SONORANT SEGMENT QUALITY IN RUSSIAN EMOTIONAL SPEECH

Veronika Makarova [1)] *and Valery A. Petrushin* [2)]

[1)] Department of Languages and Linguistics
University of Saskatchewan, Saskatoon, Canada
v.makarova@usask.ca

[2)] The Nielsen Company
Schaumburg, Illinois, USA
valery.petrushin@nielsen.com

ABSTRACT

The paper reports characteristics of sonorant segments (vowels and sonorant consonants) in Russian emotional speech. The authors describe the effects of segmental duration, energy, formants and dynamic ranges on the expression of emotion in Russian. The data come from RUSLANA, a database containing samples of neutral utterances and utterances with simulated emotions of surprise, happiness, anger, sadness and fear.

Keywords: sonorants, emotional speech, Russian.

1. INTRODUCTION

Emotion expression is one of the most fundamental characteristics of human communication [1]. While other biological species can express their emotions via color change, electrical impulses, chemicals, pheromones, touch, postures, facial expressions, gesticulation, and vocal signals, humans employ the last five and have also developed a complex system of emotion signaling in speech, which includes discoursal patterns, syntax, lexis and sound characteristics [1, 2, 3].

Many studies of the acoustic correlates of emotion in speech have focused on prosodic features [4, 5, 6, 7]. Segmental characteristics have not been given enough attention, yet some of them, e.g., formant frequencies, segmental spectra and durations are also contributing to the expression of emotions [8, 9, 10]. Studies of the segmental clues of emotions are important for the improvement of speech synthesis and speech recognition quality [6].

This paper is devoted to segmental clues of the expression of affect. Sonorant segments (stressed and unstressed vowels and sonorant consonants) were selected on the bases of commonality of their acoustic characteristics (voicing, similarities in power spectra and the existence of formants). No earlier studies have addressed the emotive characteristics of Russian sonorant consonants, and very little information is available on the segmental clues of affect in Russian [11, 12].

2. MATERIALS AND METHODS

2.1. The database

The materials for the study were retrieved from RUSLANA (RUSsian LANguage Affective speech) database [11, 12], which includes recordings of 61 speakers of standard Russian (12 male and 49 female) simulating six emotional states: neutral, anger, fear, happiness, sadness, and surprise. The subjects simulated each of the above states while reading 10 sentences representative of different syntactical types and intonational patterns. The sentences include a few types of declarative, interrogative and exclamatory sentences which contain all Russian phonemes (See [11] for the full description of the database and examples of utterances). The subjects were university students untrained in the expression of emotions. The utterances produced by the speakers were subjected to a two-step evaluation process whereby 20 listeners identified expressed emotion (stage 1) and evaluated how well a particular emotion is expressed by each utterance (stage 2). Utterances with high evaluation scores were used for the subsequent analysis, the rest were discarded. This study draws on a sample of 600 utterances produced by 5 male and 5 female subjects who ranked the best in their ability to express emotions.

2.2. Feature extraction and analysis

Acoustic features extracted from the sample are phonemic duration, average energy, average F0, average F0 derivative, average F1, F2, F3, average formant bandwidths, average power spectra (Fq) on logarithmic scale from 0 to 16000 Hz split into 16 sub-bands (see [11] for details). For the analysis of emotion expression in sonorant segments, we extracted the features for 1393 occurrences of stressed and 3891 occurrences of unstressed vowels [a, i, u, e, o, ɨ] and 3472 occurrences of sonorant consonants [m, m^j, n, n^j, l, l^j, r, r^j, j].

The extracted features were subjected to univariate ANOVA analysis to determine the effect of emotion type on each parameter variance for the groups of stressed vowels, unstressed vowels and sonorant consonants. The effects were considered significant at p values less than 0.05. Next, in cases when the effect was significant, post-hoc Tukey's tests were conducted for each pair of emotions.

The data are presented pooled together for female and male subjects due to space limitations.

3. RESULTS

The descriptive statistics and p-values for parameters that display statistically significant differences in the three groups of segments (stressed vowels, unstressed vowels and sonorants) are represented below in Tables 1-3. Dynamic ranges are represented graphically in Figures 1-3.

3.1. Duration

Duration of sonorant segments varies significantly across emotions and across the three groups of segments. Neutral segments are the shortest ones in all the three groups. In both vowel groups, all emotive states are connected with the increase of emotion as compared to neutral state, but there are no significant distinctions between any pairs of emotive states as regards their duration. By contrast, in the group of sonorants, there is a statistically significant increase in duration for the 'sad' emotive state as compared to other emotions ('angry', 'surprised', 'happy', 'neutral').

