AVATAR

Software System for Verification and Identification of Speakers

User Manual

Version 1.0

February 2021

Table of contents

1 Introduction

This version of the Avatar software package (hereinafter referred to as the System) solves the problems of identification and verification of speakers of voice messages.

Speaker identification describes the situation when there are several phonograms with voice narration and it is necessary to determine whether the speaker is the same or not.

Speaker verification describes the situation when there is a large database of phonograms with various speakers (partially identified, partially not) and it is necessary to sort the phonograms by the proximity of the characteristics of the voices for further analysis (with possible subsequent identification).

The System is context independent. Due to a special built-in module it can be applied to phonograms in any language of any language group, including speech phonograms recorded in various languages.

The System includes several special built-in databases that provide solutions for the verification and identification of speakers, depending on the duration of the phonograms and specific language groups.

The core of the System is the built-in module of the **Phonemic Machine**, which identifies the sounds of speech fragments regardless of the phonogram language and specific speaker. The Phonemic Machine module is based on modern technologies of **deep learning neural networks**.

In the current version of the System, speakers are identified by the closeness of the characteristics of two vowels [A] and [I] in two independent speech phonograms (audio files).

The basis for the **probabilistic identification of speakers** is the error graphs of the first and second kind. These error probabilities of the first and second kind are used directly in solving the problems of speaker identification and verification.

This version of the System has three localizations - English, Russian and Ukrainian. At the request of the Customer, the System can be supplemented with any language localization.

Current system parameters allow phonogram analysis for four language groups - English, mixed language group (Romano-Germanic phonogram languages), Russian, and Ukrainian.

1.1 Source Audio Requirements

- Optimal duration: 200 ms and longer
- Minimum duration: 30 ms (for sounds [A] and [I])
- Maximum duration: unlimited
- Supported file format: .wav (at least 8000 Hz)

1.2 Hardware Requirements

- Processor: from 2 GHz

- RAM: at least 4 GB (preferably 8 GB)
- OS: Windows 10 (64 bit).

Exclusive intellectual property rights to the system are protected by the copyright.

Certificate of registration of copyright rights No. 54584 dated 05/05/2014, AVATAR Software System for Verification and Identification of Speakers

2 Principles of speakers' voice characteristics identification

Speaker identification is based on automatic allocation of sounds [A] and [I] in two phonograms with fragments of speech by the software module of the phonemic machine. After fragments extraction, their average characteristics are compared for the two phonograms. Then verification and/or identification of speakers is done depending on the proximity of the characteristics of the spectra of sounds.

Note that the System applies sound identification technology, which has a different interpretation of sounds compared to classical representations. Since it is not sounds that are physiologically perceived by the hearing organs, in the System's built-in Phonemic Machine certain combinations of several "atomic" components for each sound serve as models of speech sounds. The set of "atomic" components for each sound may partially overlap with the set of "atomic" components for other sounds.

For example, the set of component structures for the sound [A] intersects with the set of structures for the sound [O]. Similarly, the sound [I] overlaps with the sound [Y]. With an averaged representation in the form of spectra, these structures can be interpreted as averaged spectra of the corresponding sounds of a particular speaker. These averaged spectra of "atomic" sounds are not completely identical to physiologically perceived sounds. Thus, in the applied model of the phonemic machine, the structural components of speech sounds are distinguished.

The System does not operate within generally accepted notions that characterize speaker identification systems - the pitch frequency and speech formants. However, these parameters are functions of the averaged integral characteristics of the spectra of "atomic" sounds and are indirectly taken into account in the identification process. It is possible to separate these speech parameters on averaged spectra and they are presented in a number of graphical outputs of the System.

The System analyses "atomic" components in small time windows of various durations, from 10 to 30 ms. The dimensions of the windows in the analysis are determined adaptively, depending on a number of factors.

The frequency resolution for constructing all spectral characteristics of the "atomic" components in time windows is 1 Hz. This resolution does not depend on the size of the time windows. The latter is achieved using non-orthogonal wavelet transforms based on the Morlet wavelet.

The interpretation of the spectra is fully consistent with the classical concepts, up to the order of resolution in frequency and specific form of window functions. Applied technological features of the system allow the analysis of voice characteristics, in some cases, for phonograms lasting from 30 to several hundred milliseconds (for such analysis, a special database of sounds [A] is used).

An important feature of the System is the probabilistic characteristics of solving verification and identification problems. All probability characteristics in the form of graphs of errors of the first and second kind are determined on the basis of special tests based on previously prepared experimental data for various language groups and phonograms of various durations. Error graphs are automatically used in the System for making decisions in the tasks of verifying and identifying speakers.

2.1 Main features

1. Verification and identification of speakers are based on the analysis of how close the characteristics of the vowel sounds spectra are for speakers in two phonograms. Currently used vowels are [A] and [I].

2. The System has at least an order of magnitude higher frequency resolution when evaluating any parameters of the spectrum of phonogram characteristics compared to other existing systems. This is achieved using non-orthogonal spectral transformations.

3. The effectiveness of the analysis of phonograms depends on the time duration of the phonograms. It is possible to analyze fragments starting from 30 ms.

For more detailed description please refer to a list of published articles [1,2,3].

3 System Installation

3.1 Operating System

The System requires Windows 10 (64 bit version).

3.2 Installation Package

The System installation packages includes the following:

1. The executable file of the system: Avatar_15_2021_01.exe,
2. Files and directories that support the operation of the System,
3. User manual.

3.3 Installation

Follow these steps to install the System:

1. Copy the Avatar's system folder with the entire System kit to your computer's hard drive.
2. Before the first launch of the System it has to be activated on this specific computer. To activate the System, open the folder with Avatar.exe file and run the file Read_Activation_08_2019_FL.exe and send the key to the System Provider.
3. After receiving the activation file File_rdkd.svi from the Provider, copy this file to the folder with Avatar.exe and replace the previous File_rdkd.svi file.
4. Launch the System by running Avatar_15_2021_01.exe file directly or by using the desktop icon.
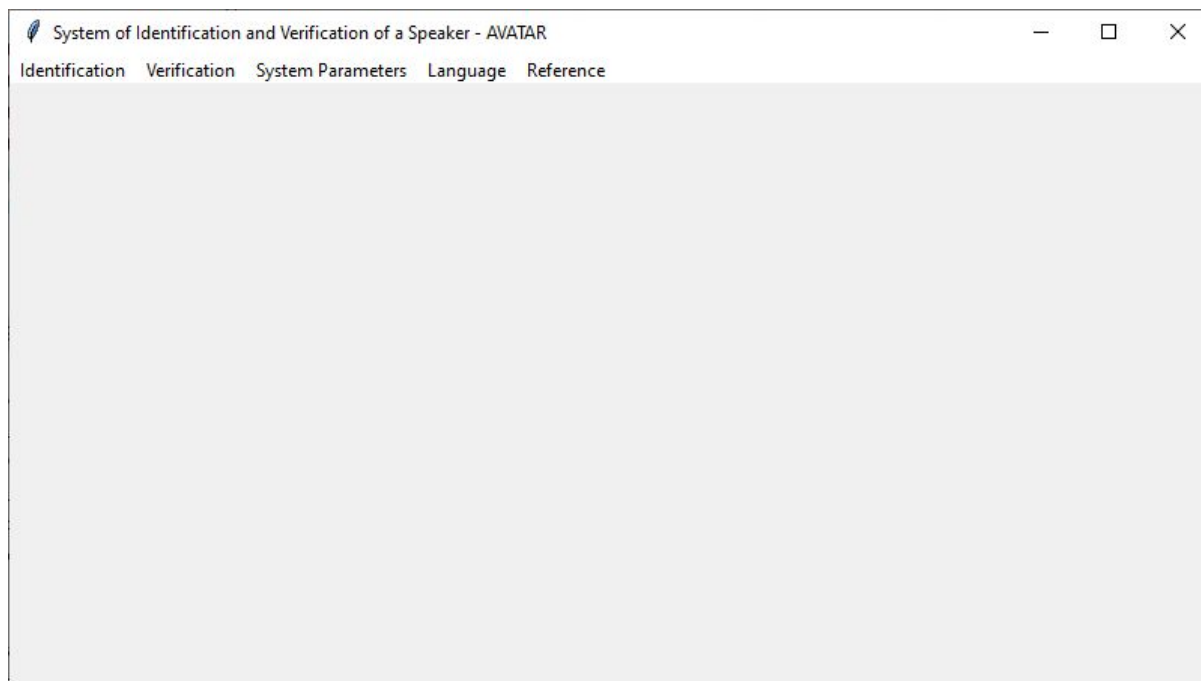
Fig. 1 Avatar System menu window

4 Main Menu

4.1 Identification

Identification menu tab has two sections: Identification and Voice Parameters.

4.1.1 Identification

This option is for recording all the relevant characteristics of the speaker's voice based on the selected phonogram. All characteristics are recorded in designated .svis files in the embedded database.

