SPECTRAL PROPERTIES OF RUSSIAN STRESSED VOWELS IN THE CONTEXT OF PALATALIZED AND NONPALATALIZED CONSONANTS

VLADIMIR B. KUZNETSOV

Moscow State Institute of Foreign Languages
Moscow, USSR 119034

ARVO OTT

Dept. of Computer Control
Institute of Cybernetics
Tallinn, Estonia, USSR 200108

ABSTRACT

Results of an experimental study of the formant frequencies at the steady-state part of russian stressed vowels in the context of palatalized/nonpalatalized consonants are presented. The analysis of the obtained data reveals that there is a considerable difference in the quality of vowels due to the consonantal environment. Traditional static description of vowel quality (F-pattern measured at the stationary portion of vowel) is basically insufficient for representation of the allophonic variation. One must take into account the dynamic properties of transitions as well.

INTRODUCTION

Specific character of the allophonic variation of russian stressed vowels is in great part attributed to the palatalization/nonpalatalization (P/NP) of the surrounding consonants. There is practically no coarticulation between palatalized consonant and the following vowel: the onset frequency of the second formant transition is primarily determined by the palatalization of the consonant [1]. Under these circumstances one should expect the quality of the vowel to undergo some changes. In fact, experiments on the identification of stressed vowels, segmented from monosyllables or words uttered in isolation, have shown that russian listeners are able to recognize some 18 vowels depending on consonantal environment [2]. On the basis of spectrographic analysis of these vowels it was concluded that the listeners' ability to distinguish so many allophones was due to the presence in the structure of vocalic nucleus of the so called "i-like" second formant transition - acoustically and perceptually reliable cue of consonantal palatalization.
Judging from the data on the second formant frequency of the vowels measured at their stationary part [3], the phonetic quality of the vowels as such does not change under the influence of the preceding P/NP consonants of different place of articulation. This observation is supported by the results obtained in a study of formant frequency patterns in russian VCV utterances [4].
Entirely different results are reported in [5,6]. It is shown that palatalization of surrounding consonants produces systematic effect on the steady-state part of the vowels uttered in isolated syllables or words.
In recent years experimental studies of vowel perception have provided some new evidence that leads us to question the correctness of the view that russian listener recognize vowel allophones on the basis of characteristic "i-like" formant transition.
As it has been shown in [7], Russian children, having yet no knowledge of foreign languages, can classify a set of 20 stationary vowels into 13 categories, some of which (for example the front vowels) are not listed in the phonemic inventory of Russian. Experimental study of perception of the vowel-like stimuli with a changing frequency of the second formant [8] has revealed that the onset frequency and the direction of the formant transition are useful perceptual cues for identification of the vowel quality. A recent experiment on the identification of the steady-state part of the vowels segmented from continuous speech has shown that these segments convey information concerning P/NP of the consonantal environment [11].
From all the facts presented above it is clear that both the spectral data and the proposed interpretations of its perceptual significance are contradictory and incomplete. To make some progress in understanding the nature of the allophonic variation in question, one must begin by collecting basic quantitative data on the spectral properties of the vowels. The present study was designed to fulfil in part this task.

SPEECH MATERIAL AND METHOD

Test vowels were uttered in a nonsense monosyllable of CVC type embedded in the carrier phrase "Say ... again". Palatalized or nonpalatalized fricative [s] was used to form a symmetrical environment. Each of the three male speakers recorded a list of 330 sentences (10 vowels * 33 repetitions). To achieve constant speech rate throughout the recording session the speaker was asked to synchronize the onset of the sentence with a periodic light pulse. The phonetic identity of the test vowels was checked up by 8 listeners in an identification experiment. Spectral analysis was computer-implemented. Prefiltered speech signal was sampled at 10 kHz by an 8-bit A/D converter and subjected to Fast Fourier transformation to compute power spectrum of the signal. The analysis window shift was 128 points. An automatic algorithm was used to compute the frequencies of the first three formants [12]. The test vowels were characterized by an F-pattern measured at the point where F2 reached its extreme value or in the middle of the vowel if there was no extreme. To enhance the reliability of formant peak location wideband sonograms were regularly made. By tracking formant trajectories on the sonograms in the vicinity of the vowel segment we were able to identify and discard spurious peaks present in the vowel spectra.

RESULTS AND DISCUSSION

The data obtained are presented in Table 1 and Fig.1. The table lists the means and standard deviations of the formant frequencies of the three speakers. Fig.1 displays the data of speaker M1 on the F1-F2 space. We use cyrillic characters to symbolize vowel allophones. The dots above transcription sign indicate the allophone surrounded by palatalized consonants. The results of the vowel identification test are as follows: speaker M1 - confusion error rate 0.30 %, speaker M2 - 4.00 %, speaker M3 - 1.40 %. The range of the vowel durations of all speakers is from 75 to 120 msec. For speaker M1 and M2 the duration means of the test vowels are reported in [10]. The vowel configuration depicted in Fig.1 is typical, exept for some minor details, to all the speakers. The most obvious conclusion to be drawn from the figure is that the F-pattern of the vowels in the context of palatalized consonants differ greatly from that one in the context of nonpalatalized consonants. The observed differences are much greater than those reported in [5,6] but in both cases they clearly go along the same lines. In the F1-F2 space the vowels [ӱ] and [ӧ] occupy the areas that are usually labelled as [y] and [øe] respectively in the languages that have front rounded vowel phonemes. Our data does not support the traditional phonetic notion that in Russian [э] is more close and front than [ӓ]. All the speakers have their second formant of the vowel [ӓ] higher than that one of [э], which indicates a more forward position of the tongue.