3.2. Energy

Energy is highly significant for the expression of emotion. Both in vowels and in sonorant consonants, the lowest energy is found in segments with 'neutral' and 'sad' emotions, medium energies are associated with 'afraid' and 'surprised' states, and the highest energies – with 'happy' and 'angry' ones. Post-hoc results reveal significant differences between 'neutral and 'sad' vs. 'happy and angry', 'afraid' vs. 'happy and angry', 'surprise' vs 'angry', 'happy'. No statistical difference in energy values is found across 'neutral' and 'sad', 'afraid' and 'surprised', 'happy' and 'angry' pairs of emotive states.

3.3. F0 and its derivative

Unstressed vowels have the highest values and the narrowest range of F0 values, whereas stressed vowels have lower values and the widest range of F0. In all segment groups, the lowest F0 values are associated with the 'neutral' state followed by the 'sad' emotion, medium values of F0 are found in segments with the 'angry, afraid, surprised' emotions, and the highest F0 values are characteristic of 'happy' emotive state.

F0 derivative is significant for stressed vowels, but not for unstressed vowels and sonorants.

3.4. Formants

F1 values are significantly affected by emotion in case of stressed vowels and sonorants, but not in the group of unstressed vowels. 'Neutral' and 'afraid' segments have lower F1 values, and 'happy' and 'angry' – higher ones. Higher F1 values correspond to more open articulations. F2 values are low in the 'neutral', 'sad', and 'afraid' segments and high in 'angry' and 'happy' segments presumable reflecting more front segment articulations. F3 values are not significant for the expression of emotions in the stressed vowels, but are significant for sonorants and unstressed vowels, whereby 'neutral' segments have lower F3 values as compared to other states.

3.5. Spectral profiles

Spectral profiles of vowels and sonorants are represented in Figures 1-3. All the frequency bands of the spectrum demonstrate statistically significant differences across emotion types. The frequency band of maximum salience is between 1500 and 10000Hz.

4. Discussion

Emotions present a considerable difficulty for analysis, identification and classification [3]. In earlier studies vowel *duration* has been found to contribute to emotion differentiation [12]. In our study, neutral utterances have the shortest duration,

Table 1: Descriptive statistics for unstressed vowels (mean/std) and *p*-value for one-way analysis of variance.

Feature	Neutral	Sad	Afraid	Angry	Happy	Surprised	p-value
Duration,ms	63.5/31.9	70.8/33.8	72.0/34.7	76.3/38.4	68.9/33.4	68.1/33.3	**<0.000001**
Energy, rms	0.027/0.031	0.032/0.023	0.039/0.033	0.075/0.056	0.063/0.043	0.044/0.031	**<0.000001**
F0, Hz	194.3/97.2	200.1/93.7	225.5/ 92.8	224.2/78.3	247.0/89.8	225.1/87.3	**<0.000001**
F0 der, Hz/s	-144.3/344	-130.1/305	-159.3/369	-190.7/448	-167.5/515	-161.2/526.5	Not signif
F1, Hz	393.4/150.2	395.1/146.8	378.6/140.8	458.4/123.3	445.8/125.9	425.4/115.4	Not signif
F2, Hz	888.0/387	899.2/369	959.6/378	1040.6/311	1014.3/312	961.8/295	**<0.000001**
F3, Hz	1643.6/611.8	1676.8/589.5	1734/574	1788.4/425.9	1802.1/421	1774.1/422	**<0.000001**

Table 2: Descriptive statistics for stressed vowels (mean/std) and *p*-value for one-way analysis of variance.

Feature	Neutral	Sad	Afraid	Angry	Happy	Surprised	p-value
Duration, ms	74.6/24.0	86.8/28.0	85.4/29.6	88.1/27.7	89.6/32.2	84.3/26.9	**<0.000001**
Energy, rms	0.028/0.016	0.032/0.022	0.050/0.040	0.069/0.034	0.077/0.050	0.051/0.026	**<0.000001**
F0, Hz	146.8/50.6	157.8/52.6	210.1/62.2	196.5/65.4	218.0/57.6	188.9/46.3	**<0.000001**
F0 der, Hz/s	-42.8/209.5	-17.7/178.2	-74.3/270.5	-31.0/328.2	-6.4/392.1	32.7/352.8	**0.00414**
F1, Hz	401.3/89.6	405.3/85.0	402.2/86.6	428.3/106.7	438.8/92.8	416.7/83.5	**<0.000001**
F2, Hz	1050.9/337	1048.5/297	1050.5/257	1107.1/297	1118.2/285	1066.8/286	**0.02574**
F3, Hz	1825.5/440	1900.2/462	1785.95/481	1886.1/413	1940.5/371	1829.7/392	Not signif

Table 3: Descriptive statistics for sonorants (mean/std) and *p*-value for one-way analysis of variance.

