МЕЖДУНАРОДНЫЙ КОНГРЕСС ПО ИНФОРМАТИКЕ: ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Материалы международного научного конгресса

Республика Беларусь,
Минск, 24–27 октября 2016 г.

INTERNATIONAL CONGRESS ON COMPUTER SCIENCE: INFORMATION SYSTEMS AND TECHNOLOGIES

Proceedings of the International Scientific Congress

Republic of Belarus,
Minsk, October 24–27, 2016

МИНСК
БГУ
2016

УДК 37:004(06)

Международный конгресс по информатике: информационные системы и технологии = International Congress on Computer Science: Information Systems and Technologies [Электронный ресурс] : материалы междунар. науч. конгресса, Республика Беларусь, Минск, 24–27 окт. 2016 г. / редкол.: С. В. Абламейко (гл. ред.), В. В. Казаченок (зам. гл. ред.) [и др.]. – Минск : БГУ, 2016.
ISBN 978-985-566-369-1.

Представлены материалы международного научного конгресса, в которых рассматриваются следующие темы: аэрокосмические информационные и телекоммуникационные системы и технологии; биоинформатика и приложения; высокотехнологичная образовательная среда (педагогические аспекты); геоинформационные системы; защита информации и компьютерный анализ данных; интеллектуальные информационные системы; информационно-коммуникационные системы; цифровые медиа технологии; информационные системы и технологии обеспечения научно-образовательной среды; компьютерные технологии в приборостроении; параллельная и распределенная обработка данных, вычислительные системы и сети; программная инженерия; распознавание образов, информационные системы управления; теоретическая информатика.

УДК 37:004(06)

ISBN 978-985-566-369-1

Секция 7. ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ СИСТЕМЫ

КОМПЬЮТЕРНАЯ СИСТЕМА АНАЛИЗА И ИНТЕРПРЕТАЦИИ ИНТОНАЦИИ РЕЧИ

Б. М. Лобанов, В. А. Житко, Ю. А. Здоронок

Объединенный институт проблем информатики НАН Беларуси
Минск, Беларусь
e-mail: lobanov@newman.bas-net.by

В этой статье описывается важность интонации в человеческом понимании речи и в создании компьютерных систем анализа, синтеза и понимания речевых сигналов. Приводятся основополагающие принципы интонационной теории речи, основанной на концепции универсальных мелодических портретов. Кроме того, описываются основные алгоритмы анализа и интерпретации интонации в произнесении, лежащие в основе развитой компьютерной системы для обучения речевой интонации. Далее показываются результаты работы системы и обсуждается ее полезность в обучении интонации иностранного языка.

Ключевые слова: интонация речи; анализ и синтез речи; мелодический портрет; анализ и интерпретация интонации; компьютерная система для обучения.

COMPUTER-BASED SYSTEM OF ANALYSIS AND INTERPRETATION OF SPEECH INTONATION

B. M. Lobanov, V. A. Zhitko, Y. A. Zdaranok

United Institute of Informatics Problems NAS Belarus
Minsk, Belarus

In this paper we describe the importance of intonation in human comprehension of speech and in creation of computer-based systems for analysis, synthesis and understanding of speech signals. We lay out fundamental principles of intonational theory of speech that is based on the concept of universal melodic portraits. In addition, we describe the main algorithms of analysis and interpretation of intonation in an utterance that underlie the developed computer-based system for teaching of speech intonation. We further show the system's output and discuss its usefulness in teaching foreign language intonation.

Keywords: intonation of speech; speech analysis and synthesis; melodic portrait; intonation analysis and interpretation; computer system for teaching.

INTRODUCTION

Intonation plays significant role during speech comprehension. Speech intonation shows a communicative intention of an utterance, its logical meaning, a prominence of the most significant theme in relation to general themes (actual division of a sentence), a distinction between semantically associated segments of speech, and an integration of speech elements within these segments.

