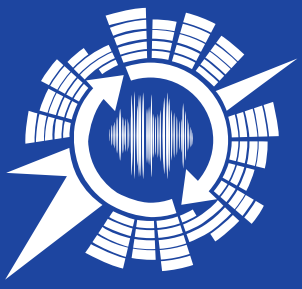
Alexey Karpov · Oliver Jokisch
Rodmonga Potapova (Eds.)

LNAI 11096

Speech and Computer

20th International Conference, SPECOM 2018
Leipzig, Germany, September 18–22, 2018
Proceedings



Springer

Alexey Karpov · Oliver Jokisch
Rodmonga Potapova (Eds.)

Speech and Computer

20th International Conference, SPECOM 2018 Leipzig, Germany, September 18–22, 2018 Proceedings

Springer

Editors
Alexey Karpov
SPIIRAS
St. Petersburg
Russia

Rodmonga Potapova
Moscow State Linguistic University
Moscow
Russia

Oliver Jokisch
Leipzig University of Telecommunications
Leipzig
Germany

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-99578-6 ISBN 978-3-319-99579-3 (eBook)
https://doi.org/10.1007/978-3-319-99579-3

Library of Congress Control Number: 2018952051

LNCS Sublibrary: SL7 – Artificial Intelligence

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

A Prototype of the Software System for Study, Training and Analysis of Speech Intonation

Boris Lobanov(✉), Vladimir Zhitko, and Vadim Zahariev

The United Institute of Informatics Problems of National Academy of Sciences of Belarus, Minsk, Belarus
lobanov@newman.bas-net.by, zhitko.vladimir@gmail.com

Abstract. A prototype of the software system presented in the paper is designed to train learners in producing a variety of recurring intonation patterns of speech. The system is based on comparing the melodic (tonal) portraits of a reference phrase and a phrase spoken by the learner and involves active learner-system interaction. The main algorithms used in the training system proposed for analyzing and comparing intonation features are considered. A set of examples of reference sentences is given which represents the basic intonation patterns of Russian, British English, American English, German and Chinese speech.

Keywords: Speech intonation · Melodic portrait · Intonation analysis
Intonation training · Pitch visualization · Language learning
Intonation assessment

1 Introduction

In world practice, Computer Assisted Language Learning (CALL) is a widespread interdisciplinary field which also includes a subarea for teaching foreign language pronunciation using speech technologies. Teaching experience manifested that a foreign accent is very noticeable in intonation and therefore it deserves our close attention. Intonation is the most ephemeral component of oral speech. Deviations in this area can lead to serious semantic changes, as well as create a wrong impression of the speaker's personality. It should be noted that intonation is especially important when studying Chinese and other tonal languages.

Accuracy in using intonation patterns when speaking and adequacy of their perception when listening is difficult for students to control (especially if they don't have a musical ear). The existing language courses and equipment provide only auditory feedback for monitoring the precision of speech intonation, which is clearly not enough.

The proposed computer trainer provides additional visual feedback, as well as a quantitative estimation of the correctness of speech intonation in the process of teaching various foreign languages [1, 2] or for an instrumental evaluation of the intonation quality of synthesized speech [3].

To date, on our website (see https://intontrainer.by) there is a prototype of the software system "IntonTrainer", available for free download. It is focused on learning the intonation of foreign languages. A set of reference sentences is given which represents

A. Karpov et al. (Eds.): SPECOM 2018, LNAI 11096, pp. 337–346, 2018.
https://doi.org/10.1007/978-3-319-99579-3_36

the basic intonation patterns of Russian, British English, American English, German and Chinese speech and their main varieties. The software package "IntonTrainer" is carried out analysis and representation on the screen of the tone pattern and spoken phrasal intonation, as well as their comparison and estimation of intonation similarity. Estimation of intonation similarity is carried out on the basis of representation of intonation in the form of universal melodic portraits (UMP) [4, 5]. Software realization of the system is written on C and C++ programming code by using Qt framework. It can be compiled under Windows platform (from XP to 10 versions), as well as under Linux platform.

