

=Speech Intonation Assistant= (SIA)

Version from 2024-05-22

USER GUIDE

1. Functional characteristics

The Speech Intonation Assistant (SIA) program is designed to analyze the statistical characteristics of the user's speech intonation parameters at fairly long intervals (about 10 seconds or more). At these intervals, two parameters that are most important for perception are measured:

1. Range of pitch changes – Pitch Variability [Hz];
2. Speed or tempo of speech – Speech Rate [wpm].

The first parameter largely characterizes the expressiveness of speech, and the second – its intelligibility.

Based on the calculation of these intonation parameters, three operating modes of the program are implemented: “Testing”, “Training”, “Monitoring”.

SIA is implemented as a separate Application for Windows and Linux.

2. Applications

This software application is intended for *testing, training and monitoring* the intonation (prosodic) characteristics of the user's oral speech. When preparing for a public speech or during the speech itself, the user must adhere to a certain style of speech. Each style of oral speech is characterized by its own characteristics of the prosodic parameters of oral speech. Various styles of oral speech are: official, journalistic, artistic, colloquial, educational, scientific, etc.

For example, SIA can become in demand for creators of such types of content as:

Audio content:

- *Podcasters*
- *Radio and TV presenters*

Video clips and YouTube:

- *Bloggers*
- *Training videos*

Advertising content:

- *Propagandists*
- *Voice actors*

Educational content:

- *Teachers and lecturers*

- *Online courses and webinars*

3. Getting started

The SIA start window, which opens after launching the program, is shown in Fig. 1.

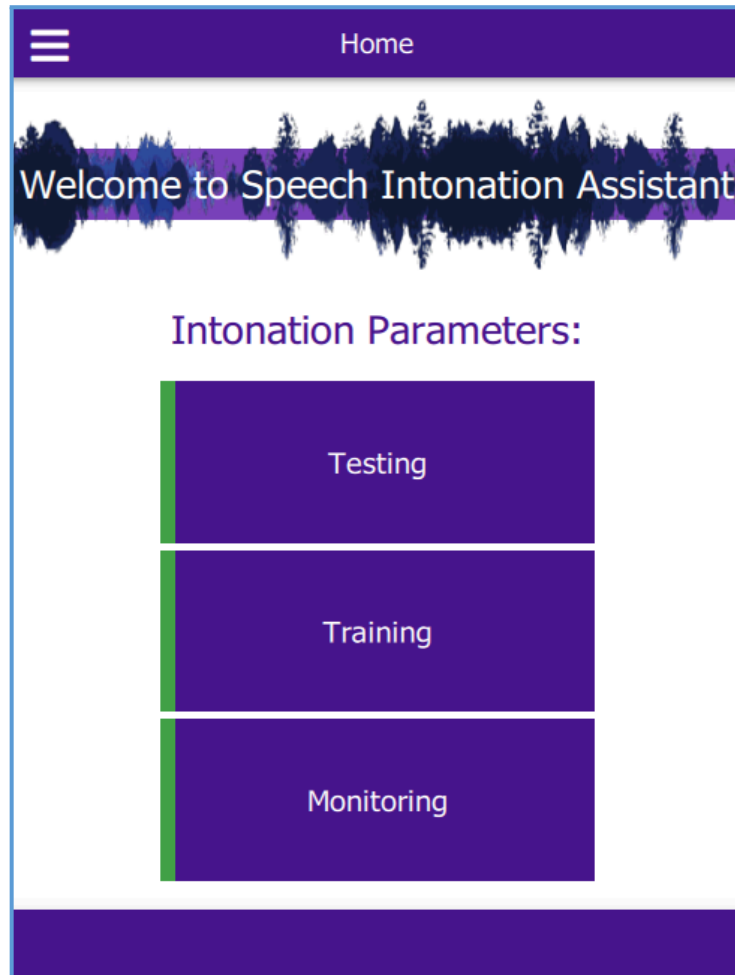


Fig. 1. Start window

After starting the program, the user is provided with functionality in 3 modes:

1. *Testing practical oral speech skills* by assessing intonation parameters: Pitch Variability [Hz] and Speech Rate [wpm].
The first mode is implemented by clicking on the “**Testing**” icon.
2. *Practical training in speech intonation skills* by comparing the current intonation parameters of the user’s oral speech with the speech parameters of the selected reference speaker and assessing the degree of their similarity.
The second mode is implemented by clicking on the “**Training**” icon.
3. *Monitoring changes in intonation parameters* of current speech in the actual process of reading a lecture, report, etc.
The third mode is implemented by clicking on the “**Monitoring**” icon.

