

= Inton[@]Trainer =

SPEECH INTONATION ANALIZER AND TRAINER: USER GUIDE (On example of British English)

1. Purpose

The software package “Inton@Trainer” (hereinafter, “Application”) is intended for analysis and representation on the screen of the pattern and spoken phrasal intonation (F0 Fundamental Frequency trajectory), as well as for 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).

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.

2. Graphical User Interface

The initial Application window (it opens after the application is started) is shown in Fig. 1.

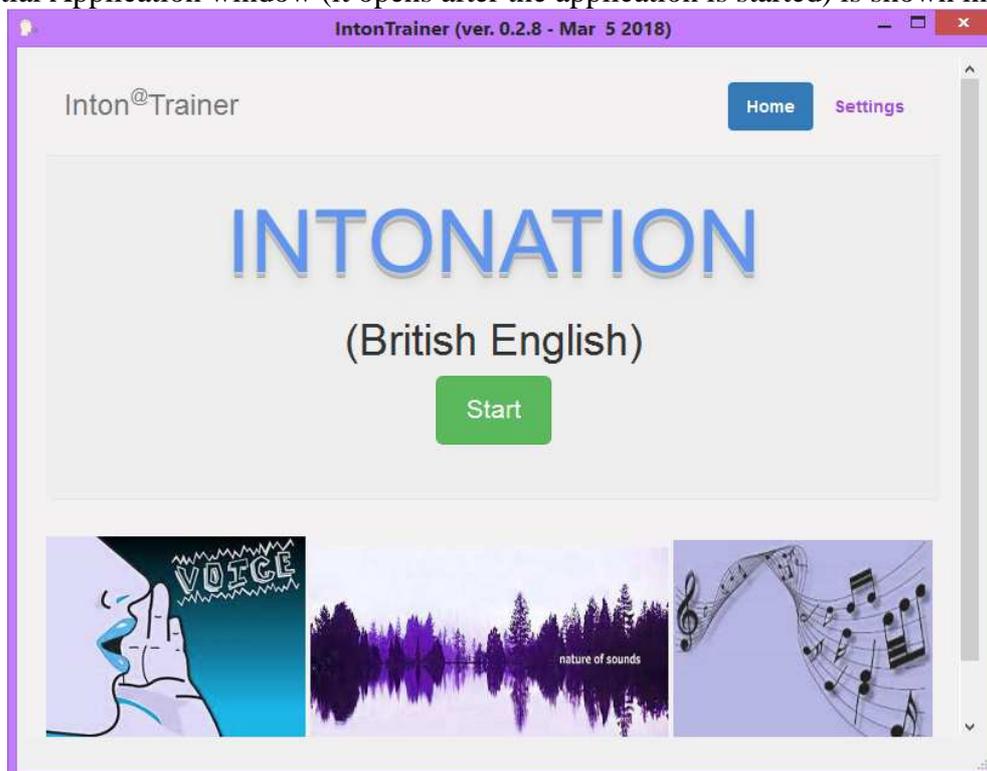


Fig. 1 The initial window of the Application

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. 2. Main window

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

2

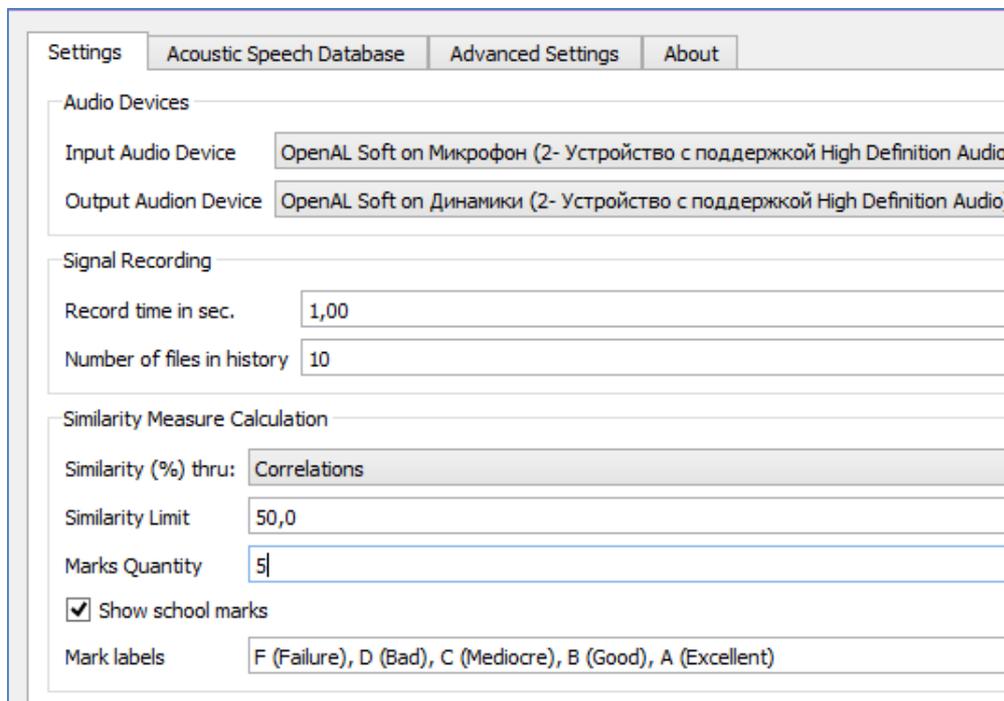


Fig. 3. The main Settings window

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:**

- 1) cross-correlation - **Correlations**,
- 2) average value of the mutual distance - **Average Distance**,
- 3) maximum of the local distance - **Maximum Local Distance**,
- 4) 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 (Figure 3) are for settings made by developers or "advanced" users of the application (recommendations on their use will be given in Sections 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 bottom, 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 (Figure 4).

