silentium systems

(604) 710-2407 Canada
(650) 319-8866 US

SILENTIUM

Software System for Audio Recording Editing Detection and Localization

User Manual

Version 1.0

March 2021

(604) 710-2407 Canada
(650) 319-8866 US

Table of contents

(604) 710-2407 Canada
(650) 319-8866 US

1 Introduction

This version of the Silentium software package (hereinafter referred to as the System) solves the problem of audio recording editing detection and localization.

To identify the distinctive features of audio tampering, a special model was developed based on a deep learning neural network. This model distinguishes fragments of pauses with and without editing with sufficient efficiency for practical purposes. The System automatically detects pauses with signs of editing using the difference in their characteristics compared to non-edited fragments. Potential cuts in pauses (or pauses with cues) are localized with a certain error probability.

The basis of probabilistic detecting and localization of editing in pauses is the graphs of errors of the first and second kind. These probabilities of errors of the first and second kind are used directly for detecting and localizing audio recording edits and cut-ins .

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.

1.1 Source Audio Requirements

- Optimal duration: 200 ms and longer
- Minimum duration: 30 ms
- Maximum duration: unlimited
- Supported file formats: .mp3, .ac3, .aac, .ogg, .wma, .aiff, .asf, .au, .flac, .mp2, .avi, .flv, .mp4, .m4a, .wav (at least 441000 Hz and 16 bit)

1.2 Hardware Requirements

- Processor: from 2 GHz
- RAM: at least 8 GB
- OS: Windows 10 (64 bit)
- GPU with at least 2 GB of graphics memory and CUDA applications installed to enable computation parallelization

2 Principles of detection and localization of audio recording tampering

The development ideology behind the System is based on deep learning neural networks. A specially developed software package analysed a set of phonograms with pauses of various lengths. These pauses were automatically cut from various phonograms to form the primary database of pauses.

At the next stage, a secondary database was formed from fragments of pauses with editing and pauses without editing. The duration of these fragments is 20 ms. Fragments of pauses with editing were prepared using a special software module. This module randomly cut out

small fragments of pauses from different pauses and then produced edited 20 ms pause fragments. Overall, the software module created a database of millions of 20 ms fragments.

This database was the original Dataset for the deep learning neural network. The training of the neural network was carried out within the framework of the problem of binary classification into fragments of pauses with and without editing. The resulting model of classification of pauses fragments became the basis of the System.

An important feature of the System is the probabilistic characteristics of solving problems of detecting and localizing audio tampering. All probabilistic characteristics in the form of graphs of errors of the first and second kind were determined based on test arrays using previously prepared experimental data. The purpose of the tests was to plot the errors of the first and second kind. Error graphs are automatically used in the System to detect and localize editing in audio recordings pauses.

2.1 Main features

The main technological features of the SILENTIUM system are:

1. Solving the problems of detecting and localizing editing is carried out on the basis of the developed model using a deep learning neural network.

2. The system automatically selects fragments of speech with no signs of editing and indicates pauses with editing along with its potential localization.

3 System Installation

3.1 Operating System

The System requires Windows 10 (64 bit version) to be installed.

3.2 Installation Package

The System installation packages includes the following:

1. The executable file of the system: Silentium_08_2021_03.exe,
2. Files and directories that support the operation of the System,
3. User manual.
4. Editing catalog containing a number of audio files with edits for testing purposes. All edits are described in Editing - Table Editing file.

3.3 Installation

Follow these steps to install the System:

1. Extract and copy the Silentium’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 Silentium_08_2021_03.exe file, run the file Read_Activation_08_2019_FL.exe and send the key to the System Supplier.
3. After receiving the activation file File_rdkd.svi from the Provider, copy this file to the folder with Silentium_08_2021_03.exe file and replace previous File_rdkd.svi file.
4. Launch the System by running Silentium_08_2021_03.exe file directly or by using the desktop icon.