4. Testing the speaker's intonation parameters (1st mode)

When you select the first mode “Testing”, a working window opens (Fig. 2).

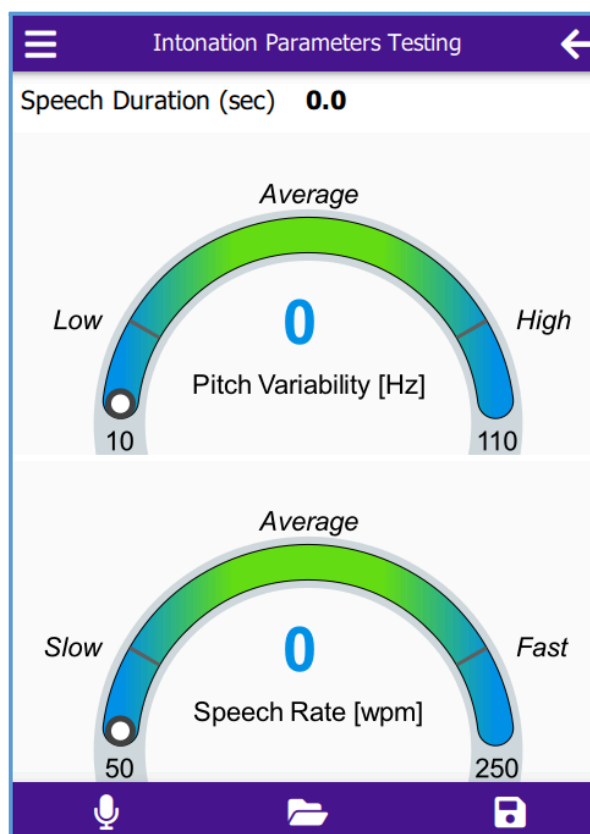





Fig. 2. Working window of the first mode


In the central part of the window there are two multi-colored arcs, which play the role of scales for the values of the intonation parameters **Pitch Variability** and **Speech Rate**. The minimum and maximum values of these parameters are indicated at the ends of the arcs. The numerical values of the parameters are specified in the settings section and can be changed by the user (see Section 7).

After opening the working window, the user is provided with the following options:

- 1) Recording through a microphone a sufficiently long segment of the analyzed speech (preferably more than 10 seconds);
- 2) Calling one of the pre-prepared test audio files stored in the “**data-tests**” folder.

The first option is implemented by clicking on the microphone icon  on the left side of the lower purple strip. As a result, a window with the icon  opens, and after its activation, a dynamic icon  appears and the recording of the spoken segment of speech begins. To end recording, you must click this icon again.

The spoken segment of speech is prepared for further processing and display of the results of intonation analysis, and is also recorded in the “**data/records**” folder.

The second option is implemented by clicking on the icon  in the middle part of the lower purple stripe. It is designed to call pre-prepared audio files from the “data/tests” folder - fragments of speech by creators of various types of content. The

“tests” folder contains various speech samples for male and female voices as examples (see Fig. 3).

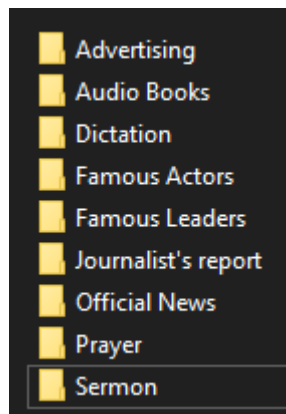


Fig. 3. Contents of the “tests” folder

Nine folders contain examples of different speaking styles in nine thematic folders:

- Sound commercials;
- Examples from Audio books;
- Examples from Dictation;
- Performances by famous actors;
- Speeches by famous leaders;
- Reports from radio and TV journalists;
- Reading by announcers of official news;
- Examples of religious prayers;
- Examples of religious sermons.