Note
All data processed during identification is entered into a database directory called Dir_Ident. To use different groups of processed phonograms in the analysis later, copy the Dir_Ident folder to another directory. Then it is safe to delete its content and to accumulate new

phonogram characteristics. All characteristics of various files in Dir_Ident have a name identical to the name of the original .wav file and .svis file extension.

Steps:

1. Click on Identification → Identification.
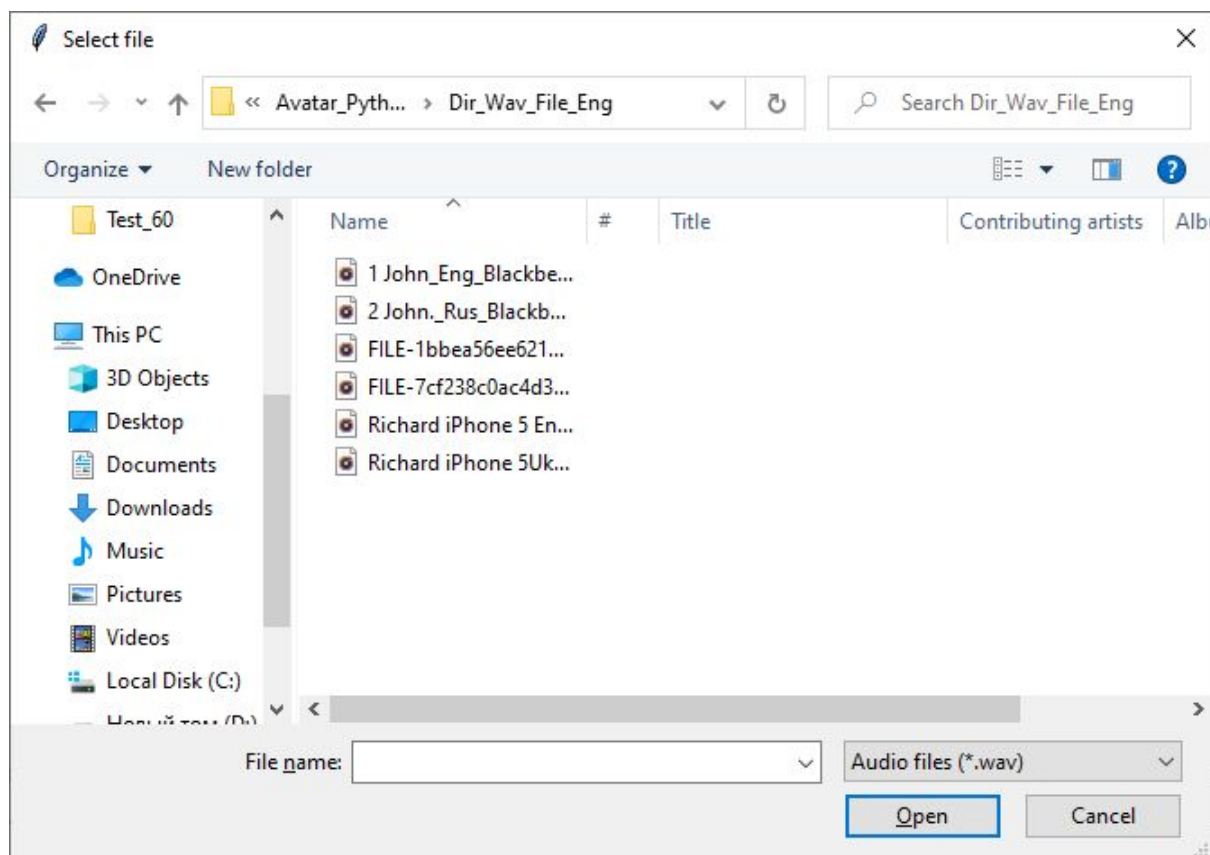2. A window for selecting a .wav file for analysis opens:

Fig. 2 Window for selecting a file for analysis

3. Select the file and click on Open.
4. Procedure for automatic analysis and data processing starts. All data is entered into the directory - Dir_Ident.

4.1.2 Voice Parameters

This menu option initiates the speaker identification process.

Steps:

1. Make sure that files with voice characteristics for the audio files in question are in the Dir_Ident directory (for more details see Identification section steps).
2. Click on Identification → Voice Parameters.
3. A window for selecting .svis files with voice characteristics of speakers for comparative analysis opens:
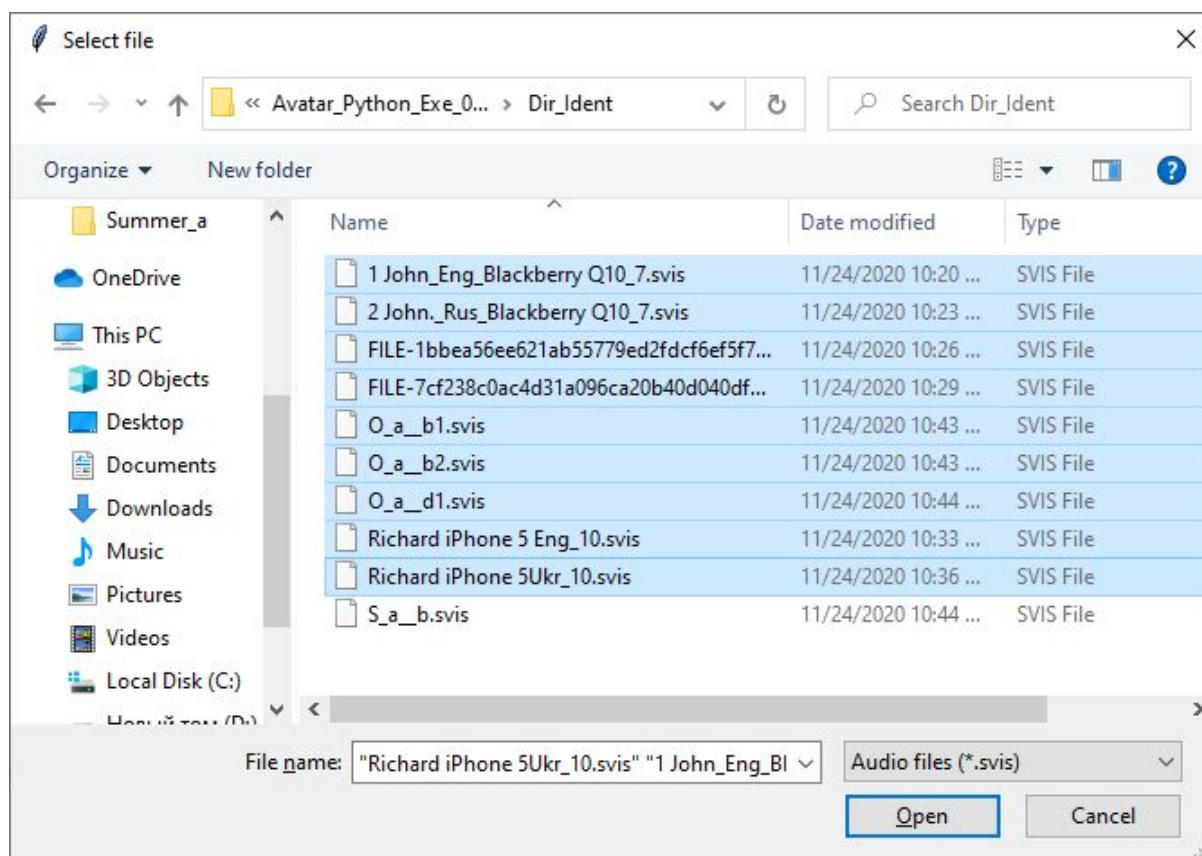
Fig. 3 File selection window for subsequent comparative analysis

Note
All files for comparative analysis are located only in the Dir_Ident (and copied) directories. it is also possible to use the verification directories - Dir_Verification (see below).

For multiple selection press Ctrl and left-click to select the files for analysis.

4. Select several .svis files for analysis
5. Click Open.
6. Phonogram Listing window with files characteristics opens:

Fig. 4 Phonogram Listing window with files characteristics

Each line includes the following information for the file:

- file name,
- number of channels,
- bit depth (bytes),
- sampling frequency,
- file size (bytes per channel),
- soundtrack duration (sec.).

Additionally, during the initial recording of the characteristics, the current date and location are entered. The location is set to "Unknown" by default. Location and date can be edited later on, and the list can be sorted by the location.

Fig. 5 Date of entry and where the files come from (default)

Phonogram Listing form contains **6 operation buttons:** Spectr[A], Freq[A], Spectr[I], Freq[I], Date, and Location.

Spectr[A] is an option for comparative analysis of the [A] sound spectra of two speakers. Whether the speakers are the same or not is calculated automatically (see Fig. 6). By default you can see one spectrum. For comparative analysis, select two characteristics files from the listing.