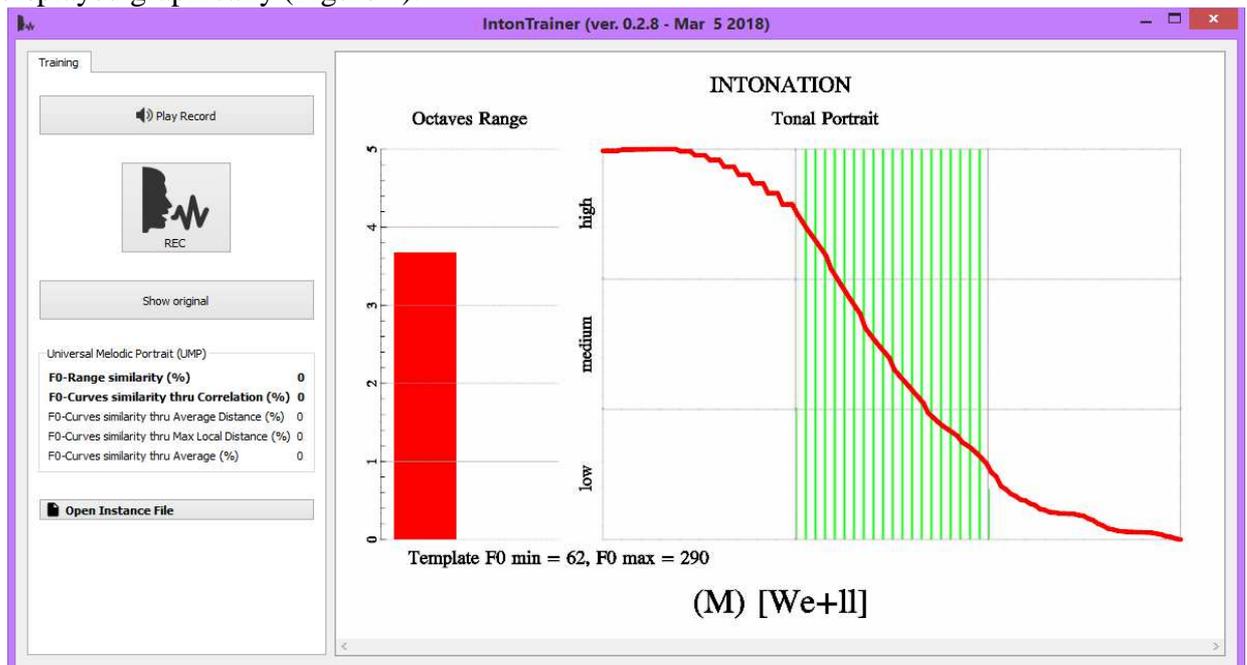


Fig. 4. Analysis results window

In Fig. 4 the red column on the left shows the range of pitch change, expressed in octaves. **Oktaves Range** = $(F0_{max} / F0_{min}) - 1$.

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.

Click “**Show Original**” button to open an additional window (Fig. 5), at the top of which you can see a waveform of the speech signal with pre-nucleus (red line), nucleus (black line) and post-nucleus (blue line) markers. In the middle part of the window you can see F0 curve in real time showing the parts of the pre-nucleus, the nucleus and post-nucleus of the UMP. The function “**Show Original**” is used only for monitoring in those rare cases when there are doubts about the correctness of the received UMP phrase.