A common linguistic idea is that foreign accent appears more evident in intonation, and therefore, prosodic aspects of speech should be explicitly taught to students who wish to communicate intelligibly in a foreign language. Foreign accent arises due to contact between two different language systems, namely, at bilingualism, because of language interference. Intonation is the most important aspect of speech that provides both linguistic and sociocultural information. Considering that functions of intonation in speech are various, and that deviations in this area can lead to significant semantic differences, incorrect intonation can make a wrong impression during the speech of a non-native language speaker.

For example, there are considerable intonational differences between Russian and English languages. American English native speakers made the following interesting observation: "Ask an average American what they are thinking about the Russian accent, and the answer will be as follows: "*Russians don't sound very friendly. I feel like they don't like me at all. I am not sure whether it comes from their language or from their culture*"*?* One of the reasons why native Russian speakers sound unfriendly in English is the so called "flat" tone. Native Russian speakers often do not use language-specific phonological representation of intonation during their conversation in English. Moreover, native Russian speakers tend to avoid using rising and falling intonations in English and, as a result, Americans may find their speech unfriendly and unpleasant. As another example, native Russian speakers often do not use question-specific intonational patterns while asking questions in English".

In this view, is it very important to emphasize intonational aspects of speech while teaching foreign language and while creating computer-based systems for speech analysis and synthesis. Previously, we've published several scientific results in the area of speech intonation analysis and synthesis which can serve as a foundation for the creation of innovative intelligent automatic systems for processing of speech signals [1–3].

1. GENERAL INFORMATION ABOUT SPEECH INTONATION

In phonetics and physiology of speech, a phrase is considered to be a comparatively independent unit in speech intonation [4]. Phrase independence manifests itself in articulatory integrity, semantic and syntactical association of linguistic units, and in presence of objective traits that allow to single it out from a speech stream. Any punctuation mark can set a boundary of intonation phrase in a written text. However, quite often the number of phrases can be greater than the number of punctuation marks. Generally, a phrase is a combination of one to five words, with a three-word phrase being the most common. A particular place of a phrase boundary is determined by the optimal satisfaction of the semantico-syntactic, phonetic and physiological requirements. The first of the named requirements prescribes a union inside a phrase between semantically connected words that cannot be split into two phrases. The second requirement denotes the tendency of language phonetic systems for definite rhythmical construction, for example, a group of two to three words combined in one phrase. Finally, the third requirement prescribes the formation of a phrase with the number of words that could be physically pronounced during the time required for one act of exhaling.

After definition of a phrase, the next step in intonational speech analysis is determination of a phrase intonation type. The main intonation phrase types include: completed and uncompleted phrases, special and common questions, exclamation phrases, and some others types. The number of intonation subtypes of the main phrase intonation types could reach several dozens [4]. The task of definition of an intonation subtype is achieved by the analysis of two factors: the position of a phrase in a text and by its semantic value. The first factor, a position in a text, is determined by the analysis of a phrase position in relation to the nearest punctuation marks. For example, its position at the beginning or the end of a text, its position before or after punctuation marks, and the type of punctuation marks in a sentence. The second factor, a semantic value of a phrase, is determined based on the meaning of a phrase and on its logically emphasized intonational center. In particular, it needs to be determined whether a phrase expresses intentions, explanations, follow up questions etc. The final decision on an intonation type and subtype of a phrase is made after considering both factors.

Speech intonation is physically realized by the set of acoustic means, named the prosodical parameters:

– melodics – the movement of the frequency of the main tone (F0);
– energetics – the current change of the force (amplitude) of the sound (A);
– rhythmics – the current change of the duration of the sounds and pauses (T).

Since melodics of speech is the most informative among these parameters, the main attention in this article is given to the melodics.

2. THE MODEL OF UNIVERSAL MELODIC PORTRAIT (UMP)

The present work is a follow up study to the previously introduced model of universal melodic portraits (UMP) of accentual units[5] (AU) for representation of phrase intonations in text -to-speech synthesis [1]. 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.

