Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Prosodic annotation in the new corpus of Russian spontaneous speech CoRuSS

Nina Volskaya, Tatiana Kachkovskaia

Saint Petersburg State University, St. Petersburg State University,
7/9 Universitetskaya nab., St. Petersburg, 199034 Russia
[volni, kachkovskaia]@phonetics.pu.ru

Abstract

This paper deals with intonation of spontaneous Russian. It contains a description of the principles of prosodic annotation used in the new corpus of spontaneous speech—CoRuSS, and statistical data derived from this corpus. The prosodic annotation system was developed specially for the purpose; it is an extension and development of a well-known system of Intonation Constructions by E. A. Bryzgunova (7 ICs). Originally intended for teaching Russian as L2, Bryzgunova system proved to be insufficient for a detailed and adequate description of spoken Russian speech intonation. The results of the study provide statistical data on the frequency of particular intonation patterns of spontaneous Russian speech and form the basis for comparison with existing data on Russian read speech intonation; they confirm previously obtained information about new tendencies in the realization of Russian non-final and question intonation by young Russian native speakers and allow us to compare the realization and frequency of particular intonation patterns across other age-groups of native Russian speakers.

Index Terms: Russian intonation, spontaneous speech, prosodic annotation

1. Introduction

Our research had two goals: first, to demonstrate that the encoding of the prosodic information used for the corpus of spontaneous speech was adequate and can provide more information about prosodic variation in spontaneous and read speech than other annotation systems, and second, to show that the significant intonational differences existing between the two speech styles can be presented quantitatively.

The present paper describes the principles of prosodic annotation used in the new corpus of Russian spontaneous speech—CoRuSS, and provides statistical data on the frequency of intonation contours in modern spoken Russian in comparison with the data obtained earlier for read speech.

The new corpus of Russian Spontaneous Speech (CoRuSS) [1] developed at the Department of Phonetics, Saint Petersburg State University, contains over 14 hours of speech recorded in a soundproof studio from 60 speakers representing 3 age groups (16–30, 31–45, 46–77), with 10 male and 10 female speakers in each group. The subjects were asked to talk freely on any topic they wished, they were not restricted in any way in this respect. They talked about traveling, science, education, family etc.

The main part of the corpus includes prosodically annotated spontaneous interactive dialogues. The annotation was performed manually by 2 expert phoneticians: by listening and using visual information of the F0 contour they ascribed the number of the intonation pattern to the intonational phrase (IP) from the pre-defined set of intonation contours.

The first version of the annotation system described here was developed for modeling intonation for Russian TTS and later proved useful for the prosodic annotation of CORPRES—a 30-hour fully annotated corpus of Russian read speech recorded from 8 professional speakers [2].

For CoRuSS the annotation system has undergone only slight modification, enabling us to compare these two corpora in terms of prosodic data—in particular, frequency of intonation patterns in spontaneous vs read speech—and at the same time to capture and encode prosodic variation characteristic of spontaneous speech of different age groups in particular.

2. Prosodic Annotation in CoRuSS

2.1. General principles

CoRuSS has been prosodically labeled by hand according to the conventions proposed in [3] and later in a modified version of the prosodic annotation system. The system differentiates between 13 basic melodic types, having rising, falling and level nuclear tones, which combine with High, Low, Mid and Falling pre-nuclear parts of the contour. The contours are described in both acoustic and pragmatic terms, each melodic type having up to four subtypes. The use of such a detailed classification for prosodic annotation enables us to obtain more information about prosodic variation of spontaneous speech: to analyze subtle individual differences and understand how prosodic marking is linked to the speaker's intent to communicate.

The CoRuSS annotation includes phrasing (into intonational phrases —IPs), prosodic prominence and tonal events. For each IP a lexical word, carrying a nuclear pitch accent, was marked, and the melodic type was assigned from the proposed inventory of the intonational patterns based on perceptual and acoustical data. Some IPs did not contain a nuclear accent—typically, if the speaker failed to finish the IP, or if there were only non-speech events or disfluencies.

Speech disfluencies—repairs, false-starts, and non-speech events (clicks, laughter, cough etc.)—were marked using special symbols.

If the intonational phrase contained more than one perceptually prominent word, such additional prominence was marked with the symbol [+].

