

COMPARISON OF MELODIC PORTRAITS OF ENGLISH AND RUSSIAN DIALOGIC PHRASES

Lobanov B.M. (Lobanov@newman.bas-net.by),
United Institute of Informatics Problems NAS Belarus, Minsk

This study is an extension of the author's works, presented at the “Dialogue 2014 and Dialogue 2015” conferences. According to the concept of universal melodic portrait (UMP), a phrase intonation can be described as a sequence of UMPs of accentual units (AUs) that make up the phrase. The present paper describes the results of pilot studies where melodic portraits for English and Russian language phrases were compared. The examined phrases were derived from simple situational dialogues and were spoken by native English and Russian speakers. The study was restricted only to phrases with a one-accent unit structure representing the three main types of phrase intonations: affirmative statements, special questions and general questions.

The described UMP model allows to investigate tonal differences within languages by applying precise quantitative assessments. The method can be used effectively for solving problems of language interference. Moreover, the UMP model could potentially find an effective application in foreign language studies. Using the appropriate software that realizes the described stages of UMP construction, a learner could be able to visually compare an intonation of the pronounced phrase with its target intonation portrait and work to eliminate a foreign accent by proper training.

Key words: intonation patterns, melodic portrait, synthesis and analysis of intonation, English and Russian intonations, English and Russian as second languages, TTS synthesis.

СРАВНЕНИЕ МЕЛОДИЧЕСКИХ ПОРТРЕТОВ АНГЛИЙСКИХ И РУССКИХ ФРАЗ ДИАЛОГОВОЙ РЕЧИ

Лобанов Б.М. (Lobanov@newman.bas-net.by),
Объединённый институт проблем информатики НАН Беларуси, Минск, Беларусь

Ключевые слова: интонационные конструкции, мелодический портрет, синтез и анализ интонации, английская и русская интонации, английский и русский как иностранный

Introduction

The present work is a follow up study to the previously introduced model of universal melodic portraits (UMP) of accentual units* (AU) for representation of phrase intonations in TTS synthesis [Lobanov et al, 2006]. According to this model, a phrase is represented by one or more of AUs. Each unit, in turn, can be composed of one or more phonetic word. If there is more than one word in an AU, than only one word bears the main stress while other words carry a partial stress. Each AU consists of *pre-nucleus* (all phonemes preceding the main stressed vowel), *nucleus* (the main stressed vowel) and *post-nucleus* (all phonemes following the stressed vowel). The UMP model assumes that topological features of melodic AU for particular type of intonation do not depend on a number or quality of phonemic content of a pre-nucleus, nucleus or post-nucleus, nor on the fundamental frequency range specific for a given speaker.

*Accent Unit often referred to as Accent Group [Ogden et al, 2000]

The UMP model allows to represent intonation constructs as a set of melodic patterns in normalized space $\{\mathbf{Time} - \mathbf{Frequency}\}$.

Time normalization is performed by bringing pre-nucleus, nucleus and post-nucleus elements of AU to standard time lengths. This sort of normalization levels out the differences in melodic contours caused by the number of words and phonemes in an AU.

For fundamental frequency normalization $F_{0\ min}$ and $F_{0\ max}$ are determined within the ensemble of melodic contours produced by a certain speaker. This sort of normalization cancels out the differences of melodic contours caused by speakers voice register and diapason.

The normalization is calculated by the formula

$$F_0^N = (F_0 - F_{0\ min}) / (F_{0\ max} - F_{0\ min}) \quad (1)$$

In certain cases it may be beneficial to use statistical normalization instead of (1)

$$F_0^N = (F_0 - M) / \zeta \quad (2),$$

where M is mathematical expectation, ζ is standard deviation. Note that M can be interpreted as a register and ζ – as a diapason of speaker’s voice.