Note. *The contents of the “tests” folder can be expanded with additional examples using existing tools (for example, Sound Forge). It should be borne in mind that this version of SIA provides the ability to analyze only signals in the following format: wav, mono, 8 kHz. Any other formats should be adapted to this format using suitable software.*

After finishing recording through a microphone or after calling a test audio file, a window opens with the results of calculating intonation parameters: the range of changes — **Pitch Variability [Hz]** and speech tempo — **Speech Rate [wpm]** (see Fig. 4).

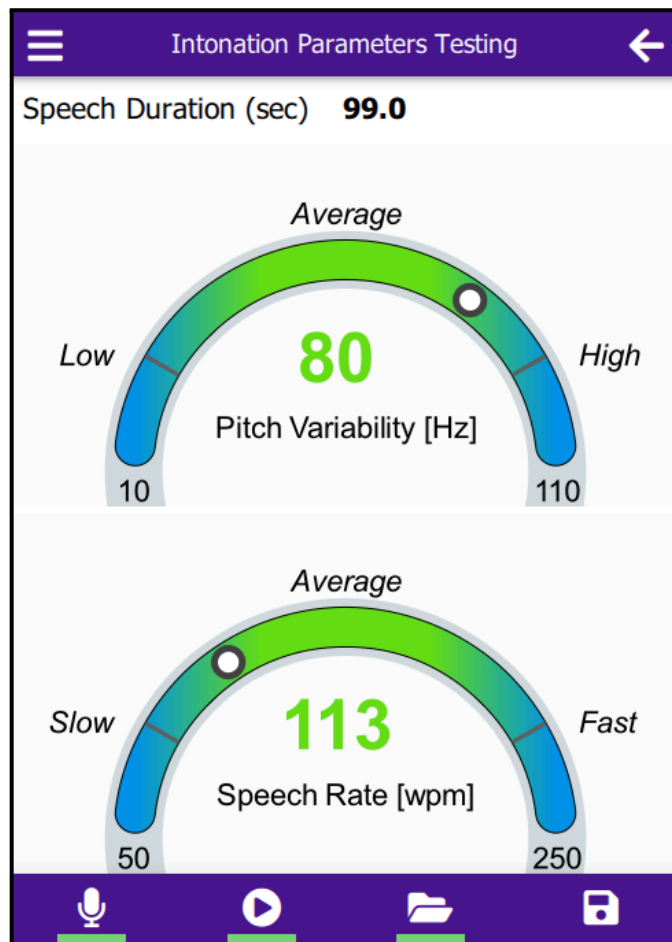




Fig. 4. Results window in the first mode

The upper part of the window displays the duration of the analyzed speech fragment – **Speech Duration (sec)**. Below are 2 scales:

- the top one displays graphically (with a white circle) and numerically (under the arc) the results of measuring the range of changes in **Pitch Variability [Hz]**,
- bottom – results of measuring speech speed – **Speech Rate [wpm]**.

The ranges of changes in the average statistical values of these parameters, obtained on the basis of examples of speech of creators of various types of content (Fig. 3), are marked in green; deviations from them are marked in purple.

In addition, an additional icon  appears in the lower purple bar, allowing you to listen to the analyzed speech fragment.

On the right side of this strip there is an icon , when clicked, the results obtained are recorded (see section 7).

5. Training in speech intonation skills (2nd mode)

When you select the second mode “Training”, a working window opens (Fig. 5).

