Phonetics of Emotion in Russian Speech

Veronika Makarova† **and Valery A. Petrushin**‡

† Meikai University, AIST, Japan

‡ Accenture Technology Labs, Accenture, Chicago, USA

E-mail: petr@cstar.accenture.com, makarova.veronika@aist.go.jp

ABSTRACT

This paper provides a description of the structure and goals of the database of Russian Language Affective (emotional) utterances. It also reports some preliminary results of an experimental phonetic analysis of the acoustic characteristics of emotional utterances (surprise, happiness, anger, sadness and fear) vs. neutral ones in Russian. The study utilizes 600 database utterances by 10 speakers. Special focus of the study is placed on interactions between prosody and syntactic patterns of utterances.

1. INTRODUCTION

Human speech processing (by either humans or machines) requires the ability to interpret multiple layers of meaning, such as word meaning and sentence structure, discoursal type of utterances, background information, cultural knowledge, etc. An important part of human message is related to rendering emotional and affective states.

It is known that emotional states are acoustically expressed via a large number of complex parameters related to segmental quality and prosody, some of which may be recurrent across languages (or even universal), and some other may be language-specific [1]. A question still open to wide debate is to what an extent and in what ways the expression of emotion can be physiologically determined, and hence universally expressed and perceived, or embedded into the system of a given language, and therefore be language-specific.

Information about acoustic ways of expressing emotions and affect is required for human-computer interaction, robotics, multilingual communication and machine translation systems, entertainment industry, alert systems, foreign language teaching/tutoring, etc. For example, it is essential for robotic pets, such as Sony's Aibo, and mechanical companions, such as Mitsubishi's robot nurse and house-sitter, to process human emotion and be able to respond to them.

In order to serve these multiple needs, it is necessary to gather and analyze data related to the expression of emotion in natural speech. This can be done, in particular, via constructing databases of emotional and attitudinally-coloured speech. Although several such databases have been created, e.g. [2], none so far have been available for Russian, which is one of the world's major languages. Moreover, only very few studies on the Russian intonation in general and on emotional intonation in particular are available [3,4].

To facilitate studies of the expression of emotion in Russian, we created a database briefly described below (see [5] for further details).

2. DATABASE

The objectives of database

The database serves for multiple linguistic, applied linguistic and engineering research purposes, and in this way encourages cross-disciplinary studies that are focused on the expression of emotions in speech. It was designed, firstly, to serve as a source for developing and training a system of emotions recognition in Russian. It also provides data for designing a new system of Russian intonation description, testing the Integrated Prosodic Notation [6], which combines the features of autosegmental and contour approaches, investigating intonation/syntax/emotion interactions and teaching Russian as a foreign language.

Structure

The database consists of 10 sentences with different syntactic, structural and discoursal types, which were read by native speakers of Russian portraying the following six affective-emotional states: neutral/unemotional, surprise, happiness, anger, sadness and fear. The sentences allowed to realize five major Russian intonation contours (IC): IC 1, IC 2, IC 3, IC 4 and IC 5 [7].

Speakers

The database includes records of utterances from 61 (12 male and 49 female) subjects, aged between 16 and 28, who are native speakers of standard Russian residing within the country.

Method

All the data were recorded on a portable Digital Audio Tape-recorder Sony TCD-D8 at 48 kHz sampling rate via Sennheiser headphone set in a sound proof recording studio of the Department of Phonetics, at St. Petersburg State University, St. Petersburg, Russia. The obtained recordings were converted into monophonic Windows PCM format at 32 kHz sampling frequency and 16 bits resolution.

Prosodic features

Along with creating the digitized files for every obtained utterance, we also extract and store a number of phonetic

ISBN 1-876346-48-5 © 2003 UAB

and prosodic features, such as F0, F0 slope, F0 derivative, energy, formants F1 - F3 and their bandwidths (BW1-BW3). For each of these parameters some statistics have been calculated, such as mean, standard deviation (SD), median, maximum, minimum, and range. We also calculated utterance duration, percentage of pauses, percentage of voiced speech, and speaking rate.