Spectr[I] is an option for comparative analysis of the [I] sound spectra of two speakers. You can view one spectrum. In a comparative analysis, a decision is automatically made to identify the speakers (see Fig. 7).

Freq[A] is an option for comparative analysis of the frequency distribution of the fundamental tone of the sound [A] of two speakers. You can view one distribution (see Fig. 8).

Freq[I] is an option for comparative analysis of the frequency distribution of the fundamental tone of the [I] sound of two speakers. You can view one distribution (see Fig. 9).

For the description of the settings for the decision making process for speakers identification see System Options section.
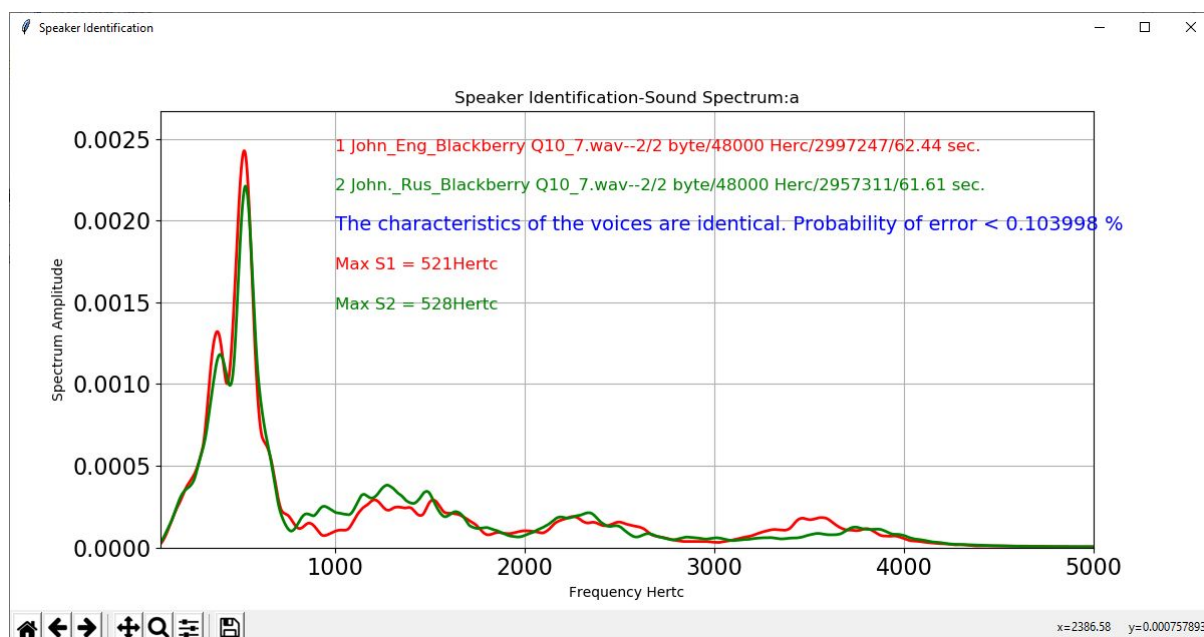
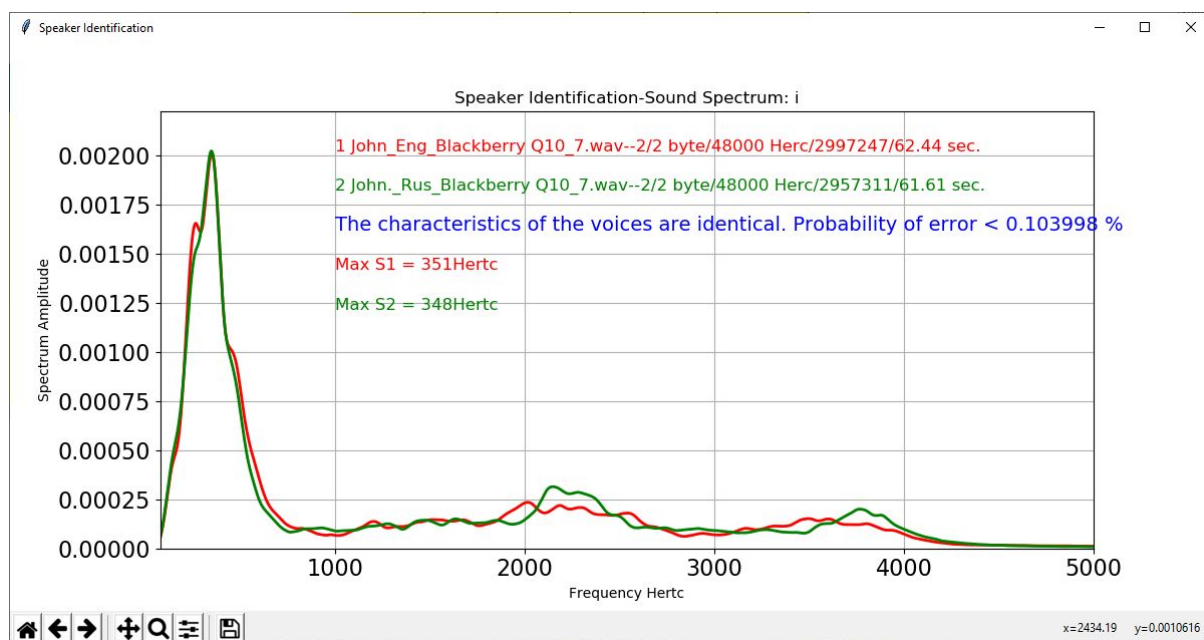Fig.6 Comparative analysis of the spectra (sound [A])

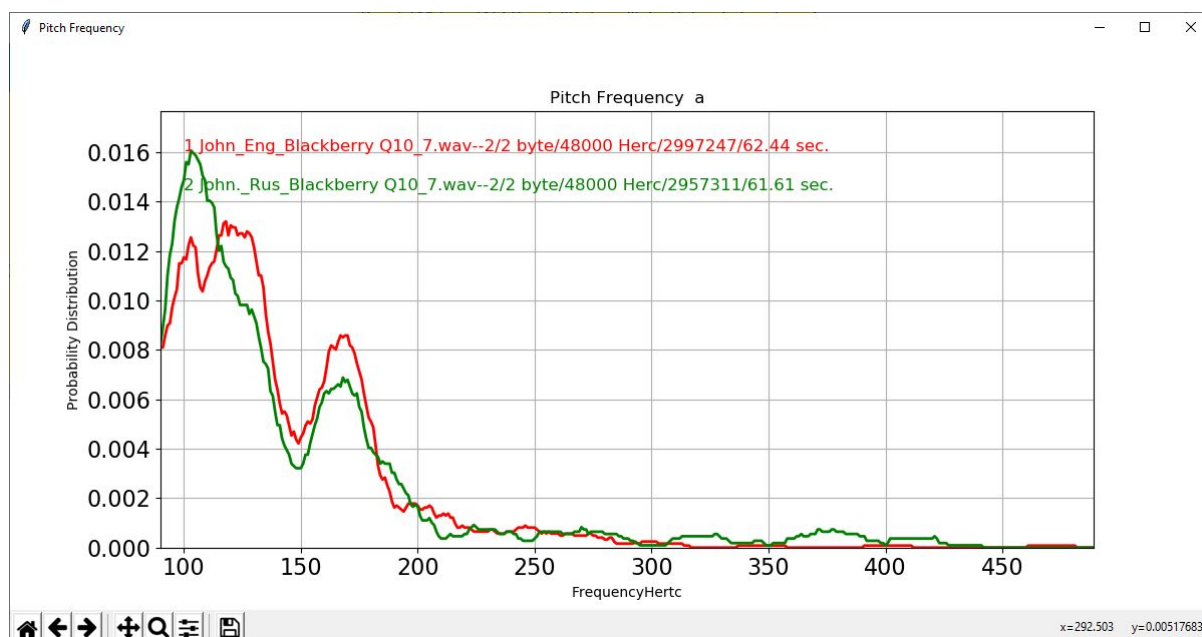Fig. 7 Comparative analysis of the spectra (sound [I])

