AUTOMATIC WORD STRESS DETECTOR

L.Zlatoustova N.Kozlenko M.Khitina L.Zakharov

Department of Philology, Moscow State University
Moscow, USSR, 119899

ABSTRACT

This paper presents the results of measurements of Russian word-stress parameters (using acoustic and statistic methods). There are also demonstrated some specific peculiarities of Russian word stress which are employed in the computer model of an automatic word stress detector.

INTRODUCTION

The analysis of the literature on the word stress reveals that the Russian stress is not distinguished in current speech by certain specific parameters. It rather aims at structuring or shaping of a phonetic word on the whole. The peculiar character of the Russian stress presents certain difficulties for automatic stress detection.

Among the acoustic correlates usually considered for stressed vowels characteristics are fundamental frequency, duration, intensity and spectrum. The absolute and relative values are of interest.

It is necessary to specify the rhythmic organization of phonetic words and frequency of occurence of phonetic words and their rhythmic structures (RS). A rhythmic structure characterizes a single word or a few words, autonomous or syntactic, forming a stressed group. RS type is designated by a fraction where the number of syllables in a phonetic word is a nominator and the ordinal number of the stressed syllable is a denominator. RS variety is designated by a succession of consonants and vowels in a RS which is shown in terms of C (consonants) and V (vowels).

The rhythmic pattern of a Russian text does not permit two or more succesive stressed syllables or a long succession of unstressed syllables. The average length of an interval between two successive stresses varies from I to 3 syllables, the most frequent being 2 syllables /I/.

In the initial and final RS there are usually no more than 2 prestressed or poststressed syllables. The RSs containing from I to 4 syllables are most characteristic for Russian syntagmas. The most frequent ones consist of 3 RS, average syntagma length being 2,8 RS.

The phonetic word of 2 or 3 syllables are predominant. The data obtained by prof. L.V.Zlatoustova show that the 6 most frequent RS types: I/I, 2/I, 2/2, 3/2, 3/3, 5/3 cover approximately 70% of any Russian text. The mentioned RS types and 3 more: 3/I, 4/2, 4/3 can cover about 90% of any Russian text (e.g. the text of a dialogue) /2/. The distribution of general RS types in different languages is demonstrated for fiction and newspaper texts /3/ (see Table I).

Table I. Frequency of occurence of RS types (%)

RS TYPE	LANGUAGE			
	RUSSIAN	BULGARIEN	ENGLISH	GERMAN
I/I	I3	I3,8	27	I7,6
2/I	I6,8	I6,5	I8,8	I6,9
2/2	2I,3	I0,6	2I,5	I6,7
3/I	6	7,8	I,4	7
3/2	I9,6	I6,6	5,5	23,3
3/3	7	2,3	3,8	3,2
4/I	I,8	I,8	0,7	-
4/2	4	I0,9	I,9	2
4/3	9,2	II,7	I,9	2,8
4/4	I,4	I,3	0,5	0,5
5/3	5	2	I,4	I

Stress in Russian is normally placed on one of the initial three syllables of a phonetic word. Preferably it is one of the central syllables of a word.

RS REALIZATION IN SPEECH

The specific character of RS realization depends on frequency of occurence of a word, its position in an utterance or syntagma. It also depends on a variety of conditions: prepared reading vs. spontaneous speech, an artistic reading by an actor vs. neutral reading by a layman, RS constituting a one-word utterance or being part of a context, the character of a text (variety and genre), normative vs. dialectical speech,etc. Extralinguistic factors are also to be taken into account.

To determine an RS type is necessary to learn the number of syllables in a RS, the stressed syllable and its position relative to unstressed syllables, consonant and vowel component markers typical of beginning and end of a RS.

To detect the stressed syllable we took a number of RSs with stressed syllables (and vowels) clearly dissimilar to the unstressed syllables of the same utterance. It should be mentioned that stressed and unstressed syllables may have similar values of such parameters as fundamental frequency, duration, intensity and spectrum. So it is preferable to choose the most typical and frequent samples as the test material for developing an automatic teaching system (ATS).

Thus our minimal text consisted of I to 4 utterances, an utterance consisted of 2 syntagmas and so on. A separate utterance also can form a text. In the utterance of 2 RSs the phonetic word of any of the 9 rhythmic types may occur as the first component. The second component may be one of the following types: I/I, 2/I, 2/2, 3/I, 3/2. 4/2. Such RS types as 3/3, 4/3, 5/3 can succeed every RS type exept 3/I, 4/2, 5/3. The third phonetic word in a three word utterance is chosen in a likely manner.