This may be illustrated as follows:

то1 есть э- [11]расчё1ты на про1чность / то1 есть [+]математи1ческий [11]расчё1т / на [10]про1чность / 9 э- / [10]сооруже1ний / [02]ну там / [+]мно1го [11]пара1метров

(that is em- [11]calculations of strength / I mean [+]mathematical [11]calculations / of [10]strength / 9 em- / for [10]construction / [02]that is / [+]many [11]parameters)

This fragment contains 7 IPs, one without lexical content

 doi: 10.21437/SpeechProsody.2016-188

(containing a non-speech event, marked with the symbol "9", and a hesitation). The melodic type is given in square brackets immediately before the word carrying nuclear stress. In this example, we observe different types of nuclear tones, indicating non-finality, emphasis and 2 cases of additional prominence marked with [+]. (Symbol "1" is used to mark a stressed vowel.)

2.2. Melodic types

The prosodic annotation system described here, is a revision and extension of a well-known Bryzgunova system of intonation constructions (IC) [4] [5, pp. 90–92] [6, p. 113], but ours is much more detailed in that it takes into account the most frequent variants of principal ICs, expressing finality, non-finality, interrogativity, and emphasis. Moreover, it includes those ICs which are absent in Bryzgunova's system: for example, a) those reflecting different degrees of finality, b) those having level tones as their nuclei, and c) those used in the parentheses and author's remarks. Our system also describes melodic types which reflect recent changes in the intonation of young Russian speakers [7].

Following Bryzgunova, a contour is defined by the type of the pitch movement (and in some cases, intensity) in the pre-nuclear, nuclear and post-nuclear parts of the intonational phrase (IP).

2.2.1. Contours with a nuclear fall

In the Bryzgunova system there are three Intonation Constructions (IC) with a falling tone in the intonation center (the nucleus) — IC-1 for complete neutral statements, IC-2 for statements with emphasis and Wh-questions, and IC-5 for exclamations. In our intonation model we propose to distinguish between several types of falls within the category IC1. Depending on how low the tone drops on the nuclear and post-nuclear syllables, they are further subdivided as follows:

- type 01: very low fall—signaling the end of a paragraph;
- type 01a: low fall—signaling the end of an utterance;
- type 01b: a fall to non-low—indicating cohesion, a link to the following utterance;
- type 10: non-low fall—as, for example, at the end of a compound sentence component.

As it follows from the definition, variants inside the category IC-1 can be compared with contextual allophones of same phoneme: for example, type 01 proposed to indicate the end of the paragraph will not be in its proper place when used instead of 01b, unless the speaker intends to abruptly finish his statement.

IC-2 is represented by three variants, having one common feature: a fall from a higher level, compared to IC-1. The first two variants are used for giving emphasis and are therefore characterized by an intensified fall, but there is the difference in timing of the pitch movement between them. Normally, it is co-terminus with the accented vowel or syllable, but in the second pattern, 02b, a stressed syllable is low pitched and there is a rise-fall in the post-nuclear syllable. This pattern signals strong emphasis and emotion.

On the other hand, the third variant, 02c—a fall from a high to mid or low level,—is normally accompanied by low intensity of the accented syllable and regular timing of the pitch movement. It is very common for spontaneous speech as it is used to convey involvement and personal contact with the listener.

We reserved two variants of falling intonation for Wh-questions, which in Russian can be pronounced either with the nuclear fall on the interrogative word—type 03,—or on some other word in the utterance—type 03a. The fall is normally from a higher level than that of the IC-1 .

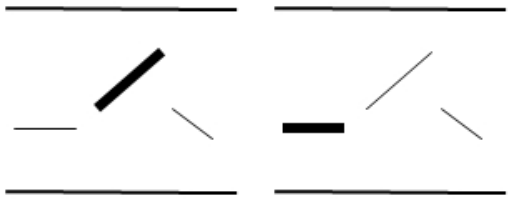
Another contour type—IC-4—is used in emotive utterances, which require a change in all the acoustic parameters: an extended pitch range, a wider interval of the falling tone, a higher level of intensity, and appropriate voice quality (timbre). Thus, 04 is used in exclamations, 04a in addressing a person, 04b in imperatives.

Type 05—a pitch rise to a very high level at some point in the pre-nuclear part, high plateau sustained up to the nucleus, and a fall on the nucleus, normally accompanied by high intensity—are characteristic of the contour type, used in another type of exclamations (IC-5 in the Bryzgunova's system).

