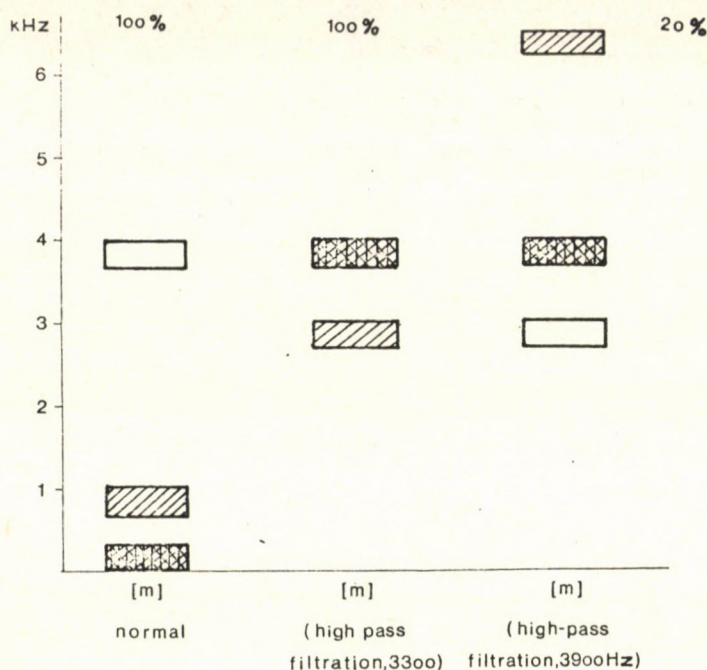


Fig. 2. Consonant [m] in word <mos 'washes'> and its identification in %

One of the questions to be asked in this respect is: why can we not use the whole information of the high frequencies in perception, why do they seem to cause difficulty? Moreover, why do these perceptual problems disappear when there are only frequency bands? (cf. the differences of the answers for sound-sequences after high-pass and pass-band filtering: 49% and 78%). This suggests that, in contrast to the main acoustic cues (below 2200 Hz), the secondary cues act alone and independently of the disturbing higher components. As to the explanation of 'disturbing higher components' let us document it with an example. The perception of the Hungarian long [a:] vowel was analyzed. The correct identification of

this sound in the sound-sequence <tádó> ([ta:do:]) after high-pass filtering is 0% and after pass-band filtering is 100% (with the cut-off-frequencies of 2200 Hz and 2200--2700 Hz). The false responses in the first case were: [tøldu, tødu, tɔdu, tøldo:]. Instead of the [a:] dominantly [ø] was perceived. Identification of this [a:] in isolation after the same high-pass filtering shows the same responses: 80% [ø] and 20% [a:] (Gósy 1986, 33). Figure 3 shows the difference in acoustic structure of this [a:] after the two types of filtering. As can be seen there are components between 2000 and 4000 Hz with very different intensities. This is assumed to cause the perceptual differences.

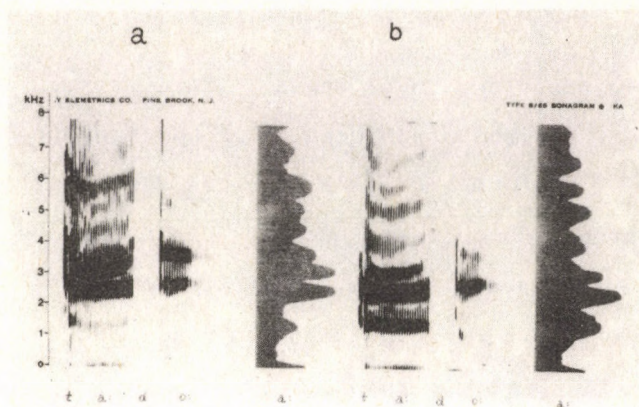


Fig. 3. Differences in acoustic structure of [a:] after high-pass (a) and pass-band filtering (b)

We made an effort to find correlations between perceptual data and spectrographic measurements. What criteria should the high frequency components fulfil in order to act as acoustic cues? (i) Identification should reach a significant level and, (ii) frequency values should be defined for correct perception. On the basis of our data it can be supposed that the components appearing in certain frequency bands correspond to the above-mentioned expectations.

The identification of speech sounds in words was analyzed, the results are summarized for vowels in Table 4 and for consonants in Table 5.

Table 4.

Vowels in words	Num- ber of words	Correct identification in % after high-pass filtration				Total
		2200Hz	2700 Hz	3300 Hz	3900 Hz	
a:	22	68.6	63.1	56.8	34.0	55.6
ɔ	7	82.8	80.0	77.0	35.7	68.8
o/o:	11	62.7	76.3	56.3	41.8	59.2
u/u:	5	68.0	84.0	68.0	40.0	65.0
e	11	74.5	63.6	74.5	51.8	66.0
e:	14	72.8	80.0	77.0	60.0	72.4
i/i:	15	55.3	76.0	64.6	60.6	64.1
o/o:	12	56.6	67.5	46.6	33.3	51.0
y/y:	5	72.0	86.0	78.0	54.0	72.5
Total	102	68.0	75.2	66.5	45.7	63.8

Table 5.

Conson- ants in words	Correct identification in % after high-pass filtration				
	2200 Hz	2700 Hz	3300 Hz	3900 Hz	Total
	CU/UC	CU/UC	CU/UC	CU/UC	CU/UC
b	58.3/76	78.6/62	68.3/62	63.3/48	66.6/62
p	56.6/85	47/8	40.7/90	20/52.5	40.9/76.8
d	45/56	77.5/70	65/58	52.5/47.5	60/57.8
t	92.5/80	82.5/95	77.5/85	62.5/40	78.7/75
g	25/55	12.5/86.6	17.5/50	5/36.6	15/57
k	63.3/59	58/74	56.6/42.8	56.6/42	58.6/54
ʒ	66.6/70	86/96.6	43.3/90	56.6/86.6	63/85.8
v	50/	82/	70/	58/	65/
f	85/	75/	97.5/	60/	79/
z	67.5/76.6	86/71	60/64	55/72	67/70.9
s	68/	58/	62/	36/	56/
ʃ	65/	75/	75/	54/	65/
ʒ	83/100	77.7/85	63.3/90	51/45	68.7/80
h	65/	85/	77.5/	60	71.8
j	70/83.3	83/73.3	66.6/83.3	60/58.3	70/74.5
ts	65/	80/	65/	50/	65/
tʃ	57.5/	67.5/	67.6/	60/	63/
l	75/70	80/75	65/70	45/42.5	66/64.3
r	50/68	82.8/77.5	47/71	25.7/44.5	51/65
m	72.5/45	80/65	80/41.2	35/28.7	66.8/45

n	92.5/50	100/48.3	90/53.3	77.5/43.3	90/48.7
ŋ	85/	80/	85/	60/	77.5/

	66.2/69.5	74/75.6	65.4/67.9	49.7/49	63.8/65.5
--	-----------	---------	-----------	---------	-----------

It seems that the vowels having the first two formants normally at the middle frequencies (at about 400--500 Hz and 1400--1500 Hz) are perceived wrongly. In case of some consonants there was no occurrence in word final position in our material. The majority of correct responses occurs between 60--70%. Four consonants remained below 50% as to their identification: [p, g] in CV and [m, n] in VC positions. Perception of [g] in VC shows an extremely low value (15%), which is the worst among the consonants.

Misperceptions of one or two sounds in the word cause incorrect understanding. For example, there are frequent misperceptions of [a:] in a similar fashion to the sound-sequences. The spectrographic analyses show an important difference depending upon the context (sound combination). The [a:] mentioned above was perceived more correctly when it occurred between fricatives or fricatives and nasal consonants (Figure 4). This can be explained by the transition phases which act as acoustic cues in this respect.

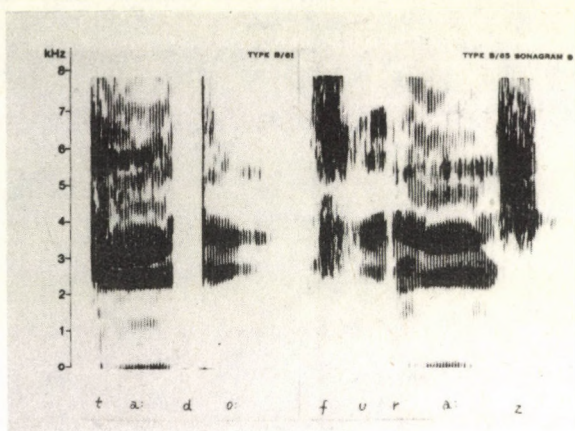


Fig. 4. Transition phases acting as acoustic cues for identification of [a:] between the vowel and the spirant in the sound-sequence <furaz> [fura:z]

Finally the role of meaning should be taken into consideration. Table 6 shows the percentage values for the correct identification of the test words. The words regarded as relatively frequent are underlined.

Table 6.

Correct identification Monosyllabic Hungarian words
of words in %

91--100	vér <u>busz</u> <u>víz</u> sín <u>kéz</u> dob fog <u>nagy</u> <u>néz</u> kép
81--90	<u>fej</u> gyűr máz nád táj nyáj <u>száj</u> zúz sün híd sor síp <u>zsír</u> tér jég mér cél les
71--80	zsák zsúr cikk csík tör pötty <u>csók</u> <u>mos</u> zab kacs <u>fal</u> rét lép <u>tej</u> <u>nem</u> csekk hegy rügy <u>szem</u>
61--70	hűt <u>gyár</u> pác bír <u>cím</u> gyík nyír díj dér tök jód bőr szőr bök sakk <u>bal</u>
51--60	<u>vár</u> <u>láb</u> <u>ház</u> sár <u>jár</u> zár lom sokk <u>sőr</u> csepp
41--50	cár rák rúg kör pék <u>seb</u> függ
31--40	bál <u>fáj</u> vád szín rög győz zöm von rom <u>csak</u> zseb
21--30	<u>gáz</u> szán rím kín menny
11--20	<u>kád</u> gúzs dög pók pok
0--10	gém géz

If we compare the frequency of use and the correct identification percentages of the test words, it seems clear that meaning generally does not play as important a role in this case as is supposed in the literature. There are

frequent words with a low understanding ratio, and rarely used words with high percentage values. There are a lot of items which cannot be used in isolation in Hungarian, e.g. <pác> 'pickle', <bök> 'digs' (61--70%) and <zöm> 'bulk', <gúzs> 'withy' (17--32%).

There are words with similar meaning or frequency and their understanding is quite different, e.g. <jód> 'iodine' (62.5%) and <géz> 'gauze' (10%), <néz> 'looks at' (97.5%) and <fáj> 'hurt' (37.5%) and <rák> 'crab' (42.5%) and <sүн> 'hedgehog' (82.5%). There are words with similar acoustic structure and different percentage values, e.g. <sín> 'rail' (90%) and <szín> 'colour' (32.5%), <csók> 'kiss' (72.5%) and <csak> 'only' (32.5%), <nád> 'reed' (85%) and <vád> 'charge' (37.5%). The grammatical category of the test words seems to be of lesser importance as well.

By way of a final conclusion the following idea will be presented. All the results have supported that perception and understanding are better in certain high frequency bands, especially in 2200--2700 Hz. This finding led us to the hypothesis that hearing-impaired people with special hearing losses can perceive/understand speech in the 2200--2700 Hz range better than in a wider band which also contains the 'disturbing' elements.

A supplementary experiment was carried out with the participation of 10 hearing-impaired adults having hearing loss of different types and extents. Table 7 shows the answers of a high-frequency hearing-impaired man for the high-pass and pass-band filtered sound-sequences.

Table 7.

Original sound sequence	Responses of a high-frequency hearing-impaired			
	after high-pass filtering		after pass-band	
sikág	sipó ^c	ʃip ^o :tʃ	sikáld	ʃika:ld
vokus	-		voszus	vosu
kete	-		<u>kete</u>	kete
türáp	-		türáf	tyra:f
szajuk	szajuz	sojuz	sajuk	sojuk
celnü	csellő	tʃel:ɸ:	felnőtt	feln ^o :t:
bató	batu	botu	<u>bató</u>	boto:
lüse	-		<u>lüse</u>	lyʃe
rivem	-		riven	riven
furáz	<u>furáz</u>	fura:z	furáld	fura:ld
siszud	-		sicuc	ʃitsuts
tádó	tádu	ta:du	<u>tádó</u>	ta:do:
csinza	csidza	tʃidzɔ	<u>csinza</u>	tʃinzɔ
natup	natu	notu	natuk	notuk
zsegém	söböl	ʃɸbɸl	segém	ʃeɣe:m
macáf	macák	moʦa:k	macák	moʦsa:k
tipa	cipa	tʃipɔ	pipa	pipɔ
bodon	bodor	bodor	<u>bodon</u>	bodon
fücesz	lücesz	lytʃes	fűszek	fysek
cátany	-		császony	tʃa:soʃ
sumpös	-		sumfes	ʃumfeʃ
deféz	nehéz	nehe:z	<u>deféz</u>	defe:z

zipod	-		zifoz	zifoz
küvecs	üvecs	yvetŝ	tüvecs	tyvetŝ
pársza	pásza	pa:so	pászta	pa:sto

Table 8 shows the responses of a mixed-type hearing-impaired woman for the words with their normal acoustic structure and after pass-band filtering.

Table 8.

Original Responses of a hearing-impaired adult
words normal sounding after pass-band filtering with
cut-off-frequencies 2200--2700 Hz

mos	moŝ	mos	moŝ	mos	moŝ
kör	køŕ	kol	kol	kör	køŕ
menny	men:	-		megy	meʃ
láb	la:b	ab	ɔd	láb	la:b
zseb	ʒeb	zse	ʒe	zseb	ʒeb
híd	hi:d	híg	hi:g	híd	hi:d
szín	si:n	szép	se:p	szín	si:n
fal	fɔl	-		fal	fɔl

The results confirm that the secondary acoustic cues can, indeed, ensure the perception/understanding of speech in case the normal decoding process cannot work because of hearing problems.

Further research should show how the above findings can be applied in audiological examinations, phoniatric work and in speech therapy.

REFERENCES

DENES, P.B.--PINSON, E.N.: The Speech Chain: The Physics and Biology of Spoken Language. New York 1973.

FANT, G.: Speech Sounds and Features. Cambridge, Massachusetts, London 1973.

GÓSY Mária: Magyar beszédhangok felismerése, a kísérleti eredmények gyakorlati alkalmazása. / Identification of Hungarian speech sounds, application of experimental data. MFF 15. 1986, 3--100.

ROSEN, S.--FOURCIN, A.J.: When less is more -- Further work. Speech Hearing and Language No. 1. 1983, 1--27.

PISONI, D.B.: Some current theoretical issues in speech perception. Memory and Cognition 10. 1981, 249--59.

PHONETICALLY BASED NEW METHOD FOR AUDIOMETRY: THE G-O-K
MEASURING SYSTEM USING SYNTHETIC SPEECH

Mária Gósy--Gábor Olaszy--Jenő Hirschberg--Zsolt Farkas
Linguistics Institute, Hungarian Academy of Sciences; Heim Pál
Children's Hospital, Budapest

INTRODUCTION

There is a close connection between the articulation and perception bases of the process of speech acquisition. The initial development of perception and understanding abilities precedes that of speech production, but this difference between them subsequently decreases: their further development is assumed to take place in a permanent interaction. The basis of speech perception and understanding is hearing; this does not mean, however, that good hearing automatically ensures the normal processing of speech perception/under- standing. That is why regular examination of hearing and understanding is very important, particularly in the early years when the acquisition of the mother tongue is in progress. The identification of speech production problems is easier than that of speech perception/understanding ones. The normal communication situations provide a better opportunity for adults to detect the speech errors of children, revealing articulatory or grammatical problems. However, perception and/or understanding/comprehension difficulties can remain hidden

because of various supplementary and compensatory strategies of children. This fact leads to delayed diagnosis and to difficulties in carrying out the appropriate corrective procedures.

The examination of hearing can be either objective or subjective (e.g. Fleischer 1976). There are a lot of well-known problems related to the hearing measurements and mass screening of children between the ages of 3 and 7. What are the criteria that a suitable method for auditory screening of these small children has to meet?

First: the sound signal that is given to a child's ear should be natural and familiar for him. Second: the measuring task should be easy to understand, that is, we should make it easy for the child to understand what he has to do during examination. Third: the measuring method should yield the highest possible amount of information about the hearing mechanism operative between two hundred Hz and eight thousand Hz.

These expectations are all met by our new screening procedure, the G-O-H system.

METHOD AND MATERIAL

On the basis of the results of a perceptual examination of Hungarian speech sounds whereby the values of their invariant cues are determined (Gósy 1986), the process of speech understanding can be further studied: the hearing mechanism and the level of recognition of words can be examined. The examination of the two processes can be

combined if we produce speech material which only involves acoustic values corresponding to invariant features (or hardly more than that). This condition is satisfied by computer-generated, artificial speech based on perceptual data.

(In Hungary attempts were made at constructing vowel generating machines in 1944 and 1960 by Tamás Iarnóczy. The technical background for speech synthesis /a PDP computer and an OVE III synthesizer/ was established at the Phonetics Laboratory of the Linguistics Institute of HAS only in 1979.) Our laboratory has developed a synthesizing system for Hungarian which has been applied to research in audiology as well.

Speech as an auditory stimulus is familiar for children, and repeating sound-sequences is a natural everyday task when the child acquires his first language, and repeats the words of his mother. Human speech is suitable for judgement of understanding level, but the speech-audiometric results cannot give exact data about the hearing capacity or about the extent and type of impairment, because natural speech is very redundant as to its frequency structure. The redundancy of speech means that speech sounds contain far more building elements than would be necessary for understanding. That is why natural speech can be understood in the case of certain hearing impairments: the redundant elements give an opportunity to guess the meaning. For example the spirants in natural speech have noise components, with different intensities, in the high frequency range and in the lower

frequencies as well. The low frequency components contain sufficient information about these consonants for hearing impaired listeners to identify them correctly. The perception of voiceless fricatives was investigated by Lawrence and Byers (1969) in five patients having steep hearing losses above 1000 Hz. The results showed that the patients could correctly identify [ʃ] in 87%, [s] in 83%, [f] in 77% and [θ] in 72% of the cases. (It must be noted that the authors were not able to give an explanation for their perceptual data.)

Our specially synthesized words contain only the necessary frequency components of each sound. The difference between the natural and the synthesized words is only in the redundancy of the frequency structure (cf. Fig. 1).

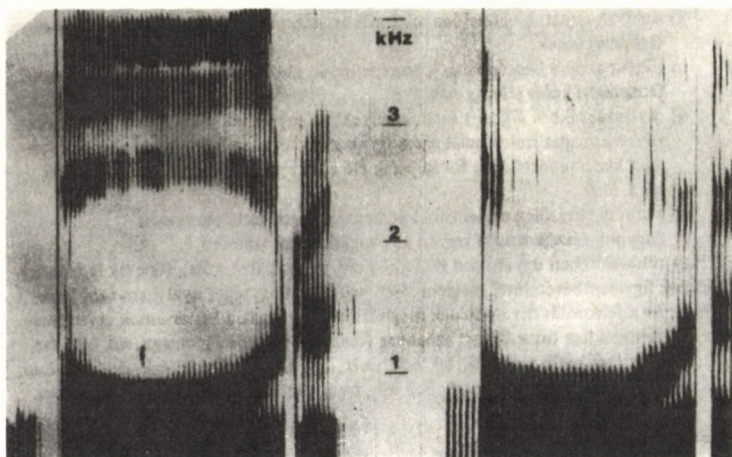


Fig. 1.

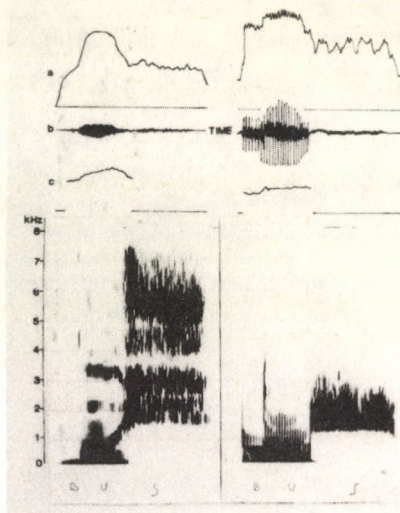


Fig. 1. Acoustic structures of natural and synthesized Hungarian, German words: <Busch> [buʃ] and <búr> [bʊʀ]

In spite of this difference, they sound very similar. The speech perception threshold proved to be the same as the normal one obtained by using natural words. Moreover, the intensity problem -- that appears when recording natural speech and applying it to speech audiometry -- could be solved (to a certain degree) by synthesized speech. Recording natural words the announcer cannot influence and control the sound intensity of the word uttered, so it varies from one utterance to the other. This is caused by the physical

properties of speech sounds, i.e. the more equalised the distribution of the first three formants of the voiced sounds is along the frequency axis, the higher the average intensity level of the sound will become. In practice it varies ± 6 dB. In our test material the intensity level of artificial words varies ± 3 dB.

When producing human speech one cannot have an influence on determining or changing the value of the frequency parameters either (e.g. place of a formant or burst of a given sound in the sequence). During the development of artificial words all components can be determined in advance.

How does the screening function of the artificially produced words work? Let us suppose that the system of speech understanding has to analyze data of quantity \underline{x} to understand the word <chair>. But the acoustic structure of natural speech is highly redundant, i.e. it contains significantly more information (data) in the speech signal than is necessary for its safe recognition. In the case of the example <chair>, it contains data of quantity $\underline{x} + \underline{y}$. The data surplus (\underline{y}) becomes stored and can be immediately called out in case of any kind of "disorder" (e.g. noise), to provide supplementary information. On the other hand the word <chair> we synthesize hardly contains more information than the necessary quantity \underline{x} . Therefore, in case there is some "disorder" at any point in the recognition process, $\underline{x} + \underline{y}$ would make identification possible, but \underline{x} itself does not, where comprehension will be mistaken (to some extent). The comprehension of a signal sequence containing information \underline{x}

requires the processing of all information in a perfectly sound fashion, e.g. by the help of normal hearing.

To provide a basis for the G-O-H method, a special test material was constructed which involved 44 meaningful monosyllabic Hungarian words. The criteria for choosing the words were as follows:

- (i) the monosyllabic words should have two or three speech sounds without consonant clusters,
- (ii) the words should contain speech sounds where the frequency parameter seems to serve as an acoustic cue for their identification,
- (iii) the test material includes three types of items: words containing only high-frequency sounds (like [sy:z]); words containing only low-frequency sounds (like [bu:]); words containing both high and low-frequency sounds (like [bus]),
- (iv) an effort was made to collect a material exhibiting almost all Hungarian speech sounds in different sound-contexts and phonetic positions,
- (v) most of the words should be familiar for children of ages 3--7; however, the sample should also include a few items that are meaningless sound-sequences for the children.

The three frequency bands generally used in the screening procedure (500, 1000, and 4000 Hz) seem to be very insufficient for the evaluation of speech understanding level. In normal hearing the acoustic information received at these frequencies accounts for some 60% of understanding.

This means that, in cases of hearing losses at other frequencies, the child -- screened and judged to have normal hearing -- cannot understand speech correctly (cf. Table 1).

Table 1.

Midfrequencies	125	250	500	1000	2000	4000	8000 (Hz)
Understanding							
of Hungarian	2	13	18	22	22	20	3
speech (%)							

Attention was also payed to the order of the words in our test material: low-frequency and high-frequency words alternate with one another. So all children have an experience of success, because they can understand and repeat correctly at least every second word. This is important for good co-operation between the child and the examiner.

Experiments were carried out with our test material in clinics and day-care-centers with the participation of 400 normal-hearing and 150 hearing impaired children.

RESULTS

People with normal hearing understand both human speech and the special synthesized artificial words equally well. But the hearing impaired patients cannot correctly understand the synthesized words, because of the lack of redundant

building elements. Speech synthesis, moreover, gives us an opportunity to define the desired frequency bands in speech sounds. These facts lead to the perceptual/understanding differences between the normal hearing and impaired hearing listeners. For example, a high frequency hearing impaired child with hearing loss above 5000 Hz cannot understand the word <szél> 'wind'. ([se:l]) if the noise component of the initial consonant of the word is above 5000 Hz. In this case, the child receives acoustic information that he identifies as a consonant like [f, h] or [t], depending on the extent of the hearing loss of the child. In the case of a slight hearing loss above 5000 Hz, the child will identify the sound-sequence <fé!> 'he is afraid of sg' ([fe:l]) which is an existing Hungarian word. In the case of somewhat more severe loss of hearing above 5000 Hz, the child will identify the sound-sequence <hé!> ([he:l]) which has no meaning in Hungarian, and with even more severe loss he will identify, instead of the spirant [s], a stop consonant like [t, p] or [k]. In the case mentioned, the child identifies the word <tél> 'winter', because it is a frequent item in children's vocabulary. (Figure 2 shows a German example.)

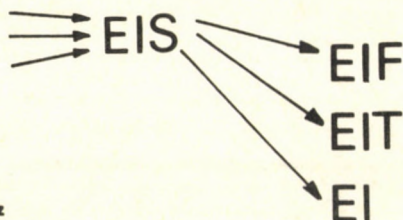
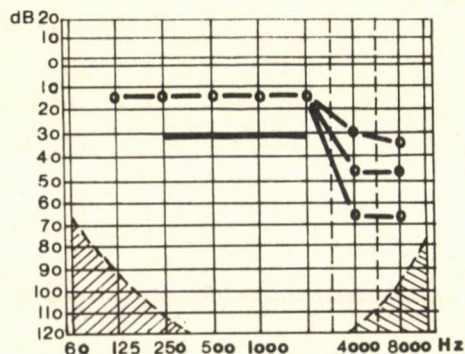


Fig. 2. Changes of possible responses in the case of different hearing losses (with a German example: <Eis> [ajs])

From the answers of the listeners judgements can be made about the type and extent of their hearing losses. The answers are regular consequences of hearing losses, they are not results of their imagination. (Experiments with filtered words confirmed us about these regular changes in perception, cf. Gósy 1986).

Mass-measurements were carried out together with a control pure-tone examination with a screening audiometer. The results of pure-tone audiometry corresponded to the results gained by our synthesized-word method. Our method,

however, gave information about the understanding level as well. In some cases the child did not co-operate in pure-tone audiometric examination, but he repeated the artificial words, so his hearing could be measured.