The purpose of this article is to give a detailed description of the structure and features of the practical use of the developed system.

2 Graphical User Interface

The initial "IntonTrainer" window (it opens after the "IntonTrainer" is started) is shown in Fig. 1. After clicking the "**Start**" button, the main window opens (Fig. 2) containing a structured list of the phrases with different tone patterns (TP): Examples, Comparizon, Ussage and others.

Fig. 1. The initial window.

Fig. 2. Main window.

Before you start, you can preview the "IntonTrainer" **Settings** (top right corner of the initial window) and change them if necessary. The main **Settings** window is shown in Fig. 3.

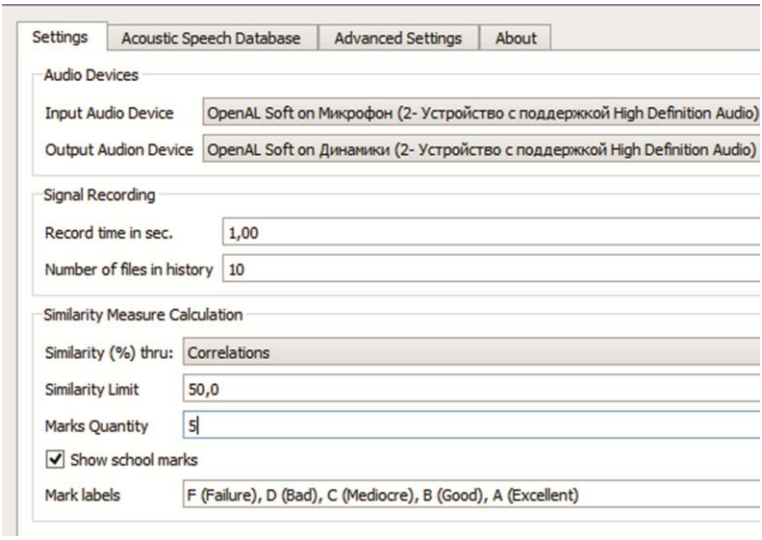


Fig. 3. The main Settings window.

In this window the user can select:

- In the **Audio** Devices section, the user can select the type of audio devices used.
- In the section **Signal Recording**, you can select the length of the recording of the signal from the microphone – **Record time in sec**. In this case, recording takes place within N seconds + the duration of the selected phrase pattern. In addition, it is possible to save in the folder "RECORDS" the specified number of phrases recorded from the microphone – **Number of files in history**.
- In the section, **Similarity Measure Calculation**, you can choose a method for assessing the similarity of the intonation of a pronounced phrase with a reference phrase, determined by one of three methods. Namely, by computing – **Similarity (%) thru**:
  a. cross-correlation – **Correlations**;
  b. average value of the mutual distance – **Average Distance**;
  c. maximum of the local distance – **Maximum Local Distance**;
  d. average value of the three above-mentioned similarity measures – **Average**.

The chosen method of calculating the intonational similarity is then used in calculating the school assessment of the intonational quality of the spoken phrase. To do this, a checkmark is made in the **Show school marks** small window. For the selected method, a **Similarity limit** is defined, which corresponds to the worst scoring score. The total number of points used to assess the intonation quality is set by the desired number of them – **Marks Quantity**. By filling in the section – **Mark labels** – digital and verbal names of points are specified.

Two additional settings – **Acoustic Speech Database** and – **Advanced settings** – in the main window (Fig. 3) are for settings made by developers or "advanced" users of the "IntonTrainer" (recommendations on their use will be given in Sects. 5 and 6 below).

The **About** button opens a window with information about the developers and a number of other information.

3 Primary Study of the Basic Tone Patterns

By scrolling the page of the **Main window** (Fig. 2) from top to button, the user is given to see the examples for the main tone patterns of English speech. For instance, by choosing the following directory:

1 – UMPs EXAMPLES – TP1 (Falling) **(M) [We+ll.wav]**, you will open the window in which the results of the intonation analysis of this phrase are displayed graphically (Fig. 4).

