COMPUTER AIDED ANALYSIS OF STRESSED AND UNSTRESSED BULGARIAN VOWELS FROM 30 MALE AND 30 FEMALE SPEAKERS

PHILIP CHRISTOV

Voice Man-Machine Communication Lab.
Mechanics & Biomechanics Institute
Sofia, Bulgaria, 1090, P.O. Box 373

ABSTRACT

In a previous paper (CHRISTOV, Proc. 11th ICA, vol. 4, pp. 161-164) an algorithm has been reported about the analysis of speech in the spectral domain by the use of phonetic knowledge. Here the results of the application of this algorithm will be discussed in the computer aided analysis of the Bulgarian vowels. The vowels are imbeded in words and uttered by 30 male and 30 female speakers in /b-b/ context. The Computer input is verified by 20 listeners.

INTRODUCTION

The research reported in this paper is aimed at the build up of an acoustic-phonetic fulcrum of the Bulgarian vowel system for the purposes of the mashine recognition of continuous speech.
Following that goal a research strategy is adopted which was implemented successfully in vowel analysis by Potter & Steinberg /1/, Peterson & Barney /2/ and others. The root of it is in using phones of comparable allophones imbeded in words uttered in equal phrases with equal intonation by 25 to 30 male and female speakers.
In addition here central allophones are used together with their unstressed contrasts and the persons who uttered them are carrefully selected among the best professional speakers in the country.

SPEECH DATA

The input speech data /3/ form a body of 614 vowel phones belonging to the central Bulgarian allophones /bV̀b/ and to their unstressed contrasts /bVb/, where V stands for a vowel from the Bulgarian vowel system. They are uttered in Standard Bulgarian by 30 male bariton and 30 female mezzosoprano professional speakers in a highly damped room and recorded with studio equipment. The vowel /b-b/ context speech fractions are imbeded in words (See APPENDIX) pronounced by the speakers with falling intonation at the end of a standard carrier sentence. A semiautomatic procedure /4/ was used for the verification of the experimental material by 20 listeners. The verification procedure rejected both false listeners and incorrect utterings so that to the computer input have been admited only the utterings for which full agreement was reached between the speaker and each of the reliable listeners.
The analog speech signal was digitalized with a frequency of 20 kHz and stored on 1/2 inch IBM-compatible magnetic tapes. This way four machine compatible sets N, where N=I, II, III, IV, of vowel utterings have been prepared to be used as machine input in this research:

I	- Set of by	UNSTRESSED FEMALE	vowels uttered speakers
II	- Set of by	STRESSED FEMALE	vowels uttered speakers
III	- Set of by	UNSTRESSED MALE	vowels uttered speakers
IV	- Set of by	STRESSED MALE	vowels uttered speakers.

ANALYSIS

The speech analysis algorithm /5/ uses phonetic knowledge to perform a pitch synchronous Fast Fourier Transform and to determine the formant frequencies in the quasi-stationary region of vowels.
The software realization of the algorithm is modular with input and output modules so that each principal module can be used either as an independent program or as a subroutine in a main program. It has the shape of an application program package including program modules for:

Pitch Extraction
Waveform Analysis
Quasi-Stationary Region &
  Representative Section Determining
Fourier Transform
Formant Tracking
Phonetic Classification & Correction

The pitch extraction module is programmed after an algorithm of B. Gold /6/.
The Fourier Transform has been limited to the 30-th harmonic of the fundamental frequency because above that harmonic no traces of higher harmonics have been detected.
The Fourier Transform module gives its output except on cards but also in the form of a computer diagram of the amplitude spectra of the representative section.
After the processing of all vowel utterings with an IBM 360/40 computer it has been established: 1) That the algorithm performes better by male voices and unstressed vowels; 2) That in 18.06% of all vowel utterings it failed to determine the position of the second formant. In all vowel utterings, not resolved by the computer, the position of the second formant has been determined manually using the machine printings of their amplitude spectra.