Fig. 8 Comparative analysis of the frequency distribution of the fundamental tone (sound [A])
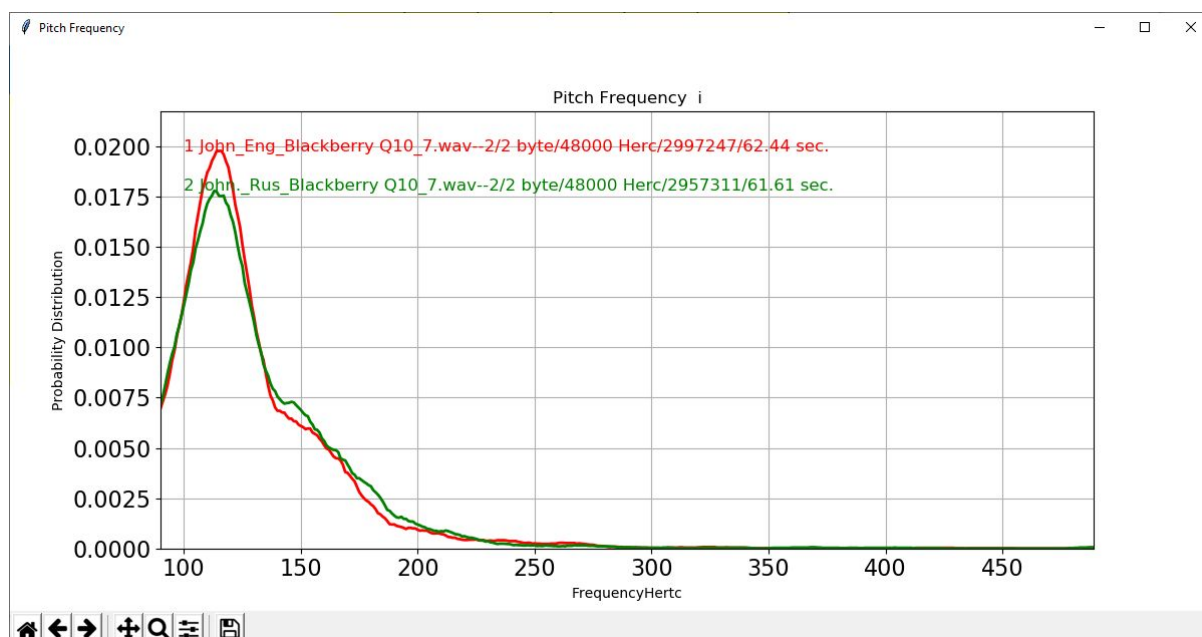

Fig. 9 A comparative analysis of the frequency distribution of the fundamental tone (sound [I])

To save any of the graphs press on the Save icon in the left lower corner of the graph window.

Date button is for editing the phonogram date. To open the calendar, select the phonogram in the Phonogram Listing window and press the Time button. Then select the correct date and click OK to save it.
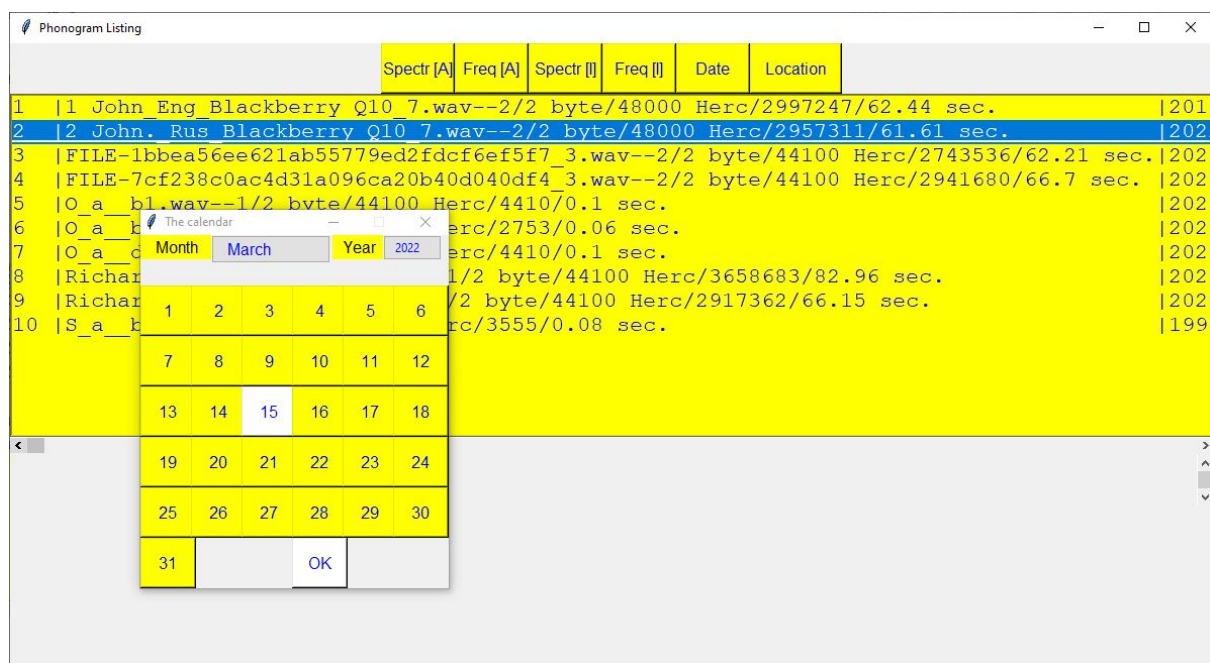
Fig. 10. Phonogram date editing

Location button is for editing the place of receipt of the phonogram. To access the Phonogram reference (location) window, select the phonogram in the Phonogram Listing window and press the Location button. Choose the correct location, then click on the Select button to save it.
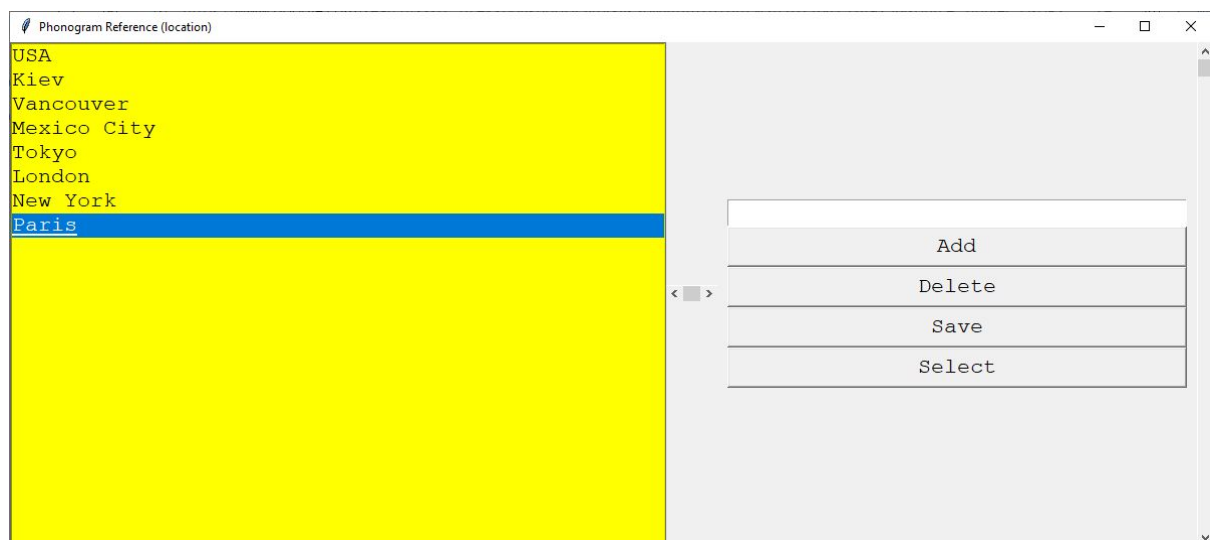

Fig. 11. Phonogram reference (location) window

The location references can be edited. You can add to the directory any entries in any language and it is also possible to delete any entries.

Note
When filters are turned on at the place of localization of phonograms, filtering collisions are possible (see Verification section). This may happen when places of receipt are different.

4.2 Verification

This option is for sorting the speakers by the proximity of the characteristics of the voices and further analysis of the sorting listing.

Verification includes two sections - Verification and Verification Listing.

4.2.1 Verification

This section allows the processing and recording of voice characteristics of speakers from a specific directory. The entire set of files of the selected directory is processed.