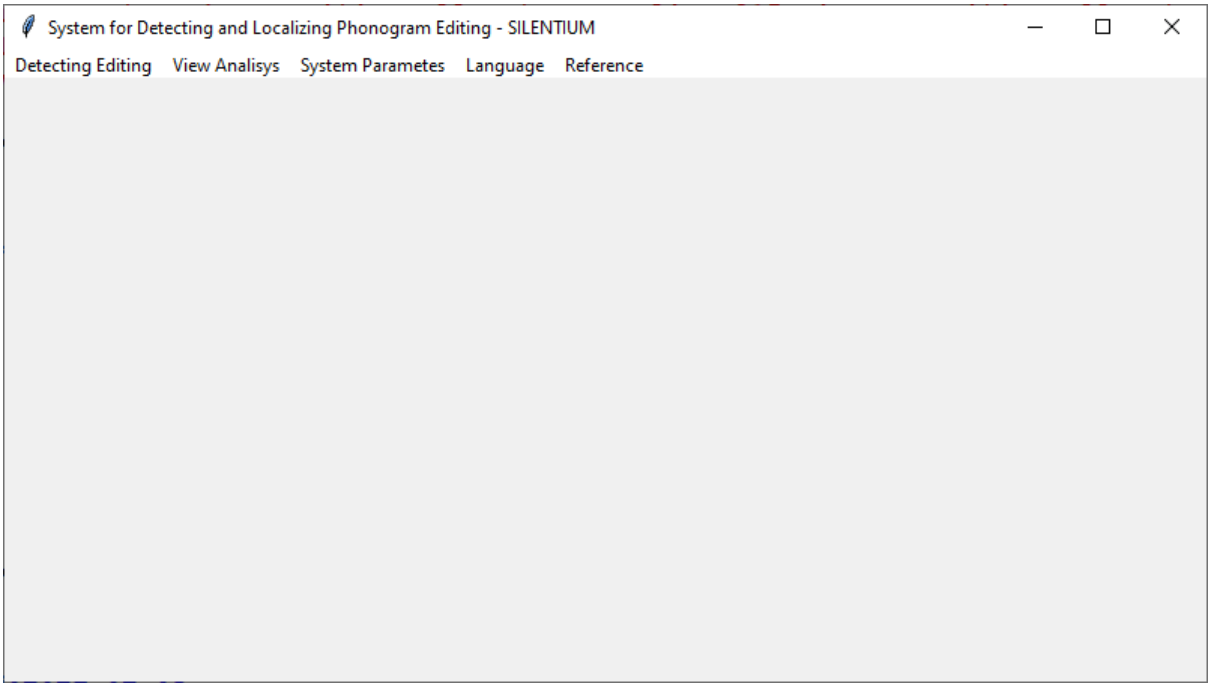

Fig. 1 Silentium System menu window

4 Main Menu

4.1 Detecting Editing

This menu includes four sections: Detection Editing, Error Graphs, Error Graphs (complete), One Pause.

4.1.1 Detecting Editing

This option determines all the necessary parameters for detecting and localizing tampering in a particular phonogram. The analysis results are recorded as files in the Analysis directory, with file names identical to the original ones and .sve file extension.

If it is necessary to further use different groups of processed phonograms in the analysis, copy the Analysis folder to another directory. Then delete all content and carry out a new accumulation of phonogram characteristics.

Steps:

1. Click on Detecting Editing → Detection of EDITING (automatic)
2. A window for selecting a .wav file for analysis opens:

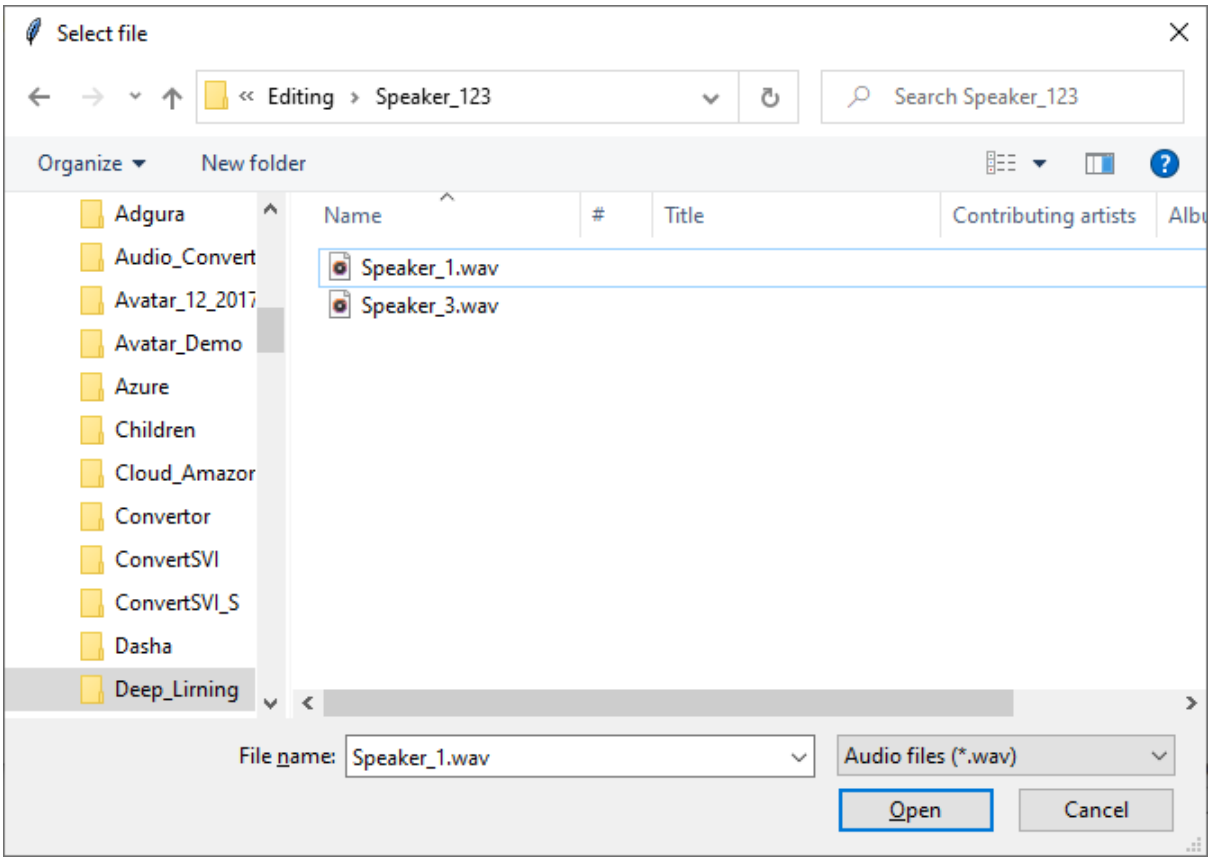

Fig. 2 Window for selecting a file for analysis

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

4.1.2 One Pause

The allocation of pauses for automatic analysis is based on the set threshold of the audio signal modulo. It is a configurable system parameter and can be changed in the System Parameters menu.

The default value is less than 0.03. The optimal threshold value for pause detection depends on the noise level. When analyzing pauses, the system skips pauses that have exceeded the threshold. This can happen for almost any threshold value and it is especially true for pauses that contain a series of high frequency pulses with a high audio signal amplitude.

One Pause option is used to analyze omitted pauses and lets users to analyze each separate pause previously cut from the phonogram.

Note

This option analyzes any part of the track which is longer than 30ms, even if this fragment is not a pause. It provides no automatic distinction between pauses and speech fragments.

Steps:

1. Click on Detecting Editing → One Pause
2. Select a phonogram fragment with a pause:

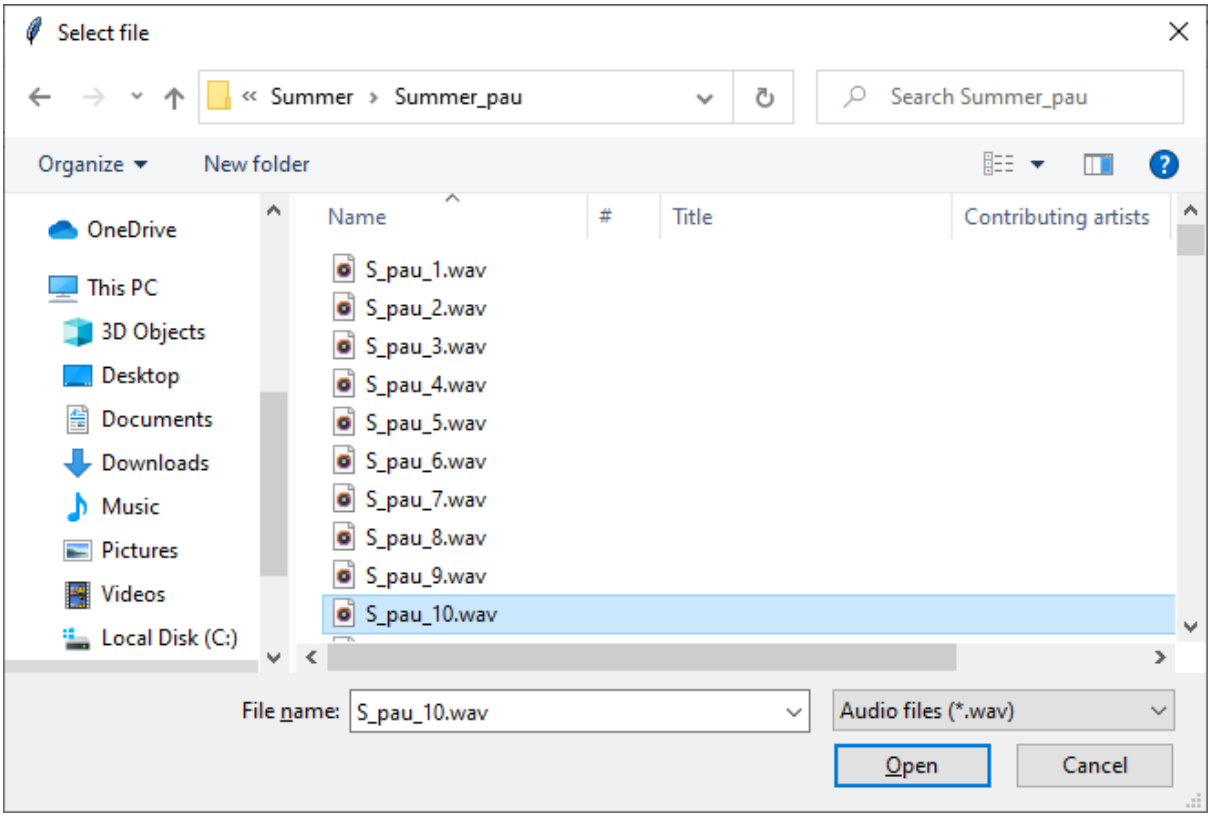

Fig 3. Selecting the file with a pause

3. The parameters of the presence of a potential editing in a pause are calculated and a graphical representation of the analysis results is displayed (see Fig 4, Fig 5)