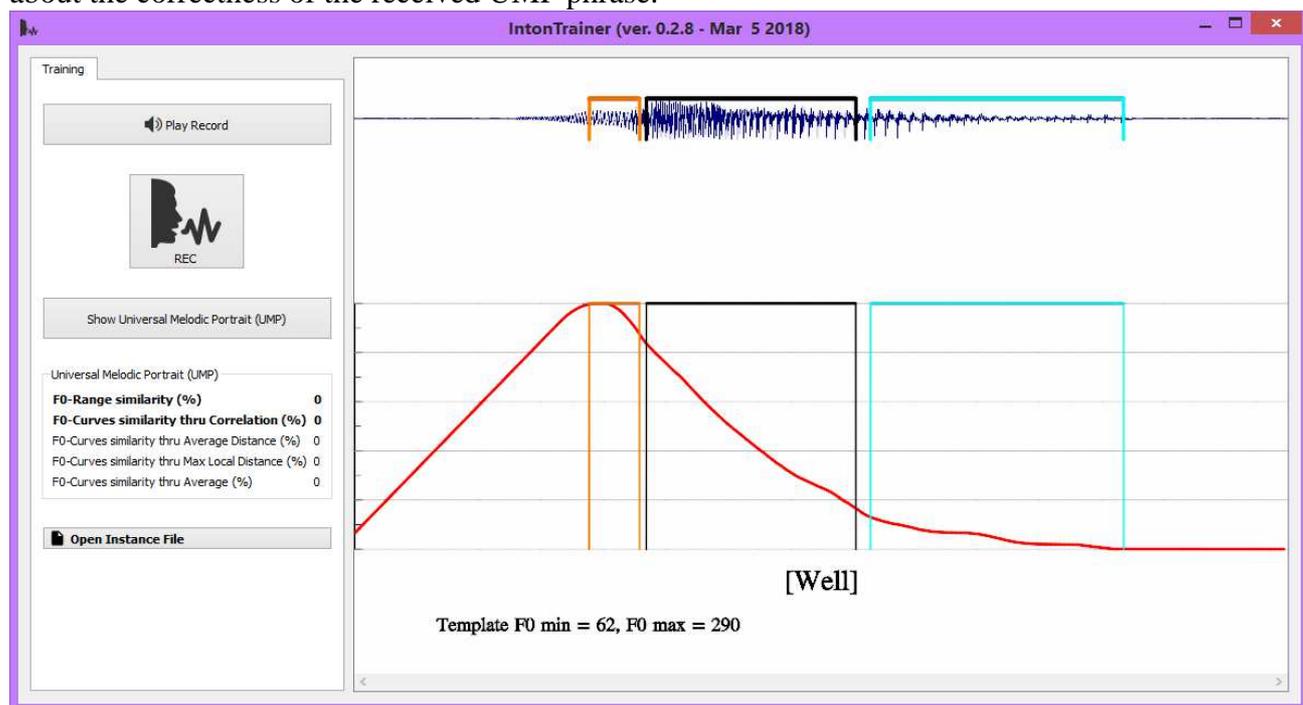


Fig. 5. Additional window displaying F0 in real time

By scrolling the page of the **Main window** (Fig. 2) from top to bottom, 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 **Application** 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. 6.

In Fig 6 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.

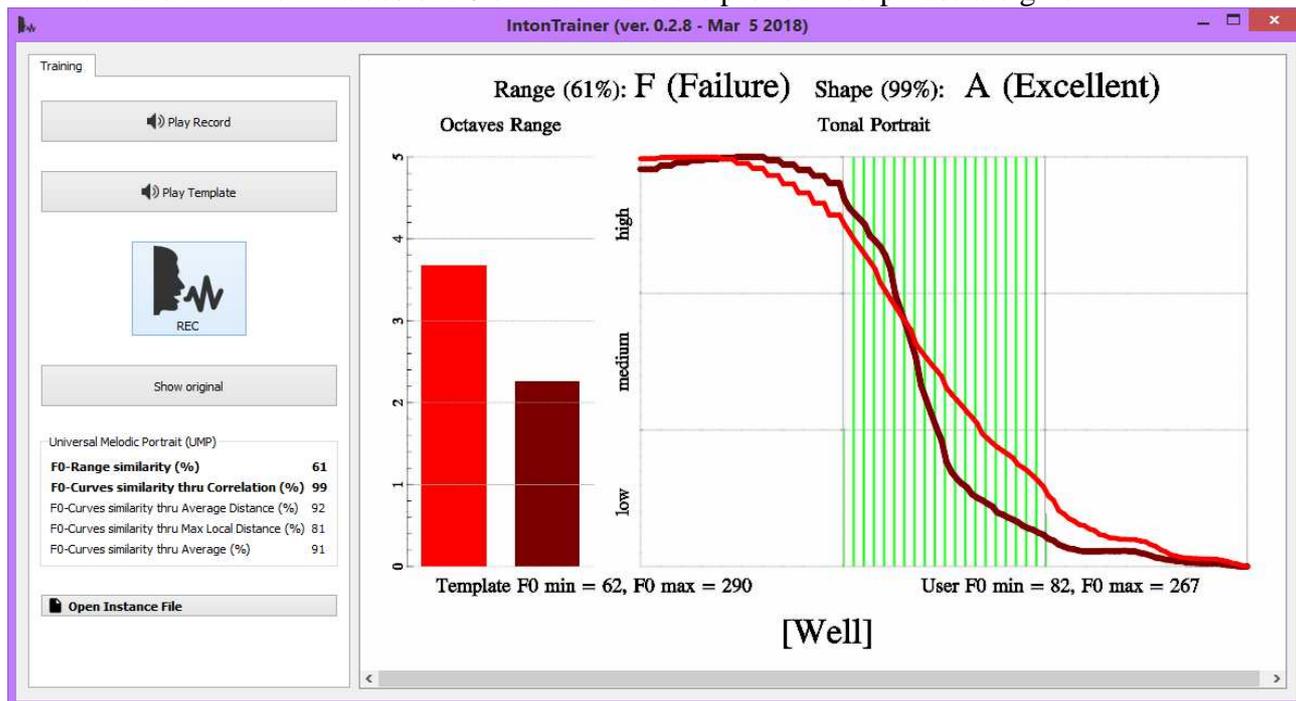


Fig. 6. The window displaying the results of analysis and comparison

To listen to the selected TP click the “**Play Template**” button and to the pronounced phrase click the “**Play Record**” button (Fig/ 7).