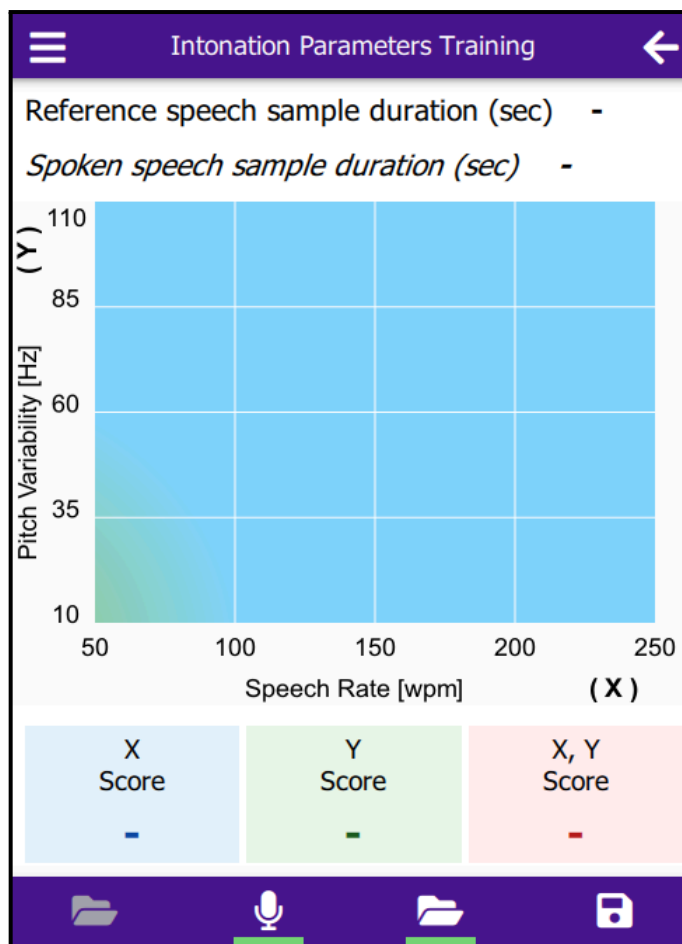

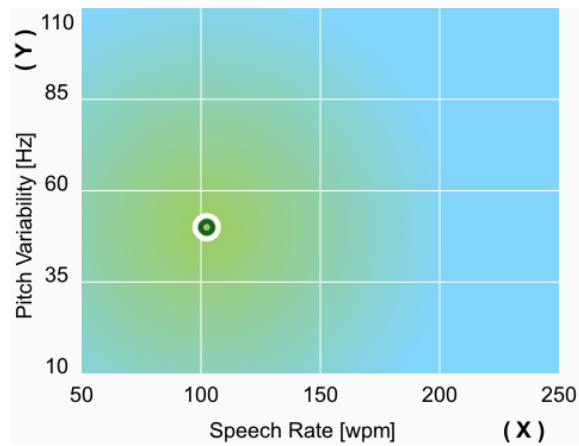


Fig. 5. Second mode operating window

The working window displays the training target in the form of a plane with coordinates (X, Y), which correspond to two measured intonation parameters, graduated in numerical values - [wpm] and [Hz].

After opening the working window, the user is given the opportunity calling from the “**data-tests**” folder one of the pre-prepared reference samples of oral speech by clicking on the icon  on the left side of the lower purple stripe. This opens the “**data-tests**” folder, in the subfolders of which (see Fig. 3) audio files of various samples of oral speech are located. These speech samples - fragments of speech from creators of various types of Internet content - can be selected by the user as suitable samples for training his own speech.

After selecting a suitable sample of oral speech, a corresponding icon in the form of a black and white ring appears on the training target of the working window (see Fig. 6). Its coordinates on the plane correspond to the measured values of intonation parameters.



Rice. 6. Training target of the working window

During the training process, the user records his speech to compare it with the desired sample. Recording is carried out through a microphone in accordance with the procedure described in paragraph 4. The spoken segment of speech is prepared for further processing and display of the results of intonation analysis, and is also recorded in the “**data/records**” folder.

After pronouncing each fragment of the analyzed speech, the values of its intonation parameters are displayed in the form of points on the training plane. As a result, the working window of the second mode takes the form shown in Fig. 7, where small dots with white circles display the result of each training speech fragment. The last (best) utterance of a fragment is displayed as a white circle with a contrasting black dot in its center.