3. PRELIMINARY RESULTS OF THE STUDY OF EMOTION EXPRESSION IN RUSSIAN SPEECH

3.1. Material analyzed

Segmented utterances of 5 female and 5 male subjects (a part of RUSLANA database) have been analyzed in the study reported in this section.

3.2. Some descriptive statistics

Proximity of prosodic features

The examined pitch contours of the phrases indicate that the overall characteristics of pitch contours in angry and happy utterances are relatively close. The expressions of fear and sadness also have (although fewer) similarities in the prosodic structure. Neutral and surprised utterances stand out by some specific characteristics. These similarities relate to the placement and type of accents, overall direction of pitch contour as well as features of intensity.

Total duration and relative speaking rate

Below we present the box plots of relative values of emotional utterance comparing to neutral. Here “relative” means that all the values are deviations in percent from the same feature of the neutral utterance of the same sentence and the same speaker.

Relative duration of utterances does not show very significant differences across emotive types. Fig. 1 shows the relative speaking rates. We can see that speaking rate for anger and happiness is an average 20% slower than for the other emotions. The variance for fear is rather high it means that there are different ways to express fear.

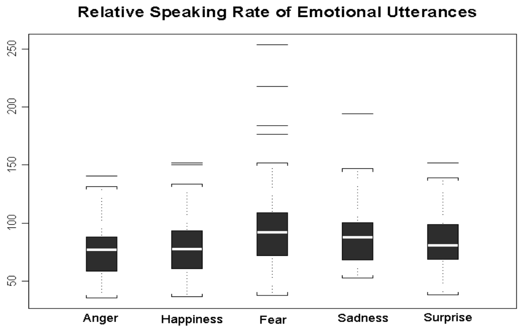

Figure 1. Relative speaking rate.

Voicing

Fear and sadness have the same amount of voicing as neutral, but angry and happy utterances have more voicing by 10 - 15% on average (Fig. 2).

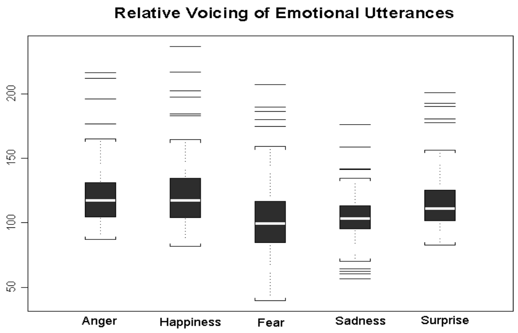

Figure 2. Relative voicing.

Intensity

Relative mean of intensity is significantly higher for angry and happy utterances. Intensity of sad and fearful utterances is close to neutral ones (Fig. 3).

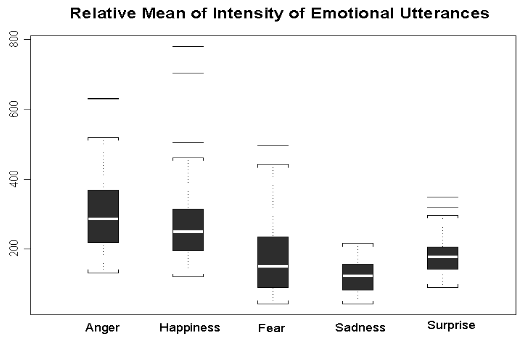

Figure 3. Relative mean of intensity.

F0 mean, maximum and standard deviation (SD)

Relative mean of fundamental frequency (F0 mean) is higher in happy, angry, fearful and surprised utterances than in neutral and sad (Fig. 4).

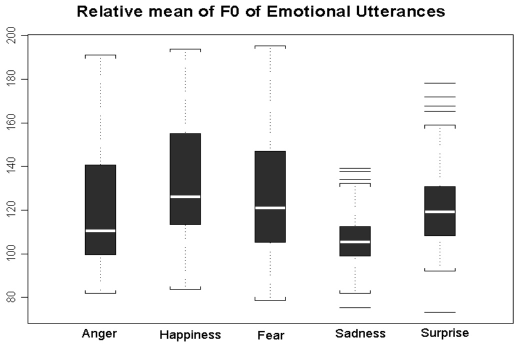

Figure 4. Relative mean of F0.