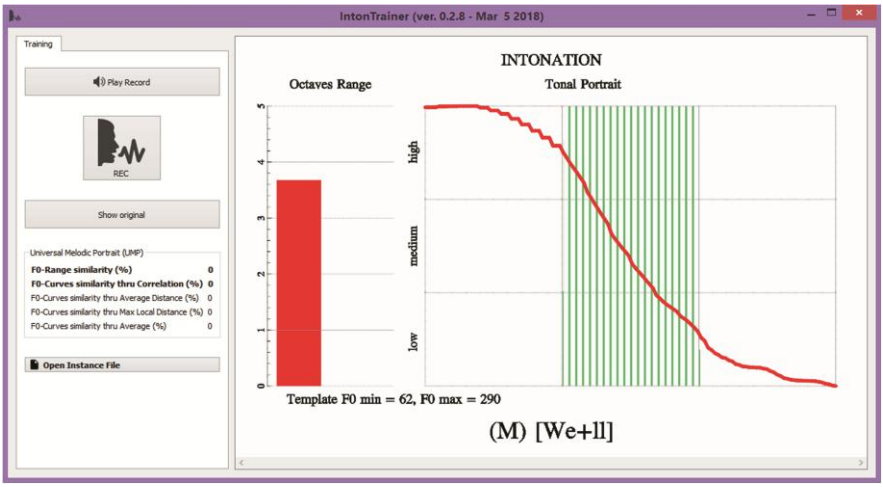


Fig. 4. Analysis results window. (Color figure online)

In Fig. 4 the red column on the left shows the range of pitch change, expressed in octaves. **Octaves Range** = (F0max/F0 min) – 1.

On the right, a linear graph of the UMP (**Tonal Portrait**) is displayed in red, the nucleus of which is marked with frequent vertical lines. Below the graphs, the minimum and maximum values of F0 for the selected phrase are listed, as well as the text of the phrase in which the nuclear vowel is indicated by the "+" sign.

To listen to the selected phrase click the "**Play Record**" button.

By scrolling the page of the **Main window** (Fig. 2) from top to button, the user is given the opportunity to see a lot of examples for the main tone patterns of English speech.

The set of examples provides audio and visual representation of the main tone patterns (TP1-TP3) presented in the form of UMPs, pairwise comparison of different TPs, shows peculiarities of TP usage, as well as TP actualization in dialogues, prose and poetry.

4 Individual Intonation Training for Correct Pronunciation

When using the "**IntonTrainer"** for individual intonation training the user must use an extended or built-in microphone.

In this case, the user should press the "**Rec**" button, wait for a short "beep-signal" and pronounce the phrase into the microphone. The text of the phrase is indicated in the lower part of the window (see: Fig. 4). After recording (to the "**RECORDS**" folder) and processing of the entered speech signal, the user will hear the second "beep-signal", and the image in the graph window shown in Fig. 4 will be replaced with the image shown in Fig. 6. The upper part of the window shows the results of comparison of the reference and spoken phrases: Range (61%) - proximity to the range of changes F0 and Shape (99%) - proximity to the shape of the trajectory F0. Near the percentage of proximity assessments, estimates **school marks** can be made (if necessary), as shown in Fig. 5.