Click the “**Show original**” button to open an additional window (Figure 7), at the top of which are the oscillograms of both phrases with the markers of pre-nucleus (red lines), nucleus (black lines) and post-nucleus (blue lines). In the middle part of the window, the combined trajectories of the F0 change of the TP (red) and the pronounced phrase (brown) are depicted in real time showing the sections of the pre-nucleus, nucleus and post-nucleus of the UMP’s. The information displayed in this window can be useful for controlling the correctness of the transfer of pre-nucleus, nucleus, and post-nucleus markers from the TP to the pronounced phrase. Errors in the transfer of the markers can greatly distort the actual shape of the UMP.

5

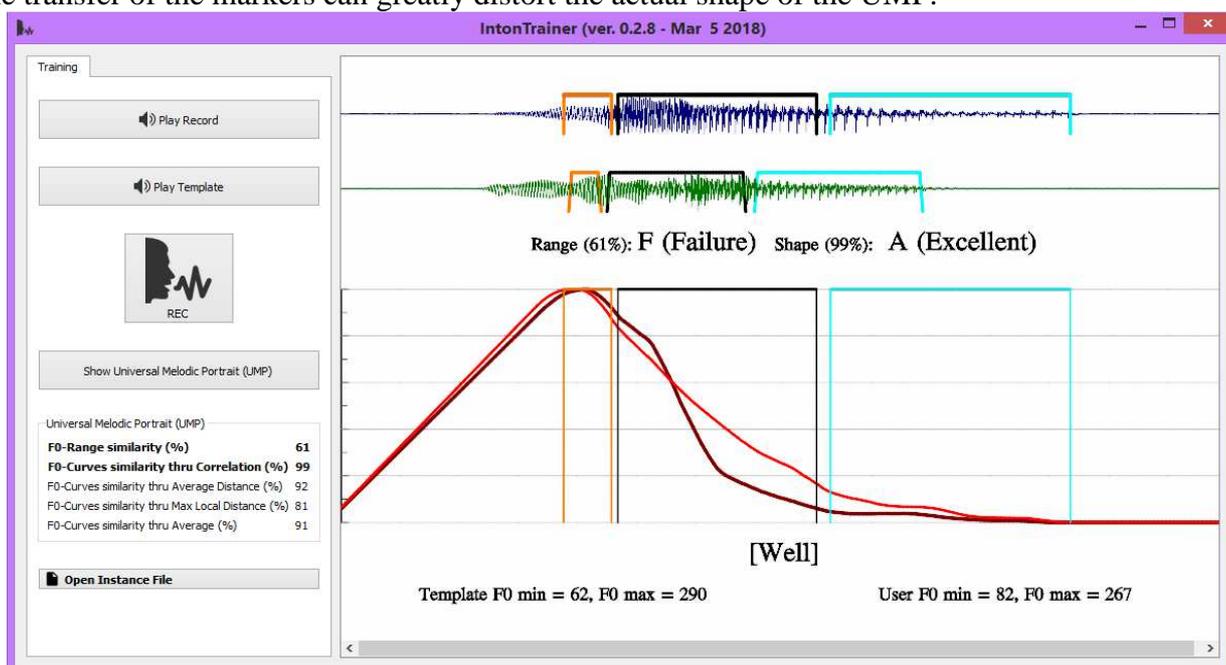


Fig. 7. An additional window for displaying the combined trajectories F0

5. Comparison of the phrase intonation from various sources

As already mentioned above, this Application can also be used as an instrument in a number of scientific and practical studies. For example, the application 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. 8.