F0 maxima group in a somewhat different way from the F0 mean: while happy utterances have the highest and sad utterances have the lowest F0 peaks (as their F0 means), surprised, angry and fearful utterances have similar medium values for F0 peaks. Sadness is the most close to the neutral state and happiness is the most distant (Fig. 5).

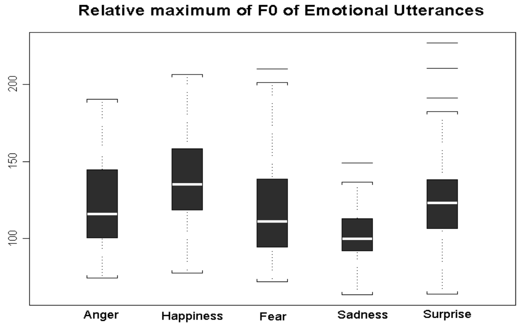

Figure 5. Relative maximum of F0.

Standard deviation values show that happy utterances have maximal change of F0 values (i.e. more pitch changes), followed by angry and surprised utterances. The least change in F0 is observed in sad utterances (Fig 6).

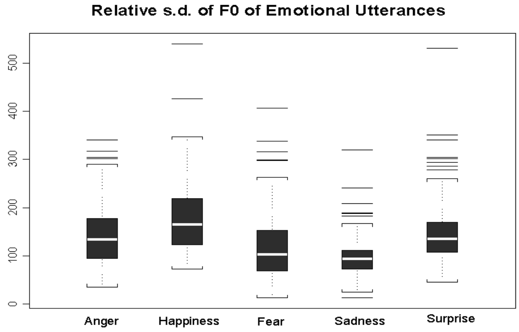

Figure 6. Relative F0 standard deviation.

3.3. Intonation contours by sentence type and emotion

For lack of space we shall only describe here the contours in statements and yes/no questions.

Statements

Neutral statements are characterized by a sentence-initial rising pitch accent (about 100 Hz on the average), overall falling (declining) direction of the pitch contour, and a very insignificant fall in pitch in the final word bearing the terminal accent (about 15 Hz on average). The terminal fall is realized mostly not within the accented vowel of the final lexically important item, but as a difference in pitch between the pre-accented syllable offset and accented syllable onset. The intensity peaks at the sentence-initial accent and then gradually declines towards the end of the utterance. These contours can be described as IC 1.

Non-finality in statements containing more than one Tone Unit is rendered by prominent rising accent followed by a fall in the post-accented syllables.

Angry and **happy** statements have a decrease in magnitude of the sentence-initial accent, and a significant increase in the magnitude of pitch movement in the final falling accent (100Hz). They are closer in shape to IC 2.

The non-final rise at the TU boundary can be replaced by a fall. An increase in the number of accents occurs for some speakers.

Surprised statements have either a small sentence-initial accent, or no sentence-initial accent, they generally have a flat contour slightly declining towards a prominent sentence-final rising accent (150 to 230 Hz in magnitude). One speaker has a succession of rising accents on every lexically important word prior to the sentence-final accent. In general, the contours in surprised statements strongly resemble the ones found in yes/no questions (IC 3).

Fearful and **sad** statements do not display a completely uniform pattern, but rather differ by the speaker. They can have either a flattened contour, or a succession of accents on all lexically important words. The pre-terminal accents are mostly falls (except for female speaker F043). The terminal accent is a fall of medium (for fear) and small (for sadness) magnitude: 50-70 Hz and 10-30 Hz respectively.

Yes/no questions

Neutral yes/no questions have a flat contour mostly without any pre-terminal accents. The contour gradually rises towards the final rising accent in the last lexically important item (IC 3).

Angry, happy and **surprised** questions have the same contour as in neutral questions, but may occasionally have an extra pre-terminal accent. The major difference as compared to the neutral questions, is the increase in the magnitude of terminal pitch rise from 50-80 Hz to 100-200Hz.

In **fearful** and **sad** question a change in the terminal accent occurs: they are falling or rise-falling (medium magnitude of about 20-80 Hz), i.e. there is a change in contour type.