The possible answers of normal-hearing and impaired listeners were predicted theoretically on the basis of perceptual investigations concerning the acoustic cues of Hungarian speech sounds. Then, laboratory and clinical experiments were carried out, and the theoretically established types were slightly revised on the basis of the data obtained. Finally, the possible answers were arranged on an answer sheet according to the degree of hearing losses. The order of expected responses on the answer sheet gives the examiner a simple method for determining the child's level of speech perception/understanding: the closer the answer is to the original word, the less deviation is found from normal performance (cf. Figure 3).

After interlingual investigations in the acoustics and perception of Hungarian and German speech sounds attempts were made to produce German synthetic speech sounds. The main similarities and deviations between Hungarian and German were defined. Twenty monosyllabic German words have been developed so far in our phonetics laboratory. The acoustic structures of synthesised words show the differences between Hungarian and German which act at the level of invariant components (they sound very similar, Tables 2 and 3).

Table 2.

ZOL (#93)

Kódszám: BÚS (Hungarian) SZINTETIZÁLÁSI ADATLAP [bu:f]

Dátum: 1985. aug.

Azonosító: 1 2A 6 15 24 24A 170 171 172

IDO	40	60	40	40	40	40	60	10	40
IDA	30	10	1	10	12	40	15	100	10
AO	2	3-2	8-10	12-14	14	14			
F0	97-	100-	122	122-	112-	109			
	92	97		116	109				
F1			400-	300	300-	252			
			300		252				
B1			60	60	60	60			
F2			654	654-	617	617			
				617					
B2			62	93	93	93			
F3	2614		2614	2614	2614	2614			
AC							10-24	28	28-18
AK							16	16	16
K1							1510	1510	1510-
									1307

Table 3.

ZOL (#1W)

KÓDSZÁM: BUSCH (German) SZINTETIZÁLÁSI ADATLAP [bus]

Dátum: 1985. VIII. 6.

Azonosító: B1 B2 BU 24 24A 170 171 172

IDO	60	40	40	40	40	60	10	40
IDA	10	1	10	10	10	2	-*	10
AO	4-2	4-10	10-12	12-10	8-2			
FO	100-	116	116	116-	122-			
	106			122	116			
F1		400-	300	300	300			
		300						
B1		60	60	60	60			
F2		654	654-	617	617			
			617					
B2		62	93	93	93			
F3		2614	2614	2614	2614			
AC						10-22	22	22-16
AK						16	16	16
K1						1510	1510	1510-
								1307

AUSWERTUNGSGEGEN FÜR G-O-H

Nr.	Normal Hören	Leichte Hörstörung	Große Wahrscheinlich- keit der Hörstörung	Sicherheit der Hörstörung
1.	STEIN	Stein stehen Stirn Skein	Sturm Stuck ein Ei fein Heim Pein kein	Tag Tu Po Pu o u -
2.	BALL	Dall Dank Bar dann Gall gar	Pall Puls Pol Kalb Tal	Tag To Tu Pu o u -
3.	SCHI	Schick Schiff Schütt Sip	Schuh Schott Futsch Fuß Fuchs Huhn voll	Pu Po u o -
4.	ZEHN	Ziel Ziem Zem Zink Sinn	See ihm Veil viel vier hin Tin	Pu Po To Tu o u -
5.	MAUS	Baus Naus	Bausch Lauf Bauch Mau Bau Rausch Maul	Ma Mu Bu Po Pu o u -
6.	OHR	Uhr Uhl Olm	uh Pott Putt	Ok Uk Ot Ut Tu -
7.	BUCH	Duch Duf Guch Guf Busch	Puch Puff Kuh Muß Mut Tuk	Po Pu To Tu o u -
8.	SCHNEE	Schbie, See Schnie Shnek	Floh froh Fluch neu Flaum flau Huhn nett	Po Pu To Tu o u -
9.	KUß	Putz Schuf Kunst Tuss	Kusch Kuff Kuh Putt Pott	Ut Ot To Tu o u -
10.	SCHMERZ	Schweiz Shnerz Schmers	Feld Fell Pferd Fenn Herz Herr Meer Fund	Pell Pet Put Pu Tu o u -
11.	BUSCH	Dusch Gusch Euc Guss Duz	Puch Puff Puh Kuh Tuk Muß Mut	Po Pu To Tu o u -
12.	MIST	Mies Misk Niß Nist Nisp	Milch mich Mief mit Muff Muh Mucks Mut Niet	Bo Bu Po Pu To Tu o u -
13.	SPIEL	Stiel Spier Skier Spind	Ski Fink Film in ihm ihr Fuchs Furcht Huhn	Pot Put To Tu o u -
14.	ZAHN	Zahl Zahn Sanft Sahn	Salz Satt Akt Amt ab an Pann Panz	Po Pu To Tu o u -
15.	SINN	Zin Zehn Sill Sim Süm Zün	Finn Funk hin hof hoch Fund Huhn	Po Pu To Tu o u -
16.	LEHR	Beer Mehl lehr	der vehr vier	Po Po Pu o u -
17.	BOOT	But Böt Dock Geot	Port Maut doch Pump Puk	Bo Po Pu To Tu o u -
18.	SCHUH	Schub Schuf Schoß	Fuß Volk voll Futch Fuchs Huhn	Pu Po To Tu o u -
19.	EIS	als ars As	auf eif eich alt eit Arsch Art Ei und	a o u -
20.	SEE	sie Seim sehr Zehn sehen	Scheß Schuh Föhn Fuß Pox Fock Mund Huhn	Po Pu To Tu o u -

Fig. 3. Answer-sheet

These synthesized words are apparently suitable for application in the German version of our G-O-H system. Experiments were carried out with German-speaking 4- and 5-year-old children in a kindergarten in Vienna (Austria). The results are summarized in Table 4.

Table 4.

German synthesized words	Understanding of the words	
	correct answers	misunderstanding of words
Boot	9%	Mond Mund Uhr Nool
Busch	9%	Nuß nusch musch mosch
Sinn	36,4%	Zin sink sim
Kuß	36,4%	Fuß Nuß puß
Buch	36,4%	uch
Spiel	54,4%	stiel skier
Schnee	54,4%	See schmee
Ohr	54%	wir ar
Schuh	82%	Stuhl
Schi	73%	schil
Ball	82%	all
Stein	82%	Schwein
Zehn	82%	ehn
Meer	82%	Beer
See	82%	Zeh
Schmerz	91%	Marz
Mist	91%	ist
Zahn	91%	-
Eis	100%	-
Maus	100%	-

For the everyday use of the procedure, a measuring set has been developed that contains a playing system, a tape with the computer generated words, an amplifier, a headphone and answer sheets (Figure 4.)

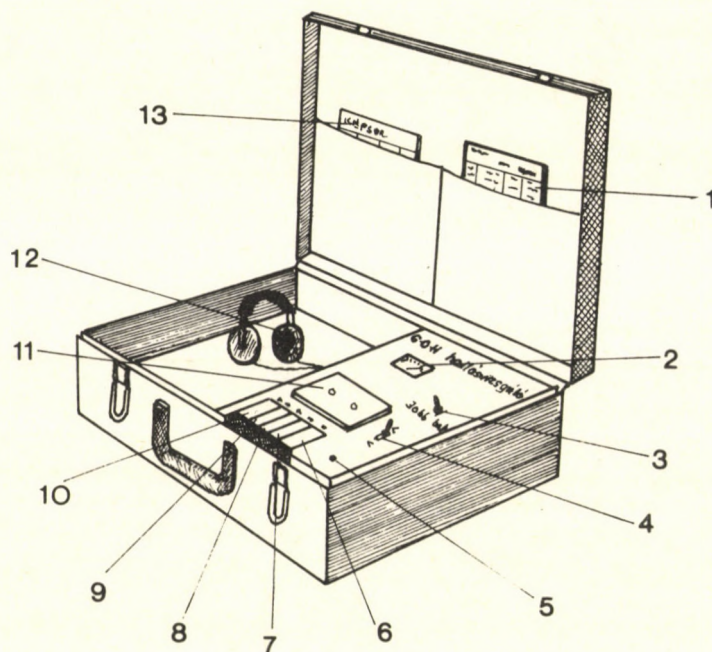


Fig. 4. The G-O-H set

(More than 150 of these sets with G-O-H system are being used in Hungary.) To carry out the examinations there is no need for any expert, physician, phonetician or audiological assistant: it can be used by nurses, kindergarten teachers, speech therapists and so on. The experiences can be

summarized as follows:

Several types of previously undetected hearing problems were found, e.g. secretory otitis media without any complaint, slight hearing losses, as well as various degrees of hearing losses of one ear.

Our method seems to be suitable for giving information about the central hearing mechanism as well. There is a sad example about a 5-year-old child who had a slight speech defect. His speech therapist could not achieve good results with him, so the child was sent to audiological examination. According to the result of pure-tone audiometry, his hearing was normal. Then this child was measured by the G-O-H system and, surprisingly, he did very poorly. The child was sent back to the clinic and after accurate examinations a brain-tumor was detected.

Our method is a good tool for speech therapists and teachers (i) to find out whether the child is mature enough to acquire writing and reading, (ii) to learn whether the child with speech error(s) has perceptual problems as well or not, and (iii) to detect dyslexia, i.e. the disturbances in writing and reading at an early age.

REFERENCES

- FLEISCHER, K.: Hals-Ohren-Heilkunde für das Kranken-
pflegepersonal. Stuttgart 1976.
- GÓSY Mária: Magyar beszédhangok felismerése, a kísérleti
eredmények gyakorlati alkalmazása. MFF 15. 1986. 3--100.
- LAWRENCE, D.L.--BYERS, U.: Identification of voiceless

fricatives by high frequency hearing impaired listeners.

J.S.H.D. D. 12. 1969, 426--34.

ON THE TONOSYNTAX OF A HUNGARIAN CHILD'S

EARLY QUESTIONS

(A Preliminary Report)

Ilona Kassai

Linguistics Institute, Hungarian Academy of Sciences

INTRODUCTION

Questions are an important means of cognitive development. Therefore the evolution of the verbal means of questioning highlights the intellectual development of the child on the one hand and its linguistic, especially syntactic development on the other.

In the present paper I give an account of a tentative analysis of questions gathered from the spontaneous speech of one child (a girl), produced in interaction with adults and regularly recorded from 1 to 3 years of age. I was particularly interested in the acquisition of the prosodic shape of questions, a topic largely neglected in child language research across the world.

THE SYSTEM TO BE ACQUIRED

In the process of language acquisition Hungarian children are faced with the following basic question types differing in form.

Wh-questions. -- They require a question-word and a specific word order characteristic of emphatic sentences in

which the emphasized element (here the question word) is obligatorily followed by the unstressed verb. The remaining constituents can either follow the unit formed by the focus and the verb as part of the comment or precede it and constitute the topic of the sentence. In case the question-word stands for the predicate it can even be the last element of the sentence. Let us see some examples illustrating the possible ordering. (Capitals indicate the location of emphatic stress.)

<MIT mond Mari a balesetről?>

what says Mary the accident-about

'What does Mary say about the accident?'

<A balesetről MIT mond Mari?>

the accident-about what says Mary

<Mari MIT mond a balesetről?>

Mary what says the accident-about

<A balesetről Mari MIT mond?>

the accident-about Mary what says

<A résztvevők száma Mennyi?>

the participants number+gen.

how many

'What is the number of participants?'

If in the neutral sentence the predicate contains some preverbal modifier (verbal prefix, predicative adjective,

etc.) the latter must follow the verb in the emphatic sentence. The statement

<Péter kinyitotta az ablakot.>

Peter out-opened the window+acc.

'Peter opened the window.'

is turned into a wh-question in the following way:

<Mikor nyitotta ki Péter az ablakot?>

when opened out Peter the

window+acc.

'When did Peter open the window?'

The statement

<Sötét van.>

dark is

'It is dark.'

gives

<Hol van sötét?>

where is dark

'Where is it dark?'

(For word order and topic-comment articulation in Hungarian see Kiefer 1967; É. Kiss 1981; Komlosy 1986.)

As in the case of wh-questions the type of utterance is signalled both morphologically and syntactically, prosodically they are not autonomous in the sense that they do not have a specific intonation. They show the same falling

contour as statements, which, however, a somewhat wider frequency range. This is a fourth or a fifth while that of statements is a third. The slight difference in the realization of the falling contour seems to provide additional support for the recognition of questions (Fónagy--Magdics 1967, 60). As for stress patterns, this question type is usually realized with a single heavy stress located on the question word. (In the Hungarian language word stress falls on the first syllable.)

Yes/no questions. -- This question type has two varieties. One of them contains an interrogative particle added to the verb on the nominal predicate:

<Olvas-e Péter újságot?>

reads-y/n part. Peter

newspaper-acc.

'Does Peter read newspapers?'

<Piros-e az alma?>

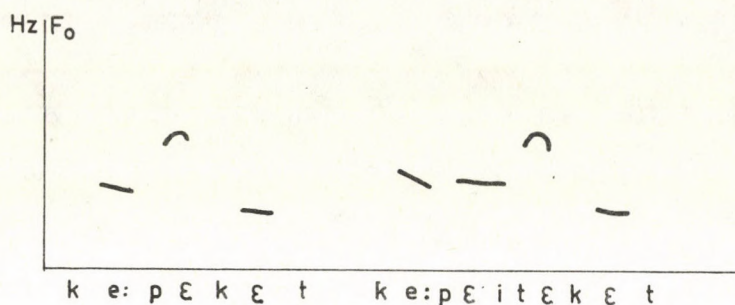
red-y/n part. the apple

'Is the apple red?'

Due to that morphological marking, the prosodic shape does not differ essentially from that of an emphatic statement. Moreover, in present-day Hungarian this variety occurs rarely in the above form, i.e. as main clause. Its use is more and more restricted to subordinate clauses.

The other, almost exclusively used variety is expressed by means of intonation. The basic pattern from which all the remaining forms can be derived seems to be a rise-fall

movement appearing on the last three syllables in questions containing only one trisyllabic or multisyllabic word. The magnitude of the rise is about a musical third while that of the fall is a fourth (Fig. 1a, b).



In questions consisting of a bisyllabic word both the rise and the fall take place on the last syllable (Fig. 2). Finally, in monosyllabic questions only the rising part of the pattern is realized (Fig. 3).

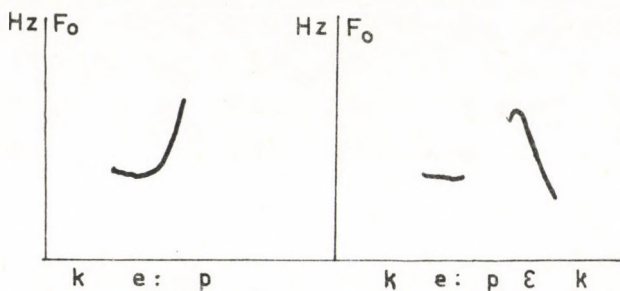


Fig. 2.

Fig. 3.

(For phonetic details see the article by Gabor Olaszy in this volume.)

The fairly simple picture presented above becomes more complicated in case the question contains more than one word and more than three syllables. The intonation of such questions is determined by the number of syllables of the last stress group regardless of the number of words it contains. The rule is as follows. If the last stress group is monosyllabic it displays the contour of monosyllabic questions (Fig. 4).

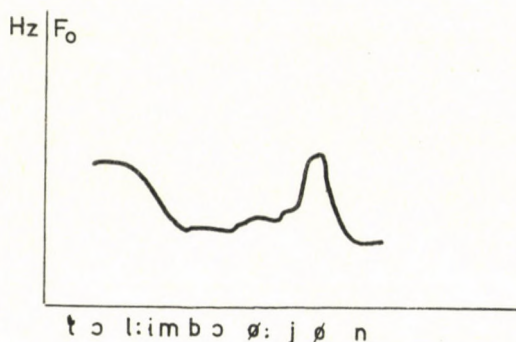


Fig. 4. <Tallinnba õ JÖN?> 'Is he COMING to Tallinn (after all)?'

If the last stress group is bisyllabic, it shows the pattern

characteristic of bisyllabic questions (Fig. 5).

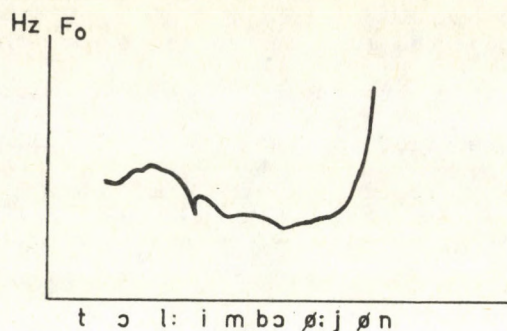


Fig. 5. <Tallinnba õ jõn?> 'Is it him who is coming to Tallinn?'

Lastly, if in the last stress group there are three or more syllables, the intonation pattern is that of the corresponding one-word question (Fig. 6).

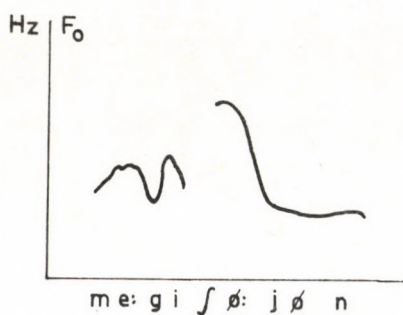


Fig. 6. <MEgis õ jõn?> 'Is he coming (contrary to expectations)?'

In other words, the prosodic shape of a morphologically unmarked yes/no question containing more than one word and more than three syllables strongly correlates with its topic-comment structure.

For the word order the following holds. The constituent bearing the main stress is usually the first element of the sentence but it can also be located in sentence-medial or sentence-final position. If the focussed element is other than the verb, wherever it stands in the sentence, it must be followed, as a rule, by the verb.

Tag questions. -- They are constructed from a statement and the interrogative morpheme <ugye 'isn't it'> added to either the beginning or the end of the statement. In the first case the overall intonation can only be falling, while in the second case there is a choice for the interrogative morpheme to be realized as a statement or a bisyllabic question. The prosodic difference conveys an attitudinal one: the falling contour means that the speaker expects a positive answer. The rise-fall of the second alternative refers to the speaker's desire to elicit a positive answer.

Elliptic questions. -- This question type shows a rising contour. Syntactically it is elliptic, i.e. only one part of the intended content is expressed. The ellipated part can be completed from the nonverbal context as either a wh-question or a yes/no question (Fig. 7).



like to have some) Wine? Coffee? Cherry brandy?'

THE PROCESS OF QUESTION ACQUISITION

I have analysed the questions occurring in the recorded material in order to find out how the interrogative system of the adult language emerges and evolves in the child.

The examination has revealed first of all that the productive use of questions is preceded by the imitation of adult models. Imitation shows two forms. In some cases adult questions were rehearsed by using the prosodic component only: the child hummed the intonation of the question. The other and greater part of imitative questions had an accurate intonation but only approximate segments. The first form of imitation decreases as the child's segmental competence progresses but the second form remains for long enough and occurs whenever the child does not understand or cannot answer the question addressed to him. The main function of imitation seems to be the learning of the verbal means of questioning. Besides, imitative questions may serve to maintain contact with the adult, thus they can perform a discourse function.

As for the order of emergence of the question types treated in the preceding section the following assumption may be made. First to appear are yes/no questions expressed by means of intonation. The first functional, not imitative question was uttered at 1;6,22 when the child, pointing at her trousers, asked: <Szép?> 'Beautiful?'. Wh-questions come next in the developmental order. The first one was registered at 1;9,18. Tag questions do not appear before 2;4. Elliptic questions come last of all, at 2;7,20. The yes/no question constructed with the particle, as expected on the basis of its adult use, did not appear as main clause in the period examined. It only occurred as a subordinate clause, at 2;8,20 for the first time.

The analysis of the formal aspect of question acquisition has revealed the following prosodic and syntactic tendencies. Prosodically wh-questions do not cause any problem to the child as their intonation is almost identical with that of statements already acquired. However, from 2;2,2 in certain utterances one can hear an extra stress on the last syllable, which is in contrast to adult realizations, but very characteristic of Hungarian children's performance (according to the common experience of adult listeners). At closer examination it turns out that extra stress occurs mainly in longer, multiword utterances beginning with the stressed question-word. Another characteristic of the child's wh-questions is the topicalization of unstressed constituents which results in shifting the question-word towards the end of the sentence. In my opinion, these

different strategies have the same goal: to give the end of the sentence perceptual prominence. The explanation for it might be the child's desire to provoke an answer or to get the partner's attention by all means.

From among word order changes required by this question type verb-subject ordering takes place at once but postposition of verbal modifiers shows inconsistencies and does not stabilize until the end of the period examined. The observed difference in the application of the inversion rule concerning subject and verbal modifiers may be given a multiple cue explanation. One seems to be based on the structure of the Hungarian language. Owing to the rich system of verbal endings each person has its own ending which makes the use of the personal pronoun as subject unnecessary in normal circumstances. Moreover, if the subject is present in the surface structure, it does not have a fixed position. According to recent research on syntax Hungarian is not a subject prominent language like English but a "topic prominent" one in which the order of elements is determined by their communicative-logical function (see E. Kiss 1981). Another explanation might be the generally observed fact that children between 2 and 3 years, independently of the word order rules of their mother tongue, are inclined place to the verb at the beginning of the sentence. For Hungarian children below 3 years the dominant word order is VS in two-term sentences and VSO in three-term sentences (S. Meggyes 1971). This means that the verb-subject ordering is already given in statements. All the child has to do is to add a question-word

to the beginning of the statement.

Yes/no questions expressed by intonation, though they appear first, have been found to cause children more difficulties than any other type. As demonstrated in the preceding section, yes/no questions without the interrogative particle show three distinct intonation patterns according to the number of syllables contained in the word constituting the question by itself. This basic distributional rule seems to be acquired early and accurately, i.e. one cannot observe any problem in its application in case the child utters a one-word question. However, when the question contains more than one word its intonation patterning becomes dependent upon the location of the emphatic stress which, in turn, is dependent on the topic-comment articulation of the question. In multiword questions one can often detect intonational mistakes: the child uses a pattern contradictory to the topic-comment structure signalled properly by one or several of the following factors: stress assignment, word order, nonverbal context. The analysis of prosodically mistaken utterances has shown some regularities in the seemingly chaotic patterning. There are utterances in which the child uses the pattern required by the number of syllables of the last word independently of its stressed or unstressed nature. E.g. the question

<Te NEM mész ide?>

you not go+2. sg. here

'Don't you go here?'

is realized with the bisyllabic pattern though the trisyllabic

one would be adequate. In a few examples the intonation mistake can be considered as the consequence of misplaced stress, as in:

<Kérsz MÉG?>

want+2. sg. still

'Do you want some more?'

which is defective because the word <még 'still'> cannot receive any stress when it occurs on its own. Finally, some of the mistakes appear to be triggered by the non-application of the obligatory word order of emphatic sentences which demand a preverbal focus. In the question

<Tied az INGed?>

yours the shirt-gen.

'Is your shirt yours?'

uttered with the bisyllabic pattern, the focussed element should precede the other one which is the predicate of the sentence.

Whatever the trigger may be, it often happens that the child corrects herself within the same discourse turn and produces the appropriate prosodic solution.

For tag questions the source of trouble is the sentence-initial position of the question-morpheme prescribing a falling contour contradictory to the semantic content of the sentence.

Elliptic questions are always produced correctly. This is probably due to the fact that this question type has a single, unambiguous intonation curve referring to non-finality.

CONCLUSION

The findings strongly suggest the one-way conclusion that children below 3 are in the process of learning the complex rule-system governing the prosodic articulation on the one hand and the topic-comment articulation on the other hand. However, two facts allow for some other explanation, too. Self-corrections and the marked tendency common to all erroneous items to shift the stress or the intonation peak to the last syllable make one think of an unconscious endeavour to ensure continuity in discourse.

REFERENCES

- É. KISS, K.: Topic and focus: The operators of the Hungarian sentence. *Folia Linguistica* 15. 1981. 305--30.
- FÓNAGY Iván--MAGDICS Klára: A magyar beszéd dallama [The Melody of speech in Hungarian]. Budapest 1967.
- KIEFER, F.: On Emphasis and Word Order in Hungarian. Bloomington 1967.
- KOMLÓSY, A.: Focussing on focus in Hungarian. In: W. ABRAHAM and S. de MEIJ (eds). *Topic, Focus, and Configurationality*. Amsterdam--Philadelphia, 1986. 215--26.
- S. MEGGYES Klára: Egy kétéves gyermek nyelvi rendszere. *NytudÉrt* 73. Budapest 1971.

BRaille-LAB,

A FULL HUNGARIAN TEXT-TO-SPEECH MICROCOMPUTER FOR THE BLIND

G. Kiss(1), A. Arató(2), J. Lukács(3), J. Sulyán(2), T. Vaspöri(2)

1. Hungarian Academy of Sciences, Linguistics Institute

2. Hungarian Academy of Sciences, Central Research Institute
for Physics

3. Eötvös Loránd University, Faculty of Natural Sciences

ABSTRACT

The authors introduce Braille-Lab, a Hungarian-speaking microcomputer developed for the blind. This 280 microprocessor-based personal computer is fitted with a Philips MEA 8000 formant synthesizer, providing for Hungarian text-to-speech conversion. The original version of the machine contains a speaking BASIC interpreter. The new version, Braille-Lab+, is also furnished with a speaking word processor and a speaking database management system running under a speaking CP/M compatible operating system. Braille-Lab has been approved and adopted by the Hungarian National Federation of the Blind, 95 sets have been installed so far.

INTRODUCTION

In the past few decades, intensive research into speech synthesis has been going on in a number of countries including Hungary. This research work has three main types of motivation.

1. Fifth-generation computers are to create a new, humanized type of man-machine-man relationship. Hence one of the main objectives of research is viva voce 'conversation' between man and machine. The various links of the man-machine-man communication chain (each constituting a research area in its own right) and the way artificial speech production fits into that chain are represented in Fig. 1.

2. Another impulse for attempts at speech synthesis was the desire to achieve a better understanding of the acoustics of speech. Indeed the principle of analysis by synthesis is more effective than any measuring apparatus, however sophisticated the latter may be: it shows what the essential components of speech really are [7]. That principle can be best implemented by formant synthesis. These considerations led to the establishment, under Kálmán Bolla's leadership, of a complex acoustic speech synthesizing system in the Linguistics Institute of the Hungarian Academy of Sciences, in the late 1970s. The hardware configuration includes an OVE III (Swedish-made) formant synthesizer [1] and a PDP11/34 computer. The effective operation of the system is guaranteed by a specially designed interactive program called FOPRO [11].