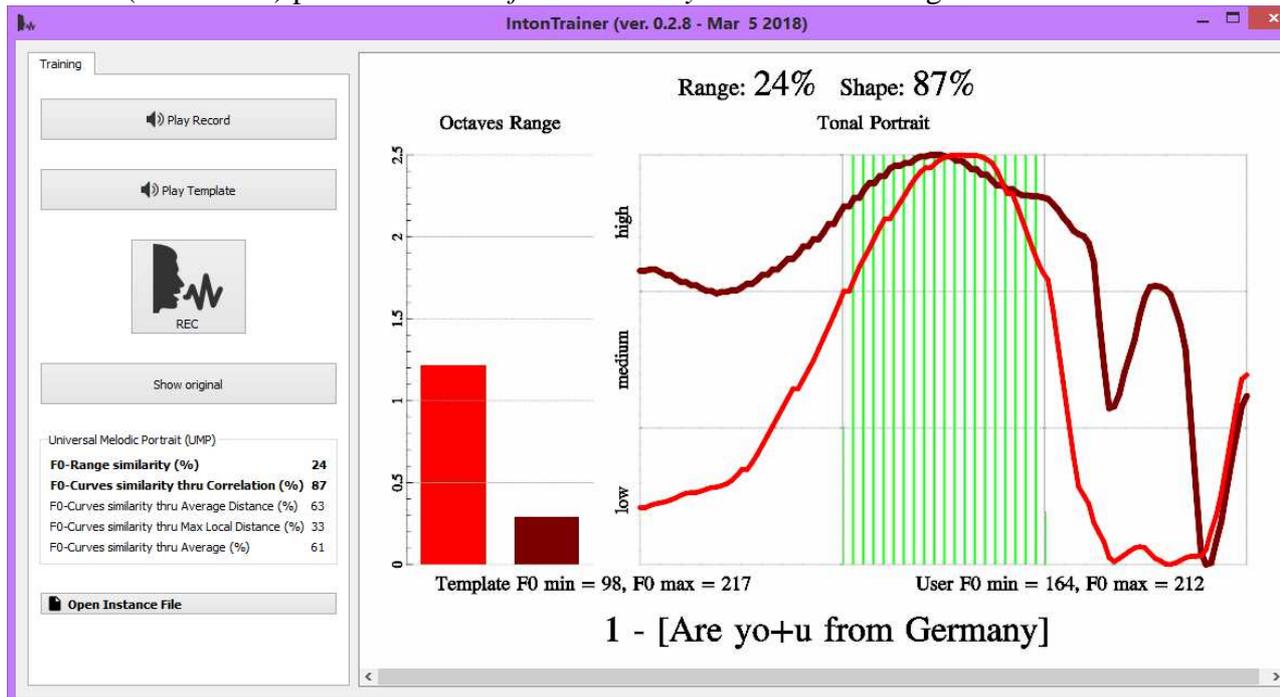


Fig. 8. The results of the UMP comparison of the natural and synthesized phrase

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. In more detail, numerical similarity estimates are given in Fig. 8 on the left side of the window, as shown below:

F0-Range similarity (%)	24
F0-Curves similarity thru Correlation (%)	87
F0-Curves similarity thru Average Distance (%)	63
F0-Curves similarity thru Max Local Distance (%)	33
F0-Curves similarity thru Average (%)	61

6. Additional settings by using the button - "Acoustic Speech Database"

The software package **"Inton@Trainer"** is an open system which allows for various modifications. First of all, this refers to the used set of reference data (patterns), which can be supplemented or formed anew in accordance with the task in hand.

When you click the **"Acoustic Speech Database"** button, a window (see Figure 8) opens, containing a structured list of reference phrases.

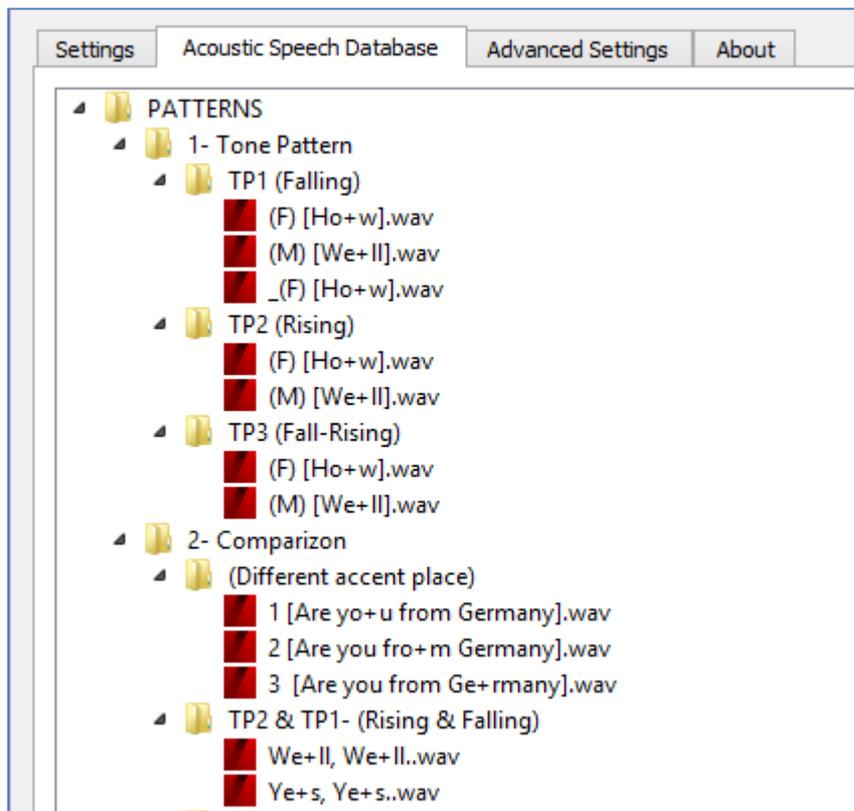


Fig. 9. Window displaying a structured list of reference phrases

Double-clicking on this list opens a file or explorer window. When you right-click a folder, the drop-down menu opens (see Figure 10):

7

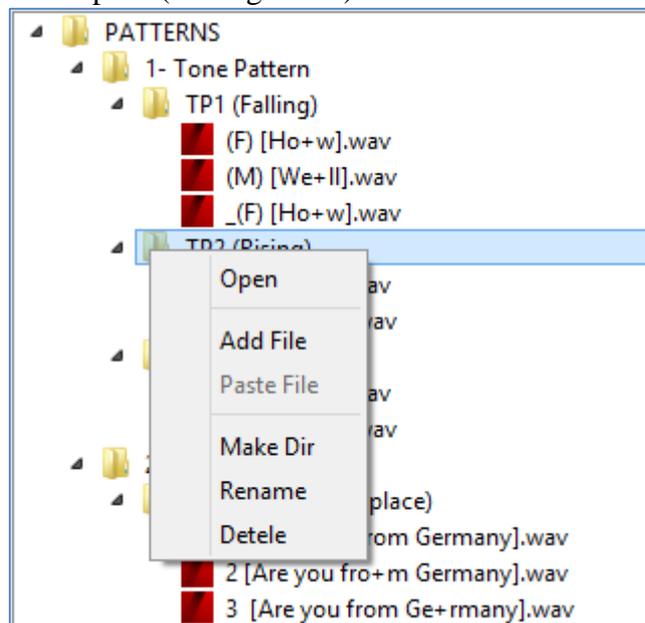


Fig. 10. Drop-down menu when right-clicking on the folder TP2

You can open the chosen folder (**Open**), add the file from the outside (**Add file**), create a new folder (**Make Dir**), rename (**Rename**) or delete (**Delete**) the folder.