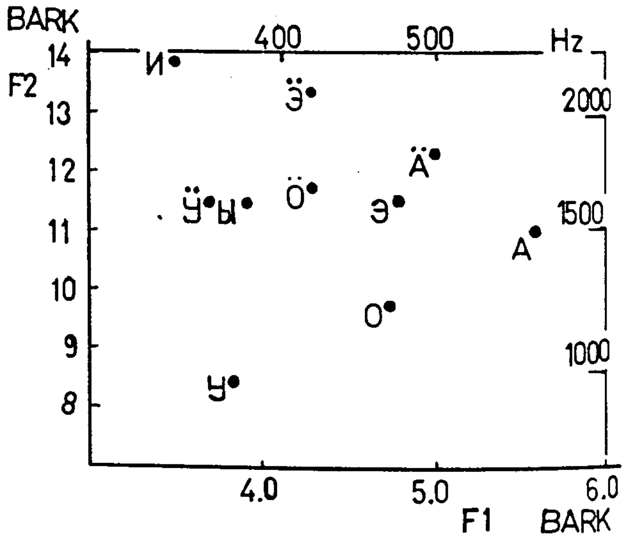

Fig.1. Vowel F-pattern of speaker M1

Examination of Fig.1 and the tabled data shows that the areas occupied by the vowel pairs [ӱ-ы], [ы-ӧ], [ӧ-э] and [э-ӓ] in the F1-F2 space may considerably overlap. Using only F-pattern at the steady-state part of the vowel there is no way to find out which of the two vowels we are dealing with.
The identification data concerning the speech materials of speakers M1,M2,M3 and two more female subjects indicate that the listeners most often confused the following vowel pairs: [ӭ-и] - 111 errors, [о-у] - 64 errors, [э-ы] - 29 errors, [ӧ-ӱ] - 28 errors and [а-э] - 26 errors.
Evidently, the vowels are confused when they are uttered in the same consonantal environment. The vowels that are the nearest neighbours in the F1-F2 space (for example, [ӱ-ы], [ы-ӧ], [ӧ-э],[э-ӓ]) but never occuring in the same consonantal context are discriminated quite easily.
The role of the P/NP consonants in the phonetic interpretation of the quality of vowel sounds was demonstrated in the investigation [9] where natural vowel in CVC syllable was replaced by synthetic stationary stimuli. The stimuli with the following coordinates in the F1-F2 space: 470/1560 Hz and 520/1560 Hz - were perceived as [ӧ] in the context of palatalized consonants and as [э] when the consonants were nonpalatalized; the stimuli with the coordinates 320/1420 Hz and 350/1420 Hz were recognized as [ӱ] in the first case and as [ы] - in the second.

Table 1. Formant frequency means and standard deviations for speakers M1,M2,M3

		F1		F2		F3	
	M1	578	24	1501	27	2558	30
а	M2	520	21	1489	50	2493	106
	M3	560	35	1521	36	2492	64
	M1	506	22	1810	36	2696	30
ӓ	M2	504	18	1943	76	2459	158
	M3	496	25	1882	98	2514	79
	M1	487	22	1633	39	2518	29
э	M2	479	26	1674	58	2569	100
	M3	452	22	1686	56	2453	32
	M1	429	26	2170	50	2801	53
ӭ	M2	403	25	2192	90	2498	324
	M3	405	19	2221	69	2621	155
	M1	470	23	1210	44	2481	48
о	M2	434	28	1192	124	2667	167
	M3	441	23	1117	56	2401	76
	M1	424	16	1652	51	2510	34
ӧ	M2	411	28	1720	112	2210	169
	M3	407	20	1611	179	2209	64
	M1	372	19	976	61	2410	40
у	M2	371	20	1064	94	2773	230
	M3	392	14	951	69	2420	69
	M1	346	16	1624	130	2391	51
ӱ	M2	372	20	1705	115	2325	130
	M3	385	13	1926	194	2297	57
	M1	375	21	1627	83	2364	50
ы	M2	385	16	1739	89	2430	176
	M3	405	19	1740	131	2398	85
	M1	337	21	2322	53	2813	47
и	M2	361	20	2329	47	2960	98
	M3	374	20	2246	70	2603	81