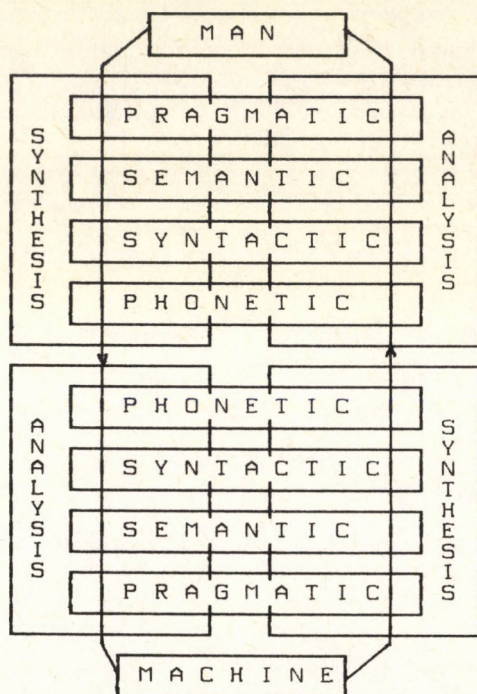


Fig. 1. The Man-Machine-Man communication chain

The utility of the system for phonetic research is demonstrated by a number of scholarly papers [5, 10]. The program was also used for designing an inventory of speech frames for a Hungarian text-to-speech (TTS) system based on the principle of formant synthesis in the early 1980s [13]. The inventory, in turn, was used in HUNGAROVVOX, a Hungarian real-time TTS system for speech synthesis [9, 12]. Later, a developing system was also made for a Philips MEA8000 formant

synthesizer [3].

3. The third type of motivation for research on speech synthesis is a desire to develop various appliances to help handicapped people (afflicted with speech disorders, blindness, etc.). The area was given a vast impetus by the appearance, in the early 1980s, of speech synthesizers contained in a single IC, e.g. UAA1003, TMS5200, SC-01, SP0256, MEA8000 [2, 4, 6, 8], since these could be built into various appliances. These considerations led to the development of Braille-Lab, a speaking computer to be used by blind people, introduced in the present paper. This Hungarian-speaking microcomputer fitted with a text-to-speech conversion system effectively helps the education of blind people in computational technology (thus creating high-qualification employment possibilities for them). Also, it accelerates their full integration into society.

THE HARDWARE OF BRAILLE-LAB

Braille-Lab is a Hungarian-made, 280 micro-processor-based personal computer. Its memory is organized on a page basis, and consists of 64 kbyte RAM and 20 kbyte ROM. The card containing the speaking module has been built into the computer with MEA8000. The TTS software is located on page 2 of ROM. The keyboard of Braille-Lab contains every letter of the Hungarian alphabet, arranged in a way almost identical with the keyboard of standard Hungarian typewriters. The built-in small loudspeaker makes it possible for the speech produced by the system to be heard without an

external loudspeaker. The built-in BASIC interpreter leaves 48 kbyte free memory capacity available for the user.

The basic version of Braille-Lab has been further developed. Braille-Lab+, the new version, runs under a CP/M compatible operating system. Along with a 64 kbyte operative memory, it is also furnished with a 192 kbyte RAM disk and a 1 Mbyte floppy disk drive. The new version further contains a speaking word processor and a speaking database management system. With these two programs, its possibilities of application by the blind have been multiplied.

THE TEXT-TO-SPEECH SOFTWARE SYSTEM OF BRAILLE-LAB

The basis for Hungarian TTS conversion by Braille-Lab is a text in Hungarian orthography, with no special symbols added. The program translates that text into a series of frame code numbers for the MEA8000 synthesizer. The frame code numbers designate the elements of a 218-member frame inventory, devised earlier. The TTS conversion is implemented in the following four steps:

1. First of all, the text to be converted to speech is transformed by the program into a series of (code numbers of) speech sounds. Hungarian orthography is a fairly accurate indicator of the series of sounds to be uttered. However, not only single letters but also combinations of two, and even three, letters may stand for single sounds. In the letter-to-sound transformation, the program basically relies on Fig. 2.:

Hungarian orthography has a unique letter or letter

1	2	3	4	5	1	2	3	4	5
1. a	1	ɔ	-		34. nn	22	n:	+	
2. á	2	a:	-		35. ny	23	ɲ	-	
3. b	10	b	-		36. nny	23	ɲ	+	
4. bb	10	b:	+		37. o	6	o	-	
5. c	11	ts	-		38. ó	6	o:	+	
6. cc	11	ts:	+		39. ö	7	ø	-	
7. cs	12	tʃ	-		40. ô	7	ø:	+	
8. ccs	12	tʃ:	+		41. p	24	p	-	
9. d	13	d	-		42. pp	24	p:	+	
10. dd	13	d:	+		43. r	25	r	-	
11. e	3	e	-		44. rr	25	r:	+	
12. é	4	e:	-		45. s	26	ʃ	-	
13. f	14	f	-		46. ss	26	ʃ:	+	
14. ff	14	f:	+		47. sz	27	s	-	
15. g	15	g	-		48. ssz	27	s:	+	
16. gg	15	g	+		49. t	28	t	-	
17. gy	16	ʒ	-		50. tt	28	tt:	+	
18. ggy	16	ʒ:	+		51. ty	29	c	-	
19. h	17	h	-		52. tty	29	c:	+	
20. hh	17	h:	+		53. u	8	u	-	
21. i	5	i	-		54. ú	8	u:	+	
22. í	5	i:	+		55. ü	9	y	-	
23. j	18	j	-		56. ũ	9	y:	+	
24. jj	18	j:	+		57. v	30	v	-	
25. k	19	k	-		58. vv	30	v:	+	
26. kk	19	k:	+		59. z	31	z	-	
27. l	20	l	-		60. zz	31	z:	+	
28. ll	20	l:	+		61. zs	32	ʒ	-	
29. ly	18	j	-		62. zzs	32	ʒ:	+	
30. lly	18	j:	+		63. sp	33		-	
31. m	21	m	-						
32. mm	21	m:	+						
33. n	22	n	-						

1= number, 2= letter (s), 3= code number
4= IPA symbols, 5= length of sound

Fig. 2. Table of letter-to-sound correspondences

combination for each speech sound (with the exception of j vs. ly). This statement is also true the other way round: any letter or letter combination always stands for the same speech sound. However, there is some problem at the internal boundary of compounds spelt as one word: the correct speech sound assignment is sometimes difficult in consonant clusters across such boundaries. For instance, in a word like <víz> +

<szegény> = <vízszegény> (arid, lit. water + poor), our program wrongly interprets zsz as zs + z, instead of the correct z + sz. To contravene that source of error, we have introduced the notion of 'bachelor' letters. A bachelor letter does not form part of a letter combination but corresponds to a speech sound on its own, irrespective of what the following consonant letter is. To produce a bachelor letter, we have to keep the F2 key pressed while pressing the appropriate key on the computer keyboard. The letter appearing on the screen will not be affected but the otherwise empty seventh bit of its ASCII code will take up the value 1. Thus, in the above example, we get the correct sound assignment if the first z of <vízszegény> is pressed with a simultaneous pressing of the F2 key. As a result of the letter-to-sound transformation, we get a series of code numbers, each code number being an integer between 1 and 33, corresponding to the thirty-two speech sounds plus pause.

2. The second step of TIS conversion is the designation of the series of frames that will realize the speech sounds of the text to be uttered. This designation is basically of a diadic nature. The 218 frames utilized are arranged in the inventory in a very special order. Each combination of sounds is realized by adjacent frames. Thus we can dispense with storing what is called a combination matrix and consequently save a significant amount of memory capacity. However, this simplified procedure results in a poorer speech quality. (For somebody whose ear is not accustomed to mechanic speech, the speech produced by Braille-Lab is seldom intelligible at

first hearing. This, however, does not represent a difficulty for regular users: experience shows that the blind users soon get accustomed to the way Braille-Lab speaks and they have no problem understanding it during regular use.) In order to further optimize the utilization of the frame inventory, various sound sequences can be realized by overlapping series of frames, as illustrated in Fig. 3.

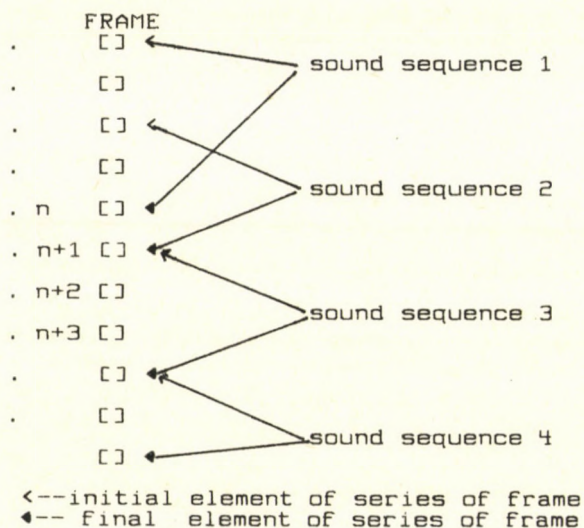


Fig. 3. The structure of the frame inventory and the way frames realizing sound sequences are specified

Long sounds are also produced at this stage by multiplying some component of the frame of the corresponding short sound (2 to 5 times, as the case may be) in the series of frame code numbers. Each element of the series of frame code numbers will be an integer between 1 and 218. That series then serves as input to the melody generating part of the

program.

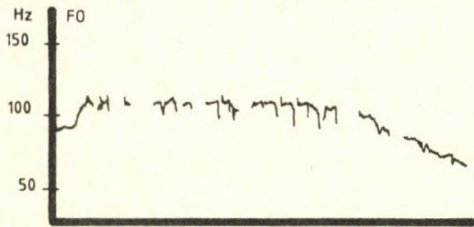
3. The melody is generated by the program by selecting the appropriate value of the PI parameter of the MEAB000 synthesizer frame-by-frame. The first step in producing the melody is the segmentation of the text into intonation units. The intonation units are marked off by .(full stop) ,(comma) ?(question mark) !(exclamation mark) or RETURN. Triggered by those punctuation marks, the program will supply the segmental structure produced so far with one of the melody patterns exemplified in Fig. 4.

The break-points of intonation curves are assigned to vowel positions. Note that the melody triggered by an exclamation mark is identical with that of a question-word question. This way of producing the relevant Hungarian melody patterns has an additional advantage, particularly for blind people, that the melody patterns unambiguously refer back to sentence-final punctuation marks. (If the intonation unit is concluded by RETURN without any punctuation mark, the melody produced is a level tone at 100 Hz.)

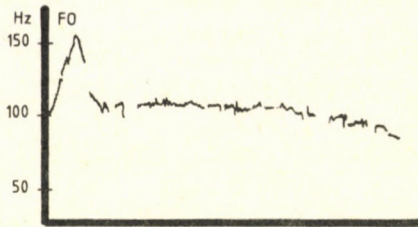
4. When the coding of segmental and suprasegmental structure is completed, Braille-Lab forwards the resulting series of code numbers to the MEAB000 speech synthesizer and the speech is simultaneously heard.

THE USE OF BRAILLE-LAB

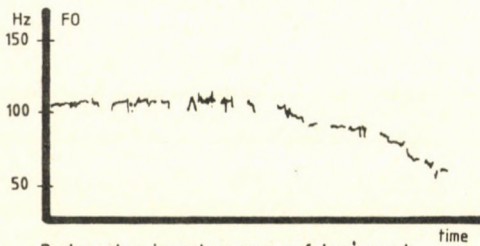
The computer is able to speak as soon as it is switched on. The following introductory words appear on the screen and are simultaneously heard [in Hungarian]: "Braille-Lab computer,



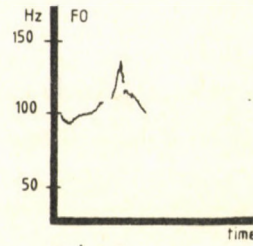
A feleségemet Budapesten ismertem meg.
'I first met my wife in Budapest'
Declarative sentence starting with
a definite article



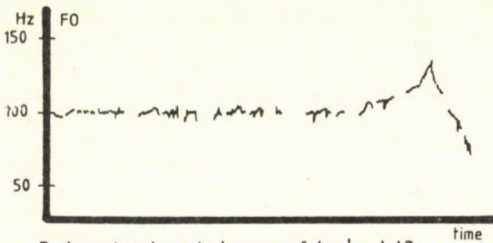
Mikor ismerted meg a feleségedet?
'When did you first meet your wife?'
Interrogative sentence starting with
a question word



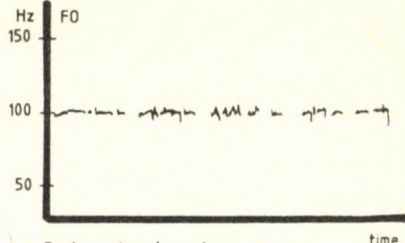
Budapesten ismertem meg a feleségemet.
'It was in Budapest that I first met my wife'
Declarative sentence not starting with a
definite article



Nyáron?
'In the summer?'
Disyllabic question



Budapesten ismerted meg a feleségedet?
'Did you first meet your wife in Budapest?'
Interrogative sentence with no question word



Budapesten ismertem meg.
'I first met her in Budapest'
Sentence without end punctuation

Fig 4. The intonation system of the Braille-lab

After that, each time a key is pressed, the system utters the corresponding speech sound, in order to make it easier for a blind person to avoid typing errors. Names of non-letter keys, including numerals, are uttered as words. E.g. on pressing % the machine says "százalék" (percent), etc. Using the cursor keys, the user can aurally check the contents of any character position of the screen.

Basically there are two situations in which Braille-Lab actually speaks: 1. during entering and editing BASIC programs; 2. at run-time when any information appearing on the screen is simultaneously said aloud.

1. During program editing, the echoing function mentioned above is in operation; in addition, at the end of each line when RETURN is pressed the computer reads out the whole line as a connected text. Numerals at this point are not read character-by-character but as wholes (e.g. twenty-five rather than two, five). The English terms of BASIC are read out according to the Hungarian value of letters, rather than in proper English pronunciation. At program listing, the list can be heard as it appears on the screen. In short, any information appearing on the screen, including e.g. error messages, is also uttered without any special command.

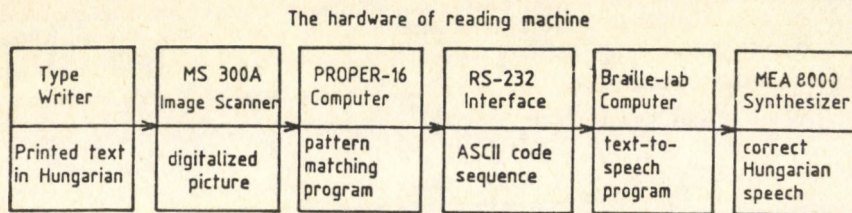
2. The information appearing on the screen during the running of a BASIC program will also be heard automatically. For instance, as a result of the running of the following short program, all Hungarian numerals between 1110 and 1125

will be heard first with a question intonation and then with a statement intonation (i.e. "Is the next number 1110? Yes, 1110." etc):

```
10 FOR I=1110 TO 1125
20 PRINT "A következő szám" I "?"
30 PRINT "Igen" I "."
40 NEXT
```

BRAILLE-LAB AS THE SPEAKING PERIPHERY OF A READING MACHINE

At an exhibition called "Hungarians in the World" held in Budapest in August 1986, the authors, assisted by researchers of SZKI (Institute for Computer Research) connected Braille-Lab with an IBM compatible PROPER-16 computer. On the other side, an MS300A Image Scanner was also connected with PROPER-16. The image recognition program developed by the SZKI people recognized printed Hungarian text. PROPER-16 forwarded the resulting ASCII code, via a standard RS232 interface, to Braille-Lab which uttered the text real time, with a proper Hungarian intonation, in an intelligibly pronunciation. To our best knowledge, this was the first time a Hungarian-speaking reading machine was presented in operation in Hungary. This presentation has proved that Braille-Lab is able to operate as the speaking terminal of a reading machine.



The software of reading machine

Fig. 5. The system of the reading machine when the Braille-Lab is speaking terminal

HOW TO MAKE BRAILLE-LAB SING

One of the special features of Braille-Lab is that it can also sing. To make the computer sing, the user has to specify the correct rhythm and the correct sequence of pitches. Rhythm can be represented by lengthening the vowels appearing in the words of the song, by entering vowel letters more than once. The length of the syllable containing the vowel will increase in proportion with the number of identical vowel letters entered. The melody has to be given in relative sol-fa letters, according to Zoltán Kodály's method. The pitch defined by a sol-fa letter assigned to a syllable will be superimposed by the program on the appropriate syllable which has been rhythmically defined as above. By that procedure, any Hungarian-text song can be produced. This special feature of the system opens up a novel area of application in the on-line representation and correction of Braille music notation [15]. Our experience shows that the fact that Braille-Lab is able to sing is a great help for

blind (or any) users in overcoming their prejudice, if any, against computers. The opening line of the Hungarian folk song "Érik a szőlő, hajlik a vessző" has to be entered in the following way:

re = "ééérik" : la = "a"

re = "szóóóó" : mi = "lőó"

BRAILLE-LAB AS AN AUTHORIZED APPLIANCE

Braille-Lab is an appliance authorized for use by the Hungarian National Federation of the Blind. By March 1987, a total of 95 sets have been installed in the schools of the Federation and by individual users. Based on the speaking BASIC of Braille-Lab, the Federation organized two beginners' courses on computation in spring 1986 and 1987. The speaking computer effectively helped the blind participants to acquire knowledge and skill in computation and to put them to creative use. The Users' Manual for Braille-Lab has been published on cassette tape and in Braille print as well.

REFERENCES

- [1] ABFONEMA: OVEIII c Speech Synthesizer Manual. Type 21001.
- [2] ANDERS, B.: Digitale Sprachsynthese für Low-Cost Anwendungen. Bauelemente der Elektronik 7. 1981, 246-50.
- [3] ARATÓ András -- KISS Gábor -- TAJTHY Tamás : A MEA 8000 beszéd szintetizátor Commodore 64 számítógépen működő fejlesztő rendszere. Magyar Fonetikai Füzetek (MFF) 15.

1986, 143--54.

[4] ~~ASTHEIMER~~, A. Sprachsyntese in LPC-Technik, Elektronik 12.
1981, 73-81.

[5] ~~BOLLA~~ Kálmán : Folyamatos beszéd szintetizáló rendszer
magyar nyelven (VOXON). MFF 10, 118--29.

[6] ~~BRUKV~~^S. H.E.--TEULING, D.J.A.: Integrated voice synthesizer.
Philips, Technical publication 48. Electronic Components
and Applications Vol. 4 No. 2, February 1982.

[7] ~~FÓNAGY~~ Iván : Utószó. In.: Laziczius Gyula.: Fonetika.
Budapest, 1963, 189--206.

[8] ~~FONS~~, K.--GARGAGLIANO, I.: Articulate Automata: An Overview
of Voice Synthesis, BYTE Publications Inc. 1981.

[9] ~~KISS~~ Gábor : Parol-sintezo kun nelimigata vortaro en la
spiegulo de la hungara lingvo. In.: Perkomputila
Teksto-prilaboro, redaktis Koutny I., Budapest,
1985, 33-47.

[10] ~~KISS~~ Gábor: A magyar magánhangzók első két formánsának
meghatározása szintetizált hangmintákat felhasználó
percepcióos kísérlet segítségével. Nyelvtudományi
Közlemények 87/1. 1985, 159--70.

[11] ~~KISS~~ Gábor --OLASZY Gábor. Interaktív beszéd szintetizáló
rendszer számítógéppel és OVE III szintetizátorral. MFF
10. 1982, 21-46.

[12] ~~KISS~~ Gábor --OLASZY Gábor.: A HUNGAROVox magyar nyelvű,

valós idejű, párbeszédes beszéd szintetizáló rendszer.

Információ Elektronika 2. 1984, 98-111.

[13] JOLASZY Gábor : A magyar beszéd leggyakoribb hangsorépítő
elemeinek szerkezete és szintézise. Nyelvtudományi
értekezések 121. Budapest 1985.

[14] JOLASZY Gábor.--PODOLECZ György : A SCRIPTOVox MEA 8000
beszédelőállító rendszer felépítése és hangelemtára. Kép-
és Hangtechnika 6. 1986, 49-61.

[15] Revised International Manual of Braille Music Notation
1956. World Council for the Welfare of the Blind, Paris.

AN INTERLINGUAL TYPOLOGICAL EXAMINATION OF VOWELS

Gábor Kozma

Phonetics Department, Eötvös Loránd University

THE PROBLEM

A special way of exploring articulatory features is a comparative investigation in the course of which similarities and differences in the articulatory and acoustic features of sounds belonging to the same type can be observed from a new aspect by emphasizing their distinctive properties. In addition, new features or new correlations among known features can be elucidated by the comparison of sound types. To compare certain sound types across languages is a further possibility. In addition to the investigation of known features from a new aspect or the discovery of new properties, that method also yields an interpretation of linguistic phenomena concomitant with the specific realizations of the analysed sound type in various languages. The realization of language in speech, the physiological mechanisms of its production, and the form of the produced acoustic result are determined by the biological possibilities of the human organism. Every language builds up its own specific sound system and set of phonological rules from co-ordinational and perceptual functional mechanisms based on this common biological foundation. The exploration

and scientific description of the similarities and differences arising from the universal biological faculty of sound production and the phonetic peculiarities of particular languages constitute the subject-matter of this paper.

THE PURPOSE OF EXAMINATION

In our analysis we have examined sounds belonging to five types ("i, e, a, o, u") in German, Hungarian, Russian, and Polish. By comparing the articulatory features, we want to answer the following major questions (in addition to the description of distinctive articulatory properties of individual sounds):

a) How do the statements describing the relationship between vowel quality and the functioning of speech organs obtain within the types and languages examined?

b) What are the limits within which we can say that a sound is articulated in the palatal, central, or velar region? How can we interpret 'more palatal' or 'more velar' articulatory character within a palatal or velar sound type?

c) How can we characterize the connection between the place of articulation of vowels and the height of the tongue?

d) How should we interpret the labial or illabial articulatory character of a sound? What are the characteristics defining labial or illabial articulation?

e) What are the articulatory processes which result in sounds of similar acoustic properties? Is the similarity of acoustic effect based upon similar articulation or upon a specific co-operation of different articulatory movements of the

speech organs?

THE COURSE AND METHODS OF EXAMINATION

The questions listed above are naturally connected with each other which partly determines the way of their examination. In the course of the analysis we state the peculiarities of sounds belonging to the type within each language first, then we compare them interlingually. Assigning them to the same type was carried out on the basis of acoustic features. Then we examined the connections between their articulatory features and acoustic characteristics on the basis of data obtained from dynamic spectrum analysis and a computerized analysis of cinelabiographic and cineradiographic recordings. The phonetic comparisons were made in terms of an exact etalon: the features given in the Universal Phonetic Standard proposed by Kalman Bolla. We also compared the corresponding sound types of the four languages with one another.

In the cineradiographic analysis we examined the distances between pairs of measurement and reference points characteristic of the position of the articulatory organs (from the standard set of 32 such reference points, only those indicated in Figure 1 are relevant for our present purposes). The area relations of the sagittal section of the supraglottal cavities can be determined by the help of the same type of examination (Bolla--Földi--Kincses 1986): labial (1), palatal (2), velar(3), pharyngeal (4), and uvular (5) regions.

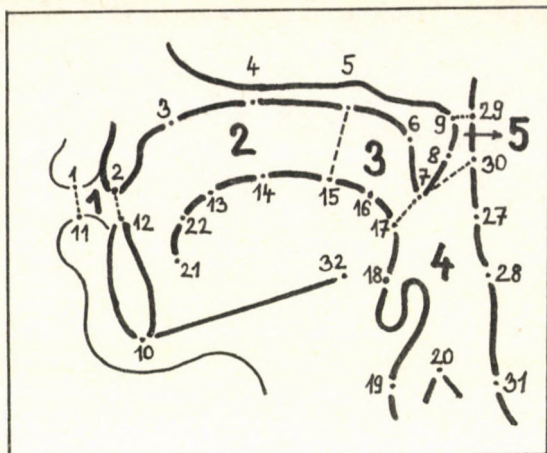


Figure 1. Radiogram scheme for the analysis of the position of the articulatory organs and the area relations of supraglottal cavities.

The movements of the lips were analysed on the basis of dynamic labiographic recordings (Bolla 1980). Our examination is connected with the interlingual phonetic research conducted in the Department of Phonetics of the Institute of Linguistics of the Hungarian Academy of Sciences for nearly a decade now, and recently also in the Ponetics Department of Eotvos Lorand University. In our interlingual comparisons we rely on the results of recent surveys of individual languages, the acoustic and articulatory data-base accumulated so far. The Hungarian, Russian, Polish, and German speech sounds are presented in independent cospectuses and papers by K. Bolla, É. Földi, and L. Valaczkai (Bolla 1980, 1982; Bolla--Földi 1981; Bolla--Valaczkai 1986). The

objective basis of comparison in our examinations was supplied by the features given in the Universal Phonetic Standard (Bolla 1984).

THE ANALYSIS OF ARTICULATORY FEATURES

The fact that the sounds analysed here do in fact belong to the same type in each case is shown by their acoustic features, i.e. by the closeness of the F1, F2 values concerned. The values can be seen in Table 1, the position of sounds in relation to each other is shown in Figure 2. In the case of the palatal types, "e" and "i", a glance at the distribution of sounds in the F1--F2 system of co-ordinates reveals that the majority of vowels belonging here are between or around pairs of reference vowels of the phonetic standard ([e] and [ɛ], respectively [i] and [ɪ]). The F1 values are very slightly scattered in the "i" type.

Table 1.