Therefore, the normalized space for UMP may be presented as a rectangle with axes (T_N, F_0^N) as schematically shown in **Figure 1**, while the interval $[0 - 1/3]$ on the absciss T_N is a pre-nucleus, $[1/3 - 2/3]$ is a nucleus, and $[2/3 - 1]$ is a post-nucleus. The intervals on the ordinate F_0^N : $[0 - 1/3]$ – low level, $[1/3 - 2/3]$ – mid-level, $[2/3 - 1]$ – high level.

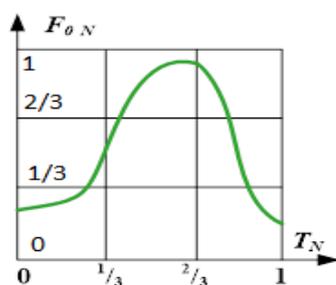


Figure 1. MPAU-representation.

Figure 2 illustrates the results of time-frequency normalization of the example one-accent-unit phrases with affirmative intonations: «It is no distance **at all**» and «It is only a couple of hundred **yards**».

The first phrase contains four phonetic words (underlined) and the second one - five. The last word in both phrases is accented (in bold font), and the nucleus is the stressed vowel in this word. **Figure 2** shows the intonograms of both phrases obtained with the PRAAT package (see: <http://www.fon.hum.uva.nl/praat/>). The figure demonstrates that phrases spoken by different speakers differ by 1.5 times in duration and 1.3 times in the maximum fundamental frequency. Despite these lexical and fundamental frequency differences, the final construction of UMPs for both phrases (the right-upper part of **Figure 2**) makes the similarity of melodic portraits evident.

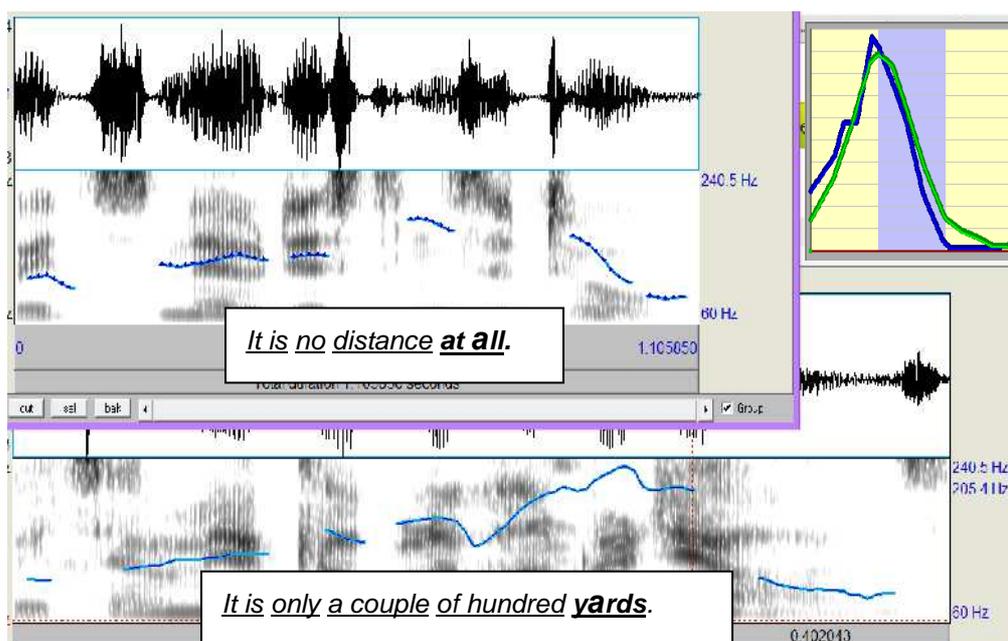


Figure 2. Illustration of time and frequency normalization.

In the earlier work [Lobanov, 2014], the efficiency of suggested approach was verified by constructing UMPs for main intonation patterns of Russian speech: IP1 - IP7. The subsequent study [Lobanov, 2015] demonstrated successful construction of UMPs for compound narrative sentences in Russian. The present study provides pilot results for comparison of UMPs of English and Russian phrases for simple dialogue spoken by native English and Russian speakers.