RESULTS

In result of the analysis the /b-b/ context vowel utterings turned into labeled vectors

$$(F_o, F_1, F_2, F_3)_{LABEL}$$

or points in the measurement space of the fundamental and the firs three formant frequencies. Each vector or point in the measurement space, together with its label, is punched on a single card. The manually corrected vectors are also punched on cards and added to the computer output to form together a set of statistical distributions of equally labeled points

$$W_{NV} \subset W_N, \; N=I,II,III,IV; \; V=i,e,a,ə,o,u$$

in the measurement space. These statistical distributions will be called here, for clarity, also "individual vowel clusters". A program for statistical processing is used to print cross-section diagrams of the statistical distributions and to compute their statistical estimates: mean, standard deviation, maxima, minima, skewness and kurtosis. The results of the statistical processing of the individual vowel clusters W_{NV} will be presented in a more technologically oriented report.
The commonly accepted two-dimensional graphic representation of three-formant analysis of vowels /7/ is adopted also in this study. The graphs are produced automatically by a computer program which first calculates the value of the approximate measure

$$F2' = F2 + \frac{1}{2} \frac{(F2 - F1)}{(F3 - F1)} (F3 - F2)$$

of the effective pitch of the higher formant group and then prints the usual F1xF2' vowel diagrams in the computer output listing.
Each of the supplemented figures (Fig.Fig. 1 to 4) shows six closed loops plotted manually around the main body of an individual vowel cluster $W_{NV} \subset W_N$ in the machine F1xF2' graph.
The amount of points inside each closed loop and of these which remain outside or enter alien closed loops is presented in attendant confusion matrixes (Tables 1 to 4).

DISCUSSION

The first impression after looking at the four diagrams in the last page of the report is that by the stressed vowel male utterings (Fig. 4) the closed loops are neatly outlined and distinctly delimited. There is no overlapping at all but many single points, which are far away from their clusters nuclei, remain outside of the closed loops (Table 4).

In contrast, by the stressed female utterings the distributions are more uniform and widely spread so that there is much overlapping (Fig. 2) but very few points remain outside the closed loops (Table 2).
By the unstressed utterings (Fig.Fig. 1 and 3) the clusters shrink around their nuclei but, nevertheless, there is some overlapping between the clusters of /o/ and /u/ for the male voices (Table 3) and between /a/ and /ə/ for the female voices (Table 1).
Despite of the overlapping the general set up of the vowel triangel is well preserved in the F1xF2'-diagrams of all four vowel cluster sets W_N shown in the figures.
A closer examination of the vowel triangles presented in Fig.Fig. 1 to 4 shows that by stressed vs. unstressed contrasting (Fig.2 vs. Fig. 1 and Fig. 4 vs. Fig. 3) there exsists a certain stretching of the vowel triangle along the F1-axis (Left to right in the F1xF2' area).
A similar stretching can be detected during the female vs. male utterings contrasting (Fig. 1 vs. Fig. 3 and Fig. 2 vs. Fig. 4) but this time along the F2'-axis (Bottom to top in the F1xF2' area).

CONCLUSION

It appears that, due to the carrefully selected and handled speech material in this study and facilitated by the simple Bulgarian vowel system, it has been possible to come within reach to some natural phenomena concerning the differences between stressed and unstressed vowel clusters produced by men and women.

REFERENCES

/1/ K. R. POTTER, J. C. STEINBERG, "Towards the Specification of Speech", JASA, 22, 807-820, 1950; /2/ G. E. PETERSON, H. L. BARNEY, "Control Methods Used in a Study of the Vowels", JASA, 24, 175-184, 1952; /3/ Ph.CHRISTOV, "A Large Bulgarian Central Allophones Data Base", 11.ICPS, Tallinn, 1-7 August 1987 (Receipt No 0653); /4/ Ph. CHRISTOV, "A Semiautomatic Speech Sound Aural Identification Procedure with Its Application to Speech Analysis", Acustica, 29, 374-349, 1973; /5/ Ph. CHRISTOV, "An Algorithm Using Linguistic Information and Its Application to the Analysis of Speech in the Spectral Domain", Proc.11.ICA, 4, 161-164, 1983; /6/ B. GOLD, "Computer Program for Pitch Extraction", JASA, 34, 916-921, 1962; /7/ G. FANT, "Modern Instruments and Methods for Acoustic Studies of Speech", The Royal Inst.of Technology: Rpt.No 8/June 11, 1957, p. 14