	UPhS					GERMAN					HUNGARIAN					RUSSIAN					POLISH				
"i" type	[i]	[i]	[I]	[i]	[i]	-	[i]	[i]	-	[i]	[i]	-	[i]	[i]	[i]	[i]	[i]	[i]	-	[i]	[i]	[i]	[i]	[i]	[i]
F1 (Hz)	259	206	280	200	230	-	265	205	-	300	291	460	238												
F2 (Hz)	1695	2770	2000	2400	2400	-	2480	2390	-	1796	2468	2200	2540												
"e" type	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]
F1 (Hz)	440	659	450	395	590	460	365	550	655	389	550	540	230												
F2 (Hz)	2400	1897	2290	2495	1900	2010	2340	2050	1640	2075	1849	1900	2350												
"u" type	[u]	-	[u]	[u]	[u]	-	[u]	[u]	-	[u]	-	[u]	-	[u]	-	[u]	-	[u]	-	[u]	-	[u]	-	[u]	-
F1 (Hz)	206	-	330	280	200	-	230	200	-	308	-	360	-												
F2 (Hz)	519	-	700	700	600	-	690	645	-	979	-	820	-												
"o" type	[o]	[o]	[o]	[o]	[o]	-	[o]	[o]	[o]	[o]	-	[o]	-	[o]	-	[o]	-	[o]	-	[o]	-	[o]	-	[o]	-
F1 (Hz)	654	449	580	445	390	-	460	390	590	449	-	552	-												
F2 (Hz)	692	582	1100	850	780	-	740	645	850	825	-	1100	-												
"a" type	[a]	-	[a]	[a]	-	-	[a]	-	-	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]	[a]
F1 (Hz)	880	-	715	765	-	-	840	-	-	734	599	645	690												
F2 (Hz)	1337	-	1290	1450	-	-	1310	-	-	1345	1745	1204	1625												

Acoustic features

Concerning the velar types "u", "o", and "a", the figure shows that of all members of the "u" type (high vowels) it is the [u] of the phonetic standard that has the lowest F2 value. In the "o" type (mid vowels) the [o] of the standard has the highest F1 and the [o] of the standard has the lowest F2. In the "a" type (low vowels) the [a] of the standard has the highest F1 with a relatively low F2.

The following general statements can be made on the basis of the known connections between the frequency values of the first and second formants, representing vowel quality, and the functioning of the speech organs (Joos 1984):

- The degree of opening between the jaws is in direct proportion to the F1 value.
- The F1 value is inversely proportional to the height of the tongue.
- The F2 value of palatal vowels is higher than that of velar vowels.

-- F1, F2 values are inversely proportional to the degree of labialization.

The four statements above are obviously interrelated. The connections between vowel quality and the functioning of the speech organs are subject to some restrictions, however. For instance, the third statement concerning the horizontal movement of the tongue, i.e. the place of articulation, does not specify whether, in the palatal region, sounds of a lower F2 value are articulated 'less palatally' (a bit further back), and conversely, whether sounds of a higher F2 value within the velar region are articulated fronter, i.e. shifted in the palatal direction. We can answer this question once we have looked at the palatal, central, and velar vowels separately.

Palatal vowels

Let us begin our analysis with the German "e" type. The radiographic articulatory data can be seen in Table 2. All four sounds ([e], [e:], [ɛ], [ɛ̃]) are articulated in the palatal region, as the radiograms show. Let us suppose that from the F2 values we can infer the changes in place of articulation within the palatal region. The sounds can be ordered as follows on the basis of their diminishing F2 value, i.e. the expected shift of the place of articulation backward (in the velar direction):

[e:] > [e] > [ɛ] > [ɛ]

(1)

P \longrightarrow U

(The German "e" type in terms of F2 values)

The actual place of articulation, i.e. the narrowing between the tongue and the palate can be characterized by distances No. 3--13 and 6--16 (Table 2). These values correspond to the prediction made on the basis of the F2 values in the German "e" type (1).

Table 2.

	„i“ type										„e“ type									
	UPhS		GERMAN			H		R		P	UPhS		GERMAN			H		R		P
	[i]	[i]	[I]	[i]	[i]	[i]	[i]	[i]	[i]	[i]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]	[e]
2-12	75	75	400	300	200	100	133	200	50	167	100	100	150	400	300	500	600	167	267	267
2-22	143	57	400	400	400	89	67	157	100	100	56	71	114	600	500	500	700	78	122	122
12-22	175	50	100	100	200	117	67	133	100	100	50	50	50	200	300	100	200	83	83	83
3-13	183	29	133	100	67	67	33	117	67	120	60	71	157	133	133	200	200	167	267	367
4-14	86	29	83	67	-	50	50	71	43	150	75	71	157	83	67	133	117	67	167	217
5-15	60	60	83	50	50	80	100	25	100	125	125	80	120	117	83	100	117	80	180	240
6-16	650	800	367	400	433	700	750	400	800	325	350	750	650	433	733	367	400	700	550	800
17-27	154	169	140	130	-	200	222	43	100	115	123	138	108	120	140	90	100	200	167	189
T3/T2	1,0	1,0	0,37	0,44	0,5	0,57	0,1	0,3	0,8	0,5	1,0	1,0	0,6	0,7	0,5	0,5	0,7	0,37	0,39	0,41
V/H	-	-	9,9	8,1	10,1	10	11	5,2	5,16	8,6	13	-	-	8,4	11,3	8,4	11,1	10	14	16

The articulatory features of „i“ and „e“ types

UPhS = Universal Phonetic Standard, R = Russian, H = Hungarian, P = Polish

Table 3.

	„u" type								„o" type									
	U	G			H		R	P	UPhS		G			H			R	P
	[u]	[v]	[u]	[u:]	[u]	[u:]	[u]	[u]	[o]	[o]	[o]	[o]	[o:]	[o]	[o]	[o:]	[o]	[o]
2-12	100	300	200	200	100	133	150	133	250	225	400	400	400	267	133	233	200	233
2-22	257	1900	2000	2000	200	244	243	200	300	314	2100	2100	2100	244	222	267	286	167
12-22	350	1700	1800	1800	250	300	233	250	350	400	1800	1800	1800	283	283	300	283	183
3-13	229	467	533	533	400	500	217	280	300	286	533	567	533	600	467	567	283	260
4-14	186	217	217	217	200	217	157	350	271	229	267	233	233	283	217	233	171	400
5-15	60	133	100	67	140	120	125	150	180	120	167	133	117	240	140	160	100	175
6-16	300	267	233	167	300	300	700	100	400	300	300	233	200	350	300	200	300	200
17-27	92	60	90	60	100	100	57	46	54	54	60	50	40	56	56	56	36	31
T3/T2	1,0	1,01	0,95	0,67	1,3	1,0	0,97	0,6	1	1	1,1	1,03	1,01	0,7	1,4	0,9	0,5	1,3
V/H	-	10,4	11,6	5,78	12,3	17	6,67	14,3	-	-	12,5	10,4	10,4	20	13,6	23,9	6,7	16,7

The articulatory features of „u" and „o" types

The following order of palatality is expected from the F2 values in the Hungarian "e" type:

$$\begin{array}{c} [\text{e:}] > [\text{e}] > [\text{ɛ}] \\ \text{P} \longrightarrow \text{U} \end{array} \quad (2)$$

(The Hungarian "e" type in terms of F2 values)

The changes in the positions of the tip of the tongue and the root of the tongue characterize the place of articulation as follows (data No. 2--22, 17--27):

$$\begin{array}{c} [\text{e:}] > [\text{ɛ}] > [\text{e}] \\ \text{P} \longrightarrow \text{U} \end{array} \quad (3)$$

(The Hungarian "e" type in terms of 2--22 and 17--27)

Comparing this order of palatality with the order under (2) based on the F2 values, we can see that in (3) [e] seems to be of the least palatal character. At the same time, the order based on values No. 3--13 and 4--14 (Table 2), measured at the prepalatal and palatal points, respectively, corresponds to statement (2), because the distances are really the smallest for [e:] and the largest for [ɛ]. The values for [e] are between the two. In addition to the discrepancy between (2) and (3), two further phenomena are also to be accounted for. The distance measured at the palatovelar border (5--15) is also the smallest for [e:] and the largest for [ɛ]; however, the distance measured at the velum (6--16) yields yet another order, differing from both (2) and (3). In answering these questions, we also have to

think of the fact that these sounds are characterized by different tongue positions, i.e. by different degrees of closeness or openness. The problem is further complicated by the fact that a sound pronounced with the same degree of jaw opening can be either palatal or velar (front or back vowel). Thus, for example, the [e] and [ɛ] members of the Hungarian "e" type, though they differ in place of articulation, are articulated with the same degree of jaw opening (267% in both cases, measured between 2--12). But the movement of the jaws is only one of the factors contributing to the degree of closeness. Although the tongue passively follows the movement of the jaws to some degree, its independent, active vertical movement is more significant. Therefore, the differences in radiographic data cannot be interpreted in themselves. Instead of discrete distances then, it makes more sense to examine the relative sizes of various areas of the sagittal section of the supraglottal tract. Being a more faithful representation of the actual changes of volume in the oral and pharyngeal cavities, this type of data reveals more about the essential properties of the process of speech. Hence, what is really relevant here is not the way the individual distances differ across vowels but the extent to which the size of the palatal (T2) or velar (T3) area predominates over other parts of the mouth cavity. If we compare the T3/T2 ratio characteristic of a sound to that of another similar sound, we can conclude that the vowel whose T3/T2 ratio is higher exhibits a strengthening of its palatal character (respectively, a weakening of its velar character) since the

relative increase of the velar area, with a corresponding decrease of the palatal area, indicates that the place of articulation is shifted in the palatal direction. Conversely, a vowel whose T3/T2 ratio is lower exhibits stronger velarity or weaker palatality, respectively.

Table 4.

	U	G	H	R	P			
"a" type	[a]	[a]	[a:]	[a:]	[a]	[ae]	[a]	[ae]
2 - 12	300	500	800	233	350	300	267	200
2 - 22	229	800	900	122	143	143	111	100
12 - 22	275	400	200	83	67	100	67	83
3 - 13	257	333	400	433	167	150	240	220
4 - 14	243	233	283	267	157	114	350	325
5 - 15	220	167	183	260	175	175	275	225
6 - 16	600	333	300	450	900	600	300	250
17 - 27	92	60	60	122	83	67	46	31
T3 / T2	1,0	1,03	0,83	0,86	0,92	0,53	1,08	1,15
V / H	-	11,1	12,3	17,4	6,67	6,86	12,3	15,4

The articulatory features of the "a" type

The ordering based on the T3/T2 ratio, for members of the Hungarian "e" type, agrees with the palatality--velarity order (3).

Table 5.

	P → V	GERMAN	HUNGARIAN	RUSSIAN	POLISH
"i" type	F2	[i]=[i:],[I]	[i],[i]	[i],[i]	[i],[i]
	T3/T2	[i:],[i],[I]	[i],[i]	[i],[i]	[i],[i]
"e" type	F2	[e:],[e],[ε:],[ε]	[e],[e],[ε]	[e],[ε]	[e],[e]
	T3/T2	[e:],[e],[ε:]=[ε]	[e],[ε],[e]	[e],[ε]	[e],[e]

The "i" and "e" sound types characterized in terms of the horizontal movement of the tongue

Hence, contrary to conclusions drawn from the F2 values (order 2), it is not [ε] but [e] that seems to be of the least palatal character. Consequently, the connection between vowel quality and the functioning of speech organs (the proportional interrelatedness of F2 values and the horizontal movement of the tongue) is modified in this case because of the dominant character of some other physiological factor or factors. The line marked F2 in Table 5 contains the relevant sounds in the order of decreasing F2, i.e. the acoustic parameter referring to the degree of palatality. The T3/T2 line shows the place-of-articulation order of the sounds on the basis of their actual articulatory features.

According to the results of our examination, the members of the Polish "e" type are also articulated at another point of the mouth cavity than it could be expected

from the F2 values. As pointed out earlier, the two kinds of order are identical in the case of the German "e" type.

The question raised at the beginning of the examination referred exactly to this point: is there a proportional connection between the F2 value of a sound and its exact place of articulation within a certain articulatory region (e.g. within the class of palatal vowels)? In the German "e" type we have found that connection. On the other hand, in the Hungarian and Polish "e" types, we have found a discrepancy between actual articulatory features and those expected on the basis of formant values. Hence, there must be some phonetic property responsible for that peculiarity of the sound types in question.

In the F1--F2 system of co-ordinates above, we have also indicated the grouping of vowels according to their articulatory features. We can see in Figure 2 that the scatter of the German "e" type approximately corresponds to the box of mid tongue position in the palatal region of the vowel chart, i.e. the sounds of that type (with the exception of [ɛ]) can be found within one degree of height of tongue. The scatter of the Hungarian "e" type is much larger; the [e:] sound is at the upper edge of the box, the [e] sound is at the lower edge of the same box, and the [ɛ] sound is in the middle of the box of low vowels (and on the borderline between palatal and central vowels). The marked differences in the degree of closeness of the members of the Hungarian "e" type can be the phonetic feature in which the Hungarian and German "e" types differ from each other, and which

results in the fact that the sounds can be characterized by some articulatory properties other than the ones expected on the basis of the F2 values. Thus, although on the basis of the F2 values we can indeed assume an order in terms of place of articulation, we can only do this with vowels of the same phonetic quality, i.e. within the limits of one height-of-tongue category. The transition between the degrees of height of tongue obviously does not constitute a sharp boundary, as demonstrated by the features of the German [] sound in relation to those of the other members of its type where only one difference can be found between the orders based on the F2 values and I3/I2 values (cf. Table 5): according to articulatory features [ɛ:] and [ɛ] are articulated in the same place, i.e. their degree of palatality is the same, whereas on the basis of F2 values a different degree of palatality is expected. The German sound can be found at the upper edge of the box of low tongue position which means a lesser difference from the other members of the sound type than in the case of the Hungarian "e" type, but the equivalent of even this slight difference can be found in the articulatory features.

This assumption is supported by the analysis of the "e" types of other languages. The orders based on the F2 values and I3/I2 values are identical for the two relevant sounds of the Russian "e" type ([e] and [ɛ]) and different in the Polish "e" type ([e] and [ɛ]), cf. Table 5. From the scatter of these sounds we can see that the two Russian sounds are within one degree of height of tongue (mid position), while

Polish [e] appears in the middle of the box of high tongue position, and Polish [e] can be found at the lower edge of the box of mid tongue position. That situation corroborates the assumption we have made on the basis of the Hungarian and German "e" types.

In the "i" type (cf. Table 5), the F2 and T3/T2-based place of articulation orders are identical in German and Russian. The sounds can be found within the box of high tongue position in both languages, though the Russian [ɨ] sound falls in the central region. We expected Hungarian [i] to be articulated fronter than [i:] on the basis of the F2 values, but according to the articulatory features the [i] sound is articulated a bit further back. It is also contrary to our expectations that Hungarian [i] and [i:] are nevertheless within the same box near each other. We also find a difference in the two kinds of orders for Polish [i] and [ɨ]: [i] falls in the box of high tongue position, and [ɨ] in the box of mid tongue position, which is in accordance with our statements made at the "e" type.

Velar vowels

The acoustic and articulatory features of the sounds will be compared on the basis of the following table.

Table 6.

	P→V	G	H	R	P
u" type	F2	[ʋ]=[u],[u:]	[u],[u:]	[u]	[u]
	T3/T2	[ʋ],[u],[u:]	[u],[u:]	[u]	[u]
o" type	F2	[ɔ],[o],[o:]	[ɔ],[o],[o:]	[o]	[o]
	T3/T2	[ɔ],[o],[o:]	[o],[o:],[ɔ]	[o]	[o]
a" type	F2	[a:],[a]	[a:]	[æ],[a]	[æ],[a]
	T3/T2	[a],[a:]	[a:]	[a],[æ]	[æ],[a]

The "u", "o", and "a" sound types characterized in terms of the horizontal movement of the tongue

The two kinds of orders of the German and Hungarian sounds agree in the "u" type. These sounds can be found within the box of high tongue position. The Russian and Polish "u" types have a single member each, so there is no possibility for comparison within those languages.

There is no such possibility in Russian and Polish in the "o" type, either. There orders based on the F2 and T3/T2 values agree in German, though German [ɔ] is at the edge of low tongue position in the central region. In the Hungarian "o" type [ɔ] is expected to be of the most palatal character on the basis of the F2 values, but according to the T3/T2 values it is articulated further back than [o] and [o:]. We can see that the [ɔ] sound is in the box of low tongue

position.

In Hungarian "a" type of 'lowest' tongue position contains a single vowel: [a:]. The F2 and I3/I2 orders of the sounds of this type agree in Polish, but they differ from each other in German and Russian. According to the articulatory data, German [a:] is produced further back than [a], and we can also see a difference in the degree of openness in Figure 2: they are on the two sides of the boundary of 'low' and 'lowest' tongue positions. The [a] and [æ] sounds of the Russian language are both in the box of low tongue position but [æ] is nevertheless articulated further back than [a] in terms of the I3/I2 values, contrary to the F2 order. Incidentally, this fact is distinctly visible on the radiograms as well.

Thus, we can observe a proportional relationship between the F2 values and the place of articulation within the palatal, central and velar articulatory regions. But this is valid only for sounds belonging to the same degree of height of tongue. On the other hand, in the Hungarian "i" and Russian "a" types, such proportional relationship between the F2 values and the place of articulation cannot be observed even within the same height-of-tongue category. But we can see in Figure 2 that, if we draw a straight line connecting any two sounds belonging to the same type in any one language, it will ascend left-to-right in the case of palatal vowels, with the single exception of the line connecting Hungarian [i] and [i:] which ascends right-to-left. Similarly, the line connecting Russian [æ] -- which is

palatal and illabial but belongs to the "a" type -- with the central vowel [a], the other member of its type, ascends right-to-left, whereas velar types (including "a") are otherwise characterized by a left-to-right ascent. (The line connecting Polish [ɛ] and [a] ascends left-to-right like all other pairs in the "a" type and in velar types in general.)

Analysis in terms of lip shape

The ratio of the vertical and horizontal diameters of the labial orifice (cf. the V/H line of the tables) is a characteristic property of the articulation of vowels but it is not sufficient for the purposes of a typological and interlingual comparison as the sole parameter for determining their labial or illabial character. This is well demonstrated by the coincidence of the appropriate values of some labial and illabial vowels. For instance, the V/H ratio is 23.89 for Hungarian [o:]; 12.29 for Hungarian [u]; 6.67 for Russian [o]; 15.97 for Hungarian [ɛ]; 13.15 for Polish [i]; 8.07 for German [i:], etc. How should we interpret these observations?

The labiality vs. illabiality of a vowel can also be inferred from the F1, F2 values: the frequency of the first and the second formants is inversely proportional to the degree of labialization. The implications of his statement are as follows:

1. This statement covers cases where F1 and F2 both change in the same direction, i.e. either both increase or both decrease.

2. It does not cover cases where the two values change in the opposite direction.

3. A lower F2 entails a more labial articulation.

4. A higher F2 entails a less labial articulation.

5. A lower F1 entails a more labial articulation.

6. A higher F1 entails a less labial articulation.

Consider the Hungarian "e" type as an example. Here the changes of F1 and F2 are of the opposite direction, so the correlation stated above for changes of the same direction applies only in an indirect form. The physiological connections predicted in statements No. 3 and 4 were found to hold true in our examinations, i.e. a sound whose F2 is lower is of a more labial character, as a comparison of U/H proportions within the Hungarian "e" type reveals (Table 2). However, for F1 whose changes are of the opposite direction, we found more labial articulation correlated with higher F1, although statement 6 above made us expect the opposite to be the case. This means that the tendency of the F1 values towards illabiality is suppressed by the tendency of F2 towards labiality. If the degree of jaw opening, i.e. the distance between the upper and lower lips correlates with the changes of F1, we can assume that the distance between the corners of the mouth, i.e. the horizontal diameter of the lip aperture is similarly linked to the F2 value. We can infer from the foregoing that the change of the distance between the corners of the mouth plays a dominant role in forming the character of labialization. (Of course, these relations are further complicated by the connections between the changes in

the form of vertical and horizontal main directions of the opening between the upper and lower lips.)

Typological comparisons

In the formant values of the sound types examined, we find considerable overlap between the "i" and "e" types. There is not too much overlap between the "e" and "a" types, and there is no overlap at all among the "o", "u", and "a" types. We can see that the Polish [e] sound is in the zone of high tongue position among the sounds of the "i" type, closest to Hungarian [i:], and German [i] and [i:]. In other words, these sounds are quite close to each other in terms of their acoustic features as expressed by F1, F2. Let us further take into consideration the fact that, according to commonly held views concerning the connection between vowel quality and the functioning of speech organs, these sounds are articulated almost at the same place with almost the same degree of height of tongue (closeness). To analyse the way these articulatory mechanisms and the acoustic form of the sounds are related to each other, we have to consider the aspects summarized in Table 7.

Table 7.

	Acoustic features		Articulation		All acoustic features
	F1, F2	other	tongue	other	
1.	+	+	+	+	+
2.	+	+	+	-	+
3.	+	+	-	-	+
4.	+	-	-	-	-
5.	+	+	-	+	+
6.	+	-	+	+	-
7.	+	-	+	+	-
8.	+	-	+	-	-

The comparison of acoustic and articulatory features of two sounds (+ = similar, - = different)

The acoustic features expressed in terms of the F1 and F2 parameters are represented in the first column of the table, all other acoustic features in the second column, the articulatory features that can be characterized by horizontal and vertical movements of the tongue can be found in the third column, all other articulatory features in the fourth column, and the overall similarity or difference of the two sounds (as defined by the first two columns taken together) can be seen in the fifth column. (The table contains all possible combinations of the above factors. Lines 5 and 6, of course, make no sense.) In those cases where the total acoustic patterns of two sounds are similar (lines 1, 2, 3, and 5 of Table 7) we have to do with different forms of realization of what is acoustically the same sound. When the total acoustic patterns of two sounds are different from each

other, we speak of different sounds, certain articulatory and acoustic features of which may coincide.

Consequently, on the basis of the first line, the Polish [e] sound should be identical with (similar to) the sounds around it in the F1--F2 scatter diagram both in its articulatory and acoustic features. Hence, one and the same sound should belong to the "e" type in Polish and to the "i" type in Hungarian and German. Since this is not the case, the assumption made in the first line of Table 7 is to be rejected.

According to the second line, the horizontal and vertical movements of the tongue play a similar role, and the rest of the articulatory factors (velum, pharyngeal cavity, volume of the mouth cavity, lips etc.) differ, but their co-operation results in a completely similar acoustic pattern (including F1, F2, and all other acoustic factors). However, on the basis of the cineradiographic data (3--13, 4--14, 5--15 values of Table 2) it turns out that the Polish [e] sound is more open than the members of the Hungarian and German "i" types. Consequently, the third line of the table should, in principle, contain a more acceptable assumption: the completely similar acoustic pattern is based on a specific constellation of different articulatory mechanisms both in terms of the movement of the tongue and in other factors. This assumption must be rejected, too, because the acoustic pattern cannot be the same in our case since we do not speak of different forms of realization of the same sound (or similar sounds).

The assumption in the fourth line is the most acceptable: we have to do with sounds agreeing in their F1, F2 values but differing in other acoustic features, i.e. sounds whose place of articulation, degree of closeness and other articulatory features are different. Incidentally, our earlier articulatory analysis has actually demonstrated that the articulation of Polish [e] involves activities of the speech organs differing from what could be expected on the basis of their F1, F2 values. These different articulatory mechanisms result in different acoustic features -- with the exception of the F1, F2 values.

As we have already pointed out, lines 5 and 6 are uninterpretable. The seventh line would be true if the Polish [e] sound and whatever is found next to it on the scatter diagram agreed in all the other articulatory features apart from place of articulation and degree of closeness. But there is at least one articulatory feature in which these sounds differ: the U/K value (cf. the last line of Table 2) referring to the degree of labialization is much lower in the German and Hungarian "i" types than with Polish [e].

The eighth line does not correspond to linguistic reality either -- since, as we have seen, the articulatory features expected on the basis of the F1, F2 values differ from the actual ones for the Polish [e] sound.

In sum, we have to accept the assumption represented in the fourth line as true. Thus, in the case of Polish [e], we cannot safely infer the articulatory features from the F1, F2

values. In addition to the horizontal and vertical movement of the tongue, other articulatory factors probably also play an important role in forming the first and second formants, and lip shape is definitely one of them. Another reason why the characteristics of the [e] sound cannot be inferred from the F1, F2 values is that both its articulatory configuration and its overall acoustic character are different from those of the other sounds found in its vicinity in the vowel chart -- a fact that points to the paramount significance of subsidiary articulatory properties and the language-specific articulation base underlying them.

By way of a summary of the analysis we have just carried out in terms of Table 7, the following guidelines can be offered for the analysis of any particular sound which resembles Polish [e] in that it can be found in the F1--F2 vowel chart area of another sound type.

-- First, it has to be found out by a sound-type-internal analysis whether the potential articulatory characteristics expected on the basis of the F1, F2 values do in fact tally with the actual articulatory characteristics of the vowel concerned. If any discrepancy is found, there is no need to compare its articulatory features with those of the sounds found in its F1, F2 environment since they will necessarily be different (cf. Table 7, line 4).

-- Then, by looking at any articulatory feature the analysis of which is comparatively straightforward -- e.g. lip movements --, it is to be established whether there is

some difference in at least one "subsidiary" articulatory mechanism (i.e. other than vertical and horizontal tongue movement). If the answer is yes, it is line 4, rather than line 7, of Table 7 that has to be accepted as valid.

Consider two examples of the type of analysis just outlined: the cases of Polish [ɛ̃] and Russian [æ̃] suggest themselves as a rewarding testing-ground. As for Polish [ɛ̃], the potential articulatory characteristics expected on the basis of the F1, F2 values do not correspond to actually attested ones, as Table 5 shows. Consequently, the articulatory characteristics of Polish [ɛ̃] will also differ from those of German [e] and [ɛ̃] which constitute its F1, F2 chart environment. However, the U/H ratios -- indicating lip shape -- of these sounds do not differ significantly. As we have seen in the discussion of the role of labiality (section 4.3), the U/H ratio is not always a reliable basis for comparing two sounds; we have also pointed out there that the change of distance between the corners of the mouth is the single most important factor in determining labiality. Expressed in the percentage of rest position distance, the relevant data are 119% and 135% for German [e] and [ɛ̃], respectively, and 82% for Polish [ɛ̃]. Hence the latter is of a more labial character. In sum, Polish [ɛ̃] is articulated with a tongue position different from what is expected on the basis of its F1, F2. The only acoustic feature [ɛ̃] shares with its vowel chart neighbours is its value for F1, F2; the rest of its acoustic properties, not shared by its neighbours, are not only based on the difference in tongue

position but also on further articulatory features (e.g. degree of lip rounding) which also contribute to what is called 'articulation basis'.