The paper is laid out in the following way: the first paragraph describes the chosen texts and audio-material as well as the method of applying MPAU model to the analysis, the second paragraph shows the results of MPAU modeling and also the analysis and interpretation of the results obtained.

1. Method

The experiment was based on English texts and audio-files from the manual [Ockenden, 2005] which included:

- 44 everyday situations, each containing four dialogues in natural conversational English;
- All dialogues consist 1051 sentences, including 704 affirmative, 325 interrogative and 22 exclamatory sentences, spoken by certain number of male and female speakers;
- Situations relevant to those studying or travelling in England, including eating out, entertainment and travel, as well as more general functions such as greetings, complaining and apologizing.

In the present study we have restricted ourselves to three major types of phrase intonation – *Affirmative statements*, *Special questions* and *General questions*. In addition, we restricted the study of intonations to the case of one-AU phrases (it is about 70 per cents of whole number of phrases). Other intonation types such as *Alternative questions*, *Tag questions*, *Commands*, *Exclamatory sentences*, *Direct address*, *Enumerating*, *Introductory phrases etc.* were not included in this study.

The comparison Russian language test material was based on direct translations of corresponding English phrases into Russian. The translated text was used to make Russian audio recordings that imitated normal conversation of two people with a standard Russian accent.

The composition of UMPs of both Russian and English phrases was performed with the aid of *PhonoClonator* and *IntoClonator* systems [Lobanov, 2014]. On the basis of a pre-marked text, the *PhonoClonator* system makes it possible to automatically segment each signal into phonemes and pitches (F0) and indicate positions of a nucleus for AU in a phrase.

Figure 3 shows the general view of the users interface of the *PhonoClonator* system for phrase processing.

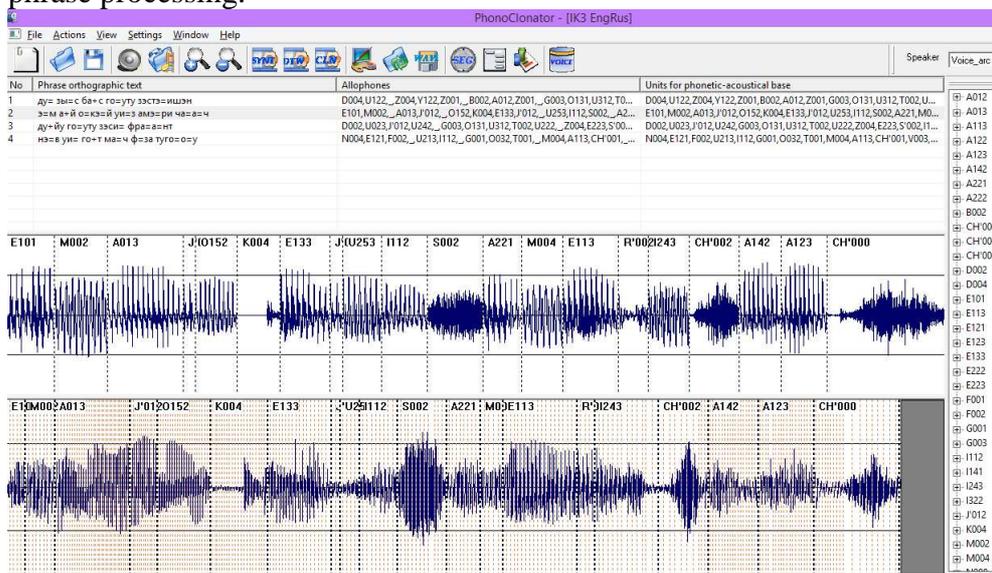


Fig. 3. PhonoClonator: the general view of the users interface

In the next step, the pre-marked audio-signals are fed into *IntoClonator* system that provides the boundaries of the nucleus, pre-nucleus and post-nucleus as well as melodic and intensity contours (Fig. 4). Minimum ($F_{0\ min}$) and maximum ($F_{0\ max}$) fundamental frequency values (F0) are determined automatically for the melodic contour of the phrase analyzed – “Am I OK for St Marys Church?”