When right-clicking the mouse on the file, a drop-down menu opens (see Figure 11):

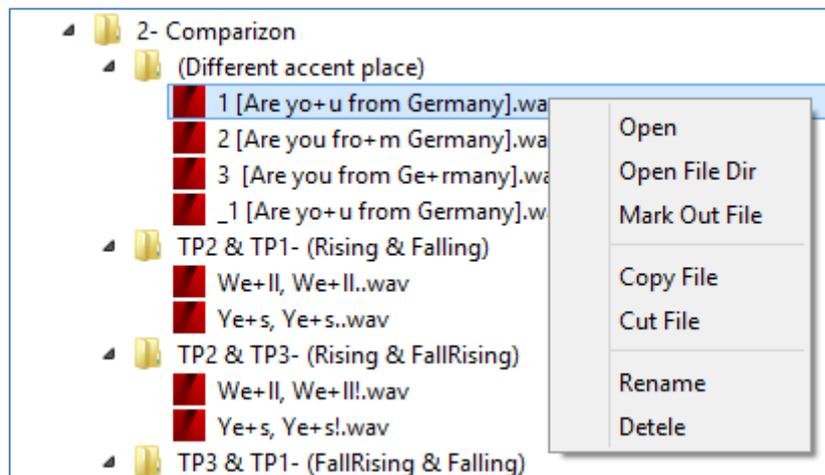


Fig. 11. Drop-down menu when right-clicking on the file: {1 [Are yo+u from Germany].wav}

In this window you can open the file (**Open**), or open the entire directory (**Open File Dir**), and you can also **copy, cut, rename** or **delete**. Thus, the set of reference data (acoustic database) used can be supplemented or formed anew in accordance with the task.

In addition, it is possible to automatically mark and design any file out the voice areas (**Mark Out File**). As a result, a new file with the same name is generated, but with the addition of the underscore to the beginning of the name: {_1 [Are yo+u from Germany].wav} (see Figure 11).

7. 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. 12 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. 13.

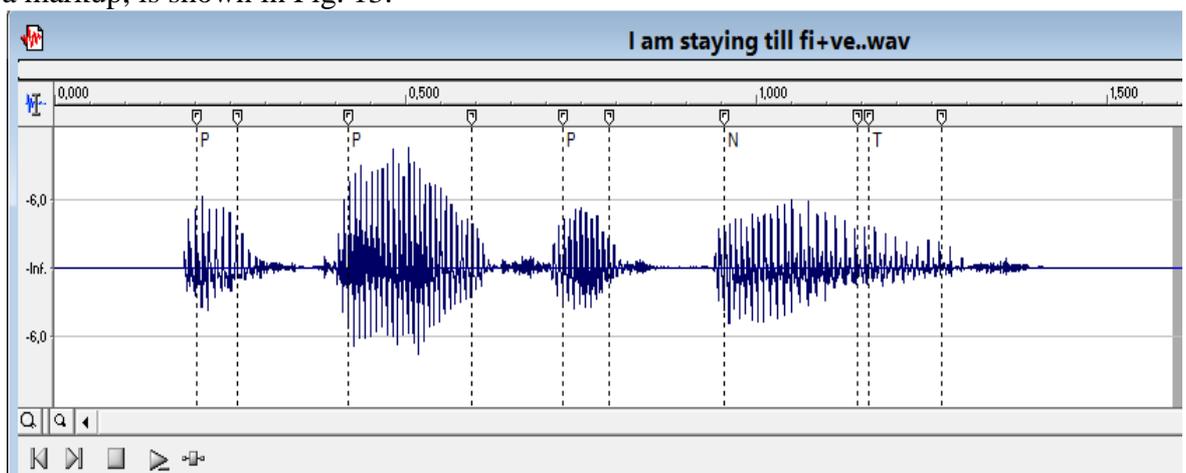


Fig. 12. Example of marking speech signal of mono-accented phrase “I am saying till fi+ve.”