The UMP model allows to represent intonation constructs as a set of melodic patterns in normalized space {***Time – 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 speaker's voice register and diapason.

The normalization is calculated by the formula

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

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

$$F_0^N = (F_0 - M) / \zeta \,, \tag{2}$$

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 abscess T_N

[6] Accent Unit often referred to as Accent Group [5].

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.

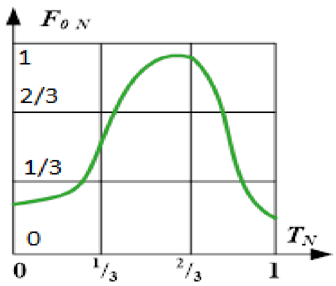

Fig. 1. UMP-representation

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.

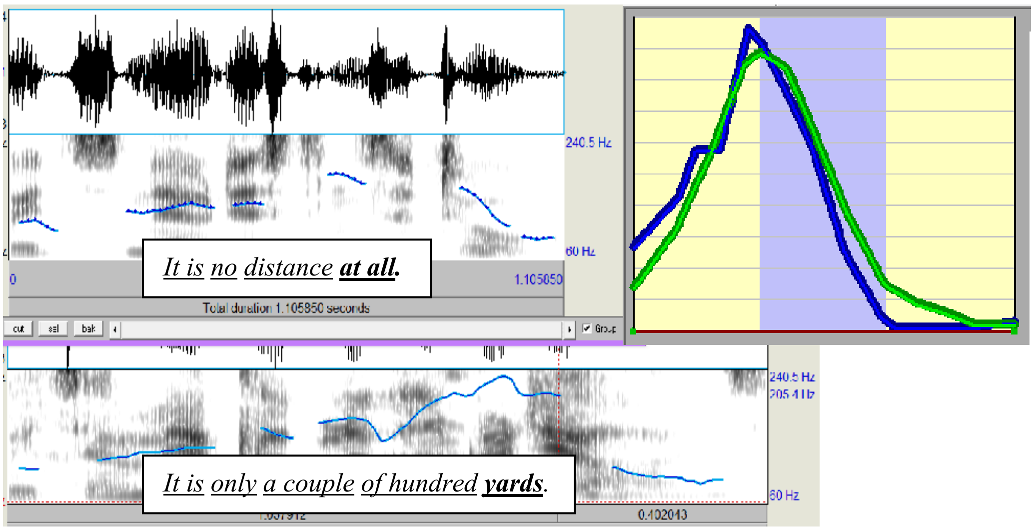

Fig. 2. Illustration of time and frequency normalization

3. THE ALGORITHMS OF INTONATION ANALYSIS IN A COMPOSITION OF THE COMPUTER TEACHING SYSTEM OF FOREIGN SPEECH

Figure 3 contains a block diagram illustrating sequence of algorithms for analysis and interpretation of speech intonation within the developed computer teaching system. The main goal of the system is to provide a student with a compact and easily interpretable image for the results of analysis of melodic and energy contours of phrases with different intonation. The system would also provide a visual, auditory and numerical evaluation of the quality of learning of a foreign speech intonation by a student.