Fig. 4. IntoClonator: the general view of the users interface

Finally, *ShapeEditor* system makes it possible to use the information processed by *IntoClonator* system for composing melodic portraits of the analyzed phrase “Am I OK for St Marys Church?” in a normalized UMP-form described above (see: figure 5).



Fig. 5. ShapeClonator: the general view of the users interface

2. Results

Here, we present the results of comparisons of melodic portraits of English and Russian phrases chosen from sample dialogues on the principle of being well-pronounced examples of the three evaluated types of intonation contours: affirmative statements, special questions and general questions.

For affirmative statements we used English and Russian phrases an example of which is listed in Table 1. The phrases were spoken by different speakers. The analyzed one-accent-unit phrases are italicized. The word that carries the main accent is printed in a bold type with its stressed vowel (nucleus) underlined. All syllables to the left of the nucleus make up a pre-nucleus and those to the right – a post-nucleus.

Table 1. English and Russian phrases spoken with affirmative intonation of statements

English	Russian
- Is it far? - <i>It is only about five minutes <u>walk</u>.</i>	- Далеко ли это? - <i>Это всего в пяти минутах <u>ходьбы</u>.</i>
- Will it take me long to get there? - <i>It is no distance <u>at all</u>.</i>	- Долго ли мне придётся идти? - <i>Это вообще <u>не расстояние</u>.</i>
- Should I take a bus? - <i>You can walk it in under five <u>minutes</u>.</i>	- Мне нужно подождать автобуса? - <i>Вы сможете <u>дойти</u> за пять <u>минут</u>.</i>
- Is it too far to walk? - <i>It is only a couple of hundred <u>yards</u>.</i>	- Долго ли придётся идти пешком? - <i>Это всего в паре сотен <u>шагов</u>.</i>

Figure 6, as well as following **Figures 7** and **8**, show melodic portrait curves obtained with the use of computational approaches described in the **Introduction**. In Figure 6 (a), thin blue lines reflect the melodic portraits of four English phrases and the bold line reflects the averaged UMP. The UMP is represented along the X-axis by the succession of three time normalized stretches – pre-nucleus, nucleus, post-nucleus, with normalized fundamental frequency relative to the phrase maximum and minimum along the Y-axis. Similarly, in **Figure 6 (b)** green lines show tone curves for the Russian phrases, and in **Figure 6 (c)** shows superimposed typical intonation contours for English and Russian affirmative statements.

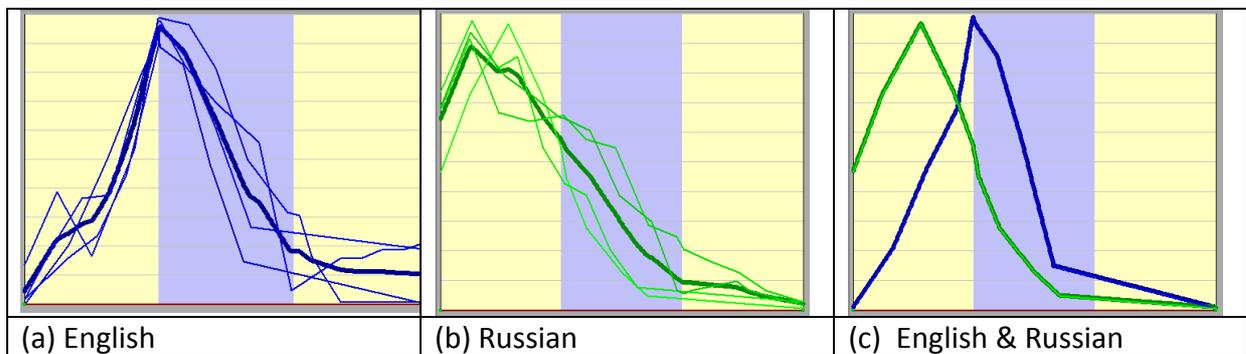


Figure 6. UMPs for English and Russian one-accent-unit phrases (affirmative statements)