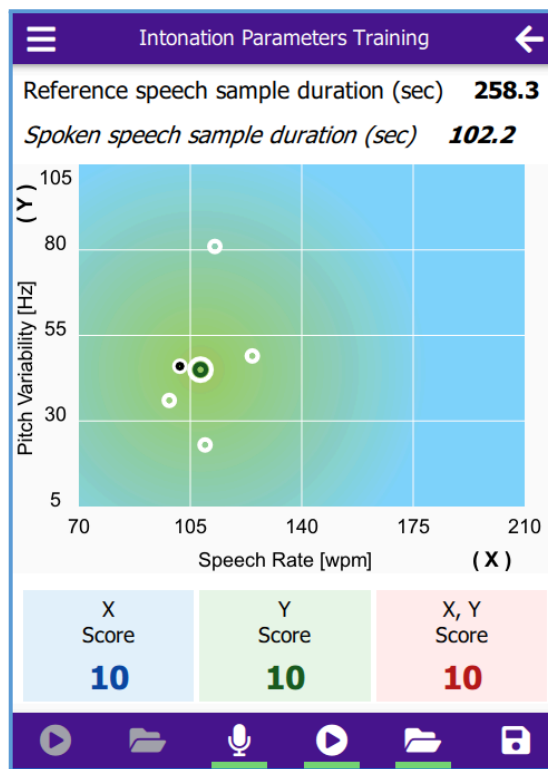


Fig. 7. Resulting view of the working window

After pronouncing a fragment of the analyzed speech that showed the greatest closeness to the selected sample, the results of comparison with the sample are displayed (see Fig. 7) as follows:

- Below the training plane are numerical estimates of proximity on a 10-point scale separately for each (X) and (Y) parameter, as well as a generalized estimate (X, Y).
- The durations in seconds of the reference sample and the spoken speech sample are shown above the training plane.

The icon in the lower purple stripe is for listening to a reference sample of oral speech, and the icon is for listening to the analyzed fragment of speech. When you click the icon, the result is recorded (see section 7).

5. Monitoring of intonation parameters of current speech (3rd mode)

When you select the third mode - "Monitoring" - the first (initial) view of the working window opens (Fig. 8).

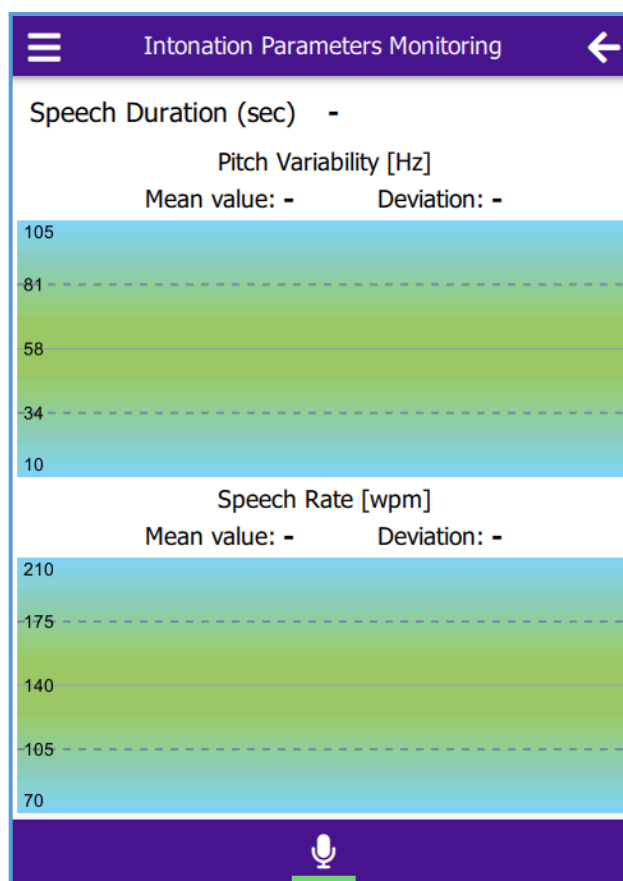



Рис. 8 Начальный вид рабочего окна "Monitoring"

The working window contains two separate fields designed to display the dynamics of development of intonation parameters of oral speech:

- **upper** – measurement results of measuring the range of changes in the frequency of the fundamental tone of speech – **Pitch Variability [Hz]**,
- **bottom** – results of measuring speech tempo – **Speech Rate [wpm]**.

The dynamics of development of intonation parameters are displayed in the form of a dotted curve on the margins of the working window in real time when the user pronounces various types of audio content.

The left side of the fields shows the scales of numerical values of the parameters and the ranges of their change. The dashed lines show the conditional zones of average values specified by the developer and marked in green. The sampling interval and ranges for changing intonation parameters can be changed by the user in the settings section (see Section 7).

Recording of the analyzed speech through a microphone is realized by clicking on the microphone icon . After clicking on this icon in the working window (Fig. 9), the current values of the intonation parameters of the analyzed speech signal begin to be displayed.