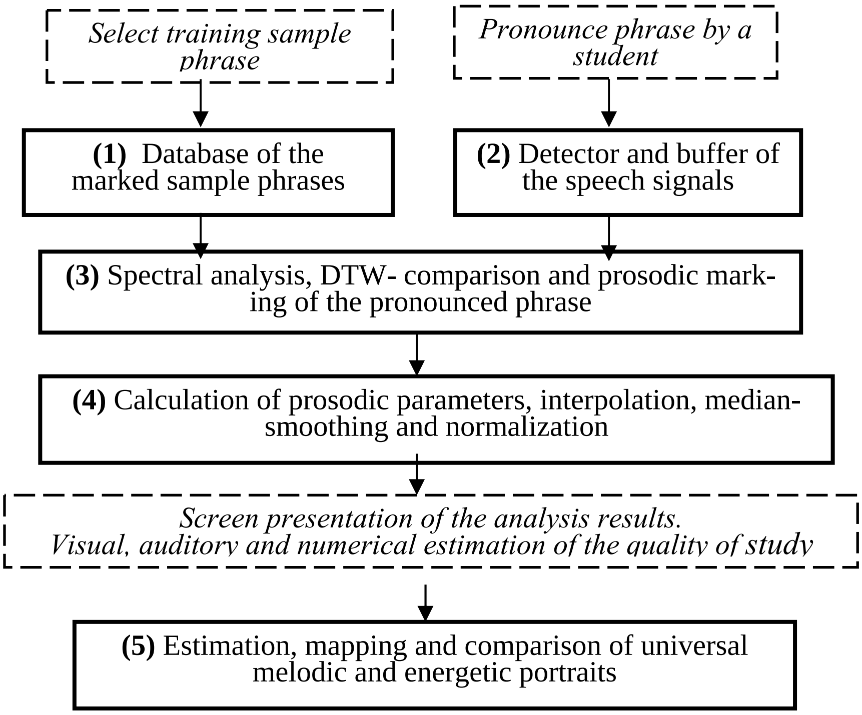

Fig. 3. The diagram of the computer teaching system of foreign speech intonation

Block 1 contains the database of sample phrases with different intonation patterns which is compiled from multimedia textbooks (see, for example, [6] for Russian language, or [7] – for English). Every sample phrase has preliminary placed prosodic marks that include phrase boundaries and placement of its nucleus.

Based on a given goal of intonation learning, a student chooses the needed sample phrase, hears it and pronounces it. The pronounced phrase is recorded on the buffer (block 2).

In block 3, the signals from both sample and pronounced phrases are spectrum analyzed and compared using the method of dynamic time warping (DTW). This is accompanied by a transfer of prosodic marks and labeling of a pronounced phrase.

In block 4, prosodic phrase parameters, such as frequency of the basic tone F_0 [8] and energy of the signal A_0 are calculated. These parameters are further interpolated on the non-vocal areas, median-smoothed and normalized.

In block 5, an estimation and comparison of universal melodic and energetic portraits are produced.

Figure 4 presents illustration of system’s output for the interrogative phrase: «Did Sasha eat the porridge»? The image shows successive processing F_0 (t) and A_0 (t) and a comparison of the sample phrase and the student-spoken phrase speech signals.

CONCLUSION

Application of an intonation mapping analyzer as a part of a speech recognition system is expected to increase reliability of recognition through the prominence of accented words and intonational segmentation of a speech flow. Intonation analysis will also be helpful for subsystems of identification of individual and emotional factors of speaker’s speech. The use of intonation system in speech synthesis systems will give an opportunity to improve intonational prominence of synthesized speech so that it will positively affect listener’s comprehension.