The comparison of English and Russian affirmative statement melodic portraits in **Figure 6 (c)** allows to establish the following differences:

- the most changes are found in pre-nucleus and nucleus regions;
- in the pre-nucleus region, maximum of the Russian UMP curve falls closer to the middle of the region, whereas the English UMP curve peaks at the end;
- in the nucleus region, the English UMP curve is characterized by a sharper decline in comparison to the Russian UMP curve.
- in the post-nucleus region, both Russian and English MP curves show an identical low level steady decline.

Next, for the study of intonation characteristics of special questions, we used the example English and Russian phrases listed in **Table 2**. The content representation and mark up in **Table 2** is similar to **Table 1**.

Table 2. English and Russian phrases spoken with the intonation of special questions

English	Russian
- <i><u>What</u> can I get you drink?</i> - A black coffee for me, please.	- <i><u>И что</u> предложить Вам выпить?</i> - Чёрный кофе, пожалуйста.
- <i><u>What</u> are you going to have to drink?</i> - I'd like something cool.	- <i><u>А что</u> Вы желаете выпить?</i> - Хотелось бы чего-нибудь прохладного

- What are you going to have? - A half of bitter, please.	- Что бы Вы хотели сейчас? - Полкружки горького, пожалуйста.
- What is it to be? - The same again, please.	- А что теперь будете пить? - То же самое, пожалуйста.

Figure 7 shows comparison of melodic portraits of special question intonations for English and Russian phrases. The figure layout and content representation is similar to **Figure 6**.

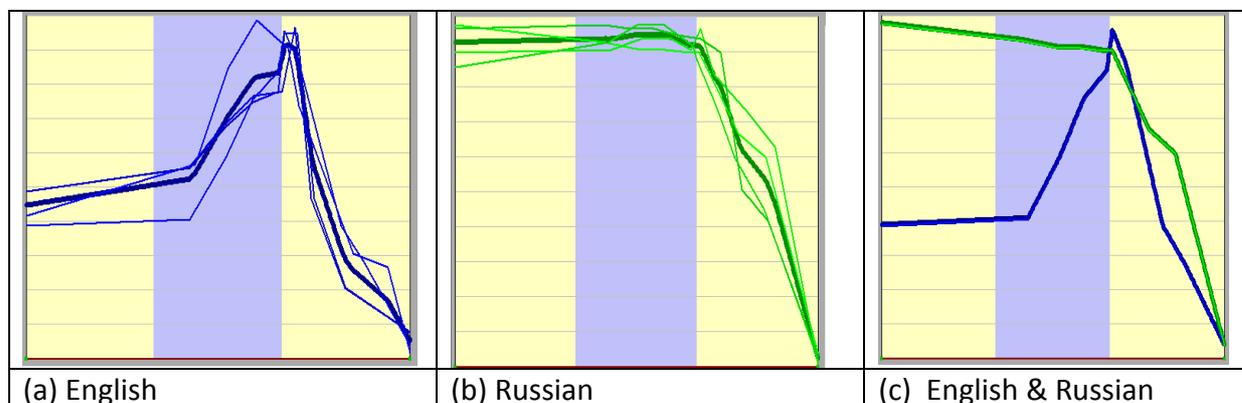


Figure 7. UMPs for English and Russian one-accent-unit phrases (special questions)

The comparison of English and Russian special question melodic portraits allows to establish the following main differences:

- the most significant changes are found in the pre-nucleus and nucleus regions;
- in the pre-nucleus region, the averaged Russian UMP is characterized by considerably higher level than the English UMP;
- in the nucleus region, the English UMP curve is characterized by a sharp rise in tonal frequency whereas the Russian curve remains on a steady high level;
- in the post-nucleus region, both Russian and English UMP curves demonstrate identical sharp interval decline.

