

THE LENGTH AND VARIABILITY IN CONNECTED SPEECH FOR RUSSIAN

R. Potapova

Department of Applied and Experimental
Linguistics. Moscow Institute of Fo-
reign Languages.

ABSTRACT - This paper presents automatic segmentation of speech in the "bottom-up" way taking into account a number of linguistic constraints: the specification of texts based on the classification of speech acts, the classification of text fragments on the basis of semantic-syntactic analysis, the segmentation of the utterance when the end-points of a phrase are determined, the segmentation into syllabic units, the specification of length and its variability in connected Russian speech.

INTRODUCTION

Of late considerable progress has been achieved in creating systems for "man-to-computer" communication on the base of a natural speech. Utilization of spoken speech in working out "man-to computer" communication systems caused the differentiation between isolated command recognition and continuous speech recognition and understanding. The solution of the task of spoken continuous speech recognition becomes more complicated due to a number of difficulties, i.e. variation of the speech input signal dependent on the speakers peculiarities of pronunciation and his emotional state as well as on the loudness, tempo and distinctiveness of pronunciation, the character of acoustic noise, the amount of lexical, syntactic and semantic information.

PARAMETRICAL DESCRIPTION OF SPEECH SIGNAL

The acoustic segmentation of speech is performed on the basis of prosodic and spectral speech data. The text in Russian is preprocessed with the speech features extractive device and sent to the computer in a string of successive readings. The reading may be represented as a vector which coordinates are definite speech signal parameters in a fixed time interval. The speech features extracting device in an analog computing device which performs automatic extraction of spectral and averaged timing characteristics and sends the digitised signal as output (input) information to a computer with the definite frequency. The time interval within the limits of which speech parameters are extracted is equal to 10 msec (accordingly the output signal frequency is 1000 Hz). For a set of low-energy frequencies the interval is equal to 20 or 40 msec. In this case the corresponding parameters can reach only each second or fourth reading. All the extracted parameters can be divided into 3 groups of timing characteristics, of spectral and noise constituent characteristics. All the digitised(non-binary) cha-

racteristics are measured in relative values. Scale coefficients, the limits of measurements and the threshold for the binary features are specially selected for speech domain. The parameters extracted with the analog device can be viewed as parametric representation of the speech signal containing the sequences of the ten measurements correlated in time: $I, F_1, F_2, F_0, N_0, N_2, N_5, O_1, O_2, O_3$.

The study of temporal characteristics was also aimed at the singling out of such parts of the temporal contour of a sentence which could be correlated with the intonation of finished - unfinished statements and questions. The main aim of the investigation predetermined many subordinate problems. It was necessary to find out the role of temporal characteristics of vowels with sentence stress to differentiate three types of intonation; to measure the degree of the change in duration for qualitatively similar vowels, which differ only in the degree of stress, to compare the duration of stressed and unstressed vowels, etc. (Potapova R.K., 1973; 1974; 1976).

Further investigation showed that we can use both data on the duration of vowels with primary sentence stress and the duration of final vowels in sentences for a reliable differentiation of the utterances in the analysed oppositions. The hypothesis, that the whole temporal contour of the vowels in a sentence is relevant, was not corroborated. The above - mentioned parts of the sentences proved to be exceptions.

Two normalizing coefficients were used for normalizing duration: T_x - the length of utterances and the mean duration of a sound T/n , where n - the number of sounds in an utterance. The duration of sounds in sentences similar in sound composition does not change with the type of intonation. It does not apply to the final position in the sentence and the vowel with sentence stress.

The main differences in the mean values of duration are observed in final unstressed vowels in an open syllable (CV) in the opposition "finished - unfinished statement". Here is an illustration of some mean relative values of vowels durations in the experimental data (Table 1).