Contour type IC-9—absent in the Bryzgunova system—is represented by three variants, depending on the tonal change in the nuclear and post-nuclear part: (low) level, (low) falling and (low) rising, with a low pre-nuclear part for all the three variants. It is common for parentheses and author's remarks and cannot be substituted by any other intonation contour.

2.2.2. Contours with a nuclear rise

Most of the rising contours (Bryzgunova's IC-3, IC-4, and IC-6)—are used either in (general) questions or in non-final IPs. The latter can be observed at IP boundaries between the subject and the predicate, at the end of grammatically dependent clauses and phrases, in enumerations, etc.



(a) 11 (b) 11a

Figure 1: Non-final rises (bold lines mark the accented syllable)

IC-3 in the Bryzgunova system is typical both for general questions and non-final (non-terminal) units. Although the tonal pattern for general questions and non-final units has some common features, namely, a rise in the intonational center and a fall in the post-central part, we nevertheless decided to concentrate on the differences between them rather than similarities:

1. the extent of F0 excursion, which is normally larger in questions;
2. the presence of F0 declination in the pre-nuclear part of the non-final units and its absence in general questions [8].

In our system, three tonal patterns are proposed for general questions, depending on the position of the nuclear accent:

- type 07: a rise(-fall) on the last word in the IP;
- type 07a: a rise(-fall) not on the last word in the IP;
- type 07b: a rise with the F0 maximum shifted to the next syllable after nucleus—this variant has become very

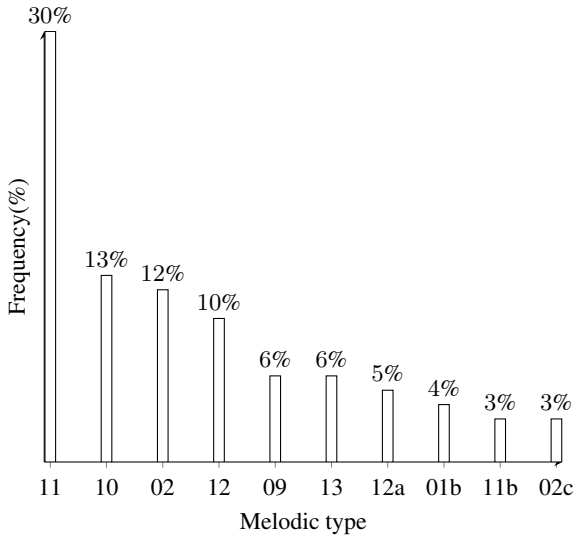


Figure 2: Frequency of melodic types in *spontaneous* speech (based on CoRuSS); models occurring in less than 2 % of cases are not included

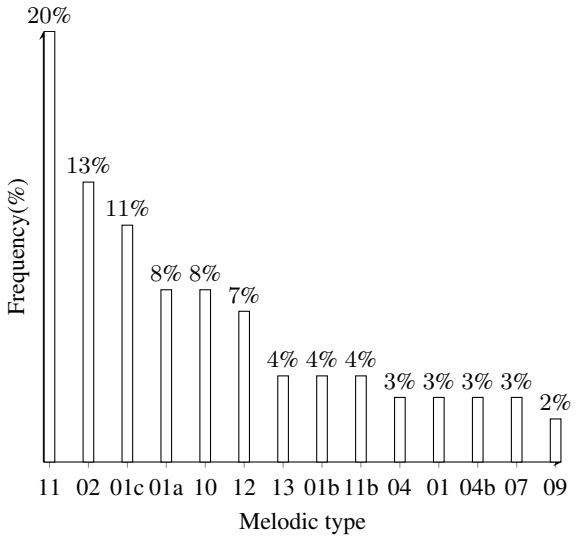


Figure 3: Frequency of melodic types in *read* speech (based on CORPRES); models occurring in less than 2 % of cases are not included (model 01c corresponds to 02c used in CoRuSS)

frequent in emotionally neutral questions produced by young Russian native speakers [7] [9].

For non-final, non-terminal, incomplete units, we proposed three tonal variants:

- type 11: a rise(-fall), normally at a smaller interval compared to type 07—see figure 1(a);
- type 11a: a rise(-fall) with the F0 maximum shifted to the next syllable after nucleus—see figure 1(b);
- type 11b: a rise(-fall) with a displaced F0 peak, often used for emphasis, but now commonly found in neutral speech of the younger generation.

Questions with an implied contrast are pronounced with a low (fall)-rise—type 08; a similar pattern for a nuclear low rise—13—is found in non-final units (both correspond to IC-4 in the Bryzgunova system).