Along the same lines, the investigation of Russian [æ] yields a similar result. This type of analysis of Polish [e] and [i], and Russian [æ], has also contributed to the interpretation of the otherwise puzzling overlaps between F1--F2 scatter areas of the "i", "e", and "a" sound types.

RESULTS

In this paper we have showed certain possibilities of realization of the known correspondences between vowel quality and the functioning of speech organs and the way individual correspondences may be interrelated. We have also presented data obtained from an exact comparison of articulatory and acoustic features of vowel sounds, reflecting specific differences within certain articulatory features of the languages examined. In addition to a better understanding of the elements of various vowel systems, the specific way articulation bases of individual languages operate has been indicated somewhat more precisely.

We have proved that there is a connection between F2 values and the forward or backward shift of the place of articulation within the palatal and velar articulatory regions. We have also pointed out, however, that this connection is only valid for vowels belonging to the same sound type and almost the same height of tongue. In such comparisons we can never predict the concrete change of the

place of articulation but rather indicate the character of the palatal or velar tendency becoming dominant. Therefore, instead of defining a certain sound in anatomically concrete place-of-articulation terms, it appears to be more correct to point out the region of the mouth cavity becoming dominant in the course of articulation. The assumption underlying this point is that place of articulation is only one of the factors of sound production which cannot be objectively evaluated in itself and which co-operates with a number of further physiological factors of various types in producing the palatal and velar areas of appropriate proportions which, in turn, interact with other areas (pharyngeal, labial, nasal, etc.) and speech organ mechanisms to produce the acoustic form of a speech sound. Our typological comparisons have shed light on the role certain articulatory features of the sounds examined have in forming phonetic properties.

Our methods of examination provide a clear and straightforward way for the objective analysis of linguistic facts. These methods improve our knowledge of the individual sounds not only by characteristics inferred from typological and interlingual comparisons, but they also make it possible to elucidate the specific interrelatedness of articulatory features from a novel aspect.

REFERENCES

- BOLLA Kálmán: Magyar hangalbum. MFF 6. 1980
- BOLLA Kálmán: Orosz hangalbum. MFF 11. 1982.
- BOLLA Kálmán: Egyetemes Fonetikai hangszabvány?
A magánhangzók. MFF 13. 1984, 71--120
- BOLLA Kálmán--FÖLDI Éva: A lengyel beszédhangok képzési és akusztikus sajátosságairól. MFF 7. 1981, 91--139.
- BOLLA Kálmán--FÖLDI Éva: A lengyel beszédhangok ajakartikulációja. MFF 8. 1981, 104--46.
- BOLLA Kálmán--FÖLDI Éva--KINCSES Gyula: A toldalékcso-
artikulációs folyamatainak számítógépes vizsgálata. MFF
15. 1986, 155--65.
- BOLLA Kálmán--VALACZKAI László: Német beszédhangok atlasza.
MFF 16. 1986.
- FANT, G.: Acoustic Theory of Speech Production. With
Calculations based on X-Ray Studies of Russian
Articulations. 's-Gravenhage 1960.
- FANT, G.: Speech Sounds and Features. Cambridge,
Massachusetts, London 1973.
- JOOS, M.: Acoustic phonetics. Baltimore 1948. (Supplement to
Language: vol. 24, No. 2. suppl.)
- LINDBLOM, B.--SUNDBERG, J.: Acoustical consequences of lip,
tongue, jaw and larynx movement. JASA 50. 1971, 1166--79.
- STRENGER, F.: Radiographic, palatographic, and labiographic
methods in phonetics. In: Manual of Phonetics. Ed. by
MALMBERG, B. 1968, 334--64.
- SZENDE Tamás: A köznyelvi magyar ejtésnorma felé. NyK LXXI,
1969, 345--85.

ON THE SPEAKING MODULE OF AN AUTOMATIC READING MACHINE

Gábor Olaszy

Linguistics Institute, Hungarian Academy of Sciences

Géza Gordos

University of Technology, Budapest

The speaking module -- called SCRIPTOVOX -- of the automatic Hungarian reading machine was developed in the years 1983--86 by a four-member research team of electrical engineers headed by Dr. Géza Gordos (University of Technology, Budapest). The members of the team were Dr. Gábor Olaszy (Institute of Linguistics), György Podoletz (University of Technology) and György Takács (Research Institute of the Hungarian Post and Telecommunication).

The speaking module uses the general purpose, programmable MEA 8000 type integrated circuit for speech generation (cf. Table 1). SCRIPTOVOX was originally developed for the real time conversion of any Hungarian text -- given in ASCII codes -- into good quality speech. One important application is its use as a speaking output module of a Hungarian Automatic Reading Machine (ARM). This ARM reads printed text (one page) automatically.

In this paper the automatic Hungarian text-to-speech (TTS) conversion is discussed. The characters of the printed text on the page are converted into ASCII codes by an optical grapheme-to-ASCII character converter developed at the Institute for Computer Research, Budapest, which is not

discussed here.

Table 1. The code table of the MEA 8000

Step	FD (ins)(Hz/B ms)	PI	AM	F1 (Hz)	F2 (Hz)	F3 (Hz)	B (Hz)	SPI
0	8	0	0	150	440	1179	726	0
1	16	1	0,008	162	466	1337	309	1
2	32	2	0,011	174	494	1528	125	2
3	64	3	0,016	188	523	1762	50	3
4		4	0,022	202	554	2047		4
5		5	0,031	217	587	2400		5
6		6	0,044	233	622	2842		6
7		7	0,062	250	659	3400		7
8		8	0,088	267	698			8
9		9	0,125	286	740			9
10		10	0,177	305	784			10
11		11	0,250	325	830			11
12		12	0,354	346	880			12
13		13	0,500	368	932			13
14		14	0,707	391	988			14
15		15	1,000	415	1047			15
16		noise		440	1110			16
17		-15		466	1179			17
18		-14		494	1254			18
19		-13		523	1337			19
20		-12		554	1428			20
21		-11		587	1528			21
22		-10		622	1639			22
23		-9		659	1762			23
24		-8		698	1897			24
25		-7		740	2047			25
26		-6		784	2214			26
27		-5		830	2400			27
28		-4		880	2609			28
29		-3		932	2842			.
30		-2		988	3105			.
31		-1		1047	3400			.
32								.
.								.
.								.
255								510

F4 = 3500 Hz, FD = Frame duration, PI = Pitch increment,
AM = Amplitude SPI = Start pitch.

The SCRIPTVOX module appears as a small size card with 280 microprocessor, 12 kbyte EPROM and 8 kbyte RAM. It is matched via an RS 232 line to other digital machines or computers.

PRACTICAL TEXT-TO-SPEECH CONVERSION

The primary requirement a text-to-speech converter system has to meet is that it should convert every character of a text in a given language (including not only letters but other characters as well) into control codes with the aid of which intelligible speech can be generated by a speech synthesizer. At the same time an important requirement is that it should recognize the different types of sentences (statements, questions, etc.). This recognition is the basis of the automatic generation of melody and rhythm. Last but not least, a fundamental requirement is the real time operation of conversion and speech generation.

The conversion of ASCII characters of the text into synthesizer control codes is realized in the SCRIPTVOX system in three steps.

1. Conversion of ASCII characters into "phoneme codes".
2. Conversion of phoneme codes into MEA control codes (speech frames) and their concatenation.
3. Realization of melody patterns by changing the so-called pitch control bits of some speech frames of the group of frames concatenated in step 2.

CONVERSION OF ASCII CHARACTERS INTO PHONEME CODES

We use thirty-three phonemes for generating Hungarian speech (cf. Table 2). Only the short versions of speech sounds are included among these thirty-three phonemes, the long versions are represented by doubling the phoneme code of

the short counterpart. When processing the graphemes of the text into phoneme codes we distinguish three types of ASCII characters.

Table 2. The phonemes, letters and phoneme code numbers in the SCRIPTOVOX system

IPA letter phoneme sym- code bol	IPA letter phoneme sym- code bol	IPA letter phoneme sym- code bol
- - 1	b B 11	h H 23
	p P 12	v V 24
a: Á 2	d D 13	f F 25
ɔ A 3	t T 14	z Z 26
o O 4	g G 15	s SZ 27
u U 5	k K 16	ts C 28
y Ü 6	ʃ GY 17	ʒ ZS 29
i I 7	c TY 18	ʃ S 30
e: É 8	m M 19	tʃ CS 31
ø Ö 9	n N 20	l L 32
e E 10	ŋ NY 21	r R 33
	j J 22	

-- The first -- and simplest -- type comprises those characters with which a phoneme code can be associated directly in one step, e.g. A, A, O, U, F, H, etc.

-- The second type of characters cannot be converted directly into code numbers: their conversion requires an examination of the neighbouring characters. Examples of such characters are S, Z, C, T, etc. For instance, the letter S occurs in the

combinations SZ, ZZS, SSZ, ZZS, CS, CCS, denoting different sounds in each case.

-- The third group of ASCII characters includes numbers, abbreviations, and other symbols like %, +, -, =, ", :, etc.

In what way are these three types of ASCII characters handled in the process of conversion? The ASCII characters obtained from the text are stored in a 1 kbyte buffer (B1). A similar buffer is reserved for the phoneme codes (B2). Apart from these, we store a lexicon containing the groups of phoneme codes corresponding to the non-letter-type ASCII characters. When performing the ASCII code-phoneme code conversion the program examines every ASCII character step-by-step and converts them, in one or more steps, into a series of phoneme codes. If a non-letter (number, abbreviation, etc.) is found in B1, the program automatically inserts the appropriate group of phoneme codes from the lexicon into the appropriate place in B2. Finally in B2 we find the original text in a form as if everything in it had been written using only letters. For example, the sentence <MOST 12 ÓRA VAN> 'It's 12 o'clock now' takes the form MOST IIZENKETIÖÖ OORA VAN.

This algorithm works as follows:

1. Setting of initial values
2. Accepting ASCII codes into B1
3. Identification of ASCII characters by scanning left to right
4. If it is a special ASCII (letter ch.) then going to the rules for setting the appropriate phoneme codes into B2

5. If it is a number then going to the number routine where phoneme codes of the number will be set into B2
6. If some other symbol, then going to the lexicon where the appropriate phoneme codes from will be set into B2
7. Identification of punctuation marks (, . ! ?) and setting the appropriate code into B2

PHONEME CODE TO MEA CONTROL CODE CONVERSION

In the next step the program converts the contents of B2 into a series of speech frames which are stored in a 4 kbyte buffer (B3). For the conversion, a collection of speech frames (data base) and a 33x33x6 element concatenation matrix (rule system) is used. The data base consists of 225 different types of speech frames. The initial contents of the 225 speech frames and the rules were defined in 1983 and have continuously been refined thereafter. The rule system includes rules for the concatenation of frames picked from the data base when converting the phoneme codes of B2 into a series of frames.

The data base

Choosing the appropriate elements of the 225-element data base every sound and sound combination (VV, CV, VC, CC), as well as all the assimilations can be realized.

As regards timing parameters, mostly 32 ms frames were devised (cf. Table 3); 8 ms frames are used only for the

stops and some other short sounds or sound parts, 16 ms frames are mostly used in the CV and VC combinations, and 64 ms ones in VV combinations and in some spirants.

Table 3. The occurrence of frame duration in the data base

Frame duration (FD) in ms	8	16	32	64
Number of such frames	21	63	109	32

In connection with amplitude levels it should be noted that formant structure (set for the MEA 8000 by the user) influences the sound level. It is a specific feature of cascade formant synthesizers. As a consequence the values of amplitude parameters of the frames may deviate considerably from the intensity values measured in live speech. Efforts must be made to use the whole range of the fifteen available amplitude values as well. When designing the amplitude values of the frames of a ITS the whole set of sounds and sound combinations of the given language must be taken into account. The higher the amplitudes, the better the signal-to-noise ratio will become. Taking all these facts into account the highest amplitude level (out of fifteen available steps) is 14 in the SCRIPTOVIX system. This value occurs in the D-A combination. The amplitude level 13 is used only in nine frames. Very low amplitude level (2) is used for example in frames realizing the voice phase of the voiced stops. The lowest amplitude level is used only in one frame used for the realization of one part of [s] (for the explanation of this low value, see the section discussing the frequency codes below). Most frames (49 in number) of the

data base contain 0 amplitude level. These frames are used for the realization of the final parts of sounds, the silent periods of stops and affricates, and the silent periods between words, phrases, and sentences (cf. Table 4).

Table 4. Amplitude levels and the number of frames set at these levels

amplitude values	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
number of such frames	0	1	9	22	30	40	18	16	10	12	2	6	3	6	1	49

The set of control codes setting the frequency response of the cascade filter in MEA 8000 is very limited. The facts that only one filter structure is used and that the excitation sources cannot be mixed represent serious further restrictions. All these result in some difficulties when designing certain speech frames but very low bit rates can be achieved. Synthesizing Hungarian we had difficulties in some sounds such as [n], [m], [ŋ], [s], [t_s], [z]. A solution for nasals is presented in Table 5.

Table 5. The frames of sound sequences MA, NA, NYA

MA

START PITCH: 114

No.	F1	F2	F3	BW1	BW2	BW3	BW4	AM	PI	FD
01.	250	1110	2047	125	125	309	309	0	0	0
02.	250	1110	2047	125	125	309	309	4	0	2
03.	250	1110	2047	125	125	309	309	6	0	3
04.	250	1110	2047	125	125	309	309	7	0	2
05.	554	988	2400	50	50	50	50	11	0	2
06.	554	988	2400	50	50	50	50	12	0	2
07.	554	988	2400	50	50	50	50	10	0	2
08.	554	988	2400	50	50	50	50	0	0	0

NA

START PITCH: 114

No.	F1	F2	F3	BW1	BW2	BW3	BW4	AM	PI	FD
01.	250	1170	2047	125	125	50	50	0	0	0
02.	202	1428	2400	50	309	309	50	4	0	2
03.	202	1428	2400	50	309	309	50	6	0	2
04.	202	1428	2400	50	309	309	50	6	0	2
05.	391	1337	2400	125	125	50	50	10	0	1
06.	554	988	2400	50	50	50	50	10	0	2
07.	554	988	2400	50	50	50	50	12	0	2
08.	554	988	2400	50	50	50	50	10	0	1
09.	554	988	2400	50	50	50	50	0	0	0

NYA

START PITCH: 114

No.	F1	F2	F3	BW1	BW2	BW3	BW4	AM	PI	FD
01.	202	988	2842	309	309	309	50	0	0	1
02.	202	988	2842	309	309	309	50	4	0	2
03.	202	988	2842	309	309	309	50	6	0	3
04.	202	988	2842	309	309	309	50	6	0	3
05.	202	2214	2842	309	309	309	309	12	0	1
06.	440	1761	2400	125	125	125	125	10	0	2
07.	554	988	2400	50	50	50	50	11	0	2
08.	554	988	2400	50	50	50	50	12	0	2
09.	554	988	2400	50	50	50	50	10	0	2
0A.	554	988	2400	50	50	50	50	0	0	0

Because of the very low upper cut-off frequency of the MEA (3400 Hz), problems occurred in synthesizing the Hungarian sounds [s] and [ts] because these sounds normally have their energy maximums at 6000-8000 Hz. We solved this problem by setting both F2 and F3 at the highest available frequency value (3400 Hz), cf. Table 6. Setting two formants at the same frequency value results a rough rising in sound intensity. Therefore the sound level for [s] can be set correctly only by giving a reduced amplitude parameter value of the frames of [s]. (cf. Table 6).

Table 6. The speech frames of sound sequences

SZA and CA

SZA

START PITCH: 114

No.	F1	F2	F3	BW1	BW2	BW3	BW4	AM	PI	FD
01.	202	3400	3400	726	309	726	309	0	noise	2
02.	202	2842	3400	726	726	726	309	2	noise	2
03.	202	3105	3400	726	726	726	309	7	noise	2
04.	202	3105	3400	726	726	726	309	7	noise	2
05.	391	1428	2400	125	125	125	125	8	0	1
06.	554	988	2400	50	50	50	50	11	0	2
07.	554	988	2400	50	50	50	50	12	0	2
08.	554	988	2400	50	50	50	50	10	0	2
09.	554	988	2400	50	50	50	50	0	0	0

CA

START PITCH: 114

No.	F1	F2	F3	BW1	BW2	BW3	BW4	AM	PI	FD
01.	202	3400	3400	726	309	726	309	0	noise	2
02.	202	3105	3400	726	726	726	309	8	noise	2
03.	391	1428	2400	125	125	125	125	8	0	1
04.	554	988	2400	50	50	50	50	11	0	2
05.	554	988	2400	50	50	50	50	12	0	2
06.	554	988	2400	50	50	50	50	10	0	2
07.	554	988	2400	50	50	50	50	0	0	2

Two Hungarian sound types, [z] and [ʒ], need mixed excitation. Moreover, there exist long versions of both these sounds where mixed excitation should be applied for

approximately 100--150 ms. Hence these sounds are composed by relying on the perceptual mechanism of the listener. In the synthesized [z] and [ʒ] a voiced frame is followed by a noisy (unvoiced excitation) one (cf. Table 7). Thus the short versions of Hungarian [z] and [ʒ] appear in the synthesized speech with a tolerably good quality.

As regards the long versions of these sounds, the only possibility was to lengthen the voiced part by concatenating two or three voiced frames. The duration of the voiced frames in long [z] and [ʒ] was determined by experiment, starting from long duration and shortening it to the minimum value which still makes the impression of a long sound in the listener. By this compromise these long sounds are well intelligible in running speech.

Table 7. The speech frames of sound sequences

ZA and ZSA

ZA

No.	F1	F2	F3	BW1	BW2	BW3	BW4	AM	PI	FD
01.	391	1528	2400	125	125	125	125	0	0	1
02.	391	1528	2400	125	125	125	125	11	0	2
03.	202	3105	2400	726	726	726	726	8	noise	3
04.	391	1428	2400	125	125	125	125	8	0	1
05.	554	988	2400	50	50	50	50	11	0	2
06.	554	988	2400	50	50	50	50	12	0	2
07.	554	988	2400	50	50	50	50	10	0	2
08.	554	988	2400	50	50	50	50	0	0	0

ZSA

START PITCH: 114

No.	F1	F2	F3	BW1	BW2	BW3	BW4	AM	PI	FD
01.	391	1528	2400	125	125	125	125	0	0	1
02.	391	1528	2400	125	125	125	125	10	0	2
03.	267	1761	2400	309	125	125	125	13	noise	2
04.	267	1761	2400	309	125	125	125	13	noise	2
05.	391	1428	2400	125	125	125	125	8	0	1
06.	554	988	2400	50	50	50	50	11	0	1
07.	554	988	2400	50	50	50	50	12	0	1
08.	554	988	2400	50	50	50	50	10	0	1
09.	554	988	2400	50	50	50	50	0	0	0

The rule system for the concatenation of speech frames

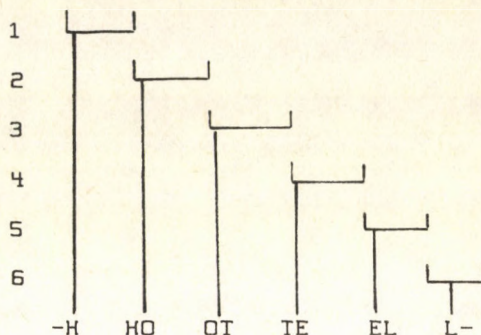
In order to get speech from the group of phoneme codes stored in B1 these codes must be converted into very many speech frames to be stored in the buffer B3. This buffer size is 4 kbyte (enough for 40 s of speech in one conversion process) in the SCRIPTOVox system. The rule system works as follows. The concatenation rules are incorporated in a 33x33x6 element matrix (cf. Figure 1) where every row and column represents a phoneme code. How is the rule matrix handled by the program? Before turning to the rule matrix the program makes a diad-like interpretation of the phoneme codes in B2.

Fig. 1. The rule matrix for concatenation of speech frames in the example of the word <HOTEL>.

Letters (ASCII codes) in B1: - H O T E L -

Phoneme codes in B2: 1 23 4 14 10 32 1

steps of conversion:



Diadic units:

-H HO OT TE EL L-

Rule matrix: (row, column):1,23;23,4;4,14;14,10;10,32;32,1

As can be seen in the example, by the step-by-step interpretation of pairs of consecutive phoneme codes a row and a column of the rule matrix are determined. This row-and-column pair points at a matrix entry, the contents of which are six bytes. These bytes represent the identifiers of speech frames that must be picked from the data base and placed into B3 one after the other. If the desired sound effect of a step during the conversion requires less than six speech frames, 0's are inserted in the superfluous bytes. If during execution the program finds 0 values it goes on to the next step. It should be noted that one complete sound combination is realised by the program totally after performing three steps: the step before the sound combination concerned, the step of its own, and the next one. This procedure will be shown in detail by taking the example of

the word <HOTEL>. The realization of the sound combination OT (cf. Table 8 solid line) requires the steps HO, OT, and IE.

Table 8. Speech frames realizing the sound combination OT

Diad	number of frames	Meaning of frames
-H =	30,225,226,226,0,0	225,226: parts of [h]
HO =	233,3,23,23,0,0	233,3,23: parts of [o]
OT =	20,40,40,30,0,0	20,40: silent periods
IE =	71,72,29,9,29,0	71,72: burst of [t]
EL =	9,207,207,0,0,0	9,29: parts of [e]
L- =	205,215,0,0,0,0	205,207,215: parts of [l]

Another important feature of the rule system is that the speech frames characterizing a sound are not necessarily included in all the diads containing that particular sound. For instance, in the HO step the frames of [h] are missing but those of [o] are present. Similarly in the OT step there are no frames for [o] and for the burst of [t], there are only silent frames. On the other hand, both IE and EL contain frames needed for the realization of [e]. These facts clearly indicate that this text-to-speech procedure is not based on a diadic interpretation in the strict sense. The rules for concatenation of speech frames are designed individually for every sound and sound combination (V, C, VV, VC, CC), as well as assimilated and long versions of sounds. When the step-by-step conversion of phoneme codes is completed, B3 contains a series of speech frames of the text to be uttered. Sending these frames to the synthesizer a monotonous, robot-like speech will be produced. Thus the realization of

TTS conversion has been accomplished at the segmental level only.

AUTOMATIC GENERATION OF MELODY

To make speech more natural melody patterns must be superimposed on the segmental realization. In spite of the complicatedness of handling the pitch control in the speech frames of MEA 8000 a fully automatic melody generation was developed for the SCRIPTIOVOX system. The melody is generated on a male voice timbre.

What are the elements of this melody generation?

1. Building microintonation into the appropriate speech frames.
2. Recognizing the articles and some conjunctions in the text and making them unstressed.
3. Recognizing comma(s) in the text and changing the intonation before the comma(s).
4. Superimposing the intonation of declarative sentences characterized by a full stop at the end.
5. Superimposing the appropriate melody patterns on the various types of questions (question mark at the end). The types of questions distinguished for Hungarian are as follows:
 - questions beginning with a question word (Q-word);
 - questions without question words are further divided into three subcases (see below).

Microintonation

Quick variations in fundamental frequency independent of context and of the speaker's will are called microintonation. The variation ranges between 10-15Hz inside a sound (cf. Figure 2). Microintonation is one of the acoustic building elements that make speech human-like.

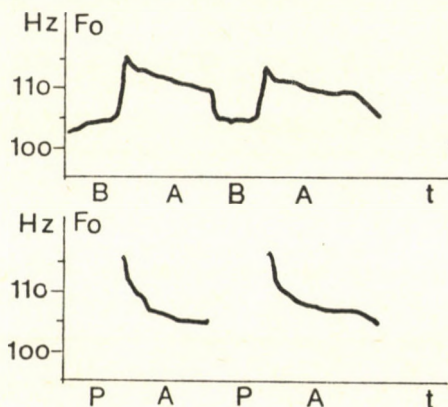


Fig. 2. The microintonation in the monotonous words, <baba> 'doll' and <papa> 'father'

Microintonation appears especially in CV and VC combinations.

Identifying articles and making them unstressed

The system identifies the words <a, az> 'the', <és> 'and', <hogy> 'that'. In Hungarian these words are uttered with a lesser degree of stress than the others. Various possibilities are available to make a word or part of it less stressed. That chosen for SCRIPTOVOX was lowering the fundamental frequency. So the pitch is decreased by 8 Hz at the beginning of the unstressed word and restored to the original value before the beginning of the next word (cf.

Figure 3). The position of the article is taken into consideration as well. An article at the very beginning of the sentence is even less stressed than an article in sentence-medial position. So the fundamental frequency of an article as the first word of a sentence is decreased by 16 Hz instead of 8 Hz.

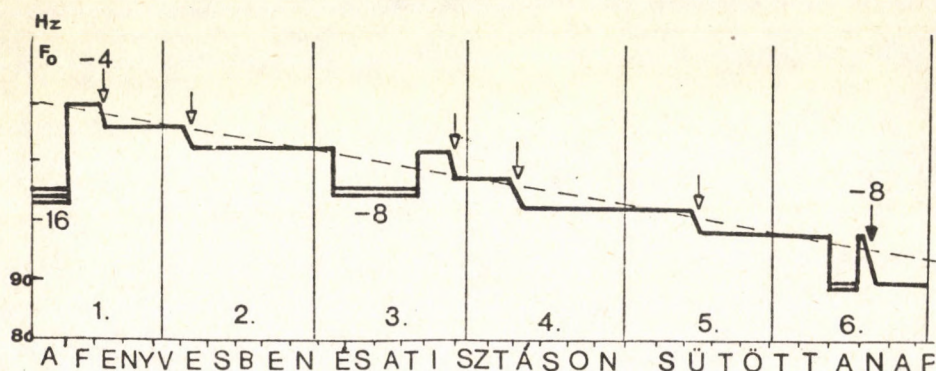


Fig. 3. The declarative intonation realized automatically in a sentence with articles and conjunction. (---- theoretical, linear falling of pitch; ——— practically realised quasi-linear falling of pitch; == parts made unstressed)

The interpretation of commas

A comma in a written text corresponds to a change in the melody and rhythm of live speech. To implement these changes the proper place of comma(s) in B3 has to be marked. How is the place of comma found in the buffer B3? For this purpose a special frame with short duration and zero amplitude is inserted wherever there is a comma in the text. How is the automatic change of intonation produced before the comma? The algorithm works as follows. The earliest 32 ms frame of the

vowel immediately preceding the comma is searched for. Then its pitch is increased by 8 Hz. The search goes on up to the next vowel before the comma and the pitch in its first 32 ms frame is increased by 4 Hz (cf. Figure 4). The pitch is then restored to the former value before the frame following the comma.