Table 1

Normalized (relative) duration (t_i) of vowels in sentences pronounced with three types of intonation in Russian				
NN Sentence	Intonation type	$t_i; \bar{t}_i$ (on vowels)		
1. Zakaz wypolnen	finished statement	0,97 (0,12)	1,17 (0,17)	1,04 (0,16)
2. --" --	unfinished statement	1,09 (0,11)	1,21 (0,14)	1,24 (0,14)
3. --" --	general question	1,10 (0,09)	1,23 (0,18)	0,94 (0,12)
				0,88 (0,12)
				0,95 (0,17)
				1,00 (0,13)
				1,15 (0,14)

As we have already said, differences in the duration of stressed vowels, which are in the initial position in the sentence, are also of great importance for the differentiation of different types of intonation if we take into account the closed construction of the temporal characteristics of sentences (Table 2).

Table 2

Values of probability (p) of the divergence of average mean duration (t_1) of the stressed vowel in the first word of the sentence

NN	Sentence	t_1 (.)	t_1 (...)	t_1 (?)
1.	Ważny zakaz wypolnen dosročno.	1,47 (p=0,12)	1,36	1,38
2.	Zakaz wypolnen dosročno.	1,22	1,27	1,33
3.	Zakaz wypolnen.	1,17 (p=0,17)	1,21	1,32

The manifestation of segment units in the connected speech as well as changes of prosodic features are unthinkable without the time aspect. At the same time it is known that the temporal pattern of utterance depends on the functioning of a number of variables such as the total length of the utterance, its rhythmic pattern, the distribution of rhythmic patterns in the utterance, the distribution of stressed and unstressed syllables within each rhythmic structure, the distribution of the stress types (word stress, syntagmatic stress, sentence stress in the utterance, the position of the syllable bearing phrase stress, the structure of the syllables constituting the utterance (VC, CVC, CCV etc), the proper duration of syllabic vowels and so on.

This enumeration of the factors that can affect the temporal pattern of the utterance in the connected speech is far from exhaustive and it demonstrates clearly the complex and intricate nature of the temporal organization of speech.

CONCLUSIONS

The obtained results make it possible to draw the following conclusions:

- the most results important values for the segmentation of the intonation variant of every member of the opposition are not absolute, but relative values of all the analysed prosodic characteristics;
- the intensity level is the least dependant parameter on the type of intonation in Russian;
- duration is more dependant on it;
- the greatest constance in the direction of duration overfalls was observed in initial and final positions in sentences;
- the calculated rhythmic pattern constructed with reference to the duration changes in phrase initial and final position is a good approximation to temporal patterns which occur in speech
- individual duration values of vowels of different quality (close or open) do not as a rule exceed the confidence interval zone (in phrase initial and final position) and repeat the general pattern of duration changes as far as their direction is concerned.

All the variability of intonation contours of the Russian language can be expressed through combinations of prosodic features.

REFERENCES

- Potapova, R.K. (1973) Ob opyte sostavleninja alfavita vremennykh konstant russkoj reči, Materialy VIII akustičeskoj konferentsii, 54-58.
- Potapova, R.K. (1974) Some timing characteristics of spoken Russian (data rules analysis-synthesis-analysis), Speech communication seminar, Stockholm, Aug. 1-3, vol. 2, 193-196.
- Potapova, R.K. (1976) O vremennykh parametrah sinteza russkoj reči, 14. akusticka konferencia akustika reči a vnimanie zvuku, Bratislava, 177-180.
- Potapova, R.K. (1977) O kompleksnom karaktere vremenennykh modifikacij glaga v potoke reči, Russkij jazyk dlja studentov-inoctrancev, Moskva, Russkij jazyk, 212-221.
- Potapova, R.K. (1984) Les microvariations temporelles en limites de syllabe, le symposium sovietico-francais sur le dialogue acoustique de l' homme avec la machine, Moskva, 122-125.

THE RHYTHMIC ORGANIZATION OF SPEECH IN
CZECH AND RUSSIAN

V.V. Potapov

Department of Philology

Moscow State University

ABSTRACT - The subject of this research consists in describing perception as revealed in the process of segmentation of Czech and Russian speech into rhythmic structures (RS); in describing peculiarities of prosodic features of segmentation in Czech and Russian. In this investigation rhythm is defined as a regular recurrence of speech units in an utterance. These units comprise syllables, rhythmic structures (phonetic words), sense-groups (syntagmas) and phrases.