Feature	Neutral	Sad	Afraid	Angry	Happy	Surprised	p-value
Duration, ms	50.89/0.90	56.1/0.90	53.34/0.90	52.21/0.89	51.23/0.90	51.02/0.91	**0.0001**
Energy, rms	0.018/0.002	0.022/0.002	0.028/0.002	0.048/0.002	0.044/0.002	0.035/0.002	**<0.000001**
F0, Hz	185.7/4.9	205.2/4.9	224.4/4.9	214.3/4.9	235.4/4.9	218.8/5.0	**<0.000001**
F0 der, Hz/s	-53.27/17.6	-21.94/17.7	-2.08/17.8	0.38/17.8	-25.8/17.7	-44.53/17.9	Not signif
F1, Hz	324.17/5.8	352.56/5.83	332.51/5.8	377.77/5.79	370.6/5.8	358.59/5.8	**<0.000001**
F2, Hz	866.9/17.02	923.16/16.9	923.9/17.03	1002.24/16.8	988.05/16.9	962.32/17.1	**<0.000001**
F3, Hz	1533.2/27.8	1635.0/27.7	1603.5/27.8	1714.8/27.5	1681.6/27.6	1686.1/28.0	**<0.0001**

and emotion, in particular sadness, is associated with increased duration. *F0 values* are of high significance for the expression of emotion [4, 15], whereas *F0 derivatives* in our study are only significant for stressed vowels, which can be explained by pitch movements occurring mostly within stressed vowels. F0 derivative in stressed vowels has higher values in ‘happy’ and ‘surprised’ states, since these emotions typically have rising pitch movements. The significant role of energy in different frequency subbands of the *power spectrum* points to the importance of the voice source signal and phonatory quality in the expression of emotions. Many studies suggest the link between the expression of effect and voice quality [16]. *Formants* F1 and F2 are directly linked to the position of articulators during vowel production. Our study shows that F1 and F2 values undergo changes in affect, which can be explained by the connection between some emotions and hyperarticulation [17], and that formant values are salient for emotion expression not only in vowels, but in sonorants as well. On the other hand, F1 values are insignificant for the expression of emotion in unstressed vowels presumably because of vowel centering caused by vowel reduction in Russian.

It will be of interest to investigate possible correlations between parameters under investigation, such as energy, F0, duration and formant values. The differences in parameter values between male and female subjects are also worth pursuing.

5. Conclusion

Our study contributes to the analysis of emotive speech by introducing segmental characteristics of sonorants in comparison with vowels. Adequate consideration of prosodic and segmental characteristics of affect will help to improve quality of speech synthesis and recognition systems for Russian. Our future work will utilize the findings in emotional speech recognizer for Russian as well as in psycho-acoustic experiments investigating the listeners’ perception of re-synthesized utterances with superimposed segmental and suprasegmental features of affect.

Figure 1: Spectral profiles for stressed vowels and the following emotions: anger – solid line and upward-pointing triangle markers, sadness – dashed line and cross markers, fear – dotted line and circle markers, happiness – dashed line and downward-pointing triangle markers, neutral – solid line and square markers, and surprise – dashed-dotted line and diamond markers.

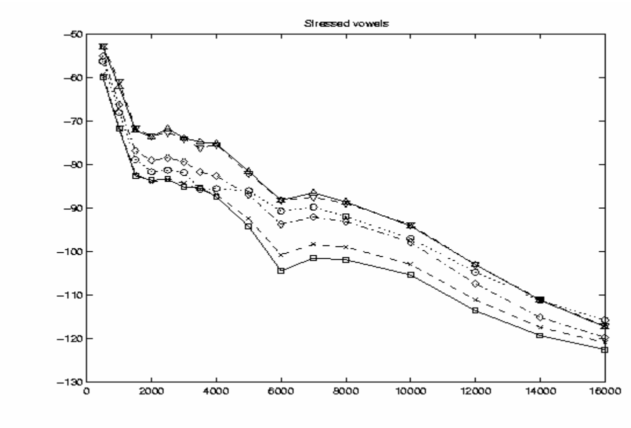


Figure 2: Spectral profiles for unstressed vowels (see Figure 1 for detailed legend).