2.2.3. Level tones

In the proposed system contour type 06 is represented by 4 variants of level tones—low 06, mid 06b; two variants of high level tones—06a and 06c—are found in neutral and emotional repeated or echoed questions (expressing disbelief or perplexity) and exclamations.

In non-final units we observe two types of level tones: a rise to a high pitch on the nuclear syllable leveled off in the post-nuclear part, is defined as the contour type 12. Slight declination can be observed, particularly if the post-nuclear part is pretty long. In Ode's transcription of Russian intonation variants with and without declination are regarded as two different pitch accents: H*H and H*M [10]. Unlike Ode, we regard them as belonging to contour type 12, since perceptually and functionally they are the same: the declining part is normally ignored by the listeners.

The second type—12a—HM*, a mid-level tone, actually realized as a step from a high pitch in the pre-nuclear part down to a medium pitch in the nucleus and sustained in the post-nuclear part.

The first type corresponds to Bryzgunova's IC-6, while the second one is absent from her system.

3. Statistics

3.1. General statistics

The total number of annotated IPs in CoRuSS is over 45,000. Of these, over 33,000 contain nuclear stress, and around 10,000 are entirely made up of speech disfluencies—such as hesitations or non-speech events.

On average, 5 % of IPs did not contain nuclear stress—i.e., were not finished due to changes in speech planning. For some speakers, this value achieved 8 % or even more.

It is commonly accepted that in spoken sentences the speaker tends to stress every meaningful word and avoids stressing functional words (in Russian, though, personal pronouns are normally stressed, which is not the case for English, for example) [11]. Thus, the average size of the intonational phrase in CoRuSS is 3.6 running words, but only 2.3 prosodic words, which is in accordance with the data obtained earlier for spontaneous Russian [12]. Depending on the speaker, the values range from 3.2 to 4.1 for running words, and from 1.9 to 2.7 for prosodic words.

3.2. Melodic types

Figure 2 shows the frequency (in percent) of intonation contours in CoRuSS. For comparison, figure 3 provides similar data for read speech, obtained from CORPRES (see Introduction).

In CoRuSS, the most frequent model across all speakers is type 11, which is a neutral rising(-falling) intonation used in non-final phrases; it is frequent in read speech as well (see figure 3). However, the choice of the model to express non-finality is speaker and sentence structure dependent: some speakers prefer other models—types 10, 12, 13, and 11b.

Other variants of model 11—11a and 11b, often realized with F0 peak shifted to the next syllable after nucleus—are rather frequent. Among young people (aged 16–30) these

melodic types occur more often, on average 18 %, while for the other two groups their frequency is 10 % [9].

Model 12a (a mid level tone, with a high tone in the pre-nuclear part lowered to a medium pitch in the nucleus and sustained in the post nuclear part) is rather common in spontaneous speech. This model is not described as a separate type in other annotation systems.

Among the falling contours, the most frequent melodic types observed in CoRuSS (see firure 2) are non-low or level (types 10, 01b); even in final IPs speakers prefer to use type 01b (a non-low fall or a fall, which levels off). The prevalence of such patterns is thus characteristic of spontaneous speech, which is completely in agreement with previous studies on Russian spontaneous speech intonation [13] [12] and with the data obtained for other languages as well [14].

This is not true for read speech (see figure 3), though, where types 01a and 01 (with the low fall as the nuclei) are more common.