APPENDIX

Word List: (In rough phonemic IPA transcription)

Stressed: /bìblija/, /bèbe/, /bàba/, /bə̀brek/, /bòbof/, /bùba/

Unstressed: /biblèjski/, /bebètʃef/, /babalə̀k/, /bəbrekovìden/, /bobòvina/, /bubàr/

NOTICE: In Tables are count F1xF2'-labeled points positions, not actual vowels number

Table 1. Confusion matrix to Fig. 1.

	i	e	a	ə	o	u	none	total
i	20	3						23
e		20						20
a			13	4			1	15
ə			3	19				20
o				1	13			14
u					2	9	1	12

Table 2. Confusion matrix to Fig. 2.

	i	e	a	ə	o	u	none	total
i	23	6						23
e	4	25		2				25
a			22	7				24
ə		2	5	21				21
o			3	4	18			23
u					2	13	2	17

Table 3. Confusion matrix to Fig. 3.

	i	e	a	ə	o	u	none	total
i	24							24
e	1	26					1	28
a			18	1				19
ə			1	13	1		1	16
o					17	3	3	20
u					5	17		17

Table 4. Confusion matrix to Fig. 4.

	i	e	a	ə	q	u	none	total
i	25						2	27
e		24						24
a		1	20	1			1	23
ə				19			1	20
o			1	2	19	1	3	26
u					2	20	2	24

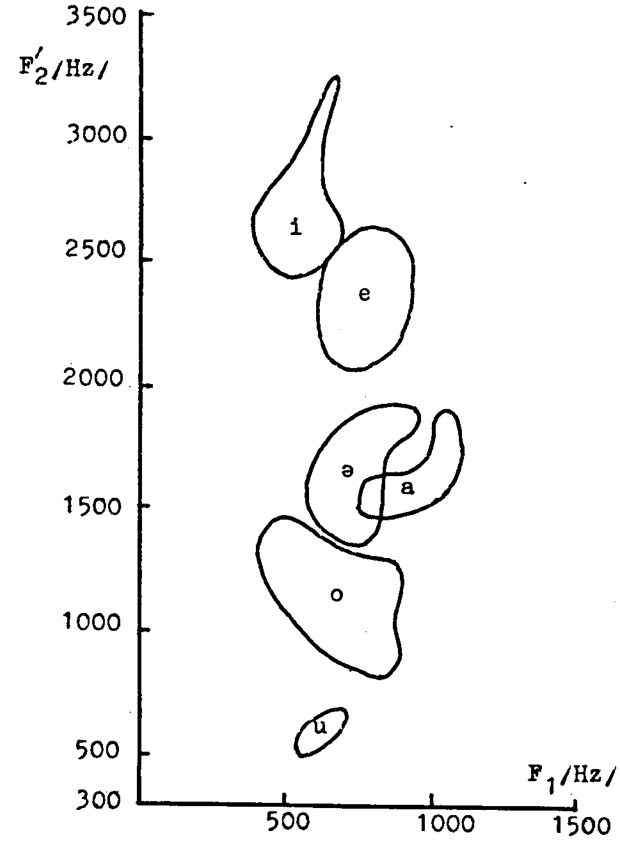

Fig. 1. First vs. second formant diagram of the Bulgarian vowels in /bVb/ context uttered by 30 female speakers