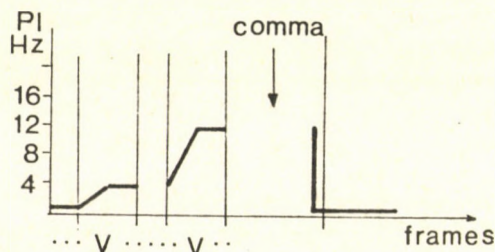


Fig. 4. The realized increase of pitch before every comma

The intonation of statements

To design an algorithm for the automatic realization of declarative intonation (falling structure) several data and rules have to be determined. One has to make decisions on the frequency value at the beginning of the decreasing structure, the degree of decrease, whether or not the decrease is linear, whether or not the degree of decrease depends on the length of the sentence, etc. The SCRIPTVOX algorithm (cf. Figure 3) allows for three types of modification in the sentence, i.e. reducing the pitch by 4 Hz, or reducing the pitch by 8 Hz, or reducing the pitch in the last word of the sentence by an additional 4 Hz.

The implementation of these types starts with the calculation of the length of the sentence. This is performed by counting the number of speech frames in B3. Depending on the length of the sentence four categories are distinguished, namely very short, short, normal and long sentences. According to these categories, in the next step the sentence is divided into three, four, five or six parts. The rules for the decrease of pitch in the different categories are shown in Table 9.

Table 9. The melody modification in declaratives

Sentence length categories	minimum number of frames	number of parts	pitch reductions		
			in every part	in the last part additionally	in the last word additionally
very short	40	3	8 Hz	-	-
short	20	4	4 Hz	4 Hz	-
normal	30	5	4 Hz	4 Hz	-
long	45	6	4 Hz	4 Hz	4 Hz

The comparatively large value for pitch reduction in very short sentences as shown in Table 9 is justified by our experiments. Similarly, the doubling of pitch reduction in the final part of the sentence helps in making the sentence sound more declarative. In very long sentences the need for an even larger fall of pitch was also indicated by the experiments. This decrease is advantageously realised at the

beginning of the last word of the sentence. It seems that the rules mentioned so far are sufficient for the realization of any declarative sentence.

The intonation of questions

The punctuation of questions is characterized by the use of question marks. In Hungarian several types of questions are used. To make an automatic realization of questions, a multi-level algorithm has to be designed. Regularities lending themselves for algorithmic procedures, as well as unambiguous orthographic forms have been found for the following types of Hungarian questions.

Questions beginning with a Q-word <(Mikor indulunk?) 'When do we start?').

Questions in which the nucleus (the word questioned) has one syllable (e.g. <Jó?> 'OK?', <A fagyi jó?> 'Do you like the ice-cream?'). This type of questions is called "one syllable questions".

Questions in which the nucleus has two syllables (e.g. <Hajó?> 'A ship?', <Ez egy hajó?> 'Is this a ship?'). This type of questions is called "two syllable questions".

Questions in which the nucleus has three or more syllables (e.g. <Hajóval?> 'By ship?', <Elindultak már a gyerekek?> 'Have the children started yet?'). This type of questions is called "three or more syllable questions".

Questions beginning with a Q-word

The system recognises twenty-one Q-words: <ki> 'who',

<kit> 'who' (acc.), <mi> 'what', <mit> 'what' (acc.), <hol> 'where', <hogy> 'how', <hogyan> 'how', <milyen> 'like what', <merre> 'which way', <hova> 'where to', <melyik> 'which', <mikor> 'when', <miért> 'why', <kiért> 'for whom', <mivel> 'with what', <kivel> 'with whom', <mennyi> 'how much', <honnan> 'where from', <meddig> 'how far', <hányadik> 'which (of a given number)'. First the intonation of the Q-word is implemented cf. Figure 5).

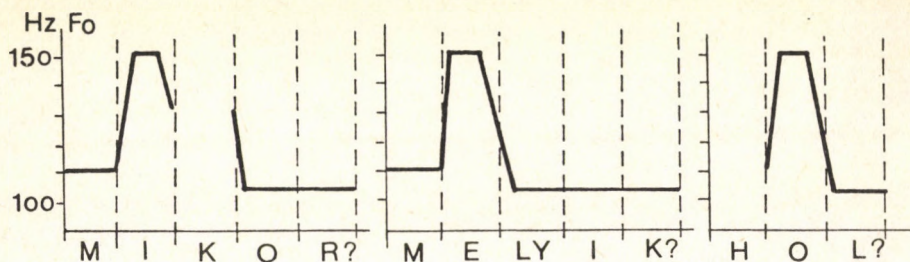


Fig. 5. Intonation realized in Q words

In the next step the algorithm of declarative sentences is applied to realise a decreasing intonation in the rest of the sentence. As regards pitch patterns for Q-words showed in Figure 5, they have proved to result in an acceptable intonation. The pattern starts with a pitch of 110 Hz (lower than in declaratives). At the very beginning of the first vowel the pitch is raised by 40 Hz to reach the peak of intonation. This value is kept for 30 to 40 ms (i.e. the duration of 1-2 frames) and then it is reduced by 45 Hz to get below the initial value.

One syllable questions

This type of question has its intonation peak at the very end of the vowel. Our experiments have shown that the sharp rising must be executed in the last one-third of the vowel and not earlier (cf. Figure 6).

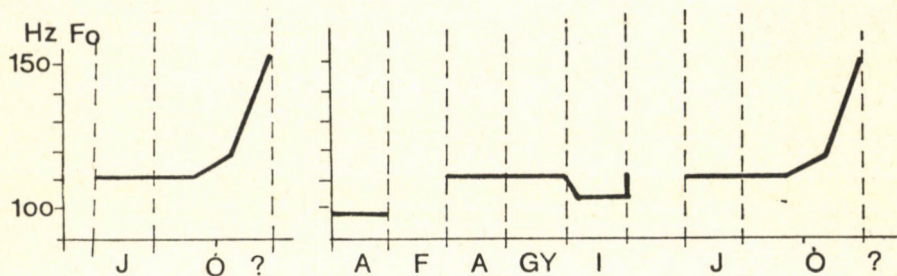


Fig. 6. Intonation realized in one syllable words questioned

The algorithm rises the pitch smoothly by 8 Hz in the second one-third of the vowel and then by 32 Hz in the last one-third. So the pitch will end up at 150 Hz. There is one more important aspect of our solution. If the monosyllabic word is situated at the end of a sentence, this pitch in the last vowel before the nucleus must be reduced by 4 Hz. This emphasizes the rising intonation of the nucleus.

Two syllable questions

Though the intonation peak in this type of question also lies in the last syllable, our experiments have shown that in this case the peak must be placed at the beginning of the vowel of that syllable and that the pitch must be reduced towards the end of this vowel. As regards the intonation of disyllabic nuclei three types of word endings are

distinguished.

- a) The last vowel is short and is followed by a voiced consonant (e.g. <Asztal?> 'A table?');
- b) The last vowel is short and it is either the last sound of the word or is followed by a voiceless consonant (e.g. <Villa?> 'A fork?', <Mászik?> 'Is it crawling?');
- c) The last vowel is long (e.g. <Hajó?> 'A ship?').

The intonation patterns for these cases differ radically in the way the pitch decreases and are illustrated by example in Figure 7.

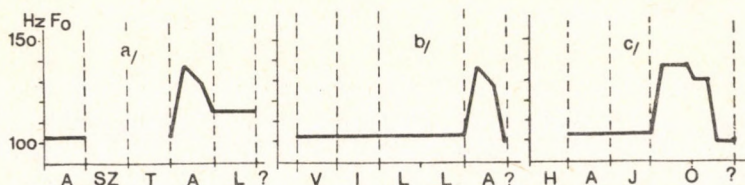


Fig. 7. Intonation realized in two syllable words questioned

Three or more syllable questions

In this type of question the intonation peak must lie in the last but one vowel followed by a decrease of pitch in the last one. The quality of the resulting pattern is further improved by distinguishing two subcases as follows (c.f. Figure 8).

- a) The intonation peak is carried by a short vowel (e.g. <Asztalos?> 'A joiner?');
- b) The intonation peak is carried by a long vowel (e.g.

<Hajóval?> 'By ship?').

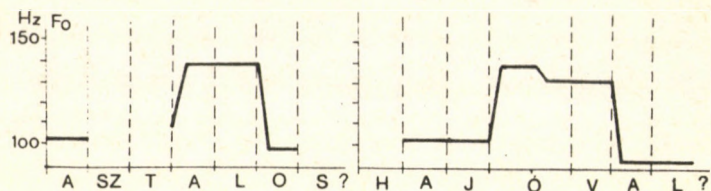


Fig. 8. Intonation realized in long questions

In both subcases the intonation peak must be short, therefore in subcase b) the decrease of pitch has to commence already in the long vowel.

THE PERCEPTUAL EXAMINATION OF SPEECH QUALITY

The complete process of designing and constructing a TTS system has to end in a scientifically based perceptual examination of the speech quality. The acceptance of the system depends on the results of this examination.

Material, method and results of the examination

The phonetically balanced (speech) material designed for the examination consisted of four groups of sound sequences: 30 syllables, 30 meaningless bisyllabic sequences, 30 one or polysyllabic words and 10 sentences (Table 10). This material was recorded by a male announcer and by the SCRIPTOVOX. Silent periods of 4--10 s were left between the sound sequences. Thirty-six, 18 year old pupils and twenty adults took part in the test procedure. The natural speech material was given for two groups of 18 pupils in a classroom. One

week later they listened to the synthesized material.

They had to put down what they thought they heard. The 20 adults listened to the material separately.

Table 10.

No.	Syllables	Meaningless bisyllabic sequences	Words	Sentences
1.	bá	acsá	száz	1965-ben születtem.
2.	ósz	kázi	mulat	Az igazgató beszédet mondott.
3.	im	makun	papír	A festőművész kiállításán voltunk.
4.	ge	szerő	feleség	A gyárkémény füstől.
5.	ős	benik	garázs	Mikor indul a vonat?
6.	la	szoge	intézet	A feladat eredménye jó?
7.	mé	niszől	hím	Galamb van a ketrecben?
8.	tu	pusók	bója	A körte a kezében van?
9.	ve	vütel	üveg	Amikor elindultak, már késő volt.
10.	zu	zsibet	autóbusz	Ez volt az utolsó fejezet.
11.	ec	gyülem	hernyó	
12.	nyé	hatyó	szín	
13.	ká	gávuc	selyem	
14.	zsü	sékid	kulacs	
15.	ji	anosz	pödör	

16. óg	vapon	kos
17. éb	summi	dió
18. su	pitez	telefon
19. dü	tufam	lusta
20. csa	csüdér	vaj
21. po	midázs	nagyközség
22. av	pahán	szelíd
23. szé	zilos	bokor
24. re	ramut	szegy
25. an	balá	kalapács
26. áf	cilü	férfi
27. győ	ören	tapos
28. at	tészir	cár
29. ni	jéba	pék
30. öd	liség	úttörő

The results of the evaluation can be seen in Fig. 9. As the figure shows, from the word level upwards the degree of the identification of synthesized speech is quite close to that

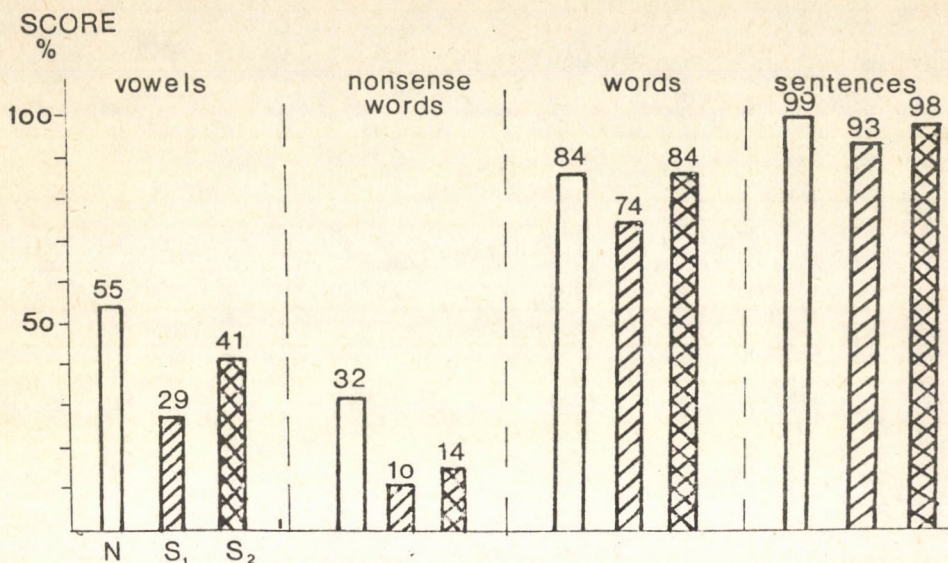


Fig. 9. The scores of identification of natural and synthesized sound sequences (N = natural; S₁ = synthesized, tested in groups; S₂ = synthesized, tested individually)

of natural speech. Further testing of the speaking system in everyday circumstances proved that the speech quality of SCRIPTOVOX is acceptable for use in industrial applications and in a reading machine.

CONCLUSION

The SCRIPTOVOX text-to-speech system displays several differences when compared with conventional unlimited vocabulary speech synthesis systems. Its data base comprises only 225 four-byte data, each representing a speech frame with 8 to 64 ms duration. The rule system converting letters and other symbols into a concatenation of speech frames uses, at one point, a novel diad-like representation. Extensive experimentation was involved in formulating the rules of

intonation for the various classes and subclasses of sentences. The memory requirement of the complete system does not exceed 12 kbytes.

The SCRIPTOVOX system seems to accomplish a good compromise among low cost, high speech quality, fully automatic text-to-speech conversion (no need for diacritics or auxiliary symbols in the text), small memory requirement and a very low bitrate (approx. 100 byte/s).

REFERENCES

MEA 8000 voice synthesizer: principles and interfacing.

Philips Technical Publications 101. The Netherlands 1983.

GORDOS Géza--TAKÁCS György: Digitális beszédfeldolgozás.
Budapest 1983.

OLASZY Gábor: A magyar beszéd leggyakoribb hangsorépítő
elemeinek szerkezete és szintézise. Nyelvtudományi
Értekezések 121. Budapest 1985.

OLASZY, Gábor: A phonetically based data and rule system for
the real-time text to speech synthesis of Hungarian.
Abstracts of the Tenth International Congress of Phonetic
Sciences. Ed. by COHEN, A. and BROECKE, M.P.R. v.d.
Dordrecht--Cinnaminson, 1983, p. 389.

ERRORS IN WORD STRESS IN THE RUSSIAN SPEECH OF HUNGARIANS

András Soproni

'Foreign accent' is a characteristic, frequently analysed phenomenon of speech in a foreign language. It can be defined as a deviation from standard pronunciation the reasons of which are to be found in the fact that the given language is not the mother tongue of the speaker. Though the phenomena of foreign accent are generally distinguished from pronunciation mistakes -- the former term referring to phenomena connected with the level of spontaneous speech, whereas the latter referring to those characterizing earlier phases of language learning -- we will disregard this distinction and all deviations from the standard will be uniformly called errors. The errors I shall deal with in this paper can often be observed in the Russian speech of Hungarian learners. In the course of communication between two persons not speaking in their native language, deviations from the phonetic standard on both sides, compounded by the deformations introduced by the speakers' different perceptual bases, seriously impede comprehension, especially if communication takes place by the help of technical devices (radio, telephone) in the midst of numerous disturbing factors of the speech situation. Somebody who speaks in a language foreign to him to a person whose mother tongue is the given language is not in an easy position, either. We

have conducted the following experiment to examine the phenomena concerned. Hungarian listeners had to evaluate the Russian speech of five persons of various mother tongues (Hungarian, Russian, Polish, German, Viet-Nameese) from the point of view of intelligibility. The results have unanimously demonstrated that Russian speech with a Hungarian accent is considered to be more intelligible for Hungarians than genuine Russian pronunciation, and the Russian speech of speakers of other mother tongues is felt to be even harder to understand.

The reason for foreign accent lies in the differences between the mother tongue and the foreign language. If the corresponding phenomena of the two languages are compared, we can establish a linguistic model of native interference (Samuilova, 1971). This model helps to explain the errors and even makes it possible to predict them, therefore it has a great significance for language teaching, and especially for writing text-books.

If we try to establish a model of interference between Russian and Hungarian with respect to stress, we will come up against some difficulties: there are a lot of unsettled, not properly understood questions. Can Russian and Hungarian stress be considered phenomena of the same order at all? Indeed, the only property they have in common is that both are (suprasegmental) factors of the acoustic organization of words. At the same time, there are significant differences between them, in physical parameters and placement.

The physical nature of Russian word stress is a hotly

debated issue. According to experimental research, the main factor is the length of the stressed vowel, but intensity, pitch, and the quantitative and qualitative reduction of vowels, i.e. the phonological rule resulting in the occurrence of different types of vowels in stressed and unstressed syllables, also play a role. This latter results in the formation of what is called 'qualitative accent', characteristic of the Russian language (cf. Sovremennyj russkij jazyk 1985, 102). The main means of the implementation of word stress in Hungarian, on the other hand, is intensity, while length is a phonemic feature of vowels.

Stress placement in Russian is 'free' and 'moving'. Regularities determining the place of word stress are tendencies, rather than strict rules. In the majority of cases these tendencies can be attached to morphological markers, but some of them are of a phonetic nature. In Hungarian, by contrast, word stress always falls on the first syllable of the word.

On the basis of the foregoing, we can generally determine the points where Hungarians will commit errors in their Russian speech. But such a theoretical model can be misleading if it does not properly take the different nature of word stress in the two languages into consideration. In that case practice will falsify our predictions. Lebedeva writes for example: "In the Hungarian language stress cannot fall on even syllables; besides, primary and secondary stress can be distinguished (primary stress falls on the first

syllable, the other odd syllables are more intensive than the even ones, the stress of the fifth syllable is stronger than that of the third one). It is known from practice how difficult it is for Hungarian students to learn the rhythm of a Russian word and the place of word stress." (Lebedeva, 1975 pp. 13--4.).

As regards the features of Hungarian word stress, the above characterization is valid; but it would be difficult to find a Hungarian person who applies word stress in his Russian speech according to these rules.

Kalman Bolla determines the peculiarities of a Hungarian accent more exactly: "When pronouncing Russian words, the typical mistake of Hungarian learners of Russian is that they utter the first syllable, regardless of the place of stress, with a tenser articulation characteristic of the Hungarian language, and they lengthen the stressed syllable. Thus the components of Russian stress -- intensity and length -- are separated from each other, resulting in a characteristic type of foreign accent". (Bolla, 1977, p. 122.).

This formula is basically adequate, but it still can be refined in some respects on the basis of empirical observations. The formula is valid primarily for those words in which the 'Russian stress' -- or its vestige, the long vowel -- is on the third, fourth, etc. syllables. Let us add that the placement of the 'Russian stress' can also be incorrect. (In the examples below, a colon (:) placed above the vowel will denote a Hungarian-type stress, and to represent the Russian stress we will use a vertical line (|)

placed above the vowel. If the Russian stress is substituted by length, we will mark the length of the vowel by writing a colon (:) after it.)

КОМАНДИР	+kámand'i:r
ОБЕСПЕЧЕНИЕ	+ábespetʃe:n'ije
ОТСУТСТВОВАЛ	+átsutstvova:l

What is more, in such words it is often the case that the Hungarian stress appears together with a full-fledged Russian stress, expressed by a long and prominent vowel.

КОМАНДИР	+kámand'í:r
ОБЕСПЕЧЕНИЕ	+ábespetʃé:n'ije
ОТСУТСТВОВАЛ	+átsutstvová:l

But sometimes even the remnants of the Russian stress are missing.

ДЕВЯНОСТО	+d'ívinosto
ПОТОМУ ЧТО	+pátamuʃto

On the other hand, if the Russian stress falls on the second (and not on the third, fourth, etc.) syllable it will be inconsistent with the Hungarian stress. The following incorrect varieties are possible:

a) Hungarian stress on the first syllable, and a rudimentary Russian stress on the second:

второй +ftáro:ɟ_n

двадцатый +dvát͡sa:tiɟ

откуда +átku:da

b) Hungarian stress on the first syllable, with not even the remnants of the Russian stress on the second:

Ленин +lénin

Москва +mòskva

тринадцать +trinatsat'

The complete absence of the Russian stress is more probable in disyllabic than in polysyllabic words.

It can be missing from the first syllable:

ВОИНОВ +vóɟinov

ОПЫТ +ópit

ОТДЫХ +ód:ih

It can be missing from the second syllable:

ВСТУПИТЬ +fstúpit'

второй	+ftaroj
двенадцать	+dvenat'sat'
ему	+jimu
земля	+zemlja
морской	+marskoj
огнём	+agn'om
прямой	+primoj
разгром	+razgrom
судьба	+sud'ba

Hungarian learners usually substitute the [ʌ] sound occurring in an initial unaccented position by a short [a] sound, and the [ɨ] sound by [i]. If the first syllable is unstressed in a Russian word, we will mostly find these vowels substituted for the qualitatively and quantitatively reduced ones in the syllable pronounced with Hungarian stress.

второй	+ftaroj
--------	---------

ему	+jimu
земля	+zimlja
прямой	+primoj
частей	+tʃist'ej

But we also find forms in which there is only quantitative reduction, without any qualitative reduction. It is an especially frequent phenomenon in reading since Russian orthography does not indicate qualitative reduction, which can be misleading. The mistake is mainly characteristic of the syllables containing the phoneme /a/, it is less frequent in the case of syllables containing the phoneme /e/, and in the case of /o/ it is extremely rare. It seems that reductions of the type [o] → [ʌ] and [e] → [ɨ̞] are more reasonable for Hungarian because of the closer relationship in tongue position than the [a] → [ɨ̞] reduction. The qualitative reduction is often omitted in international loanwords, too, but of course the influence of the mother tongue is obvious here. We can find the above-mentioned regularities of sound-substitution in any syllable, i.e. regardless of the position of the reduced vowel. What is important for us here is that the Hungarian stress on the first syllable does not influence the qualitative reduction.

Examples:

морской	+morskoj
---------	----------

земля	+zēmlja
-------	---------

прямой	+prjāmoj
--------	----------

In international loanwords:

момент	+mōment
--------	---------

метро	+mētro
-------	--------

январь	+jānvar
--------	---------

New consider words in which the stress falls on the first syllable. If the Hungarian placement of stress influenced the way Hungarians distribute stresses in their Russian speech, words with initial stress would be the easiest for them. In contradiction to this, experience shows that these items are the most difficult of all. We have no opportunity to fully analyse those analogies which result in these rhythmic errors, we will only mention a few examples here: +транспорт, +выполнить + некоторый, +танковый

More important from the point of view of our present purposes are the cases when even the trace of initial Russian stress disappears, i.e. a short vowel with Hungarian stress takes its place. Though not very frequently, such cases also occur:

воин	+vōjin
------	--------

опыт	+ōpit
------	-------

НЕСКОЛЬКО

ː
+n'eskolko

We have the following guess to offer for the explanation of the first two incorrect forms above. The majority of errors involving stress shift are due to morphologically-based analogies. But the two words above are difficult to analyse, they do not provide sufficient cues for any analogy, so if the speaker does not put the stress on the first syllable, he does not dare to put it on the second one either. In the case of the third word -- the typical incorrect form of which is: <НЕСКОЛЬКО> -- the intention of the right placement of stress conflicts with the supposed prohibition of stress on the first syllable.

Very rarely, and exclusively in the course of reading, it can also happen that the speaker pronounces a short vowel showing the intention of qualitative reduction in the place of the stressed initial syllable. Such forms show a more complicated instance of the above-mentioned confusion:

ОПЫТ ː
 +äpit

ВОИНОВ ː
 +väjinov

Monosyllabic words deserve special treatment. All Russian content words are stressed. In these two sentences for example: <Весь день шёл дождь. Всю ночь рос гриб.>

there is no direct opposition of stressed and unstressed words, yet all of them qualify as stressed, a fact that shows

the importance of segmental elements in the implementation of Russian stress, i.e. the phenomenon called 'qualitative accent' (Matusevic 1976, p. 223).

We can find the most surprising errors in monosyllabic words. The words <ПОЛК> and <ЧАСТЬ> are pronounced:

- with short Hungarian [o] and short [a] sounds -- [ˈpɒlk, ˈtʃast'];
- with [a] and [i] sounds substituting reduced [ʌ] and [ɪ] used under the influence of the forms [palkof] and [tʃast'ej] -- [ˈpalk, ˈtʃɪst];
- with long [a:] and long [i:] (mainly in reading) which is a special hybrid of qualitative reduction and stress, and the frequent practising of the reduced form plays in its formation -- [ˈpalk, ˈtʃi:st'].

The appearance of Hungarian stress is relatively frequent in certain words. They are as follows:

a) Russian proper names also used in Hungarian:

Москва, в Москве +mɔskva, +vmɔskve:

Владимир Ильич Ленин +vlɔdimir ɪljitʃ lɛnin

b) in common nouns also used in Hungarian as Russian loanwords:

спутник +spʉtn'ik

большевик +bɔlʃevik

колхоз +kɔlhoz

c) in certain names of months (especially in dates, in the genitive case):

января +jɪnvarjə

февраля +fɛvraljə

d) in cardinal and ordinal numbers:

один +ad'i:n

тринадцать +trina:tsat'

e) in certain frequently occurring phrases, in stereotypes, especially those that became fixed in the first phase of language learning:

никто не отсутствует +n'ikto: n'e ätsu:tsvujet

состав класса +sasta:v klässa

во втором эшелоне +vaftaro:m eʃɛlo:n'e

выявление цели +viɟivle:n:ɟe tʃɛ:li

достиг рубежа атаки +däst'ig rübeʒa: äta:ki

заместитель командира по +zämešt'it'el kāmānd'ira pa

технической части +t'ehn'itʃeskoj tʃa:st'i

ЛИЧНЫЙ СОСТАВ

+lʲit͡ʃnɨj sʲasta:v

С ОГНЕВЫХ ПОЗИЦИЙ

+sʲagn'evɨh pʲozitsij

ОГНЁМ ПРЯМОЙ НАВОДКОЙ

+ʲagn'om pʲimɔj nʲavotkoj

ПЕРВОЙ МИРОВОЙ ВОЙНЫ

+pʲervojuj mʲiravɔj vʲajni

НЕ ИМЕЕТ ЗНАЧЕНИЯ

+n'imejet znʲat͡ʃe:n'ija

What conclusions can be drawn from the foregoing?