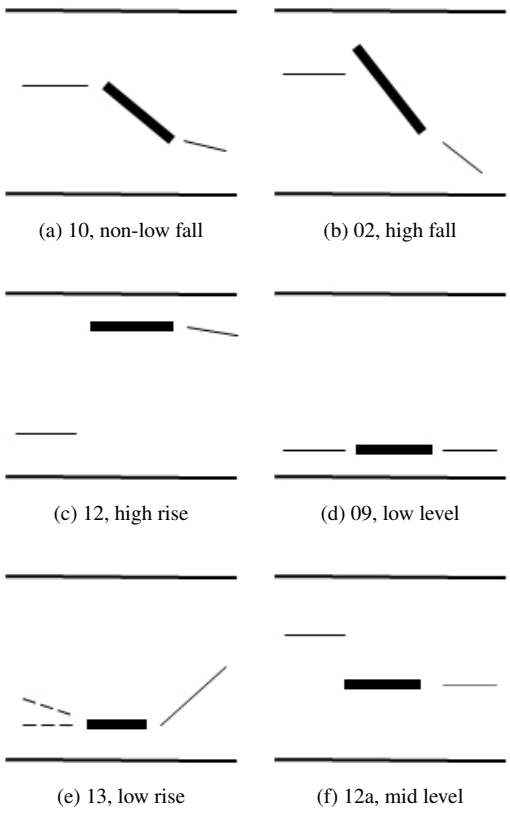
Melodic types used for parenthesis—mostly, 09 and 09a—are also much more common in spontaneous speech compared with read speech. This is due to the material. In read speech these models are used mostly in direct speech to separate author's words (e.g., "she said quietly"). In spontaneous speech these types are observed in a broader set of situations, including various types of comment, often longer than one IP. The intonation of parenthetical phrases and comments in our material displays similarity to the one described for English in [15]: they were all uttered in a low pitch within a compressed pitch range and practically no f0 modulation in the pre-nuclear part; as for the final pitch accent, the same set of nuclear accent types—level, rising and falling—was observed.

4. Conclusions

The data obtained from CoRuSS show that melodic types used in read and spontaneous speech are the same, but their frequency and distribution is different: this parameter can be used to differentiate spontaneous and prepared (read) speech along with other intonational characteristics which may serve as prosodic cues to the speaking style [16] [17] [18] [19]. The results of our study show that IP-final intonation contours in particular can be used to differentiate spontaneous and prepared (read) speech: the prevalence of non-low, level and particular types of rising tones in spontaneous speech reveal differences between the two speech types in the degree of planning and pre-planning involved, the effort to keep the listener's attention and the information the speaker intends to communicate in each case. In read speech, where syntactical and lexical content is prescribed for the speaker, final (low) falling tones, signaling the end of a sentence or a paragraph, are more frequent.

These results are consistent with the data reported in previous studies on intonational variation in speaking styles changes in Russian [13] [12], and in other languages as well [20] [21] [22]. We can tentatively suggest that they seem to reflect common universal human strategies in speaking in general—from intention (to support interaction with the partner) to performance (reflected by preference of particular intonation contours).

The results allow us to get more information about phrase non-final intonation patterns in spontaneous teenagers' speech: in particular, the use of rising-(falling) intonation with a displaced F0 peak—the pattern whose frequency of occurrence is spreading very fast and which has become particularly common with young Russian speakers [23].



(a) 10, non-low fall (b) 02, high fall

(c) 12, high rise (d) 09, low level

(e) 13, low rise (f) 12a, mid level

Figure 4: Most frequent model types in spontaneous speech (bold lines mark the accented syllable); for type 11 see figure 1

Further research is required to estimate the degree of "popularity" of this intonation pattern and to extend the study to other age groups, including children. The analysis and comparison of children, teenagers and students' vs. adults' spontaneous and read speech intonation of interrogative and non-final patterns may shed light on the source of such changes, which may influence the Russian intonation system as a whole.

5. Acknowledgements

The research is supported by the Russian Science Foundation (research grant # 14-18-01352).

6. References

[1] T. Kachkovskaia, D. Kocharov, P. A. Skrelin, and N. B. Volskaya, "CoRuSS - a new prosodically annotated corpus of Russian spontaneous speech," in *Proceedings of LREC 2016 [to be published]*, 2016.

[2] P. Skrelin, N. Volskaya, D. Kocharov, K. Evgrafova, O. Glotova, and V. Evdokimova, "CORPRES - Corpus of Russian Professionally Read Speech," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science, P. Sojka, A. Horák, I. Kopeček, and K. Pala, Eds. Springer Berlin Heidelberg, 2010, no. 6231, pp. 392–399.

[3] N. B. Volskaya and P. A. Skrelin, "Prosodic model for Russian," in *Proceedings of Nordic Prosody X*. Helsinki: Frankfurt am Main: Peter Lang, 2009, pp. 249–260.

[4] E. A. Bryzgunova, "Intonation [intonacija]," in *Russian Gram-*

mar, N. Shvedova, Ed. Moscow: Nauka, 1980, vol. 2, pp. 96–122.

[5] P. Cubberley, *Russian: A Linguistic Introduction*. Cambridge University Press, Oct. 2002.