Steps:

1. Click on Verification → Verification.
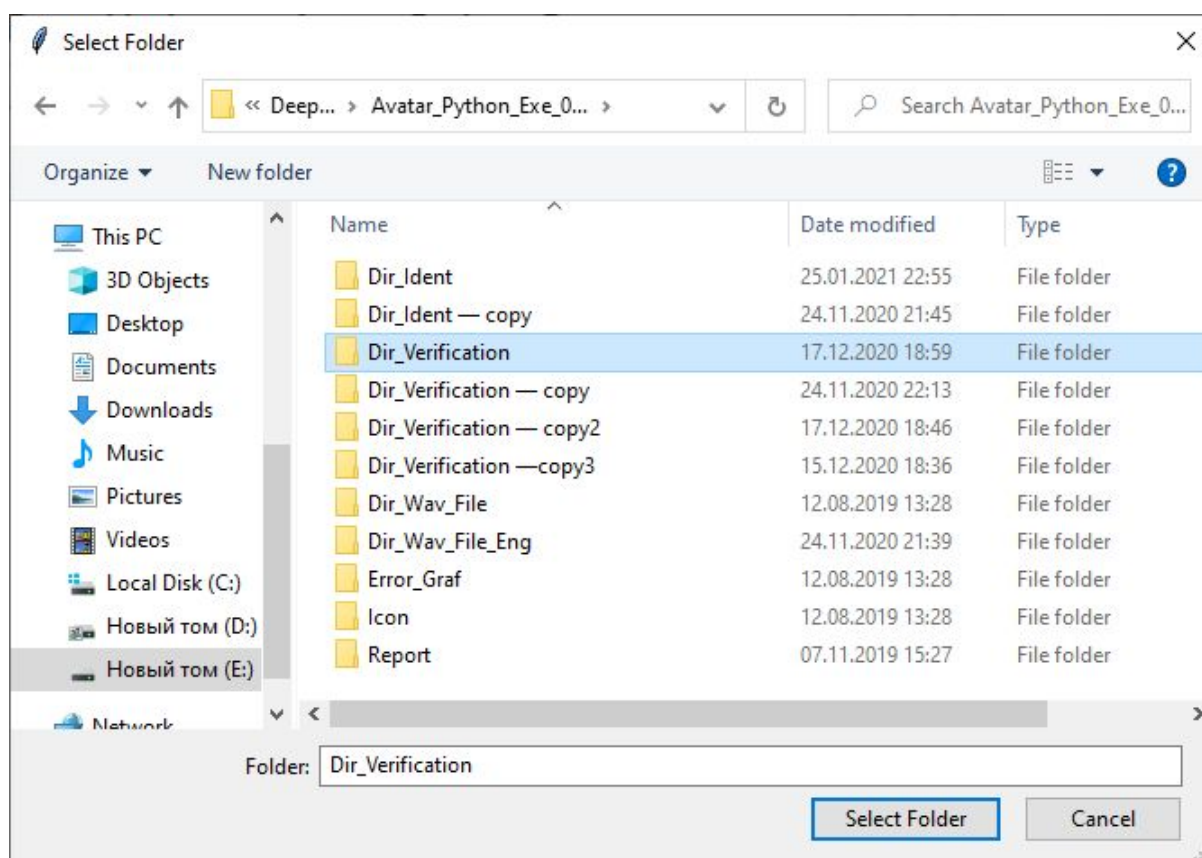2. A window for selecting a directory with .wav files opens:

Fig. 12 Selecting the file directory for verification

3. Select the directory, the file analysis process follows.

Note
If there is a large number of files, this process can take a lot of time. 60 sec. file takes about 2-3 minutes to process on a computer with a processor with a clock frequency of 3 GHz. It is better to select a directory with the appropriate number of files. Records for verification can be added in stages from different directories.

All analysis files are written into a common directory - Dir_Verification.

To create several directories with different verification options, copy the Dir_Verification directory after each analysis session. Then it is possible to use different directories when sorting speakers.

4.2.2 Verification Listing

This menu option initializes the sorting of the speakers by the proximity of the characteristics of their voices.

Steps:

1. Click on Verification → Verification Listing in the main menu.
2. A window for selecting a directory for analysis and sorting opens
3. Select the directory that contains files with the results of the analysis for identification and verification. For example, Dir_Verification, Dir_Ident or various copies of these directories.
4. Verification Listing window with files available for sorting opens:


```
Verification Listing                                                              —  □  ×

    Sort [A+I]      Sort [A]      Sort [I]      Date    Location   Filter by Date   Filter by Location    Delete File

1   |1 John_Eng_Blackberry Q10_7.wav--2/2 byte/48000 Herc/2997247/62.44 sec.                    |2010
2   |2 John._Rus_Blackberry Q10_7.wav--2/2 byte/48000 Herc/2957311/61.61 sec.                   |2022
3   |FILE-1bbea56ee621ab55779ed2fdcf6ef5f7_3.wav--2/2 byte/44100 Herc/2743536/62.21 sec.|2020
4   |FILE-7cf238c0ac4d31a096ca20b40d040df4_3.wav--2/2 byte/44100 Herc/2941680/66.7 sec. |2020
5   |O_a__b1.wav--1/2 byte/44100 Herc/4410/0.1 sec.                                      |2020
6   |O_a__b2.wav--1/2 byte/44100 Herc/2753/0.06 sec.                                     |2020
7   |O_a__d1.wav--1/2 byte/44100 Herc/4410/0.1 sec.                                      |2020
8   |Richard iPhone 5 Eng_10.wav--1/2 byte/44100 Herc/3658683/82.96 sec.                 |2020
9   |Richard iPhone 5Ukr_10.wav--1/2 byte/44100 Herc/2917362/66.15 sec.                  |2020
10  |S_a__b.wav--1/2 byte/44100 Herc/3555/0.08 sec.                                      |1999
11  |Д1_mono.wav--1/2 byte/48000 Herc/1768773/36.85 sec.                                 |2020
```

Fig. 13 Verification Listing window of files for sorting

The parameters of each record are the same as parameters of records in the Identification section.

Verification Listing form contains **8 operation buttons**: Sort[A+I], Sort[A], Sort[I], Date, Location, Filter by Date, Filter by Location, Delete File.
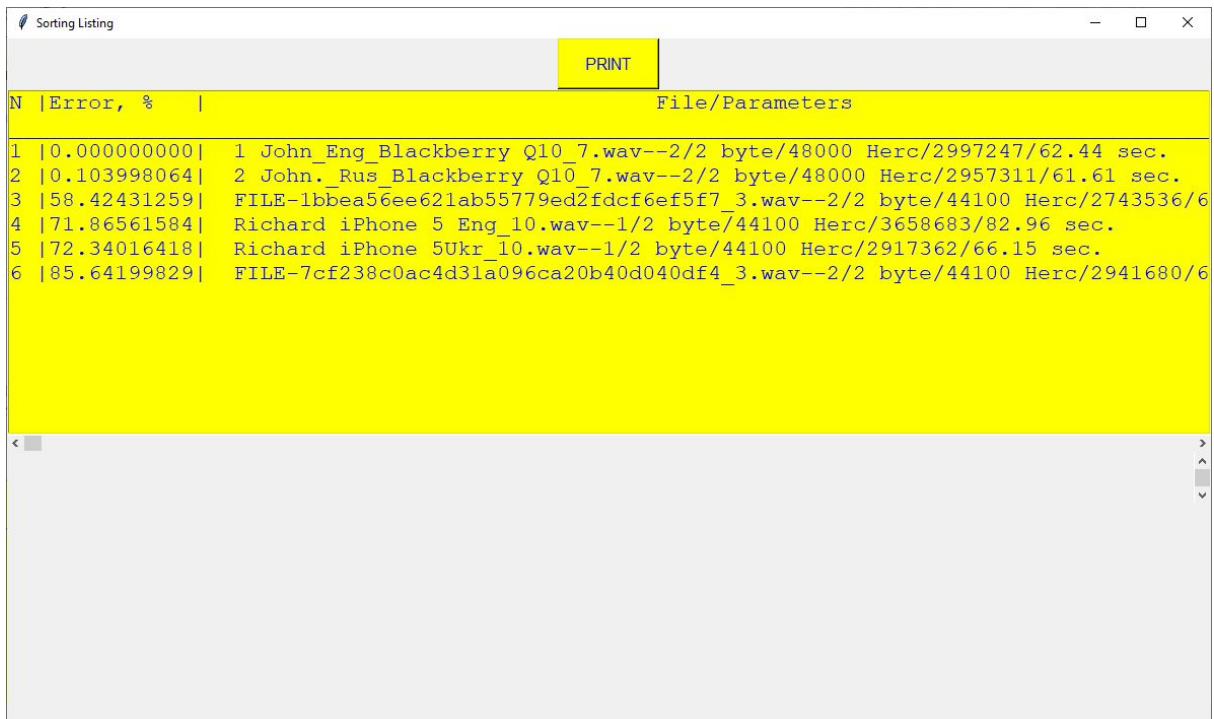
Sort[A+I] button sorts the speakers by the proximity of the integrated characteristics of the set of sounds [A] and [I] (see Fig. 14, Fig. 15).