Finally, for the study of intonation characteristics for general questions, we used the example English and Russian phrases listed in **Table 3**. The content representation and mark up in **Table 3** is similar to **Table 1**.

Table 3. English and Russian phrases spoken with the intonation of general questions

English	Russian
- Does this bus go to the station? - No, you'll have to get off at the bank.	- Этот автобус идет на вокзал ? - Нет, он идет к банку.
- Am I OK for St Marys Church? - No, we only go as far as the park.	- Я правильно иду к церкви? - Нет, вы только дойдёте до парка
- Do you go to the sea-front? - No, you're going the wrong way.	- Вы идёте к приморскому бульвару ? - Нет, Вы пошли неправильным путём.
- Have we got much further to go? - It's the next stop.	- Должны ли мы ещё дальше ехать? - Ваша остановка – следующая..

Figure 8 shows comparison of melodic portraits of general question intonations for English and Russian phrases. The figure layout and content representation is similar to **Figure 6**.

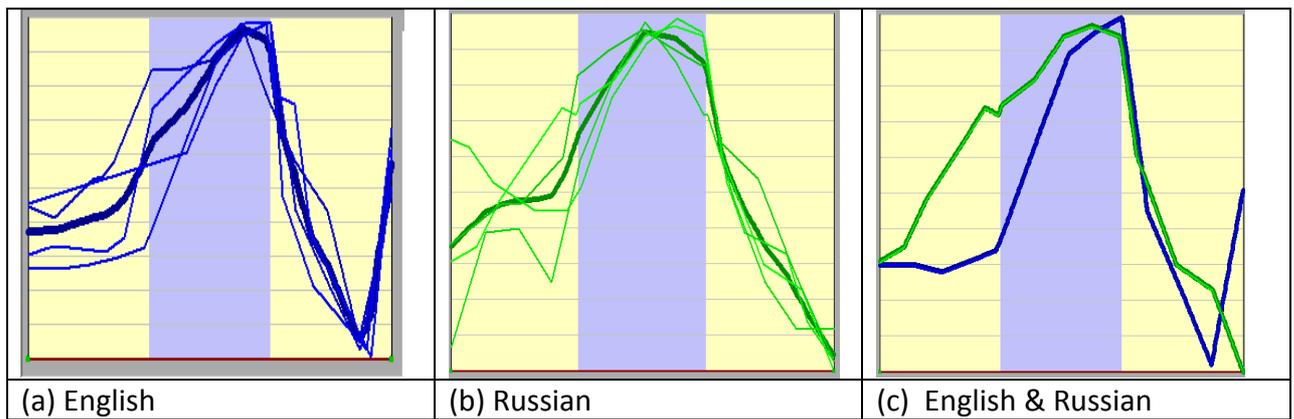


Fig. 8. UMPs for English and Russian one-accent-unit phrases (general questions)

The comparison of English and Russian melodic portraits for general questions allows to establish the following main differences:

- the most significant changes are found in the pre- and post-nucleus regions;
- in the pre-nucleus region the English UMP curve is characterized by a steady low level, whereas the Russian UMP follows a substantial rise;
- in the post-nucleus region the English UMP curve is characterized by sharp decline with a subsequent rise towards the end of the phrase. On the other hand, the Russian UMP curve shows only steady decline;
- in the nucleus region, the English UMP curve shows a sharper rise in comparison to the Russian one.

Conclusions

The present paper describes the results of pilot studies where melodic portraits for English and Russian language phrases were compared. The study was restricted only to phrases with a one-AU structure representing the three main types of phrase intonations: *affirmative statements*, *special questions* and *general questions*.