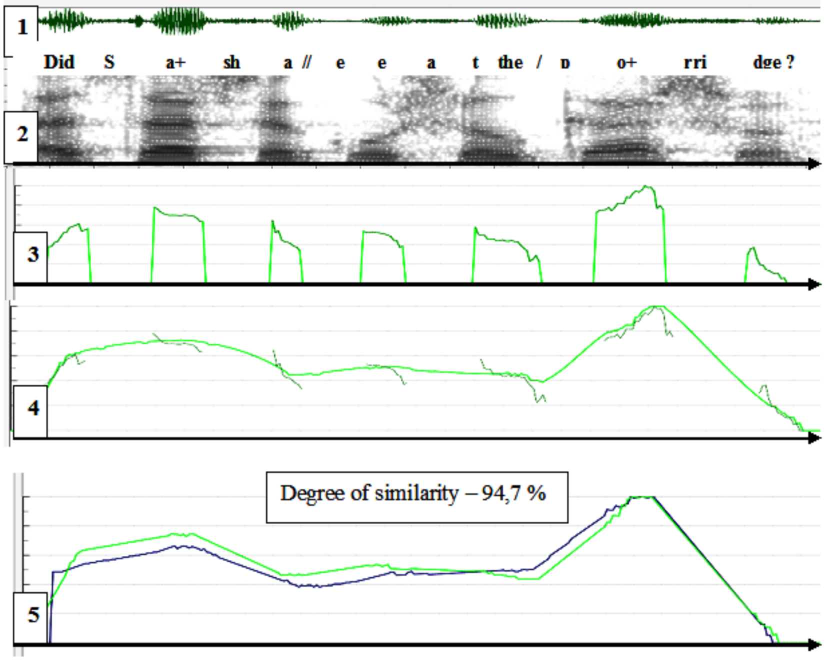

Fig. 4. The illustration of speech signals processing: *1*– oscillogram; *2* – spectrum; *3* – $F_0(t)$ (original); *4* – $F_0(t)$ (after interpolation and median smoothing); *5* – comparison of two melodic curves $F_0(t)$ – sample and spoken phrases

We believe, that there is a great potential in both domestic and international markets for a new and innovative product such as the proposed computer system for intonation training integrated into a foreign language educational courseware. To our knowledge, there is no satisfactory software available for such teaching system and, therefore, such system appears to be of great relevance. In presented work, we describe the design and the output of the computer-based system for analysis and interpretation of speech intonation.

LITERATURE

1. Language and speaker specific implementation of intonation contours in multilingual TTS synthesis / B. Lobanov [et al.] // Speech Prosody : proceedings of the 3-rd International conference, Dresden, Germany, May 2–5, 2006. Dresden, 2006. Vol. 2. P. 553–556.
2. Lobanov B., Okut T. Universal Melodic Portraits of Intonation Patterns of Russian Speech // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference «Dialogue». Iss. 13 (20). M. : RSHU, 2014. P. 330–339.
3. Lobanov B. Comparison of Melodic Portraits of English and Russian Dialogic Phrases // Computational Linguistics and Intellectual Technologies: Papers from the Annual Inernational Conference «Dialogue». Iss. 15 (22). M. : RSHU, 2016. P. 382–392.
4. Bryzgunova E. Intonation // Russian grammar // Science. M., 1982. P. 96–122.
5. Prosynth: an integrated prosodic approach to device-independent, natural-sounding speech synthesis / R. Ogden [at al.] // Computer Language and Science. 2000. P. 177–210.
6. Odintsova I. Sounds. Rhythmic. Intonation // Flinta-Science. M., 2011.
7. Ockenden M. Situational Dialogues // The English Centre, Eastbourne ; Revised Edition. Longman, 2005.
8. Shimamura T., Kobayashi H. Weighted Autocorrelation for Pitch Extraction of Noisy Speech // IEEE Transactions on Speech and AudioProcessing. 2001. Vol. 9. P. 727–730.

Научное издание

МЕЖДУНАРОДНЫЙ КОНГРЕСС ПО ИНФОРМАТИКЕ: ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Материалы международного научного конгресса

**Республика Беларусь,
Минск, 24–27 октября 2016 г.**

INTERNATIONAL CONGRESS ON COMPUTER SCIENCE: INFORMATION SYSTEMS AND TECHNOLOGIES

Proceedings of the International Scientific Congress

**Republic of Belarus
Minsk, October 24–27, 2016**

На русском и английском языках

В авторской редакции

Корректоры *Е. И. Кожушко, Е. В. Демидова, Е. В. Гордейко*
Компьютерная верстка *И. В. Остаповец*

Ответственный за выпуск *Е. А. Логвинович*

Электронный ресурс 32,6 Мб

Белорусский государственный университет.
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/270 от 03.04.2014.
Пр. Независимости, 4, 220030, Минск.