INTRODUCTION

Among diverse problems of oral speech, studied by comparative linguistic research the problem of rhythm is attracting more and more attention, i.e. not only the rhythmic structure on the whole, but also the rhythmic structure of units of rhythm in different languages and in different types of speech activity.

METHOD

Two native speakers of Czech took part in the experiment. After a preliminary preparation the text was read in a soundproof studio and recorded (the speed was 38 cm/sec). In three months' time the same subjects made the auditory analysis of the text. There were two programmes in the auditory analysis: the recording of the text was written in spelling and then segmented into sense-groups and RS, the subjects did this while listening to the recorded text; copies of the text, typed without capital letters and punctuation marks were to be segmented into RS without listening to the recording of the text. After a statistical analysis RS were classified into groups and types.

THE GENERAL DESCRIPTION OF ACOUSTIC-PHONETIC ANALYSIS

The conclusions after the 1st part of the analysis are as follows: in Czech speech there are six classes of RS; the number of the types of RS is limited, there are only 2: the first type with the first stressed syllable (this type dominates because the position of stress in Czech is fixed), the second type has stress on the second syllable, this type is rare; the analysis of the frequency of the types of RS has shown that the most frequent type in the analysed texts is RS which consists of 2 syllables, RS of 3 syllables come next and RS of 1 syllable rank third; RS of 2 and 3 syllables cover up the greater part of the text (85%), the rest (15%) is covered by RS of 4-, 5- and

6-syllabic RS; the data of the analysis of Czech speech has been compared with the analogous data of the analysis of Russian speech (Zlatoustova, 1981). This comparison has revealed some common features as well as some peculiarities in the segmentation of speech into RS typical of the Czech and Russian languages: 1) for both these languages RS of 2- and 3-syllables constitute the most frequent types; 2) but in Czech RS of 1-syllable is also one of the most frequent which is not characteristic of the Russian language; 3) the scope of variation of the types of RS is very limited in Czech but in Russian it is quite the opposite.

The second stage of the experiment consisted in the analysis of the prosodic features of stressed and unstressed syllables in different RS in Czech speech (Potapov, 1987). The results of the experiment showed that RS of different classes (ex. 2/1, 3/1, 4/1), which differ in their position in the phrase, are marked by the following prosodic features. RS of 2-syllables: 1) at the beginning of the sense-group the duration of stressed syllables is bigger than that of unstressed ones, but unstressed posttonic syllables are marked by a higher level of intensity; 2) in the middle of the sense-group stressed syllables have two patterns of the distribution of duration. Sometimes the duration of stressed syllable prevails, sometimes both the syllables are equal in their duration; 3) at the end of the sense-group stressed syllables have 2 different prosodic structures: stressed syllables may be marked by both increased duration and intensity, sometimes they are marked only by duration. RS of 3-syllables: 1) at the beginning of the sense-group all the syllables tend to be equal, the first syllable, as a rule, is marked by greater intensity; 2) in the middle of the sense-group prosodic features vary: intensity increase on the first stressed syllable, on all the posttonic syllables. As for duration, it is either distributed among all the 3 syllables equally, or the last syllable is marked by more duration. RS of 4-syllables: 1) at the beginning of the sense-group there is a tendency to increase the intensity of the first stressed syllable; 2) the duration of RS in the same position as indicated above has equal parameters; 3) syllabic sonorants are marked by a significant increase in duration and intensity; 4) there are no significant changes of F_0 at the juncture of the syllables and within RS. There are minor changes of F_0 parallel to a steady insignificant rise in pitch or fall in pitch within syllables in RS. Thus in Czech stressed and unstressed syllables are equal in their duration. Intensity and F_0 at the juncture of stressed and unstressed syllables changes insignificantly.

CONCLUSIONS

The results of the analysis of Czech speech have shown: 1) the most variable parameter is the duration of sonorants. At the end of a syllable after a vowel or a fricative the duration of a sonorant considerably exceeds the duration of the same sonorant placed at the beginning of a syllable before a vowel; 2) according to intensity one can differentiate between the syllabic and the non-syllabic sonorants. And the ratio of a syllabic