The described results of comparisons of UMPs of English and Russian phrases are consistent with the observations of linguists involved in comparative studies of intonation in order to provide guidelines for mastering foreign languages. These guidelines often tend to have rather vague and descriptive language, for example:

“The melody of an English phrase differs markedly from a Russian one:

- a). The English voice range is much wider meaning that the beginning of the phrase is higher and the end of the phrase is lower in tone than in Russian.*
- b). English is characterized by the tonal movement within a vowel at a perceptibly longer time stretches which gives an impression of ‘singing’ stressed vowels.*
- c). The reference point of tone modulation in English is the lowest tone level while in Russian it is the average level.*
- d). The English cadence reaches the lowest point of the range, as well as tone rising from the lowest level.*
- e). The English phrase is characterized by the centralized accent. It is within the stressed syllable that the widest and longest voice cadence is exercised.”*

(see: <http://xreferat.com/71/1238-1-uprazhneniya-v-obuchenii-ritmu-i-intonacii-angliyskogo-yazyka-v-osnovnoiy-shkole.html>).

The described normalized UMP model of the phrase intonation allows to investigate the tonal differences between different languages by applying precise quantitative assessments. The method can be used effectively for solving problems of language interference. Moreover, the UMP model could potentially find an effective application in foreign language studies. Using the appropriate software that realizes the described stages of MP construction, a learner could be

able to visually compare the intonation of the pronounced phrase with its target intonation portrait and work to eliminate a foreign accent by proper training.

The importance of mastering proper intonation in language instruction is emphasized by many authors:

«Intonation, the “music” of a language, is perhaps the most important element of a correct accent. Many people think that pronunciation is what makes up an accent. It may be that pronunciation is very important for an understandable accent. But it is intonation that gives the final touch that makes an accent correct or native. Often we hear someone speaking with perfect grammar, and perfect formation of the sounds of English but with a little something that gives her away as not being a native speaker». (See <http://www.goodaccent.com>)

Another example. When talking about a Russian accent in American English some native speakers make interesting observations:

«Ask your average American what they think about the Russian accent and they say;

“Russians don’t sound very friendly. I never feel as if they like me. I’m not sure if that’s because of their language, or if it’s a cultural thing. “

One reason that Russian English speakers don’t sound friendly is their flat tone.

You simply don’t use enough intonation when you speak.

Russian English speakers don’t use the rising-falling intonation that Americans find friendly and engaging. You don’t use sufficient intonation when asking questions».

(see: <http://www.confidentvoice.com/blog/russian-english-speakers-5-reasons-why-americans-dont-understand-you/>)

The author is grateful to **Dr. Anna Osipovich** for the useful discussions and for the help in preparation of English version of this paper.

Referencies

1. Lobanov B., Tsurulnik L., Zhadinets D., Karnevskaya E. (2006) Language- and Speaker Specific Implementation of Intonation Contours in Multilingual TTS Synthesis // Speech Prosody: Proceedings of the 3rd International conference. Dresden, Germany: Vol. 2. – pp. 553-556.
2. Lobanov B., Okrut T. (2014) Universal Melodic Portraits of Intonation Patterns in Russian Speech [Universalnye melodicheskie portrety intonacionnyh konstrukcij russkoy rechi]// Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2014” [Komp’yuternaya Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2006”], Bekasovo, pp. 330-339.
3. Lobanov B.M. (2015) An Experience of Melodic Portraits Creation of Complex Declarative Sentences of Russian [Opyt sozdaniya melodicheskikh portretov povestvovatelnyh predlozheniy russkoy rechi] // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2014” [Komp’yuternaya Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2015”], Moscow, pp. 414-426.
4. Ockenden M , (2005) Situational Dialogues // The English Centre, Eastbourne / Revised Edition. - Longman - 98 pp.
5. R. Ogden, S. Hawkins, J. House, M. Huckvale, J. Local, P. Carter, J. Dankovicova, , and S. Heid, (2000) “Prosynth: an integrated prosodic approach to device-independent, natural-sounding speech synthesis,” Computer Language and Science, pp. 177–210, 2000.