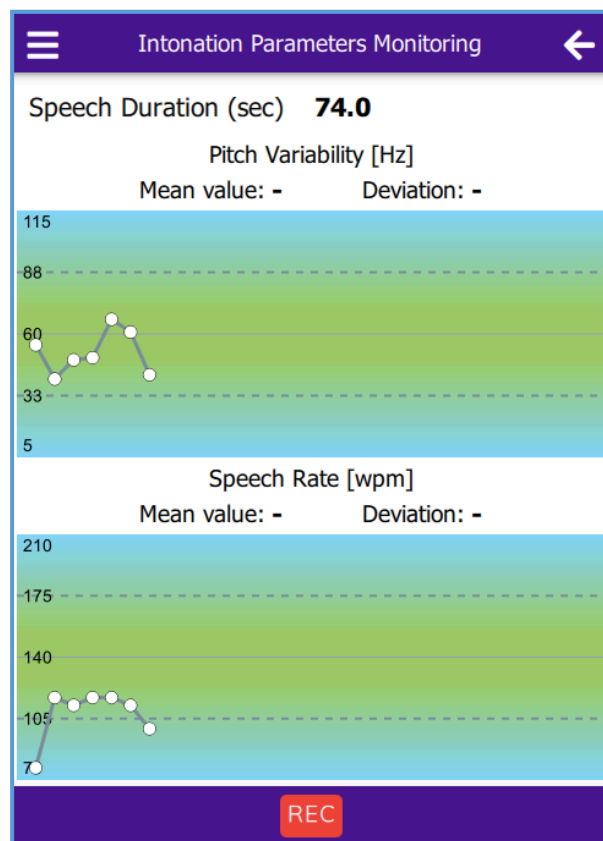



Fig. 9. Display of current values of intonation parameters

After completing the speech and clicking on the icon , the recording of the analyzed speech signal ends and the third (final) view of the working window of the “Monitoring” mode opens (Fig. 10).

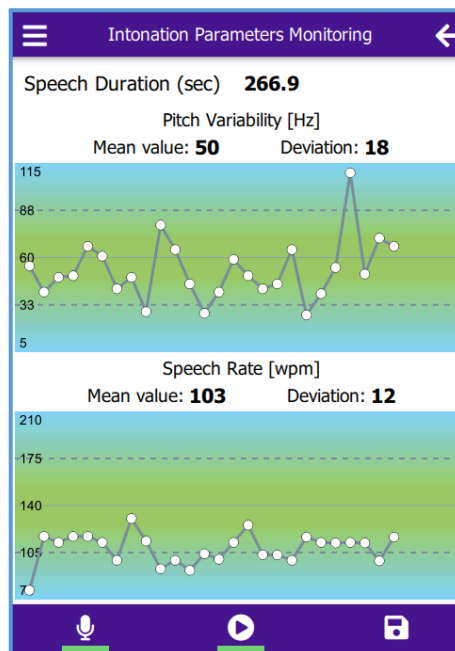




Fig. 10. Final view of the working window of the “Monitoring” mode

The upper part of the working window shows the duration of the spoken speech signal – Speech Duration (sec). The lower and upper fields show the trajectories of changes in the corresponding intonation parameters, and above each field: their **Mean Value** and **Deviation** for the specified time.

The icon  in the lower purple stripe is intended for listening to the analyzed speech fragment. When you click the icon , the results are recorded (see section 7).

In Fig. 11 shows, as an example, the results of measuring the intonation parameters of 2 speakers at a conference for the same speaking time. There are obvious differences in the dynamics of changes in these parameters during speech, as well as in their **Mean Value** and **Deviation** estimates for the same five-minute period of time.