[6] I. Nesterenko, "Analyse formelle et implemetation phonetique de l'intonation du parler russe spontane en vue d'une application a la synthese vocale," Ph.D. dissertation, Universite Aix-Marseille I - Universite de Provence, France, 2009.

[7] N. B. Volskaya, "Melodic type and its variants [sootnoshenije intonatsionnogo tipa i variantov]," in *Materialy XXXVII Mezhdunarodnoj filologicheskoj konferenctii Formaljnye metody analiza russkoj rechi*. Saint-Petersburg: SPbGU, 2008.

[8] D. Kocharov, N. Volskaya, and P. Skrelin, "F0 declination in Russian revisited," in *Proceedings of the 18th International Congress of Phonetic Sciences*, T. S. C. for ICPhS 2015, Ed. Glasgow, UK: the University of Glasgow, 2015, pp. 293.1–5.

[9] L. Demidchik, "Phonetic realization and functional load of rising-falling intonation in the speech of young Russian adults [Foneticheslaja realizatsija i funktsionaljnaja nagruzka voskhodjascheniskhodjashchei intonatsii v rechi russkoj molodezhi]." Saint Petersburg, 2009.

[10] C. Ode, "Transcription of Russian intonation, ToRI, an interactive research tool and learning module on the internet," in *Dutch Contributions to the Fourteenth International Congress of Slavists*, ser. Linguistics, H. P. Houtzagers, J. Kalsbeek, and J. Schaeken, Eds. Netherlands: Rodopi, 2008, vol. 1, pp. 431–450.

[11] J. Vaissière, "Language-Independent Prosodic Features," in *Prosody: Models and Measurements*, ser. Springer Series in Language and Communication, A. Cutler and D. R. Ladd, Eds. Springer Berlin Heidelberg, Jan. 1983, no. 14, pp. 53–66.

[12] N. B. Volskaya, "Prosodic features of Russian spontaneous and read aloud speech," in *Phonetics of Russian and Finnish*, V. d. Silva and R. Ullakonoja, Eds. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang, 2009, pp. 133–144.

[13] N. D. Svetozarova, Ed., *Phonetics of Spontaneous Speech [Fonetika spontannoj rechi]*, Leningrad, 1988.

[14] A. I. Mata, H. Moniz, and F. Batista, "Comparing phrase-final patterns across speech styles and age groups in European Portuguese," *Cahiers de linguistique française*, no. 31, pp. 171–176, 2014.

[15] B. Grosz and J. Hirschberg, "Some intonational characteristics of discourse structure," in *ICSLP 92 Proceedings*, Banff, Alberta, Canada, 1992, pp. 429–432.

[16] K. Silverman, E. Blaauw, J. Spitz, and J. Pitrelli, "A prosodic comparison of spontaneous speech and read speech," in *Proceedings of ICSLP*, Banff, 1992, pp. 1299–1302.

[17] E. Blaauw, "On the perceptual classification of spontaneous and read speech," Ph.D. dissertation, LEd, Utrecht, 1995.

[18] J. Hirschberg, "Prosodic and other acoustic cues to speaking style in spontaneous and read speech," in *Proceedings of ICPhS 95*, vol. 2, Stockholm, 1995, pp. 36–43.

[19] B. Megyesi and S. Gustafson-Capkova, "Production and Perception of Pauses and their Linguistic Context in Read and Spontaneous Speech in Swedish," in *Proceedings of ICSLP'2002 - 7th International Conference on Spoken Language Processing*, Denver, USA, 2002, pp. 2153–2156.

[20] L. Trask, R. L. Trask, and R. M. Millar, *Why Do Languages Change?* Cambridge University Press, 2010.

[21] H. Mixdorff, H. R. Pfitzinger, and K. Grauwinkel, "Towards Objective Measures for Comparing Speaking Styles," in *SPECOM 2005 Proceedings: 10th International Conference Speech and Computer*. Greece: University of Patras, 2005, pp. 131–134.

[22] G. Bruce, "Modelling Swedish Intonation for Read and Spontaneous Speech," in *Proceedings of ICPhS 95*, vol. 2, Stockholm, 1995, pp. 28–35.

[23] N. Volskaya, "Generation Gap Mirrored by Russian Intonation [Konflikt pokolenij v zerkale russkoj intonacii]," in *Festschrift in honour of O.F.Krivnova [Sbornik statej k jubileju O.F.Krivnovoj]*, A. Arkhipov, I. Kobozeva, and K. Semenova, Eds., Moscow, 2013.