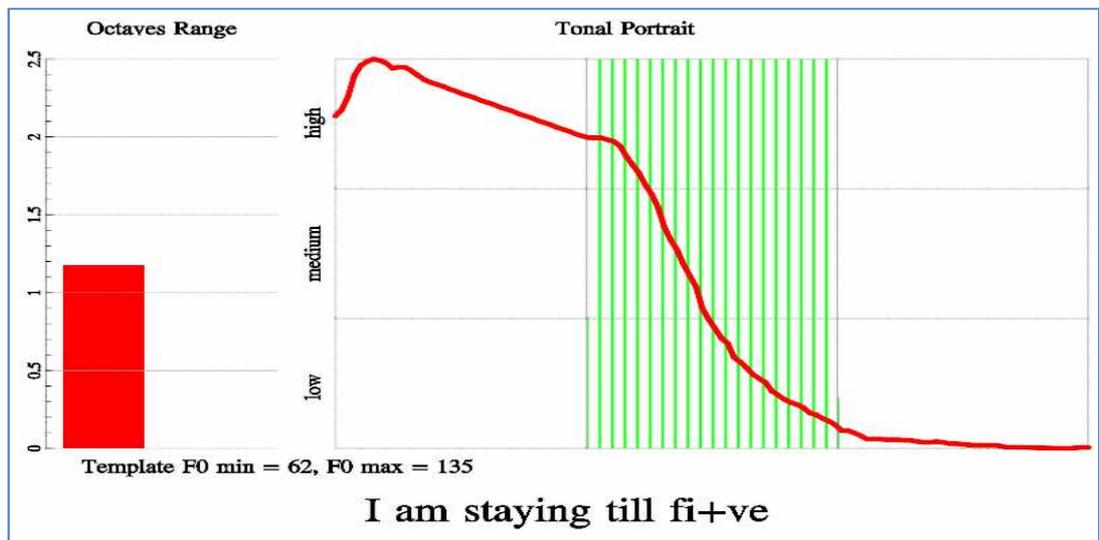


Fig. 13. The result of the construction of the UMP phrase:
“I am saying till fi+ve.”

In figure 14 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. 15.

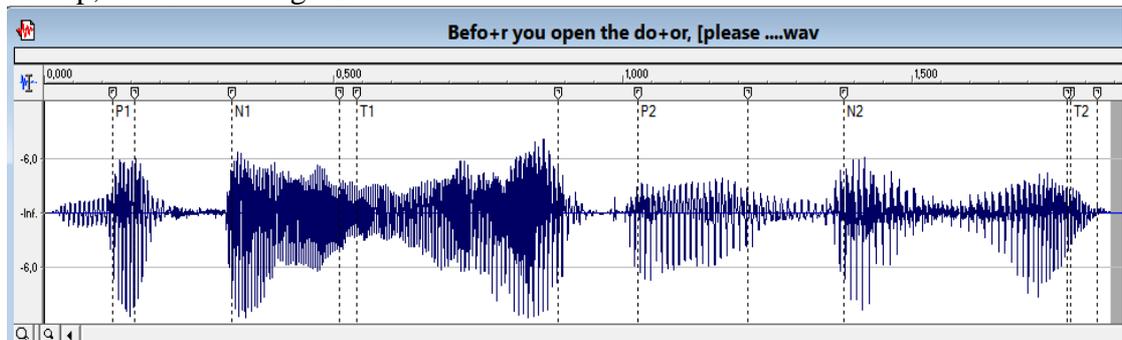


Fig. 14. Example of marking of two-accented phrase: *“Befo+re you open the do+or, ...”*

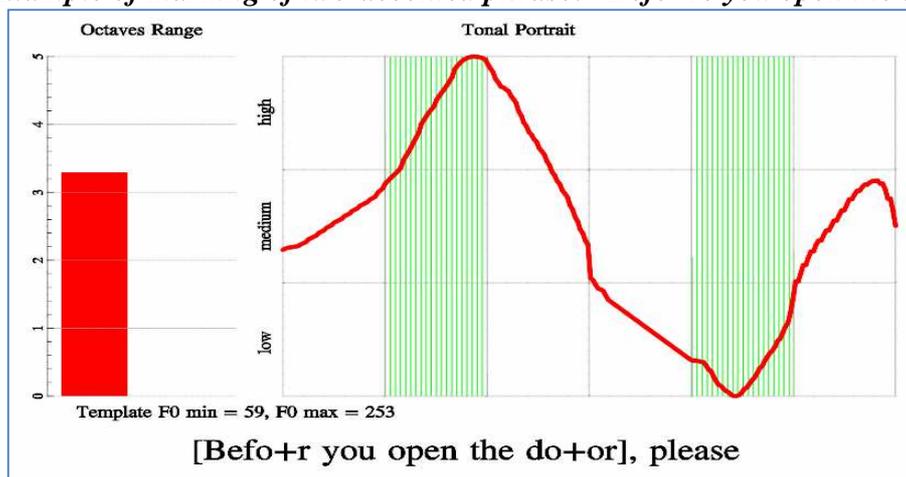


Fig. 15. The result of the construction of the UMP phrase:
“Befo+re you open the do+or, ...”

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. 16 shows the UMP phrase *“Are Yo+u from Germany?”* built in the presence of markup on P, N, T - regions, and in Fig. 17 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**" (see Figure 11).

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).