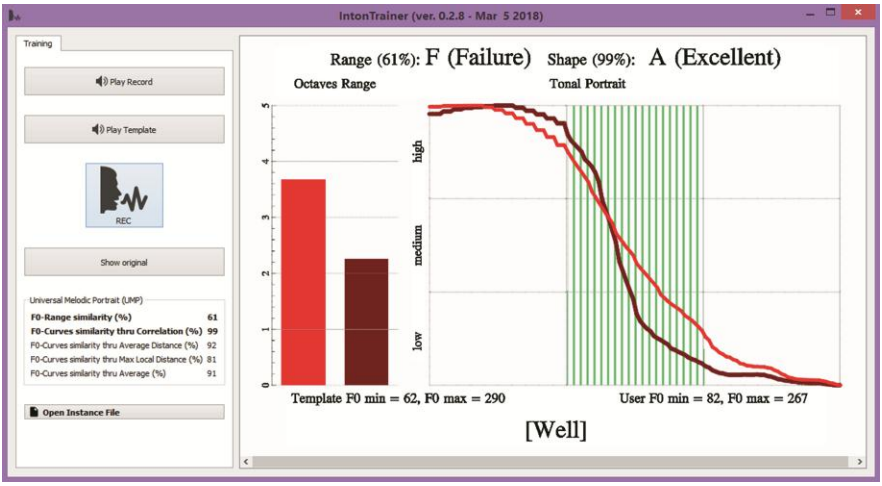


Fig. 5. The window displaying the results of analysis and comparison. (Color figure online)

In Fig. 5 the red column on the left shows the range of F0 change of the TP phrase, and the brown one – the pronounced phrase. On the right, the linear graph of the UMP of the TP is displayed in red, and the brown one of the pronounced phrase. Below the graphs are the minimum and maximum values of F0 of the TP and the pronounced phrase are given.

To listen to the selected TP click the "**Play Template**" button and to the pronounced phrase click the "**Play Record**" button (Fig. 5).

5 Comparison of the Phrase Intonation from Various Sources

As already mentioned above, the "IntonTrainer" can also be used as an instrument in a number of scientific and practical studies. For example, the "IntonTrainer" can be

successfully used in experimental phonetic studies, during which it becomes necessary to compare the reference intonation with the intonation of the phrases studied from various sources (for example, when comparing the intonation of natural and synthesized speech). In this case, instead of using an external or built-in microphone, the "**Open Instance File**" button in the left section of the window is used. When using this button, a speech signal of the same content is selected from a specially created "TEST" folder, but obtained from another source, for example, from a speech synthesizer. The graphical and numerical result of the UMP comparison of the natural (red line) and synthesized (brown line) phrase "*Are You from Germany?*" is shown on Fig. 6.

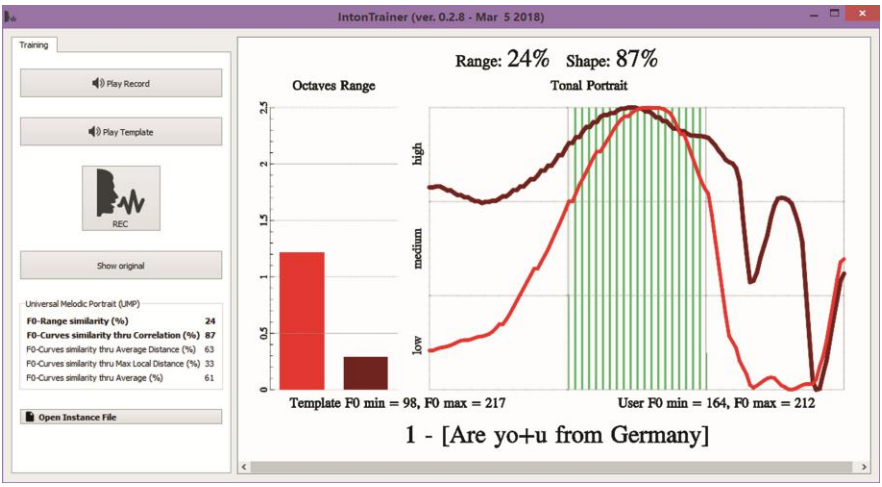


Fig. 6. The results of the UMP comparison of the natural and synthesized phrase. (Color figure online)

In the upper part of the window the results of comparison of the natural and synthesized phrase are shown: **Range** (24%) – proximity to the range of changes F0 and **Shape** (87%) – proximity to the shape of the trajectory F0.