Sort[A] button sorts the speakers by the proximity of the integrated sound characteristics [A] (see Fig.16)

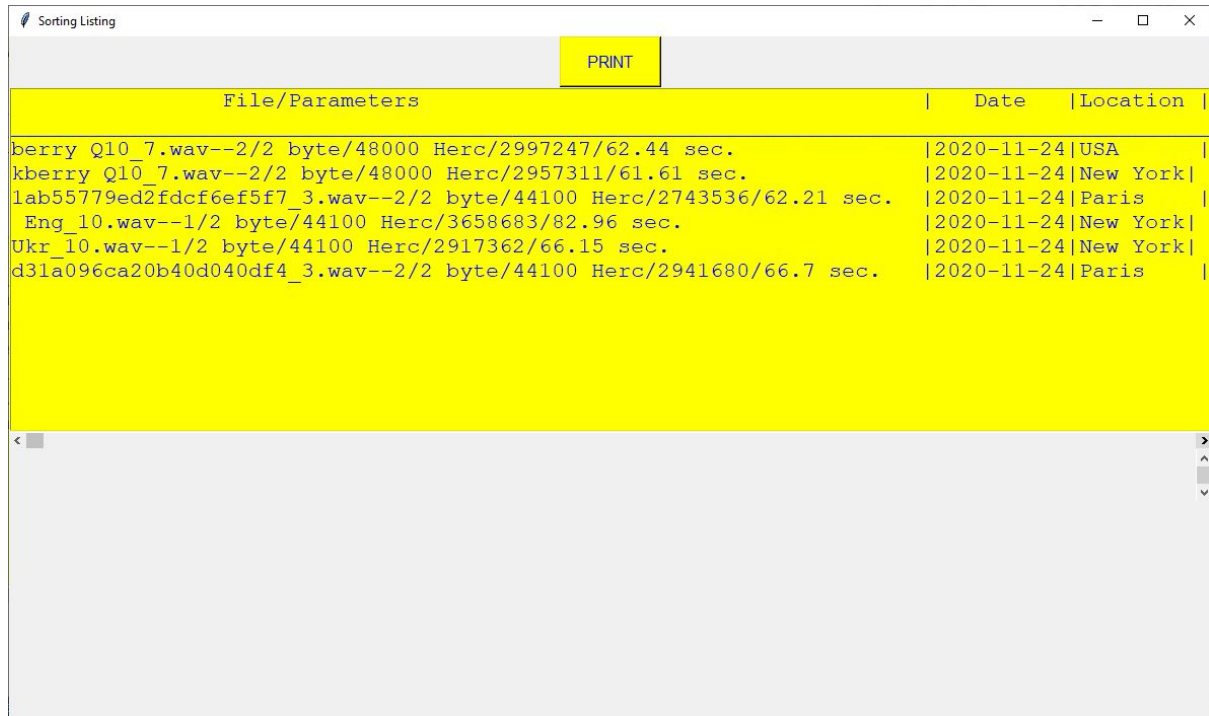**Sort[I]** operation button is not used in the current version.

Note
To sort the records, select the record relative to which the sorting is carried out and click on one of the sorting buttons.

Fig. 14 Sorting Listing window: sounds [A] and [I]

Fig. 15 Sorting Listing window: sounds [A] and [I]

Fig. 16 Sorting Listing window: sound [A]

The use of two sorting options is determined by the analysis options for short records. For recordings lasting less than a second or tens of milliseconds (for example, "cut out sound"), it is possible that either [A] or [I] sound is absent or not representative enough.

The System solves this issue automatically. If there is only [A] sound in the selected recording, the program will not allow sorting. To check if this is the case, try sorting in this embodiment only by the sound [A]. If there are no components of both sounds in the record, the System will give a notification that sorting relative to this record is impossible.

In the current version of the System the greatest efficiency of sorting records is achieved by the combination of two vowels.

Note

When sorting by the selected record, some short records may be skipped if they do not contain statistically significant components of one or another sound. For example, Fig. 14-16 show sortings by the same record, but the number of sorted files is different. This is due to the presence in the sorting listing of several files with a duration of 60-100 ms, which contain only the [A] sound. However, in the sorting listing for [A] sound only, these files are still included.

The second column in the sorting listing is the probability of a mistake of the first kind when comparing two speakers. This value depends on the sorting option, the duration of the phonogram, the language of the phonogram, and a number of other factors. The system automatically makes decisions on these parameters. For the description of the settings for the decision making process see System Options section.

IMPORTANT If there are files in the identification and verification tasks with a sampling frequency of less than 11025 Hz (for example, 8000 Hz), the system automatically implements a recalculation of the characteristics comparison at a frequency of 8000 Hz. Recalculation is carried out by limiting the range of integrated spectra of sounds to 4000 Hz.

This is also applicable for files with different sampling rates for the task of identifying speakers. For the verification task, when you select a characteristics file with a sampling frequency > = 11025 Hz in the listing, the system skips all sorted characteristic files with a sampling frequency of 8000 Hz when sorting. If a file is selected for sorting with a sampling frequency of 8000 Hz, then all files with a higher sampling frequency will be taken into account in the sorting listing.

Sorting Listing Fields:

N - index number

Error,% - the probability value of the error of the first kind in percent. Using this value, records are sorted by the proximity of the characteristics of the voices.

File / Parameters - file name, number of channels, bit depth (bytes), sampling frequency, file size (bytes per channel), duration (sec.).

Date - date of the phonogram recording

Location - the place of receipt of the phonogram.

Sorting listings can be saved in an editable MS Word document template. Click the PRINT button to print or save the sorting listing.

IMPORTANT In most cases, the value of the error during verification cannot serve as a measure of the probability of the error of identification of two speakers in the sorting list to make a decision. If a sorting list contains several thousand or even several hundred records, the number of similar characteristics of voices can be significant.

To automate the decision making in case of speaker identification a decision threshold for the magnitude of the error is used. However, this threshold in most cases will be subjective.

LISTING VERIFICATION

N	Error, %	File/Parameters	Date	Location
1	0.000000000	1 John_Eng_Blackberry Q10_7.wav--2/2 byte/48000 Herc/2997247/62.44 sec.	2021-01-26	Unknown
2	0.103998064	2 John._Rus_Blackberry Q10_7.wav--2/2 byte/48000 Herc/2957311/61.61 sec.	2021-01-26	Unknown
3	58.42431259	FILE-1bbea56ee621ab55779ed2fdcf6ef5f7_3.wav--2/2 byte/44100 Herc/2743536/62.21 sec.	2021-01-26	Unknown
4	71.86561584	Richard iPhone 5 Eng_10.wav--1/2 byte/44100 Herc/3658683/82.96 sec.	2021-01-26	Unknown
5	72.34016418	Richard iPhone 5Ukr_10.wav--1/2 byte/44100 Herc/2917362/66.15 sec.	2021-01-26	Unknown
6	85.64199829	FILE-7cf238c0ac4d31a096ca20b40d040df4_3.wav--2/2 byte/44100 Herc/2941680/66.7 sec.	2021-01-26	USA

Fig. 17. Printing Verification Listing

Date and **Location** operational buttons are the same as corresponding buttons in the Identification section.

Filter by Date option is for preliminary filtering before sorting phonograms by time.

Fig. 18 Filtering phonograms by time

Filter by Location option is for preliminary filtering by the place of receipt of phonograms. To filter the records select the place of receipt, then click on Select apply the filter.
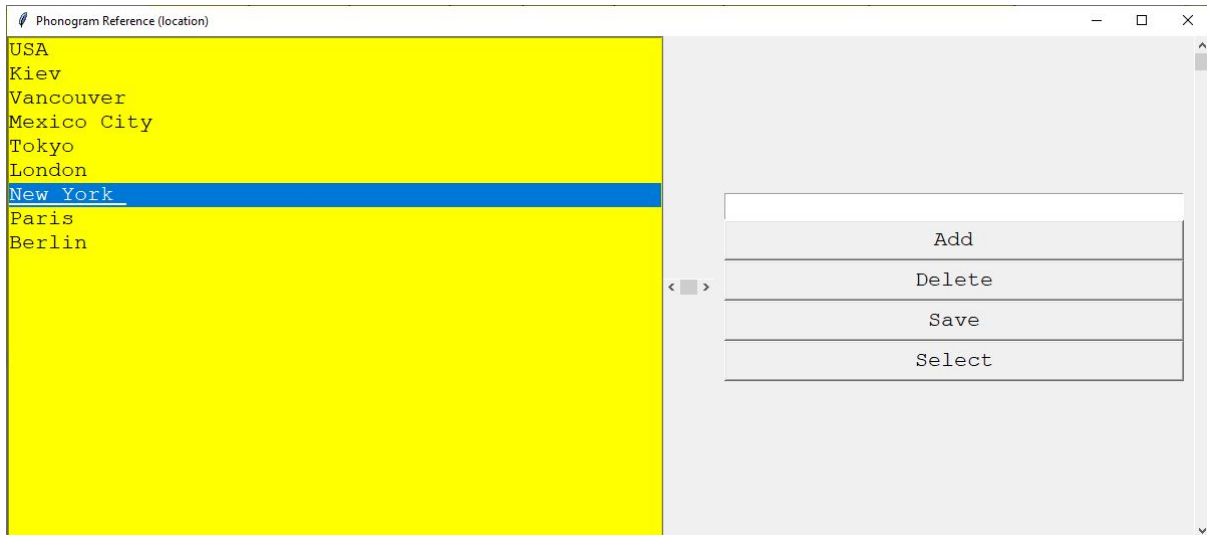

Fig. 19 Filtering the phonograms by the place of receipt

Delete File button deletes files with identification and verification parameters from the corresponding directories. Files in the identification and verification directories are part of the built-in relational database. Removing them directly (with the Windows options) can lead to a violation of the integrity of the system.