An identification experiment [8] with the vowel-like stimuli having F1=300 Hz throughout stimulus duration and F2 consisting of symmetrical initial and final transitions and steady-state part varing from one stimulus to another in the range of 1200-1600 Hz, has shown that the stimuli were perceived as [ы] when the slope of the initial F2 transition was positive and as [ӱ] when it was negative.
In Russian, as it is known [1], the second formant transition from palatalized consonant into the stationary part of the vowel [ӱ] is falling and it is rising from nonpalatalized consonant into [ы]. Thus, it might be concluded, that in Russian the second formant transition from consonant into the following vowel not only conveys information concerning the place of consonant articulation or its palatalization, but is a useful cue for the phonetic classification of vowels as well.
A complex analysis of the data on inherent vowel duration [10], vowel spectra and confusion errors leads us to the assumption that vowel identity is coded not only in its spectral parameters but in its duration as well, though in Russian the duration of vowel is not phonologically significant. There seems to be an inverse relationship between the spectral and the perceptual similarity of the vowels, on the one hand, and their difference in inherent durations, on the other. For example, speaker M1 provides for the poorly discriminated vowel pairs the following duration contrasts, expressed in percent to the range of inherent vowel durations, that in his case is 33 msec: [и-ӭ] - 88 %, [ы-э] - 64 %, [у-о] -45 %, [ӱ-ӧ] - 35 % and [э-а] -20 %. An interesting aspect of these data is that large differences in duration are established for those vowels that are spoken, judging from the spectral data, with the same tongue height - the most significant and linguistically universal factor determining inherent vowel duration. It is tempting to make a suggestion that in Russian the duration of some vowels does not result automatically from the universal physiological limitations of the articulatory mechanism, but is generated at a higher control level.
It is necessary to touch upon the subject of the vowel production in the context of P/NP consonants. In the conventional view it is assumed that in both cases vowels are produced by the same articulatory gesture. As to the observed differences of the F-pattern at the steady-state portion of the vowels they are ascribed to the effect of the "undershoot". It is believed that the target articulation could be achieved if the duration of the vowel is made large enough. We think that the amount of spectral differences found in this study for the vowels in question could not be explained within the framework of the "undershoot" theory. We support the view developed by Lindblom that the undershoot observed for vowels during faster speech is programmed into the gesture and is not a result of a too fast succession of motor commands. The essence of the speech production process is not an inefficient response to invariant control signals but elegantly controlled variability of response to the demant for a relatively constant end [13].

CONCLUSION

In the present work it has been shown that there is a considerable difference in the spectral properties of russian stressed vowels at the steady-state part depending on the P/NP of surrounding consonants. The allophonic variation can not be adequately represented if one uses only static characteristics (F-pattern at the stationary part). The dynamic properties of the second formant transition should be taken into consideration as well. Data from the present study together with [10] indicates that there is a tendency to compensate for the lack of spectral contrast between vowels by increasing their difference in duration.

ACKNOWLEDGEMENT

The authors would like to thank T.Timmerman for his helpful assistance in stimuli measurements.

REFERENCES

[1] O.F.Krivnova, V.B.Kuznetsova "Phonetic nature of consonantal feature palatalization/nonpalatalization in Russian" In: "Theoretical and historical problems in Russian" (in russian), Moscow, p.37-67, 1980

[2] L.V.Bondarko, L.A.Verbitskaya et al. "Recognizable sound units in Russian speech" In: "Production and perception mechanism of complex sounds" (in russian), Moscow, p.165-179, 1966

[3] L.V.Bondarko "Phonetic representation of language and phonological representation of speech" (in russian), Leningrad, 1981

[4] E.T.Purcell "Formant frequency patterns in Russian VCV utterances" J.Acoust.Soc.Am., vol.66, p.1691-1702, 1979

[5] B.M.Lobanov "Classification of russian vowels spoken by different speakers" J.Acoust.Soc.Am., vol.49,p.606-608, 1971

[6] M.Romportl "On the vowel system of Russian" In: "Studies in phonetics", Prague, p.37-58, 1973

[7] V.B.Kuznetsov "Classification of steady-state vowels by russian listeners" (in russian) In: "Sensory systems", Leningrad, p.105-116, 1979

[8] V.V.Lublinskaya, N.A.Slepocurova "Perception of vowel-like sounds with a changing spectrum" (in russian), Human psychology, Moscow, 3(1), p.77-84, 1977

[9] N.V.Puhir, L.P.Sherbacova "Perception and phoneme classification of phonetically ambiguous vowels" (in russian) In: "Proc. of seminar on automatic recognition of auditory patterns", Lvov, v.2, p.77-80, 1974

[10] V.B.Kuznetsov, A.Ott, A.V.Ventsov "Inherent vowel duration in Russian: production and perception data" (contribution to the present congress), Tallinn, 1987

[11] V.I.Garbaruk "Feature perception of speech sounds" (unpublished dissertation, in russian), Leningrad, 1984

[12] J.D.Markel "Digital inverse filtering, a new tool for formant trajectory estimation", IEEE Trans. Audio Electroacoust., vol. AU-20, p.129-137, 1972

[13] B.Lindblom "Chairman's comment on session 2" In: "Speech Communication Proc. of SCS", Stockholm, vol.2, IX-XXV, 1974