6 Prosodic Marking of Reference Phrases of the Acoustic Database

An important factor in the formation of the acoustic database of the studied phrases is their prosodic marking by the areas (regions) of pre-nucleus, nucleus and post-nucleus. Currently this operation is performed manually using the standard application "**Sound Forge**", but in the future it is supposed to be automated.

The speech signal of the phrase is recorded in a "wav" format with a sampling of 8 kHz, 16 bits and is labeled into regions P1 (pre-nucleus), N1 (nucleus), T1 (post nucleus) as shown in Fig. 7 for a single-nucleus (mono-accented) phrase: "***I am saying till fi+ve.***", pronounced by a male voice. The result of the construction of the UMP of this phrase, obtained on the basis of such a markup, is shown in Fig. 8.

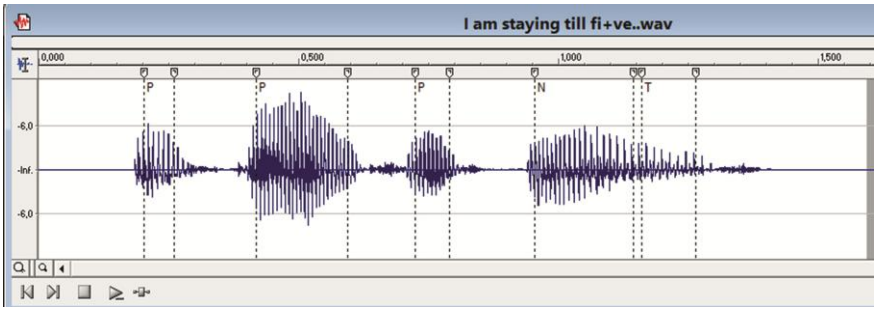


Fig. 7. Example of mono-accented phrase "I am saying till fi+ve."

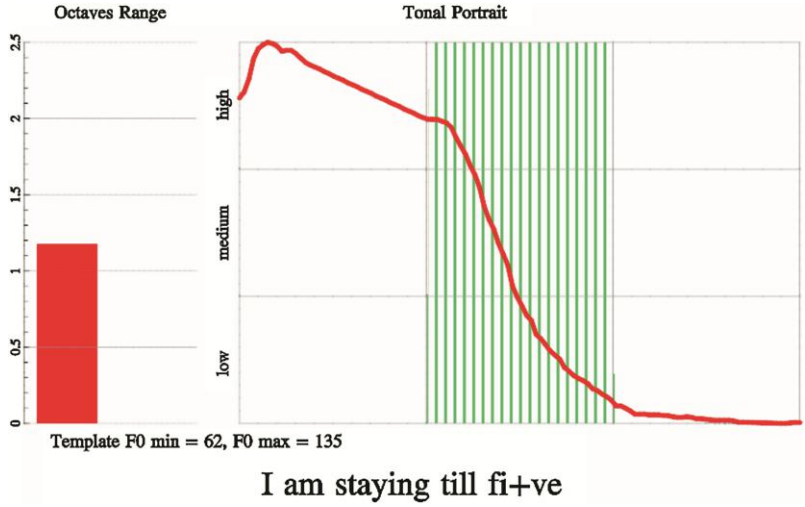


Fig. 8. The result of the construction of the UMP phrase: "I am saying till fi+ve."

In Fig. 9 an example of marking of two-accented phrase: "***Befo+re you open the do +or, …***" is shown. The result of the construction of the UMP of this phrase, obtained on the basis of such a markup, is shown in Fig. 10.