1. A conflicting co-existence of the Russian and Hungarian types of stress can be observed in the Russian speech of Hungarians. Sometimes they can be reconciled with each other, sometimes one of them replaces the other. It is in the nature of things that numerous individual variations and transitional forms can be observed.

2. The possibility of co-existence of the two kinds of stress within a word repeatedly draws the attention to the significant difference between the two phenomena. In the mind of the Hungarian learner, the opposition of Russian stressed vs. unstressed vowels comes to be identified with the opposition of Hungarian long vs. short vowels. The graphic symbol of Russian stress seems to reinforce that fallacy. On the other hand, the fact that Hungarians in their Russian speech do not normally pronounce two long vowels within a word demonstrates the intuitive acquisition of the fundamental laws of Russian rhythm.

3. As we have seen, intensity and length are not the

only components of Russian stress that can get separated from each other; the third component, qualitative accent can also live a separate life. Since in the case of this last type of error the sound shape of the word is seriously distorted, the prevention and correction of such errors is of utmost importance.

4. The fact that Hungarian stress falls on the first syllable does not influence the way Hungarian learners distribute their Russian stresses. It can be proved by numerous examples that the source of errors is an incorrect extension of the regularities of Russian stress placement, i.e. interference within the same language. It is remarkable that Arto Mustaioki drew a similar conclusion from an examination of the errors committed by Finnish learners of Russian. He writes: "Mistakes made in the placement of stress in various forms of Russian nouns cannot be explained by the influence of the mother tongue of the experimentees, but by false analogy, the overgeneralization of the phenomena of the Russian language" (Mustaioki, 1980, p. 29).

REFERENCES

- BOLLA, K.--PAPP, F.--PALL, E.: Kurs russkogo jazyka. Budapest 1977.
- LEBEDEVA, J.G.: Zvuki, udarenije, intonacija. Moskva 1975.
- MATUSEVIČ, M.I.: Fonetika. Moskva 1976.
- MUSTAIOKI, A.: Ősibki v udareniji sušestvitel'nyh russkogo jazyka u finskih studentov. Aspekt 1980/2.
- SAMUJLOVA, N.I.: K voprosu akcenta. In: Pamjatnik akademika

U.U. Vinogradova. Moskva 1971.

Sovremennyj russkij jazyk. Teoretičeskij kurs. Fonetika.
Moskva 1985.

ON THE ACOUSTIC STRUCTURE OF GERMAN PLOSIVES AND NASALS

László Valaczkai

József Attila University

This paper explores the acoustic structure of (oral and nasal) stops segmented from continuous German (female) speech. In principle, the realization of the phoneme /r/ as a tremulant would also constitute part of the subject-matter of this paper; however, the informant pronounced it as a velar-postdorsal voiced fricative rather than as a trill. The regional variants of /r/ and their relation to Standard German would deserve separate study.

The investigations reported here cover

1. total articulation time and internal temporal organization;
2. frequency and intensity structure of components;
3. frequency assimilation; and
4. intensity assimilation.

The purpose of the investigations is to provide material for contrastive phonetic research and to contribute to a clarification of the perceptual relevance of acoustic components -- especially F2 -- of speech sounds.

The method followed here involved an instrumental investigation of temporal organization, intensity relations, and frequency structure as part of a complex contrastive physiological/acoustic research project. Since frequency and intensity structure both appear in the same temporal

framework, the recordings will appear in the figures below in a common system of co-ordinates with time represented along the horizontal axis. The vertical axis shows component frequency values alongside the spectrogram of a word containing the speech sound under investigation, and intensity values alongside the intensity curve. On the time axis, 125 mm = 1.00 s; the calibrations of the vertical axis extend as far as 8 kHz and 40 dB, respectively.

All recordings and visual displays were made in 1986 in Budapest, with the instruments of the Phonetics Department of the Linguistics Institute of the Hungarian Academy of Sciences. The intensity curves were made by an F - J Electronics IM 360 intensity meter. A special feature of these curves is their commensurability, i.e. the fact that they allow us to determine the relative order of intensity of the speech sounds investigated. The spectrograms were made by a dynamic sound spectrograph No. 700 of Voice Identification Incorporation. Along with spectrograms of full words, sections have also been made to represent the frequency and intensity structures of pure phases of the sounds in question, as a function of the time of articulation. 300 Hz (wide-band) sections are supplemented by 45 Hz (narrow-band) sections, the latter being more detailed and thus lending themselves to a minute analysis. Finally, amplitude sections have been made to furnish frequency and (relative) intensity data over a period of 8 ms in the pure phase of articulation of the speech sound concerned.

An unfortunate limitation of the methods of this paper

is that our analysis was not supplemented by synthesis, according to the method of 'analysis by synthesis' as applied e.g. by Olaszy (1985) and Kiss (1985). Indeed, the function of apparently insignificant components on the recordings yielded by our analysis could only be clarified by synthesis, and conversely, synthesis could shed light on certain components that cannot even be seen on the recordings. The procedure known as 'analysis by synthesis' will hopefully constitute part of a major project we intend to embark on in 1987.

The corpus of investigation. Following a clarification of some matters of principle concerning informants' pronunciation and competence (cf. Littmann 1965; Lotzmann 1974; Valaczkai 1978), we have recorded words and connected texts containing the speech sounds to be investigated as spoken by a female speaker, a university student from Thuringia. For the purposes of the present investigation, we have compiled the following minimized inventory from that linguistic corpus:

<Backe, Bäume, bese, Bibel, Ebbe, Übel, Paar, Hypothek, Üppig, Dieb, Dame, Ödem, Adam, Idee, Tat, utopisch, Mutter, Auto, gegen, Garage, Kegel, malmen, Magen, Masse, Name, ihm, Napf, Zahn, Hang, Cognac>. (The full inventory of words, excluding connected texts, contains 190 items.)

THE RESULTS OF INVESTIGATION

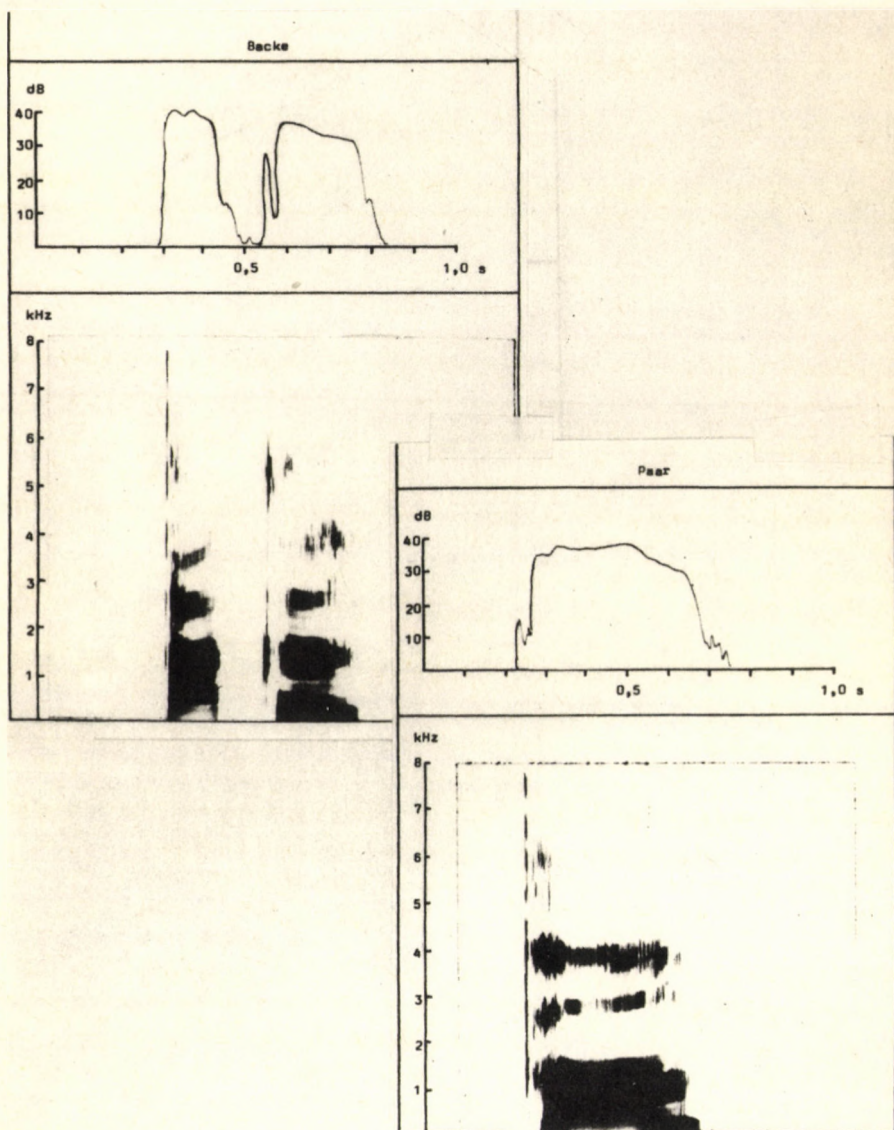
[b]

1. Temporal organization. In word initial position the

duration of closure cannot be measured since one of the reference points is missing. The time span of plosion, depending on the degree of muscular tension, is between 10--14 ms. The longest plosion, incidentally, occurred before a rounded long vowel. In word internal intervocalic position the full temporal organization is analysable. The total time of articulation of [b] varies between 90--105 ms, the average length of closure is between 80--90 ms.

2. Frequency and intensity structure. In word initial prevocalic position, in the phase before plosion there is no voicing, not even what is called muffled voicing, a phenomenon observable in Hungarian [b] in a similar position (cf. Olaszy op.cit. 34, 37; Olaszy comments that in the pre-plosion phase there is voicing but, for lack of air flow, it has no formant structure). Muffled voicing has also been observed by Bolla in initial [b] in Russian (Bolla 1981, cf. the glottogram of Table 38 and comments in English on p. 74 and in Russian on p. 118) and in American English (Magyar Fonetikai Füzetek /Hungarian Papers in Phonetics/, henceforth MFF 9. 1981, pp. 86--7). Initial German [b] has no voiced formant structure in the plosion phase, either. That voiceless realization excludes an opposition of (strictly) voiced vs. voiceless in that position; the opposition is based on other acoustic factors here. A similar phenomenon was also observed in male German speech (Valaczkai 1984). In an initial [b], continuous and intensive noise concentration areas (NCA's) are formed in the frequency bands between 470--785, 1145--1715, 2290--2850, 3470--3830, and 5000--5500

Hz. The intensity values of NCA's are as follows: the intensity peak is around 665 Hz with a value of 21 dB, the next highest peak is around 1320 Hz, with an intensity value 14 dB lower.



3. Frequency assimilation. Intensive NCA's are located depending on the frequency domains of the formant structure of the adjoining vowel. The first formant of the adjoining vowel bends to the pure phase value from around 500 Hz, and the second formant from around 1500 Hz.

4. Intensity assimilation. On plosion, the intensity value abruptly increases and in approx. 30--32 ms, via a characteristic 'overshot peak', reaches the intensity value of the adjoining vowel. Hence, amplitude assimilation between [b] and the adjoining vowel takes place in the plosion phase.

Intervocally, low intensity muffled voicing can be detected in the closure period as well. As can be seen in the spectrograms of <beben> and <Bibel>, some two-thirds of the total closure time has this voicing, but in the last one-third (and especially just before plosion) it cannot be found on the recordings. However, the spectrogram of <Ubel> did not exhibit such devoicing. The reason for that may be assimilation across the strongly reduced unstressed vowel; however, we did not examine that instrumentally.

[p]

1. Temporal organization. In word initial position the duration of closure cannot be measured since one of the reference points is missing. The time span of plosion is roughly 8--10 ms, this is the phase of the characteristic burst noises, followed by a period of intensive aspiration of approx. 28--30 ms. In intervocalic position the total time of articulation gives an average value of 100 ms. Of that, some

65--68 ms is the time of closure, 8--10 ms that of plosion, and the rest is the period of intensive aspiration. We have found 'over-articulation' (like in the case of the [b] of <Ebbe>) in <Üppig>: total time of articulation approx. 162--164 ms, of which closure takes some 120--122 ms, plosion 16 ms, and aspiration 24--26 ms.

2. Frequency and intensity structure. In the phase of plosion continuous energy zones of various intensity values can be found in the frequency bands of 785--4750, 5000-6900, and 7200--7750 Hz. The turbulence effects of aspiration create NCA's of extra intensity in the frequency domains of 1000--1575, 2360--2900, 3570--4300, as well as 5000--5500 and 5800--6100 Hz. The lowest NCA around 1215 Hz is the most intensive: 19 dB; the intensity of the second NCA around 3930 Hz is only some 2.5 dB less; that of the third peak, around 4000 Hz, is already 20 dB weaker.

3. Frequency assimilation. It is between the aspiration phase and the adjoining vowel, rather than between the plosion and the vowel, that assimilation can be found: the vowel formants, especially F2, get assimilated to the second intensive NCA of turbulence effects. If the intensity peak of that NCA is above 2500 Hz, the second formant of the adjoining vowel will bend to its pure phase from between 2700--2750 Hz (more exactly, from approx. 2720 Hz).

4. Intensity assimilation. There is no assimilation between the plosion phase and the formant structure of the adjoining vowel. Amplitude link can be observed between the aspiration phase and the adjoining vowel. At release, there

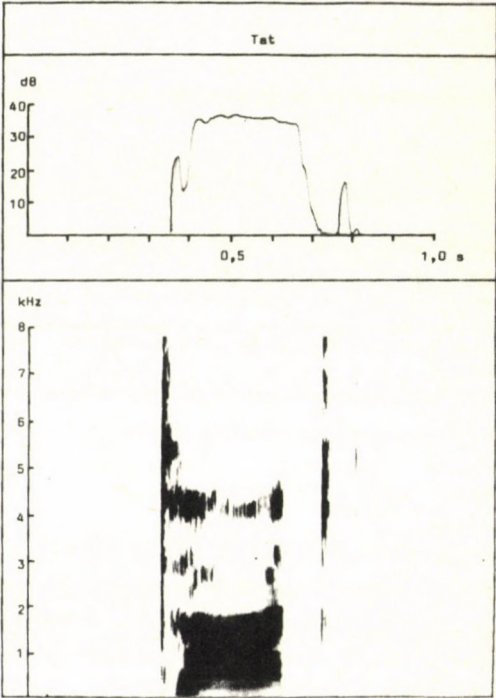
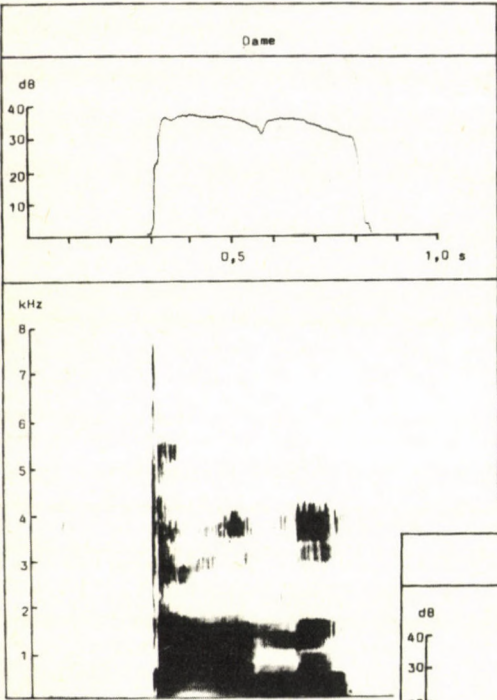
is a sudden increase of intensity from zero to 16 dB; then for a period of 15 ms it keeps decreasing until it becomes some 8 dB weaker; then there is some stagnancy and finally intensity reaches the level of the adjoining vowel by another sudden increase.

[d]

1. Temporal organization. In word initial position the duration of closure cannot be measured. The time span of plosion is approx. 10 ms. Intervocally, the total time of articulation is 90--110 ms. The [d] in <Adam> deserves special mention for its uncommon shortness: overall duration 72--74 ms, plosion time about 8 ms.

2. Frequency and intensity structure. As in the case of [b], word initial [d] is pronounced voiceless even if followed by a vowel. In an intervocalic position, the recordings exhibit voicing; but the temporal distribution and intensity structure of that voicing deserves special attention: its intensity, as compared to that of the preceding vowel, shows a gradual decrease in roughly two-thirds of total articulation time; then, after a short period of stagnancy, it falls abruptly to a minimum intensity level which is 17 dB less than that of the preceding vowel. In the phase of plosion intensity suddenly increases and reaches that of the adjoining vowel by the end of that phase. Due to its voicelessness, the phase of plosion is well separable from the adjoining vowel; in its frequency structure continuous noises are found up to 8 kHz.

Formant-like NCA's cannot be detected in it.



3. Frequency assimilation. The formants of the adjoining vowel are clearly assimilated to the more intensive NCA's that take shape on plosion. That assimilation is especially well-attested if the adjoining segment is a front vowel. If the vowel's second formant is above 4.5 kHz, the frequency of the second intensive NCA is between 4.3--4.5 kHz in the phase of plosion.

4. Intensity assimilation. The intensity of [d] starts from zero and reaches that of the adjoining vowel by an abrupt increase. On the borderline between plosion and the adjoining vowel, an 'overshot peak' can be clearly observed. A similar phenomenon has also been attested by Olaszy (1985, 43 and elsewhere).

[t]

1. Temporal organization. The duration of closure in initial [t] cannot be measured. The phase of plosion is made up by two distinct parts: the intensive burst noises lasting 8--10 ms in word initial position and the aspiration period whose duration can be as much as 80 ms (word finally, 72 ms). The length of closure in word final [t] was 100 ms. Intervocally, total articulation time varied between 116--137 ms; closure time between 90--95 ms; and the phase of plosion took 8 to 10 ms. In one case -- where [t] was preceded by an unstressed vowel and followed by a stressed vowel -- the length of closure was significantly shorter: 66

ms. However, the phase of plosion was 10 ms here, too.

2. Frequency and intensity structure. The intensity maximum of the vowel was 35--36 dB, the intensity of plosion 12--13 dB less. In the intervening aspiration period, intensity dropped a further 10 dB or so. The intensity of final [t] in the phase of plosion is almost 18 dB less than that of the preceding vowel and 7--7.5 dB less than the intensity maximum of an initial [t]. The intensity of the aspiration phase of final [t] is only slightly above zero. In the frequency structure of initial [t], intensive noise components are found in the frequency bands of 1300--2100, 2640--3220, 3860--4650, and 5050--5930 Hz. Within those frequency domains, intensive peaks appear around 1570 Hz, 2860 Hz, 4290 Hz, and 5430 Hz. The third NCA is the most intensive of all; the first is some 3 dB less, the second and the fourth are both 6--7 dB below the third.

3. Frequency assimilation. The frequency band of the lowest NCA of [t] assumes the frequency of the second formant of the adjoining vowel. The frequency of the aspiration phase slightly bends to that of the second, third, and fourth formants of the vowel. The compliance of frequency values of the adjoining vowel shows a pattern similar to that observed by Olaszy (op.cit. 47) for Hungarian: the vowel formants assimilate "...to the theoretical formant pattern determined by the articulatory configuration of [t]". However, the degree of bending to the pure phase as expressed in absolute values is not significant in the case of the female German speech investigated here.

4. Intensity assimilatoion. On plosion, there is an abrupt increase of intensity starting from zero. No direct amplitude link has been found between release and the adjoining vowel. The intensity pattern of the release phase exhibits a sudden decrease from maximum to half that value and then leads over to the phase of aspiration. Following the lower intensity of that phase, there is another abrupt increase leading on to the adjoining vowel with a characteristic "overshot peak".

[g]

1. Temporal organization. Since the first 10--12 ms portion of the closure phase of initial [g] exhibits low-intensity muffled voicing, total articulation time can be measured here: it is approx. 90--92 ms. The time of plosion is 10--12 ms. Intervocally the total time of articulation is shorter: 62 ms or so, of which the plosion phase takes relatively longer, presumably because of the assimilation due to the reduction of [ə]: 16 ms.

2. Frequency and intensity structure. The frequency structure of [g] also displays multiple complexity. The low-intensity voicing observed in the eearly stage of closure phase subsequently vanishes from the recordings, but the phase of plosion shows voicing again. Its frequency structure is characterized by formant-like concentration areas interdependent with the formant structure of the adjoining vowel. This interdependence is the most obvious in the case of F2. Frequency values of the concentration areas are as

follows: F1 up to 785 Hz, F2 1070--1790 Hz, F3 2070--3290 Hz, F4 4570-5220 Hz. Within those domains, intensity peaks are found around 430, 1430, 2720, and 4860 Hz. The intensity values of the first and third peaks are roughly equal: 10 dB, those of the second and fourth peaks are 9 dB less than that. The intensity of initial [g] starts from zero, abruptly increases in the phase of plosion and reaches the intensity of the adjoining vowel. In the closure phase of intervocalic [g], voicing can be observed up to approx. 400 Hz, varying across words, but since there is no air flow, formant-like frequency structure is not observable. Its intensity starts decreasing around one-third of articulation time; at that point it is some 16 dB less than in the preceding vowel and forms an intensity minimum. After that, there is approx. 7 dB increase, but immediately before plosion intensity drops by 3 dB again. In the moment of plosion, intensity abruptly grows and reaches that of the adjoining vowel accompanied by the well-known phenomenon of "overshot peak".

3. Frequency assimilation. Of the formants of the adjoining vowel, it is above all F2 and F3 where compliance with the appropriate concentration areas of [g] can be attested. Its degree is approx. 10%.

4. Intensity assimilation. The intensity maximum of plosion falls behind the intensity of the adjoining vowel by a mere 3-4 dB. The difference is equalized in an approx. 40--50 ms portion of the articulation time of the vowel at the end of which the vowel's amplitude level is reached.

1. Temporal organization. The duration of closure cannot be measured in an initial [k]. The time span of plosion is 11--13 ms. The length of the subsequent aspiration phase is approx. 47 ms. In a word internal intervocalic position, total articulation time is 140--142 ms of which closure takes 105--106 ms, plosion 11--12 ms, and the rest is aspiration.

2. Frequency and intensity structure. The intensity of [k] starts from zero and increases abruptly in the moment of plosion. The intensity maximum of the plosion phase is 24 dB, i.e. 14 dB less than that of the adjoining vowel. The most intensive frequency domain of initial [k] is between 2570--3280 Hz. There is an intensity peak around 3110 Hz, of approx. 14dB. The frequency band of the next NCA is 3700--4430 Hz. The intensity maximum of this area is around 4290 Hz, with a value 6 dB less than the previous one. The third intensive NCA is found between 4720--5720 Hz, with an intensity peak around 5070 Hz which is only 3-4 dB weaker than the first peak. Frequency values of NCA's of the aspiration phase correspond to those of plosion NCA's. The only difference in end values is that the frequency patterns of the aspiration phase show obvious assimilation effects of the formant structure of the adjoining vowel.

3. Frequency assimilation. Strong assimilation can be observed between the first and third intensive NCA's of [k] and the formant structure of the adjoining vowel. The values of NCA's are in general 6--10% higher than the frequency

values of the corresponding formants of the vowel.

4. Intensity assimilation. There is no direct amplitude link between plosion and the adjoining vowel. Following a practically steady decrease of intensity in the aspiration phase (the intensity values of this phase is some 12--14 dB less than the intensity maximum of plosion) intensity shows an abrupt increase on the adjoining vowel, reaching its maximum values directly.

[m]

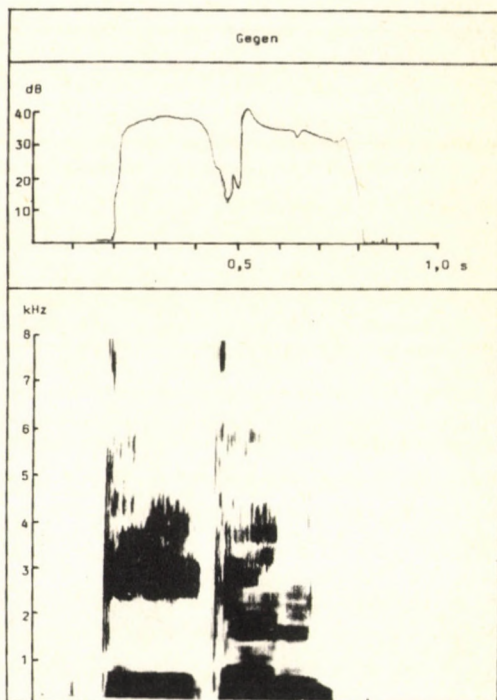
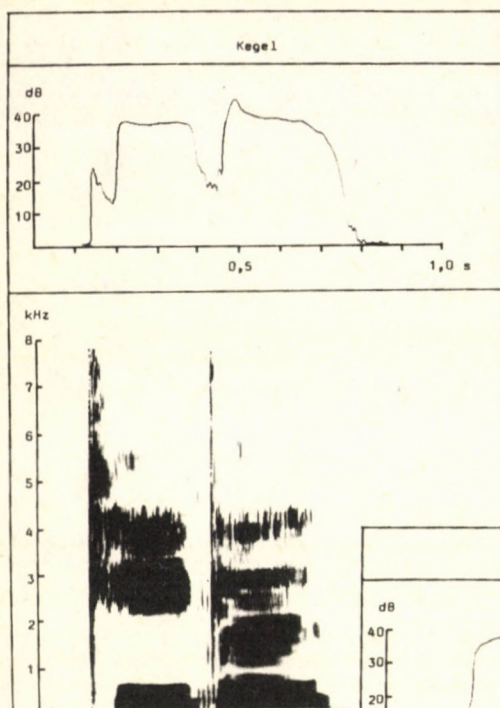
1. Temporal organization. Total articulation time varies considerably, depending on phonetic context. Word initially it is usually much shorter than e.g. in Hungarian, and it is only in an intervocalic position that it reaches the lowest limit of Hungarian duration data (cf. Olaszy op.cit. 68). The articulation time of initial [m] in <malmen> is approx. 60 ms, whereas that of the word internal nasal is 90 ms. The initial segment of <Magen> takes 68--70 ms; the intervocalic [m] in <Name> is 100--102 ms long.