To properly delete the files manually using Windows File Explorer, delete all three files with the same name and three different extensions (.svi, .svis, .svid).

4.3 System Parameters

System parameters menu Includes three sections: System Parameters, Phonogram Reference (location), and Error Plots.

4.3.1 System Parameters

Steps:

1. To open System Parameters, click on System Parameters → System Parameters in the main menu.
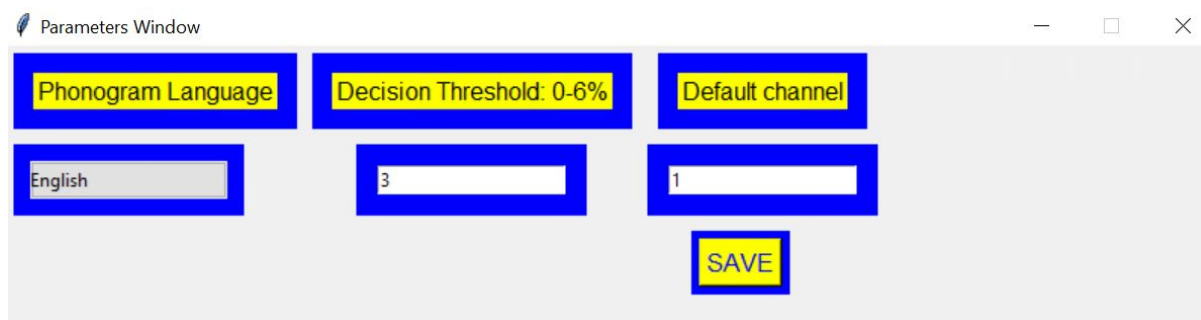2. Parameters Window opens.

Fig. 20. System Parameters window

System Parameters form sets three important system parameters: Phonogram Language, Decision Threshold for identification, and Default Channel.

3. Make the changes to the System parameters.
4. Click on the Save button.

4.3.1.1 Phonogram Language

In the current version of the System, the phonogram database was tested for four language options: English, Russian, Ukrainian and a mixed language group. For each of these options, specific error graphs of the first and second kind were created and used to automate the decision making process for speakers identification.

The probability of an identification error depends on the language group. This parameter has to be selected to optimize the precision of identification.

During the verification process this parameter affects the absolute value of the error only , which is indicated in the sorting listing.

4.3.1.2 Decision Threshold

To better understand how the Decision Threshold operates, refer to the example below.

Consider a graph of errors of the first and second kind for the Ukrainian language phonograms, with a duration of phonograms from 5 to 15 seconds, the sounds [A] + [I] (see Fig. 21)
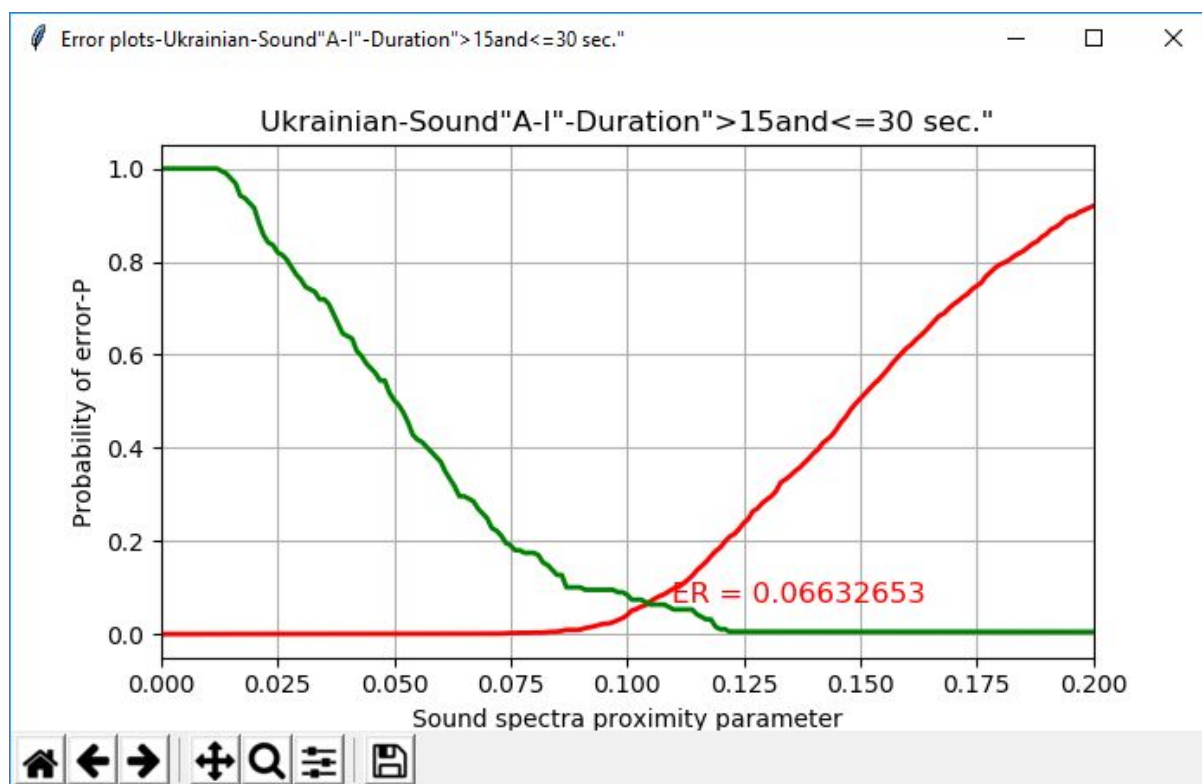
Fig.21. Graph of errors of the first and second kind.

The intersection point of the error graphs is 6.66%.

The System provides four types of decisions:

1. **The characteristics of the speakers' voices are identical.** This decision for two audio recordings is made when the proximity of their characteristics is less than the decision threshold and the probability of a decision error is indicated.

The probability of error depends on the proximity of the vowel spectra of the two phonograms. The error can be very small - less than 0.1% (0.001 in absolute terms of probability) or even less than 0.01%. All probabilities in this case are determined from the graph of type I errors (red graph).

2. **The characteristics of the speakers' voices may be close.** This decision for two audio recordings is made when the closeness of the vowel spectra is more than the decision threshold, but less than the intersection point of the error graphs. For example, the error value is 3% (0.03). All probabilities in this case are determined from the graph of the type I errors (red graph).

3. **Voice characteristics may vary.** This decision for two audio recordings is made when the proximity of the sound spectra is more than the decision threshold, but less than the intersection point of the error graphs. All probabilities in this case are determined from the graph of the second type errors (green graph).

4. **The characteristics of the voices are different.** This decision for two audio recordings is made when the proximity of the sound spectra is less than the decision threshold. All

probabilities in this case are determined from the graph of the second type errors (green graph).

Thus, the decision whether the speakers' voices are identical or different is based on the subjectively accepted value: the threshold of decision-making. The subjectivity of setting the threshold value is due to the physical nature of the identification problem..

The preliminary recommended threshold value is 3% (0.03).

Recommended threshold values based on the duration of the phonograms:

- Over 30 seconds - 3% (0.03)
- From 15 to 30 seconds - 4% (0.04)
- From 5 to 15 seconds - 5% (0.05)
- From 1 to 5 seconds - 6% (0.06)
- Less than 1 second - 6% (0.06)

If the length of the phonograms is different, it is recommended to set the threshold for decision-making by the length of the phonogram with minimum duration.

The second value of probability - the point of intersection of graphs - is a more objective factor in decision making, because it is based on extensive testing of hundred thousands of phonograms in the specifically prepared data sets.

4.3.1.3 Default Channel

When analyzing phonograms, one recording channel is analyzed. If there are several recording channels, it is necessary to indicate which recording channel will be analyzed by default.