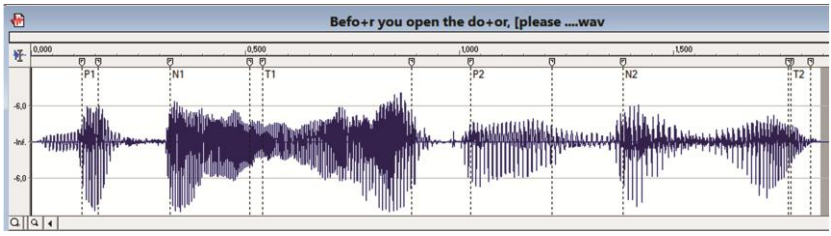


Fig. 9. Example of marking of two-accented phrase: "Befo+re you open the do+or, …"

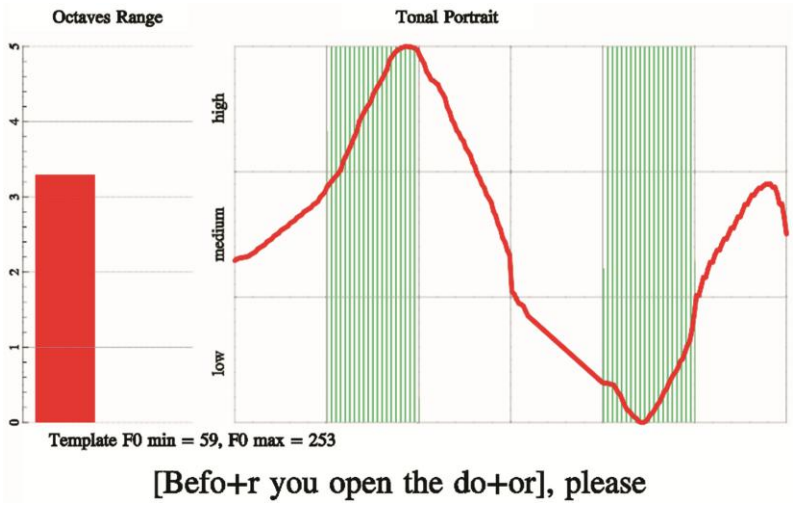


Fig. 10. The result of the construction of the UMP phrase: "Befo+re you open the do+or, …".

If, for some reason, it is difficult to determine the position of the elements of the accent structure of the phrase (P, N, T), then all the voice regions of the phrase can be assigned one the same index N. In this case, it is considered that each of the voice regions represent as nuclear. For example, in Fig. 11 shows the UMP phrase "***Are Yo+u from Germany?***" built in the presence of markup on P, N, T – regions, and in Fig. 12 shows the trajectory of F0 in the case when each of the voice regions of the phrase is assigned the same index N. Speech signal of the phrase was created by selecting the **Mark Out File** in the **Advanced Settings section** – "**Acoustic Speech Database**".

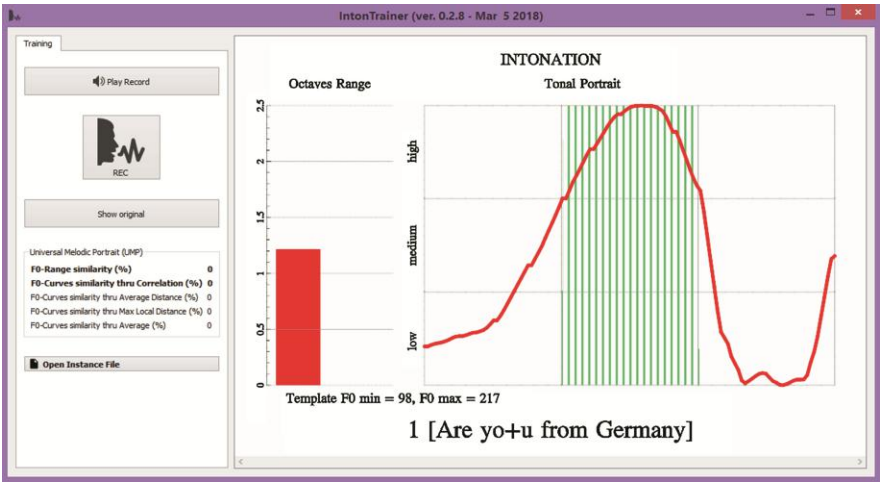


Fig. 11. Example of displaying the trajectory F0 (UMP) in the presence of markup phrases on P, N, T – regions.