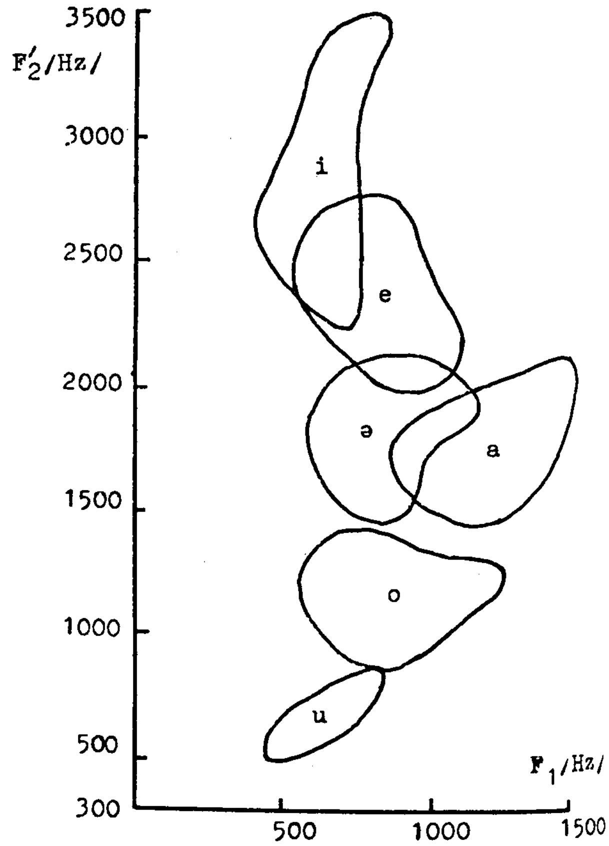

Fig. 2. First vs. second formant diagram of the Bulgarian vowels in /bV̌b/ context uttered by 30 female speakers

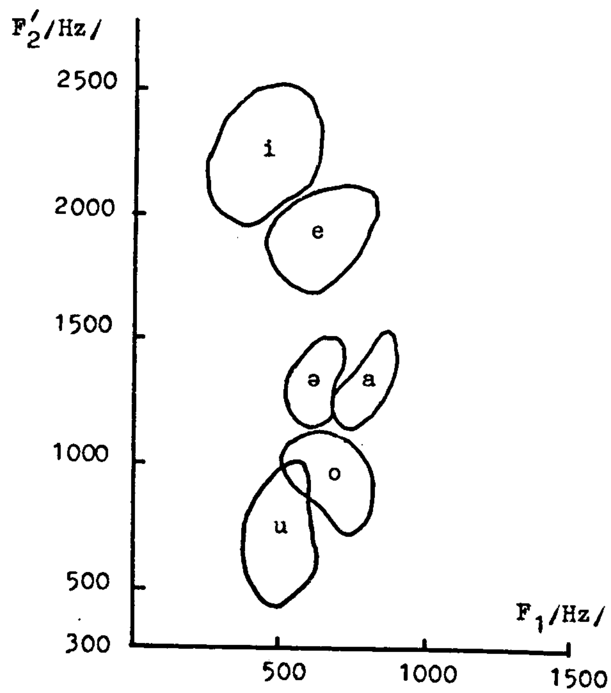

Fig. 3. First vs. second formant diagram of the Bulgàrian vowels in /bVb/ context uttered by 30 male speakers

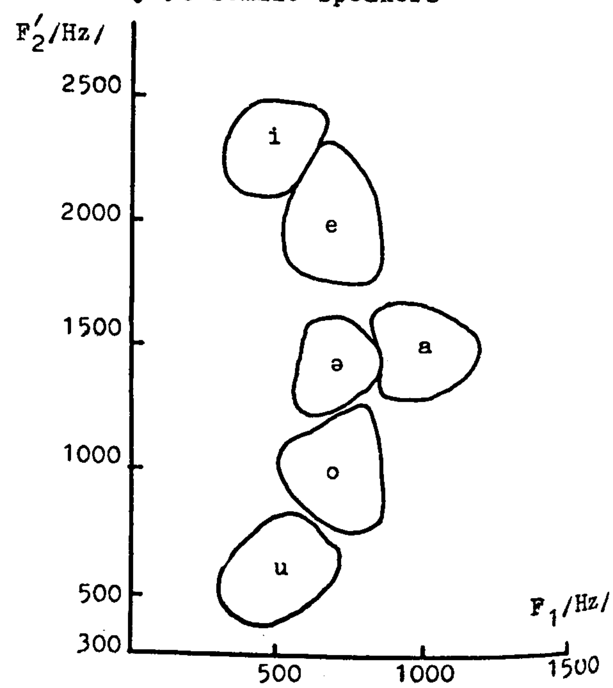

Fig. 4. First vs. second formant diagram of the Bulgarian vowels in /bV̌b/ context uttered by 30 male speakers