The most frequent words are preferable to be chosen to form utterances. The comparison of the structural types of word entries in word counts shows that independently on the material used (written vs. oral text) and the size of analysed selection (400 thousand vs. I million occurences). One can note a certain similarity of rhythmic types and varieties in both selections.

The important finding was that the most frequent structural models found in word counts are the most frequent in the texts. Thus, there is a limited number of basic structural types of words and phonetic words in Russian /4/.

One of the structurally important variables of a phonetic word is the relative duration of its vowels in strong and weak positions. It has been repeatedly mentioned that vowel length change accounts for phonetic word duration variance due to conditions such as separate vs. contextual occurence, initial vs. final position in an utterance, emotional vs. neutral content, whether or not a RS bears the phrase accent,eto.

The temporal structure of a phonetic word is essentially conditioned by the relationship of broad and narrow vowels in strong and weak positions. Finally, it is important whether the initial syllable of a word is covered and the final syllable is open.

RESULTS AND DISCUSSION

The comparison of durations of the stressed and the first prestressed broad vowels in the RS of VCV́CV(C) and CVCV́CV(C) varieties revealed that in the final position the first RS vowel is always shorter than the stressed vowel irrespective of its being open or covered.

In the RS of the CVCV́CV(C) variety with the covered initial syllable in the beginning of an utterance the first prestressed vowel is shorter than the stressed vowel. The first prestressed vowel that starts an utterance is always longer than the stressed vowel.

Initial position (one-syntagma utterance)

The unstressed vowel in the absolute beginning of a RS of the 3/2 type and VCV́CV(C) variety is in 89% instances longer than the stressed one. In case of a covered initial unstressed vowel (RS of the CVCV́CV(C) variety) is in 90% instances shorter than the stressed one.

Final position (one-syntagma utterance)

In the final position of a RS in an utterance both the initial vowel of an open prestressed syllable and the vowel of the first prestressed covered syllable (i.e. RSs of the VCV́CV(C) and CVCV́CV(C) varieties) are in 96% instances shorter than the stressed vowel.

Table 2. The relationship of durations of prestressed and stressed vowels connected to narrow vs. broad types of vowel sounds /2/

VOWELS	RS	COEFFICIENT
prestressed narrow stressed broad		0,535
	2-syllable	0,486
	3-syllable	0,59I
	4-syllable	0,454
prestressed narrow stressed narrow		0,867
	2-syllable	0,82
	3-syllable	0,926
	4-syllable	0,827
	5-syllable	I,043
prestressed broad stressed broad		0,796
	2-syllable	0,826
	3-syllable	0,722
	4-syllable	0,827
	5-syllable	0,88

The mean intensity of RS components have been analysed as function of RS position in an utterance, qualitative nature of the components in a RS, and of syllable types. One could justifiably expect that in final position of a RS in an utterance the intensity of all the components have been lower than in utterance initial RSs. This regularity is connected with the phrase intensity contour and has been repeatedly mentioned recently.

The mean intensities allow to estimate intensity changes in strong and weak elements of a RS.

The question of the absolute prominence of stressed vowels (stressed syllables) especially in the RS with a narrow stressed vowel (in a closed syllable) and a broad prestressed vowel is of special interest.

Our experiment shows that a tendency exists for the stressed vowel (independently on its position) to be more prominent in a RS in the absolute beginning of an utterance with similar vowels and consonants. If the first syllable is stressed it is most certain to be marked by increased intensity.

In the absolute beginning in a RS of the VCV́CV variety the usual distribution of mean intensity is like this: the poststressed vowel is the least intensive, the most intensive is the stressed one, and the final vowel, however short, is more intensive than the preceding one. A different distribution is revealed in the absolute ending of a RS: the most intensive is the stressed vowel, the least intensive - second poststressed. In some cases the intensities of the stressed and the poststressed vowels are equal (in the initial position of a ES. However, the stressed vowel may be as intensive as the following unstressed if the former is front, high and the latter – low.

The utterance final stressed vowel in ES of [illegible] or [illegible] varieties is not, as a rule, marked by increased intensity, the stressed vowel, however, is more intensive if a ES is under phrase intensity contour termination /5/.

In the beginning of an utterance there is a marked tendency for the initial syllables to be stronger. This is probably due to their greater importance as information carriers.

The syllable intensity may alter the observed relationships if compared syllables have different sound structures. Thus the analysis of mean intensity suggests the tendency for the stressed vowels in the utterance initial and the utterance final ESs (the latter with the first syllable stressed) to be marked by an increase of energy. Though the case may be that the intensities of stressed and unstressed vowels would not differ due to some sound dissimilarities.