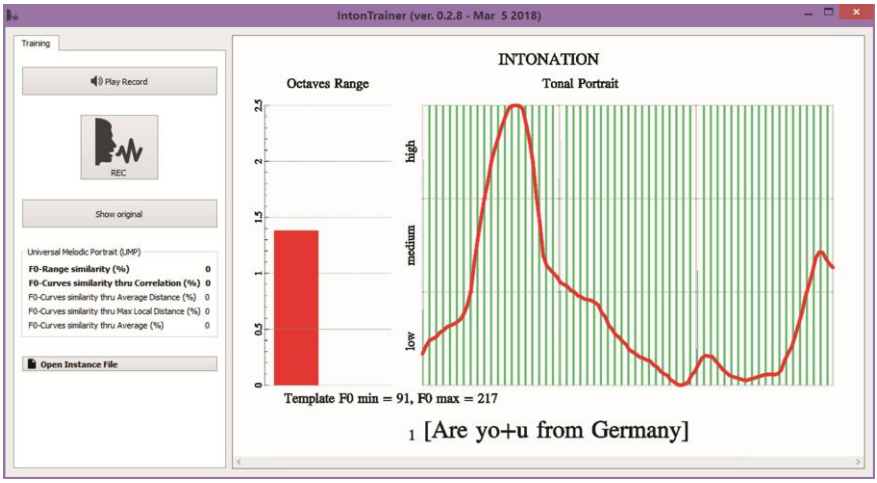


Fig. 12. An example of mapping the trajectory F0 in the case when each of the voice sections of the phrase is given the same index N.

Note that the automation of the process of marking phrases to voice regions is constantly improving. In the final analysis, it is assumed that the process of determining the position of each of the elements of the accent structure of the phrase is complete: P1 (pre-nucleus), N1 (nucleus), T1 (post nucleus).

7 Conclusions

At present, using the "IntonTrainer" system, experiments are conducted to learn by students the intonation of Russian and English. Preliminary results indicate a significant effectiveness of its use.

The software package is recommended for use in the following popular fields:

- In linguistic education used as a means of visualizing intonation.

Primary introduction and study of the basic tone patterns (TPs) of oral speech their pairwise comparisons, peculiarities of their usage as well as their actualization in dialogues prose and poetry.

- In individual intonation training for correct pronunciation used as a means of feedback.

Individual training for correct pronunciation of TPs when studying a foreign language or improving intonation skills of one's native language in some professions: call center operators, radio and TV announcers, etc.

- In scientific and practical research used as a means of comparing intonation from different sources.

Study of individual, emotional and stylistic features of intonation. Comparative evaluation of speech intonation in norm and pathology. Estimation of the intonational quality of synthesized speech.

References

1. Lobanov, B., Karnevskaya, Y., Zhitko, V.: On a way to the computer aided speech intonation training. In: Karpov, A., Potapova, R., Mporas, I. (eds.) SPECOM 2017. LNCS (LNAI), vol. 10458, pp. 582–592. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66429-3_58
2. Lobanov, B.M, Zhitko, V.A., Kharlamov, A.A.: A computer system of teaching intonation patterns of Russian speech. In: Proceedings of the International Conference "Dialog 2017» , Moscow, pp. 287–302 (2017)
3. Lobanov, B.M., Solomenik, A., Zhitko, V.: An experience of the objective estimation of intonation quality of the synthesized Russian speech. In: Proceedings of the International Conference "Dialog 2018» , Moscow (2018)
4. Lobanov, B., Okrut, T.: Universal melodic portraits of intonation patterns of Russian speech. In: Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", vol. 13(20), Moscow, pp. 330–339 (2014). (in Russian)
5. Lobanov, B.M.: Comparison of melodic portraits of English and Russian dialogic phrases. In: Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", vol. 15(22), Moscow, pp. 382–392 (2016)