Sentence type and accentual structure constraints

In general, sentence type and accentual structure of an utterance appears to place some constraints on the prosodic expression of emotions. We find most emotion-related changes in statements, and the least in questions and exclamations (because of restrictions these types place on the pitch in the pre-terminal part of the contour). Utterances having the major sentence or logical accent close to the beginning display fewer changes of pitch parameters across emotions again due to limitations in pitch movements in post-accentual parts of the contour.

4. DISCUSSION: EXPRESSION OF EMOTION IN SPEECH

Descriptive frameworks for emotional states

No fixed frameworks for the analysis of emotions exist, and various competing approaches to classifying emotional and affective states have been proposed within the linguistic, psychological, biological, and speech engineering traditions [8]. Our database includes 5 emotional states (contrasted with neutral (unemotional) utterances) which are listed in most descriptions of emotional speech [8], however, we consider it possible to identify more emotional-affective states, attitudes and moods in Russian (or any other language) speech.

Emotional categories

It has been suggested that emotional states can be divided into categories [8]. Speech processing research indicates that some acoustic parameters can be similar across certain emotional states, [9], whereas speech perception studies show that some emotions also get misconstrued for one another by human listeners [10]. Our study demonstrates the proximity of the acoustic characteristics of anger and happiness -- higher intensity, slower speaking rate, higher voicing, greater magnitude of pitch movements, larger number of accents. Anger and happiness also share some features in common with surprise, such as higher F0 maxima, higher F0 SDs. Fear and sadness have respectively lower values for the above features and also display flattened contours with smaller magnitudes of pitch movements and fewer accents. These results suggest that a new classification of emotions is possible based on the proximity/distance of their acoustic expression and on their correct/incorrect identification by human listeners.

Universal and language-specific in the expression of emotions in speech

Russian language data in our study agree with the results for many other languages which display that happiness is associated with increased intensity and higher pitch, whereas sadness has low values for these parameters [8].

On the other hand, we found some specific features related mostly to interactions between the emotional state and sentence types as well as structure and type of intonation patterns.

Emotion/syntax/intonation interactions

It appears that certain emotional states are more closely linked with certain sentence types and patterns. E.g., the expression of surprise is connected with interrogative pitch contour. On the other hand, adding emotions of fear and sadness to the yes/no question changes the terminal accent and contour type from rising to falling, i.e. resembling the ones found in declaratives.

5. CONCLUSION

We have reported some preliminary results of the study investigating the expression of anger, happiness, surprise, fear, sadness in contrast to neutral/unemotional utterances in Russian speech drawing on the data from RUSLANA database. We have reported some descriptive statistics data related to the acoustic cues of the six emotive-affective states and provided some details of intonation contour variation across emotions and sentence.

REFERENCES

[1] Kappas, A., Hess, U., Scherer, K. R. 1991. Voice and emotion. In: *Fundamentals of Nonverbal Behaviour*, R.s. Feldman & B. Rime (Eds). Cambridge: CUP, 200-238.

[2] Campbell, N. Towards a grammar of spoken language: Incorporating paralinguistic information. *ICSLP 2002*, 673-676.

[3] Holden, K. T. & Hogan, J. T. 1993. The emotive impact of foreign intonation: An experiemnt in switching English and Russian intionation, *Language & Speech,* 36 (1), 67-88.

[4] Makarova, V. 2000. Acoustic cues of surprise in Russian questions. *J.Acoust.Soc.Jpn (E),* 21,5, 243-250.

[5] Makarova, V., Petrushin V. RUSLANA: A database of Russian emotional utterances. *ICSLP 2002*, 2041-2044.

[6] Makarova, V. 2002. Prosodic feature perception by human subjects. *9th AICSST*, Melbourne, Austr, 2-5 Dec. 2002.

[7] Bryzgunova, E. A. 1977. *Zvuki i intonatsiya russkoy rechi.* Moscow: Russkij Yazyk.

[8] Cowie, R. Douglas-Cowie, E., Tsapatsoulis, N., Votsis, C., Kollias, S., Fellenz, W., Taylor, J.C. 2001. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, Jan 2001, 33-79.

[9] Murray, I. & Arnott, J. Synthesizing emotions in speech: Is it time to get excited? *Proc. 4th Int Conf. Spoken Language Processing*, Philadelphia, PA, 1996, 1816-1819.

[10] Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76-92.