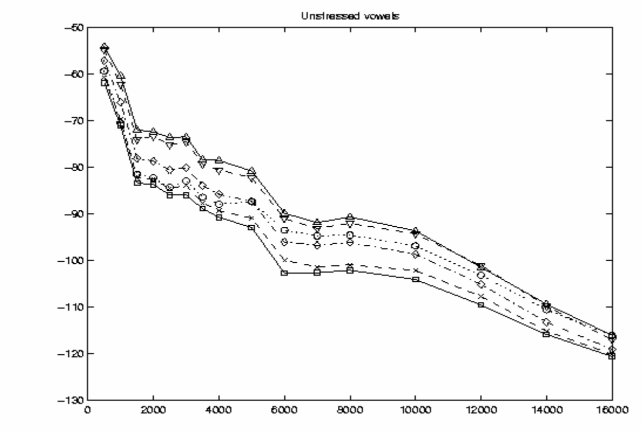
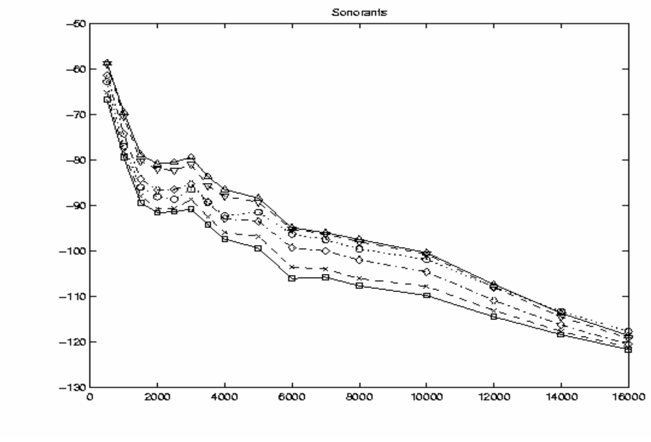


Figure 3: Spectral profiles for sonorant consonants (see Figure 1 for detailed legend).



6. REFERENCES

[1] Hauser, M. *The evolution of communication.* Cambridge, MA: MIT Press, 1996.

[2] Bright, M. 1984. *Animal language.* London: BBC.

[3] Plutchik, R. *Emotion: A psychoevolutionary synthesis*. New York: Harper and Row, 1980.

[4] Banse, R. and Scherer, K. 1996. Acoustic profiles in vocal emotion expression. *Journal Personality Social Psychology*, Vol 70, N3, pp. 614-636.

[5] Arnfield, S., Roach, P., Setter, J., Greasley, P., and Horton, D. 1995. Emotional stress and speech tempo variation*. Proc. of ESCA-NATO Workshop on Speech under Stress*, Lisbon, pp. 13-15.

[6] Cowie, R., Douglas-Cowie, E. Tsapatsoulis, G., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. Emotion recognition in human-computer interaction, *IEEE Signal Processing Magazine*, vol 18, N 1, 2001, pp. 32-80.

[7] Min Lee, C. and Narayanan, S.S. 2005. Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing*, vol 13, N 2, pp. 293-303.

[8] Fernandez, R. *A computational model for the automatic recognition of affect in speech*. PhD thesis, MIT, 2004.

[9] Kienast, M. & Sendlmeier, W. F. 2000 Acoustical analysis of spectral and temporal changes in emotional speech*. Proc. ISCA Workshop on Speech and Emotion,* Belfast, 5-7 September, pp. 92-97.

[10] Tickle, A. 2000. English and Japanese speakers' emotion vocalisation and recognition: A comparison highlighting vowel quality. *Proc. ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, Belfast, 5-7 September, pp. 104-109.

[11] Makarova, V., Petrushin, V. RUSLANA: A database of Russian emotional utterances. *ICSLP 2002*. Denver, USA, Sept., 2002.

[12] Petrushin, V. A. and Makarova, V. Parameters of Fricatives and Affricates in Russian Emotional Speech, *Proc. SPECOM 2006*, Saint-Petersburg, Russia, 2006, pp.73-80.

[13] Montero, J.M.., Gutierez-Arriola, J., Palazuelos, S., Enriques, E., Aguilera, S., Pardo, J.M. Emotional Speech Synthesis: From speech database to TTS. *ICSLP 98.*

[14] Arnfield, S., Roach, P., Setter, J., Greasley, P., and Horton, D. Emotional stress and speech tempo variation. *Proc. of ESCA-NATO Tutorial and Research Workshop on Speech under Stress*, Lisbon, 1995, pp. 13-15.

[15] Paeschke, A. and Sendlmeier, W. F. Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements, *Proc. ISCA Workshop on Speech and Emotion*, Belfast, 5-7 September, 2000, pp 75-80.

[16] Ní Chisaide, A. N. & Gobl, C. Voice source variation. In: Hardcastle, W. J & Laver, J. *The Handbook of Phonetic Sciences.* Oxford: Blackwell

[17] Fischer, K., Batliner, A. What makes speakers angry in human-computer conversation? *Proc. of the 3rd Workshop on Human-computer conversation*, Belleegio, Italy, 3-5 July, 2000.