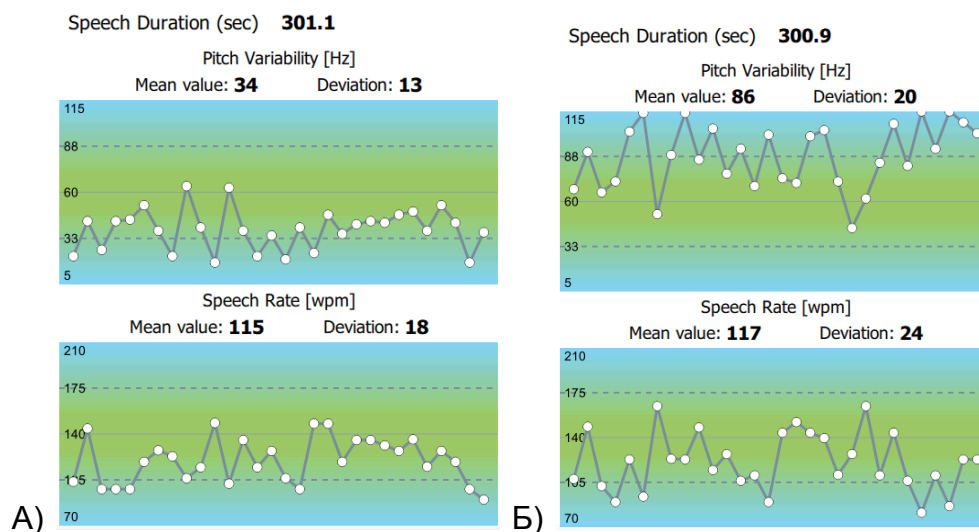



Fig. 11. Measurement results for speakers A) and B)

6. Helper icons

In the upper bar on the left of the working window of each mode there is an icon , which, when clicked, opens an additional information window (see Fig. 1).

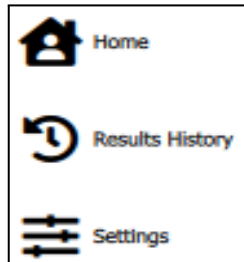



Fig. 12. Information window-1

The settings icon  **Settings** allows the user to change the maximum possible values of intonation parameters recommended by the developer (see Fig. 13), simultaneously for all modes.

Pitch Variation Maximum

115

Maximum Value [Hz]

Pitch Variation Minimum

5

Minimum Value [Hz]

Speech Rate Maximum

210

Maximum Value [WpM]

Speech Rate Minimum

70

Minimum Value [WpM]

Fig. 13. Information window-2

Additionally, only for the third mode, it is possible for the user to change two characteristics of the recording process and display parameters (see Fig. 14).

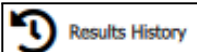
Visible monitoring points number

30

Sliding monitoring window in Sec

10

Fig. 14. Information window-3

By selecting the icon:  **Results History** the contents of the “results” file are called up (see Fig. 15A for the first mode, Fig. 15B for the 2nd mode and Fig. 15B for the third mode). The data is also saved in the **results.csv** file in the folder with the SIA program.

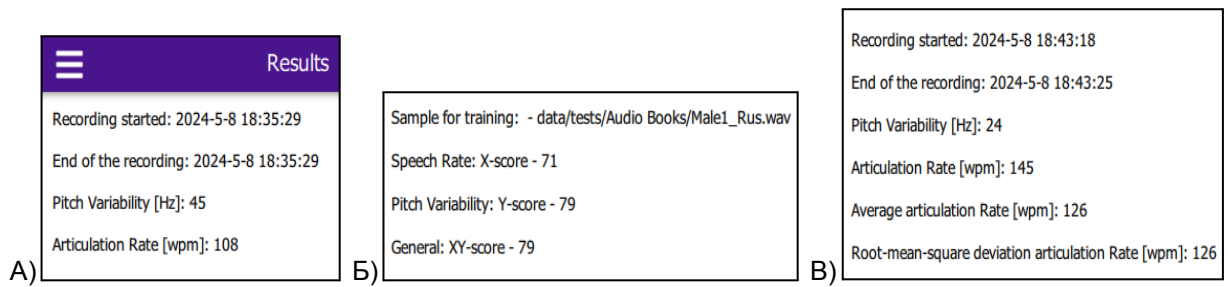
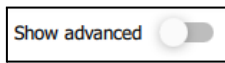


Fig. 15. Contents of the “results” file”

ATTENTION!

Internal information, called up by the mark in the upper right square



, is intended for the developer only. Changing its contents by the user is highly undesirable and can lead to incorrect operation of the program.

CONCLUDING REMARKS

We encourage users to check our website from time to time for useful SIA updates.

To learn more about the theoretical background and code for developing SIA: see additional information posted on this site.