4.3.2 Phonogram Reference (location)

To add new locations or delete locations from the list of available options for phonogram editing:
1. Click on System Parameters → Phonogram Reference (location).
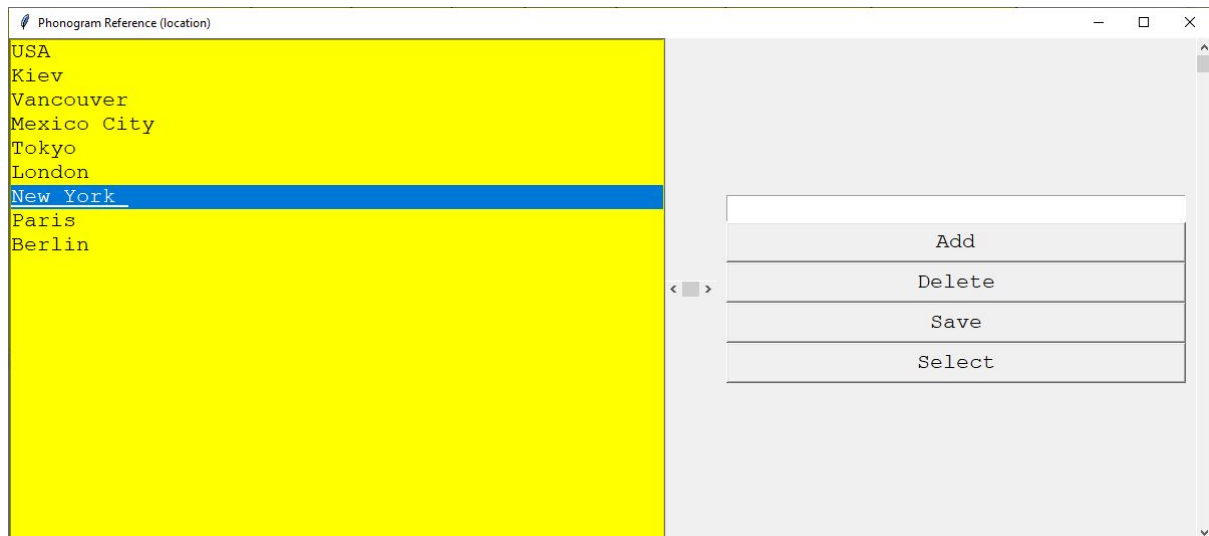2. Phonogram Reference (location) window opens:

Fig. 22. Phonogram Reference - Location

3. To delete the location select it from the list and click on the Delete button.
4. To add the location type in the white space and click on the Add button.
5. Click on the Save button to apply the changes.

4.3.3 Error Plots

Error Plots menu option shows error graphs for the following languages:

- English
- Mixed Language Group
- Russian
- Ukranian

The sound options are:

- Sounds [A-I]
- Sound [A]

To view the error graphs:

1. Click on System Parameters → Error Plots
2. List_Fono window opens.
3. Select the phonogram language and sound from the corresponding lists and click on the Save button.

Fig. 23. Error graphs.
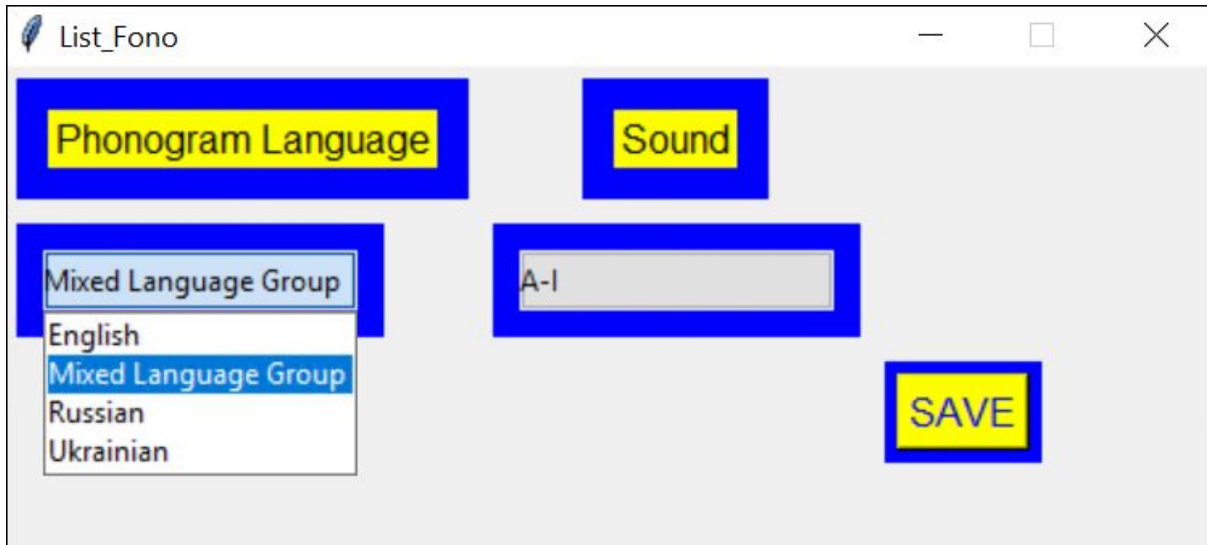
4. The error graphs open, each in a new window. Available time ranges are:
    - < 0.5 sec.
    - > 0.5 <= 3 sec.
    - > 3 and <= 5 sec.
    - > 5 and <= 5 sec.
    - > 15 and <= 30 sec.
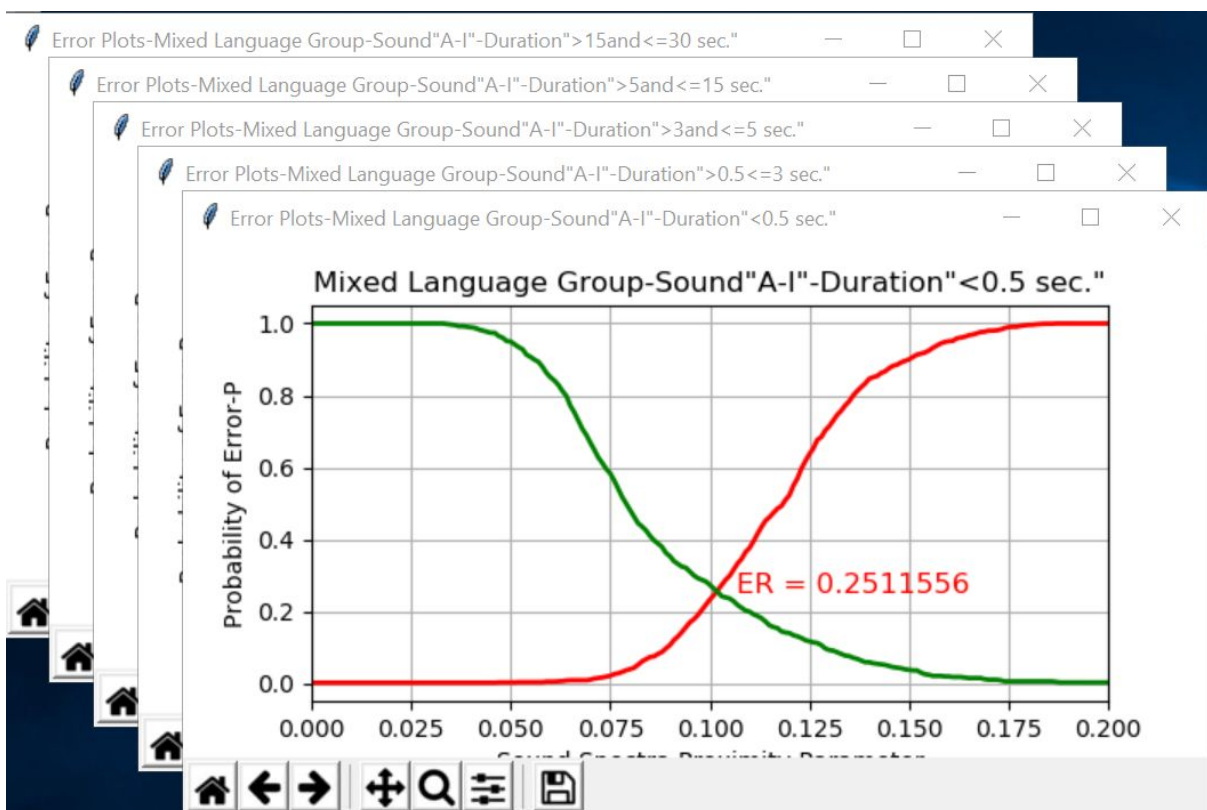5. To save any of the graphs click on the Save icon at the bottom of the individual graph window (see Fig. 24):

Fig. 24 Error Plots graphs

4.4 Language

This menu option changes the language localization of the system.

Current version of the system provides three localizations: English, Russian, and Ukrainian. All the graphs and forms are displayed in one of these three languages.

At the request of the Customer, any localization of the system can be added.

To select the localization:

1. Click on Language in the main menu.
2. Click on one of the languages in the list.

5 Publications

1. Soloviev V. I. Identification of voice characteristics based on the maximums of the wavelet transform. Registration. Zberigannya i drobik Danikh, National Academy of Sciences of Ukraine, Institute of Problems of Information Technologies, Volume 15. No. 2, 2013. p. 39-48

2. Solovyov V. I. A multifractal approach to the problems of voice recognition of a speaker Scheme. nat. un-t ime V. Dahl. - 2013.-№ 18 (207) part 1 - p.203-209

3. Soloviev V. I. Identification of voice characteristics based on multifractal analysis. Newsletter of the Ukrainian National University V. Dahl. No. 6 (195), 2013. - S. 77-85.