Table 3. Mean intensity /dB/

POSITION IN AN UTTERANCE

V́ beg.	V́ end.	V – I poststressed beg.	V – I poststressed end.	V – 2 poststressed beg.	V – 2 poststressed end.
43,6	34,0	33,8	24,8	35,0	I9,0

V – I prestressed beg.	V – I prestressed end.	V́ beg.	V́ end.	V – I poststressed beg.	V – I poststressed end.
44,0	36,0	45,0	34,8	4I,7	2I,3

V – 2 prestressed beg.	V – 2 prestressed end.	V – I prestressed beg.	V – I prestressed end.	V́ beg.	V́ end.
4I,8	35,8	42,7	35,4	43,28	30,7

Thus the mean intensity estimates of the vowels allow to determine the position of a ES in an utterance, characterize its sound structure but cannot be used as a reliable stress detector in Russian.

THE STRESS DETECTION ALGORHYTHM

In accordance with the above-mentioned we propose to use in stress detection algorhythms the most important acoustic parameters directly related to the phonetic word structure.

To determine the ES type in any position only one variable (duration) or two (duration, intensity) are sufficient. This is true for a one-word utterance with the first stressed syllable (ES types: I/I, 2/I, 3/I – with the covered initial and closed final syllables).

We suggest the stress detection algorhythm based on the method of loudness measurement. The objective methods of subjective loudness /Ψ/ measurements are available now as the international standard N 532)

In addition to such variables as intensity /I/, duration /t/, energy /E/, fundamental frequency /Fo/, etc. loudness allows to take into account the spectrum shape /S/ and the sound field shape /P/. Thus the loudness is a functional of a series of variables $\Psi = \psi(I, t, Fo, S, P)$ and as a tool of measurement is more adequate than most of physical (acoustic) parameters. As far as energy /2/ is a function of intensity and duration, the loudness is a complex function of energy. In the present study both I and E were used and a comparison has been made.

The test material has been tape-recorded and re-recorded by means of an analog-to-digital converter on a digital tape which has been computer analysed by the Automatic System of Scientific Research (ASSR). The system is able to measure loudness and loudness level (in accordance to international standard N 532) and possess the necessary service software /6/, /7/.

The result of such an analysis are the automatic graphs: oscillograms with the I/20 msec time marks, multiparametric graphs reflecting intensity, loudness level and fundamental frequency as function of time.

The algorhythm has been tested on a limited material where the compared vowels were prosodically different. It demonstrated that loudness integral maxima corresponded to stress vowels in all cases. After parameter optimization the algorhythm has been tested on a more varied material with a new variable of the word position in an utterance. A new phrase test was compiled to compare vowels:

I. Stressed and strong unstressed vowels (non covered in the first prestressed syllable and open in the first poststressed syllable).

2. Vowels of different proper duration and intensity.

3. Vowels in different phonetic environments.

After the linguistic correction the reliability of the algorhythm increased to 92,3%.

CONCLUSION

The attained reliability justifies the use of loudness along with other parameters in working out algorhythms of automatic word stress detection in current speech. It's use as a part of the ATS is recommended.

REFERENCES

/I/ Kagarov E.G. O ritme russkoy prozaicheskoy rechy. Nauka na Ukraine, I922, N 4, s.324-332.

/2/ Zlatoustova L.V. Foneticheskie edinitsy russkoy rechy. M.G.U., I98I. – Io6 s.

/3/ Zlatoustova L.V. Universalii v prosodicheskoy organizatsii teksta (na materiale slavianskikh, germanskikh i romanskikh jazykov). Vestnik MGU, ser. 9, Philologia, I983, N 4, s. 69-78.

/4/ Khitina M.V. Edinitsy ritma russkoy rechy i ikh ispolzovanije v razlichnykh tekstakh: Avtoreph. diss. ... kand. philol. nauk. M., I986. – I9 s.

/5/ Zlatoustova L.V. Phoneticheskaja priroda russkogo slovesnogo udarenija: Avtoreph. diss. ... kand. philol. nauk. L., I953. – 23 s.

/6/ Kozlenko N.P., Rezviakina Z.N. Izmeritel' urovn'a sluchainykh protsessov. B'ulleten' isobretenii i otkrytii i tovarnykh znakov. BIOTZ,N I9, I968.

/7/ Kozlenko N.P. Avtomaticheskoje raspoznavanije slukhovykh obrazov na osnove modeli integralnogo vosprijatija kachestva zvuka. Materialy Vsesojuznogo seminara "Avtomaticheskoje raspoznavanije slukhovykh obrazov" (ARSO-I3). Novosibirsk, I984, s. 75-76.