2. Frequency and intensity structure. [m] has a voiced formant structure similar to that of vowels. The frequency values are as follows: F1 up to approx. 500 Hz, F2 1000--1640 Hz, F3 1930--3000 Hz, F4 3430--4290 Hz. The resonance points and values of intensity maximums of the formants are the following: F1 around 140 Hz, 29 dB; F2 around 1280 Hz, 23 dB less; F3 at 2140 Hz, a further 8 dB; weaker. The intensity of initial [m] starts from zero; when the oral closure is released, it increases abruptly and almost reaches the

intensity of the adjoining vowel. The difference is as little as 3--4 dB.

3. Frequency assimilation. Since there is coarticulation between [m] and some elements of the adjoining vowel, the formant values of the latter influence the frequency of the formants of [m]. Its formant values vary according to the harmonic type (i.e. frontness vs. backness) of the vowel. The formants of [m] do not bend to those of the adjoining vowel; vowel formants, on the other hand, show some degree of compliance with the formants of [m]. That compliance is levelled off quickly, in approx. 20--24 ms. With F1, even that slight compliance can be missing.

4. Intensity assimilation. On the plosion of the oral closure of [m] the intensity of the adjoining vowel shows an abrupt increase; their amplitude link is characterized by a sudden growth of the intensity of the vowel.



[n]

1. Temporal organization. Word initially, the time of articulation is 80 ms before a short vowel, and 90 ms before a long vowel. Word finally it is 110 ms. If the reduction of [] results in an assimilation between two nasal consonants, the length of [n] may reach 136--150 ms.

2. Frequency and intensity structure. [n] has a voiced formant structure similar to that of vowels. F1 up to 500 Hz, F2 1430--1930 Hz, F3 2930--3360 Hz. These values vary according to phonetic environment. (E.g. in <Napf>: F1 up to 250 Hz, F2 1860--2290 Hz, F3 2430--2720 Hz. In <Zahn>: F1 up to 300 Hz, F2 2720--3150 Hz.) Intensity peaks: F1 around 145 Hz, 26 dB; F2 around 1790 Hz, 36 dB less; F3 around 2360 Hz, its intensity is identical with that of the second peak.

3. Frequency assimilation. Of the formants of the adjoining vowel, it is especially F2 and F3 that yield to the appropriate formants of [n].

4. Intensity assimilation. Word initially it is identical with that of [m]. Word finally there is an abrupt decrease of intensity but one that lasts longer than initially.

[ŋ]

1. Temporal organization. Total articulation time 144 ms. The oral closure remains unreleased word finally. In an intervocalic position, plosion time is 8--10 ms.

2. Frequency and intensity structure. The facts that the closure phase is voiced throughout and that there is nasal

air flow make it possible for a formant-like frequency structure to take shape. F1 up to 570 Hz, F2 1145--1575 Hz, largely depending on the second formant of the adjoining vowel, F3 2145--2720 Hz, F4 4145--5290 Hz. Within the frequency band of F1 the intensity maximum is around 250 Hz with a value of 36 dB. The most intensive point of the frequency band of F2 is around 1430 Hz, with an intensity value 28 dB less than that of the first. The intensity peak of F3 is around 2645 Hz, its value is 41 dB lower than that of the first.

3. Frequency assimilation. If the second formant of the adjoining vowel is below 1600 Hz, the F2 of [ŋ] will also be lower. If the third formant of the vowel is below that of [], the vowel formant will comply with the appropriate formant of [ŋ]. However, the formants of [ŋ] do not bend.

4. Intensity assimilation. Word finally, the intensity of [ŋ] shows an abrupt decrease from that of the preceding vowel to zero. The process takes some 80 ms.

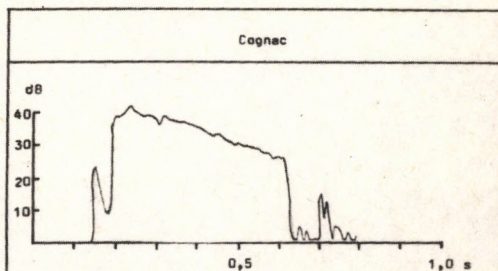
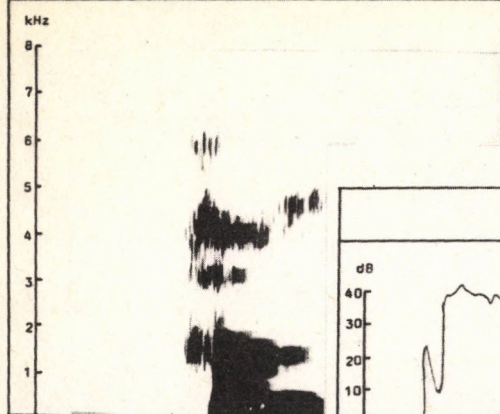
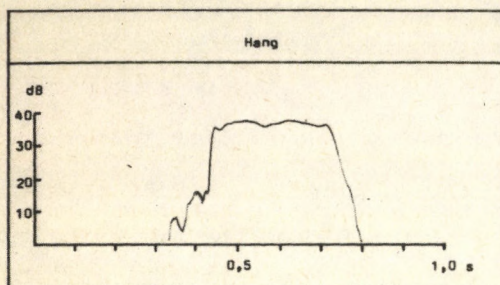
[n̩j]

This is a complex speech sound in German, made up by elements of the phonemes /n/ and /j/.

1. Temporal organization. Total time of articulation 120 ms.

2. Frequency and intensity structure. [n̩j] has a clearly observable formant structure. F1 285--465 Hz, F2 1290--1715 Hz, F3 2415--2540 Hz. Within the frequency band of F1, there is an intensity peak of 22 dB around 420 Hz. The

intensity peak of F2 is around 1500 Hz, 20 dB weaker than the first.



3. Frequency assimilation. F2 and F3 bend to pure phase from the appropriate formants of the preceding vowel. F1 does not bend.

4. Intensity assimilation. Intensity shows a gradual decrease from the closure phase of [n]; by the end of articulation it is some 6 dB weaker than the intensity maximum of the preceding vowel. Intensity keeps on decreasing on transition to the following vowel.

REFERENCES

- BOLLA, K.: A conspectus of Russian speech sounds. Budapest 1981.
- BOLLA Kálmán: Az amerikai angol beszédhangok atlasza. MFF 9. 1981.
- BOLLA Kálmán: A magyar magánhangzók akusztikai analízise és szintézise. MFF 1. 1978, 53--67.
- BOLLA Kálmán--VALACZKAI László: Német beszédhangok atlasza. MFF 16. 1986.
- EZAWA, K.: Die Opposition stimmhafter und stimmloser Verschlusslaute im Deutschen. Köln 1969.
- FANT, G.: Analysis and synthesis of speech processes. In: Manual of Phonetics. Ed. MALMBERG, B. Amsterdam, 1968. 173--277.
- FÓNAGY Iván--SZENDE Tamás: Zárhangok, réshangok, affrikáták hangszíneképe. NyK LXXI, 1969, 281--344.
- GÓSY Mária: A [b, d, g] mássalhangzók percepció vizsgálata. MFF 10. 1982. 84--109.
- KISS Gábor: A magyar magánhangzók első két formánsának

- meghatározása szintetizált hangmintákat felhasználó
percepciók kísérlet segítségével. NyK 87/1. 1985.
- LADEFOGED, P.: Elements of Acoustic Phonetics. Edinburgh and
London 1962.
- LINDNER, G.: Grundlagen und Anwendung der Phonetik. Berlin
1981.
- LITTMANN, A.: Die Problematik der deutschen Hochlautung. In:
Deutschunterricht für Ausländer 15. München, 1965,
65--89.
- LOTZMANN, G.: Sprechwissenschaftliche Aspekte für Aussprache-
normierung des Deutschen. In: Sprach- und Sprechnorm.
Tagungsbericht der 7. Wissenschaftlicher Regionaltagung.
Inzigkofen 1974. Heidelberg, 1974, 65--83.
- OLASZY Gábor: A magyar beszéd leggyakoribb hangsorépítő
elemeinek szerkezete és szintézise. NytudÉrt. 121. 1985.
- TARNÓCZY, I.: Die akustische Struktur der stimmlosen
Engelaute. ALinguH IV, 1954, 313--49.
- VALACZKAI, L. : Über deutsche Aussprachenormen... In:
Budapester Beiträge zur Germanistik. Hrsg.: MADL,
A.--JUHÁSZ, J.--SZELL, Zs. 1978, 323--30.
- VALACZKAI, L.: Untersuchungen zur Funktion der akustischen
Faktoren der distinktiven Perzeption im Deutschen.
ALinguH XXXIV. 1984, 261--70.
- VÉRTES, O. András: A magyar beszédhangok akusztikai
elemzésének kérdései. In: Fejezetek a magyar leíró
hangtanból. Szerk. BOLLA Kálmán. Budapest, 1982, 71--114.
- WANGLER, H.-H.: Grundriss einer Phonetik des Deutschen.
Marburg 1960.

ON NASALITY IN FRENCH

Domokos Vékás

Department of Phonetics, Eötvös Loránd University

This paper mentions (1) some articulatory and acoustic characteristics connected with the phenomenon of nasality in the literature. Then (2) it examines whether the "opposition" $\tilde{U} \sim UN$, having an important morphonological role in French, can be considered firm in a phonetic context where the difference is mainly to be found in the timing of nasalization. Finally, in a more general perspective (3), it calls the attention to certain characteristics of the French sound system (in particular, relations of duration) that probably have a stabilizing effect on the existence and autonomy, often considered to be shaky, of French nasal vowels. On the basis of the second and third points, we can establish the range of parameters whose role or importance is to be clarified for a better understanding and more precise description of French vocalic nasality.

(1) From the point of view of articulation, nasality is a relatively simple, well-defined and widely known phenomenon. But it is worth emphasizing that the total surface of the nasal cavities is quite large as compared to their volume (approximately 50 cubic centimetres, with individual variations), and moreover acoustically it can be considered absorbent (Ohala, 1975); hence it represents a

certain obstacle for the sound-producing air stream which, by its nature, moves forward in the direction of the least resistance. In the case of phonetically nasooral sounds this means moving towards the lips, and if the speaker wants to keep up a relatively strong nasality, he has to open the velopharyngeal aperture the widest possible. This is not unrelated to the fact that, whereas for the pronunciation of French nasal consonants the soft palate moves a little forward but only slightly descends, in the case of nasooral sounds the descent is much more significant (Rochette, 1973); on the other hand, the pharynx wall can draw further back so that the velopharyngeal passage becomes much freer.

It may sound surprising, but this mechanically simple and on the perceptual level positively homogeneous phenomenon is quite complicated acoustically. The spectrum of nasalized vowels is formed by an interaction of formants of oral origin and of formants and antiformants of nasal origin. (Antiformant: "At certain frequencies the standing wave dominating in the nasal cavity is contrary to the wave in the oral cavity and so it results in the extinction or at least drastic reduction of acoustic energy" Liénard, 1977, 81.) The constancy of the size of nasal cavity and the underestimation of the mutual effect of the connected resonating cavities made numerous researchers, in analysing their spectrograms, look for the characteristics of vocalic nasality in terms of invariant features -- without succes (Curtis, 1970.). Experiments made by the help of an electric vocal tract analogue have proved that the intensity and frequency values

of formants and antiformants mainly depend on two factors: on the oral configuration and the degree of coupling of resonating cavities (Curtis, 1970, 70--2; Ohala, 1975, 293--4). For instance: changes of different types can be observed in the spectrums of two vowels of different oral configurations even in the case of a nasalization of completely the same degree.

Theoretically (Mrayati, 1975), the following formants and antiformants can be found up to 3000 Hz or so in the spectrum of a French type (that is, nonhigh nasal) vowel (\tilde{U}):

F_{1n}, A₁, F'₁, F'₂, F_{2n}, A₂, F'₃.

However, the spectrum of a natural nasal vowel does not make the identification of all the enumerated components possible in consequence of its complexity. We know it from the experiments described by Delattre that the most important parameter of vocalic nasality is a 10--15 dB reduction of intensity of the first formant; this reduction is sufficient in itself to produce a nasal vowel by synthesis. But it is more correct to say (Lonchamp, 1979) that the reduction in intensity of F'₁, corresponding to the first formant of an oral vowel, is the consequence of the closeness of an antiformant (A₁). Increasing nasalization, F'₁ takes up higher and higher values, but the increase of A₁ on the frequency axis is more radical: in the case of strong nasalization F'₁ can be almost completely annihilated. But F'₂ is hardly reactive to the degree of coupling between the oral and nasal cavities; the mean frequency value of F'₂ is very similar to the F₂ of an oral vowel of the same

configuration.

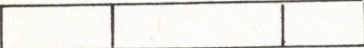
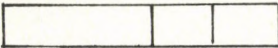
The acoustic characteristics of a nasal vowel are not constant along the time axis: a short initial phase of a few cs at most can be considered oral: even if nasality is present in that phase, it is not dominating at all. The degree of nasalization then quickly increases and an overwhelming portion of the vowel can be considered nasalized. This process is well illustrated on the articulatory level by the cineradiograms of Brichler (1970). On the other hand, only a short phase of a French oral vowel is nasalized in a nasal environment, as opposed to e.g. Portuguese (!) oral vowels (Clumeck, 1975).

(2) It is well known that vowels adjacent to nasal consonants are often at least partially nasalized. Naturally such vowels are phonologically oral, since their nasality is conditioned by the phonetic environment, hence it cannot be considered an inherent feature of the vowel (Ruhlen, 1978). However, phonetically nasooral French vowels can be opposed to oral vowels, and even to combinations of oral vowel plus nasal consonant (VN). According to a number of scholars, the latter oppositional possibility (\tilde{U} ~VN) is also a necessary condition for positing autonomous nasal vowel phonemes (Vachek, 1964). Be that as it may, the opposition \tilde{U} ~VN plays a morphological role in French which is considered to be very important, for example, by Herman (1984, 83); consequently, it is useful to examine what phonetic features differentiate the members of a "quasi minimal pair" like <plein ~ pleine> [pl $\tilde{æ}$] ~ [pl $\tilde{ɛ}$ n].

A heavily nasalized 'nasal' vowel of the French type is relatively easy to tell apart from its oral "counterpart", even for somebody whose mother tongue is not French; provided the latter is not nasalized by an adjacent nasal consonant (since there is none). In certain positions, for example, before a pause, the $\tilde{U} \sim UN$ difference can also be noticed relatively easily by e.g. Hungarians studying French, but there are contexts (as we shall see) where only the French can make the correct decision. It was in connection with such problems that it first came to my mind that a closer look at the issue of vocalic nasality in such environments might be useful.

Though in principle a (Parisian) French nasal vowel has no consonantal extension, there is a type of phonetic context where one is formed by assimilation; in such cases the $\tilde{U} \sim UN$ opposition is phonetically based mainly on a difference in the timing of nasalization: in both cases a more or less nasalized vowel is followed by a shorter or longer nasal consonantal phase. For example, in the sentence <Dominique est plein de bonne volonté> [...plẽdɔbɔn...] (the phonetic transcription is relatively broad) the [ã] strongly nasalizes the following plosive which is homorganic with the dentalalveolar nasal, hence its first half essentially corresponds to a short [n]. The nasalized and oral parts of this plosive can be well separated from each other on a dynamic spectrogram. (This heavy assimilation is not surprising: think of the fact that the soft palate is a relatively less mobile organ and it has not yet closed the

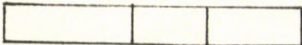
velopharyngeal opening when the oral stop is already formed. Incidentally, a nasal consonant nasalizes its environment to a somewhat lesser degree since in such cases the lowest position of the soft palate is not as low as it is in the case of a French nasal vowel, cf. Rochette, 1973.) Thus, a vocalic part is followed by a nasal consonantal phase in the same way as it is in the case of the adjective in the feminine: <...pleine de bonne...> [plɛ̃dɔ̃b n]; the difference is to be found primarily in the temporal proportions of the components, i.e. in the timing of nasalization:

	A.	B.	C.	total duration
	vocal phase	nasal consonantal phase	nonnasal consonantal phase	(cs)
{nd(ə)}				
	8	12	4	24
æ̃d(ə)				
	11	3.5 3.5		18

In the case of the masculine form the nasalization of the vowel begins early, and the nasalized consonantal phase is short. Only the end of the vowel is nasalized in the feminine form, and the nasal consonantal phase is relatively long since the nasalized first part of [d], homorganic with [n],

is added to it.

The sentence containing the masculine form has another, much more frequent pronunciation without [ə]: [...plæ̃dbɔ̃h...]. In this case (i.e. when U is followed by two oral voiced consonants) progressive nasal assimilation is the strongest possible (cf. Malécot, 1972; Rochette, 1973); it also appears from my spectrograms that about half of the whole length of [db] is nasalized: one could almost say that [nb] is a more precise phonetic transcription:

	A.	B.	C.	total duration
	vocal phase	nasal consonantal phase	nonnasal consonantal phase	(cs)
æ̃db				
	10	6	7	23

We have recorded the sentences containing masculine and feminine forms referred to above in all the possible and some intentionally incorrect pronunciations for a perception test; the informant (a trained phonetician from Paris) read out the sentences at a relatively fast tempo and in an informal style, and spectrograms were made of all the sentences. The preliminary results show that the Hungarian subjects who speak French well and (even though their mother tongue lacks distinctive vocalic nasality) are quite able to differentiate a \tilde{U} from a UN sequence e.g. before a pause, made numerous mistakes in the test; they tended to classify the sentences containing a masculine form (i.e. \tilde{U}) with those containing a

feminine form (i.e. VN), probably because they perceived the consonantal nasal phase as a nasal consonant, and they could not distinguish the early-onset nasality of the vowel before it from that of an oral vowel under the influence of a nasal context. The French subjects correctly perceived the timing difference that was too subtle for Hungarians: they classified all the masculine forms correctly and, in the case of feminine forms, they could hear the VN sequence properly (unlike Hungarians) when the informant artificially omitted the [ɔ] which is obligatory in feminine forms, and even when the [d] and the [ɔ] were both left out via an artificial and arbitrary elision, even though in that case the total duration of [...&nb...] agreed with that of [... \tilde{a} db...] (or rather [... \tilde{a} nb...]). In sum, the French subjects correctly interpreted VN sequences even when we added some artificially-produced phenomena normally concomitant only with \tilde{U} , though they expressed reservations concerning the acceptability of those unnatural forms.

So we can say that the $\tilde{U} \sim VN$ difference has proved to be firm even in the most critical situation. Along with the difference in the timing of nasality, this unequivocal result is also due to the fact that the oral configuration of nasal vowels somewhat differs from that of their oral "counterparts".

(3) The acoustic model of vocalic nasality raises the question of the stability of nasal vowel systems. On the basis of the foregoing it is by no means to be considered accidental that we do not find nasal vowels in the

overwhelming majority of languages in the world; and where they do exist, their number does not always reach, and never exceeds, that of the oral vowels. The only significant stable point of nasal vowels is the second formant, and the frequency values of second formants have to differ from each other as much as possible for two such vowels to be distinguished. Since the frequency value and intensity of the first formant considerably vary or get weakened depending on the degree of nasalization, it can help the identification of the vowel only to a lesser extent.

The system of French nasal vowels was examined by Walter (1982; 1984) from the point of view of regional variants and she found three main types of systems (in addition, she found a few more types with a very small number of informants each). If we have a closer look at the systems she surveyed, we can make the following remarks: In the case of one of the four-member systems the oral configurations of the vowels concerned (in terms of Walter's articulatory characterization), hence presumably also their second formants, are very similar to those of the "corresponding" oral vowels. It is just in this system that we can find short consonantal extensions which can compensate for the occasional insufficiency of nasality at the perceptual level (Ohala points out that a basically oral vowel will be perceived as nasalized in direct proportion with the shortness of the following nasal consonant -- 1975, 298, 303). In another type of four-member systems the configurations (and second formants) of nasal vowels differ

from those of their oral "counterparts", occurring in a VN sequence, to various degrees, and there are no consonantal extensions. This system traditionally represents the phonetic standard of the language, even today. The values of the second formants of the four members were measured by Lonchamp. One of his most interesting remarks is that the frequency value of the second formant of [æ̃] is not more than 1400--1500 Hz, although it used to be 1600--1700 Hz according to measurements taken a few decades earlier.

So [æ̃] is acquiring a greater autonomy with respect to its oral "counterpart", but the value of its second formant now comes very near to that of the second formant of [œ̃] (about 1250 Hz). It is exactly the latter differentiation that is missing in the third type of (three-member) systems, where the realizations are something like [æ̃], [œ̃], [õ̃]. It is in this system that the acoustic difference (i.e. the difference between the second formants) is the greatest both among nasal vowels, and between nasal vowels and their oral "counterparts". Since this nasal vowel system is increasingly gaining ground to the expense of the others, it seems that the autonomy of nasal vowels, denied by some linguists and considered to be shaky by many others, can become stronger in the French vowel system.

Nasal vowels are somewhat longer than oral vowels in every position, therefore it has been suggested (e.g. by Delattre and Monnot, 1968, 287) that length may outlive nasality when the latter ceases to exist one day, taking over

the distinctive role. However, that assumed tendency is contravened by the fact that nasal vowels are significantly longer only in close syllables and especially when followed by a voiceless consonant; also, the length of a French vowel depends primarily on stress, a fact that would somewhat impede the development of a short vs. long correlation in vowels. We also have to take into consideration that nasal vowels, if further lengthened, could reach the total duration of a VN sequence -- but this is to be "avoided" since it would impede the differentiation. For the time being, the length of a VN sequence approximately corresponds to the average length of a $\tilde{V}C$ sequence; according to the measurements of Uihanta (1978), in stressed syllables they are $128 + 104 = 232$ msec, and $178 + 65 = 243$ msec, respectively.

It further appears from the above data (which are also supported by my own measurements) that the average length of French nasal consonants is considerable, it exceeds the average length of oral consonants. This can be important in several respects: on the one hand, we perceive a vowel followed by a too short nasal consonant as more nasalized (Ohala, 1975) and it is essential that a French oral vowel should remain as oral as it possibly can; and on the other hand, a relatively long duration makes it possible that the opening and closing of the velopharyngeal passage should not overlap too much with the production of the neighbouring sounds, and consequently that the preceding vowel should be as little nasalized as possible. (Incidentally, if the slight

nasalization of oral vowels is insufficient for the recognition of the consonant that follows, an increase of the length of the latter can make up for the missing information.) Moreover, a longer nasal consonant differs more significantly from the short consonantal extension of nasal vowels formed in the course of progressive nasal assimilation, and in some contexts (as we have seen) this can have an important role. We can assume that from the point of view of distinctive vocalic nasality it is a favourable, i.e. stabilizing, condition that there is no short vs. long correlation in the French consonant system either, and therefore the relatively considerable length of nasal consonants is not in direct danger.

In order to determine more precisely the types and relative importance of the various parameters that play a role in vocalic nasality, it would be expedient to establish the nasal and oral, vocalic and consonantal (!) phases of ideal quality and duration of oral and nasal vowels and VN sequences in various phonetic contexts, and on the basis of this to make up a perception test with synthetically produced stimuli and to analyse the influence of various deviations. It could be useful to carry out such experiments with subjects of various linguistic backgrounds; the results might contribute to a better understanding of the perceptory and articulatory bases of various languages.

REFERENCES

- BRICHLER-LABAEYE, C.: Les voyelles françaises, mouvements et positions articulatoires à la lumière de la radiocinématographie. Paris 1970.
- CLUMECK, H.: A cross-linguistic investigation of vowel nasalization: an instrumental study. In: Nasalfest, Papers from a symposium on nasals and nasalization. Ed. by FERGUSON, C.A.--HYMAN, L.M.--OHALA, J.J. Stanford, California 1975.
- CURTIS, J.F.: The Acoustics of Nasalized Speech. CPJ 1970. vol. 7. 380--96.
- DELLAITRE, P.--MONNOT, M.: The Role of Duration in the Identification of French Nasal Vowels. IRAL VI, 1969. 295--325.
- HERMAN, J.: Phonétique et phonologie du français contemporain. Budapest 1984.
- LIENARD, J.S.: Les processus de la communication parlée. Introduction à l'analyse et à la synthèse de la parole. Paris--New York--Barcelone--Milan 1977.
- LONCHAMP, F.: Analyse acoustique des voyelles nasales françaises. Verbum II, 1. Institut de Phonétique--Université de Nancy II 1979.
- MALECOT, A.--METZ, G.: Progressive Nasal Assimilation in French. Phonetica 26. 1972. 193--209.
- MRAYATI, M.: Etudes des voyelles nasales françaises. Bulletin de l'Institut de Phonétique de Grenoble 1975.
- OHALA, J.: Phonetic Explanations for Nasal Sound Patterns. In: Nasalfest ... 1975. 289--317.

- ROCHETTE, C.E.: Les groupes de consonnes en français. Etude de l'enchaînement articulatoire à l'aide de la radiocinématographie et de l'oscillographie. Paris 1973.
- RUHLEN, M.: Nasal Vowels. In: Universals of human language. Ed. by GREENBERG, I.H. Stanford 1978.
- VACHEK, J.: On some basic principles of 'classical' phonology. Zeitschrift für Phonetik 17. 1964. 409--31.
- VIHANTA, U.U.: Les voyelles toniques du français et leur réalisation et perception par les étudiants finnophones. Jyväskylä 1978.
- WALTER, H.: La dynamique des phonèmes dans le lexique français contemporain. Paris 1976.
- WALTER, H.: Rien de ce qui est phonique n'est étranger à la phonologie. In: Discussion Papers for the Fifth International Phonology Meeting. Wiener Linguistische Gazette 1984. 276--80.

Címünk:

MAGYAR FONETIKAI FÜZETEK

A Magyar Tudományos Akadémia

Nyelvtudományi Intézete

Fonetikai Osztály

Budapest I., Szentháromság u. 2. Pf. 19.

1250

Address for communications:

HUNGARIAN PAPERS IN PHONETICS

Department of Phonetics,

Institute of Linguistics,

Hungarian Academy of Sciences

Budapest I., Szentháromság u. 2. Pf. 19.

H--1250