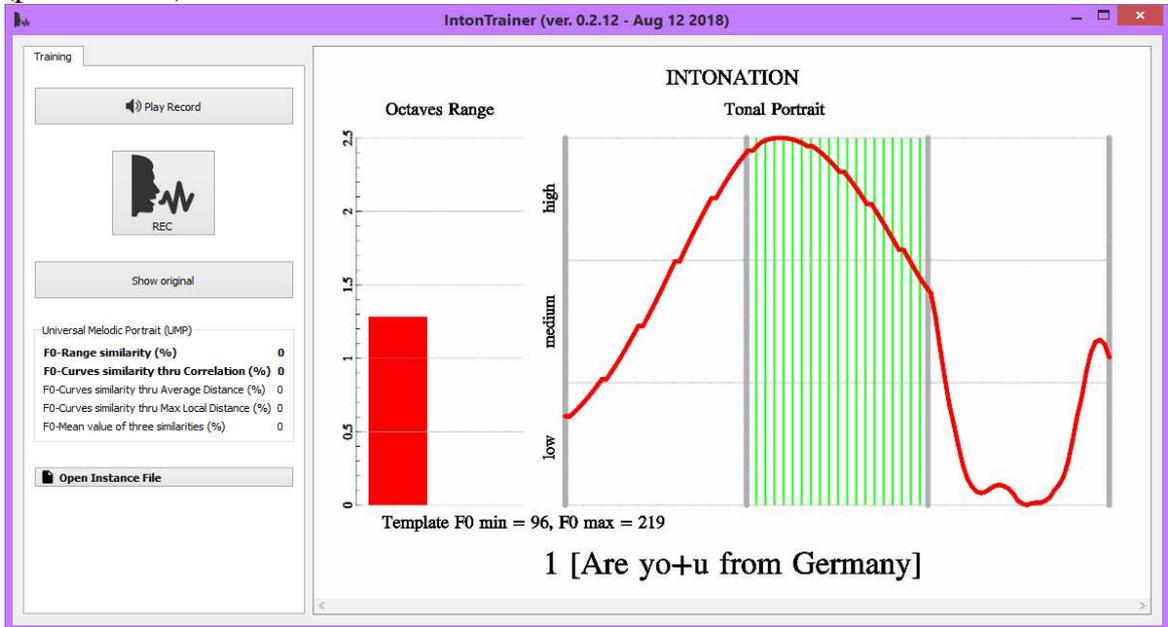


Fig. 16. An example of displaying the trajectory F0 (UMP) in the presence of markup phrases by hand on the P, N, T - regions

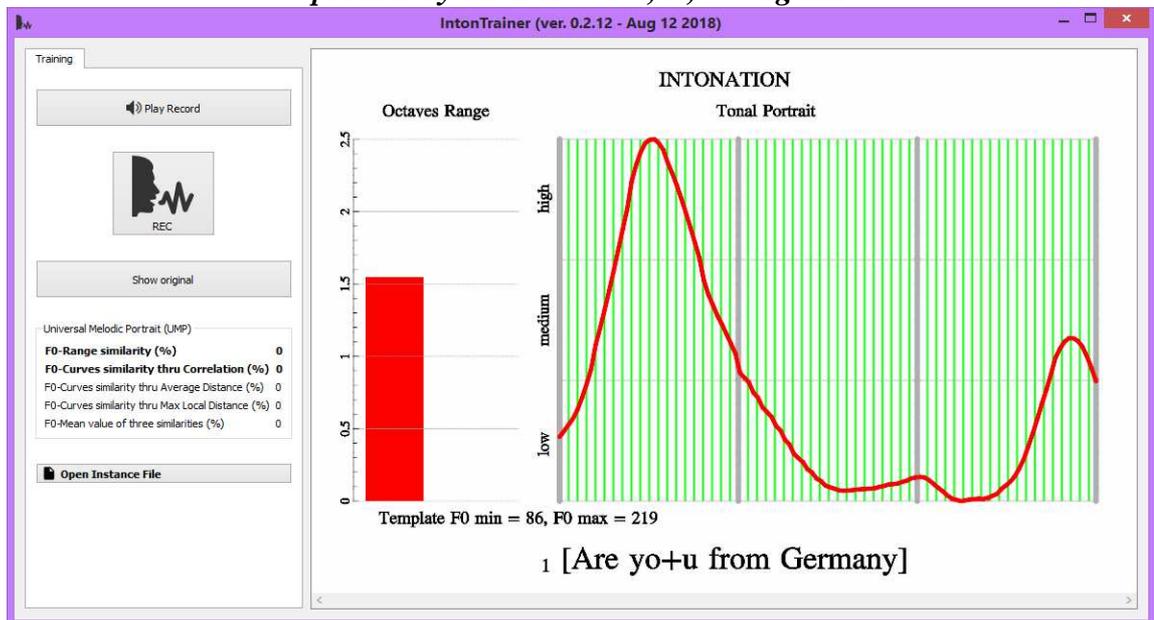


Fig. 17. An example of displaying the trajectory F0 in the case of markup phrases automatically given the same index N for each of voiced regions

8. CONCLUDING REMARKS

On this site the user is given the opportunity to download free of charge three demos of the "**IntonTrainer**", intended for training the basic intonational constructions of the **Russian** language and 2 variants of the **English** language (British and American). The site also presents prototypes of demo versions for **German** and **Chinese**. In addition, there is a demo version with

examples of intonation training for **call center** speakers in Russian and English, as well as a demo version for teaching **emotional** intonation in English.

As already mentioned, the "**IntonTrainer**" software package is an open system and the user is provided with ample opportunities for updating, updating and creating reference databases for various languages and various practical applications. In this case, we recommend using the option of automatic marking of reference sound files by selecting the operation (**Mark Out File**) in the additional settings section - "**Acoustic Speech Database**".

Finally, we recommend that users from time to time look at our site, which can constantly appear useful updates to the program "**IntonTrainer**"

To learn more about the theoretical basis of the **Application**, see the additional information posted on this site.