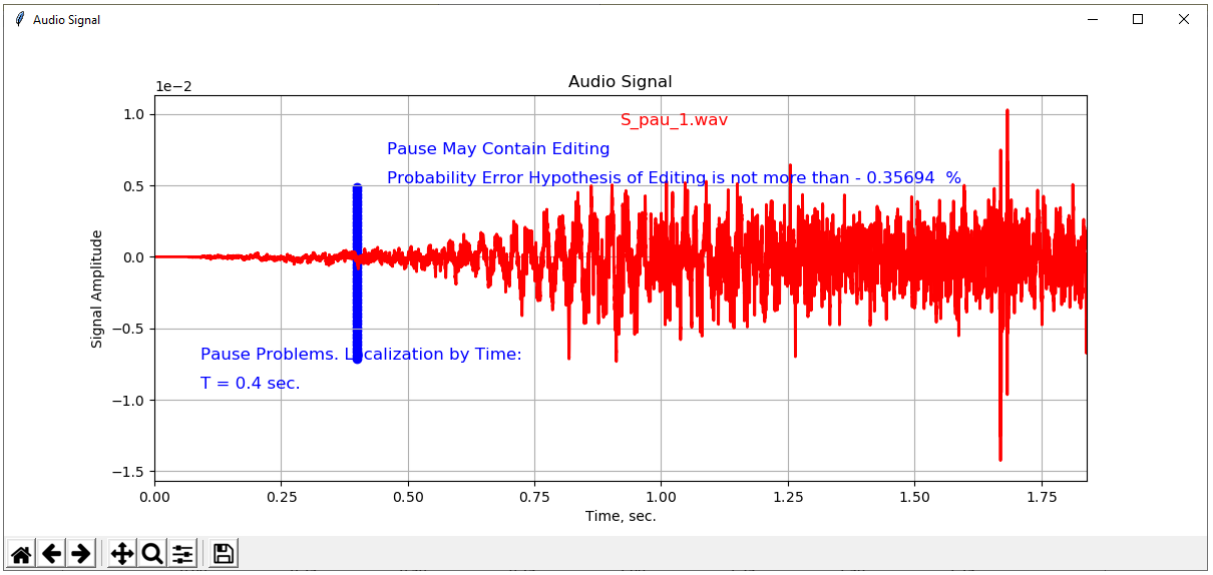

Fig. 4 Audio signal graph for a pause with evidence of tampering

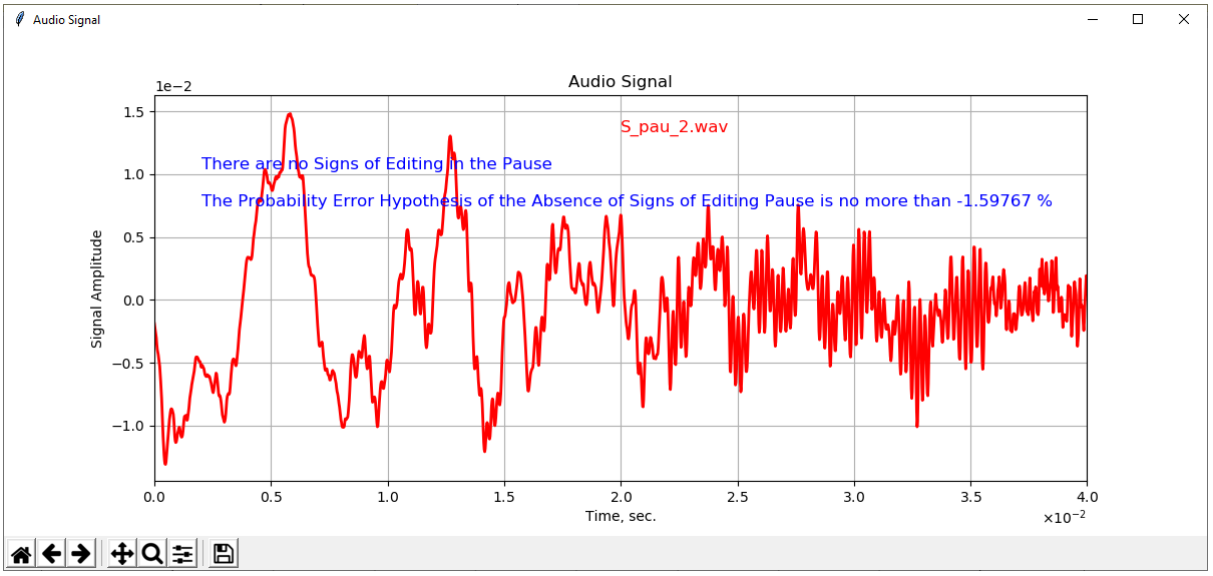

Fig. 5 Audio signal graph for a pause with no signs of editing

(604) 710-2407 Canada
(650) 319-8866 US

4.2 View Analysis

View Analysis menu tab includes three sections - Viewing fragments without editing, Problematic Pauses, Full Soundtrack.

4.2.1 Viewing fragments without editing

The low probability of editing in the 20 ms pause fragment is a very reliable evidence of the tampering absence. The error probability can be very small: less than 0.01% - 0.0001%. For pauses with a duration of up to 100 ms this makes it possible to conclude reliably that there was no editing.

This is the method the System uses to determine which fragments of the phonogram are free from tampering. Note that such unedited fragments can be quite lengthy.

Steps:

1. Click on View Analysis → Viewing fragments without editing
2. A window for selecting a file with .sve extension opens:

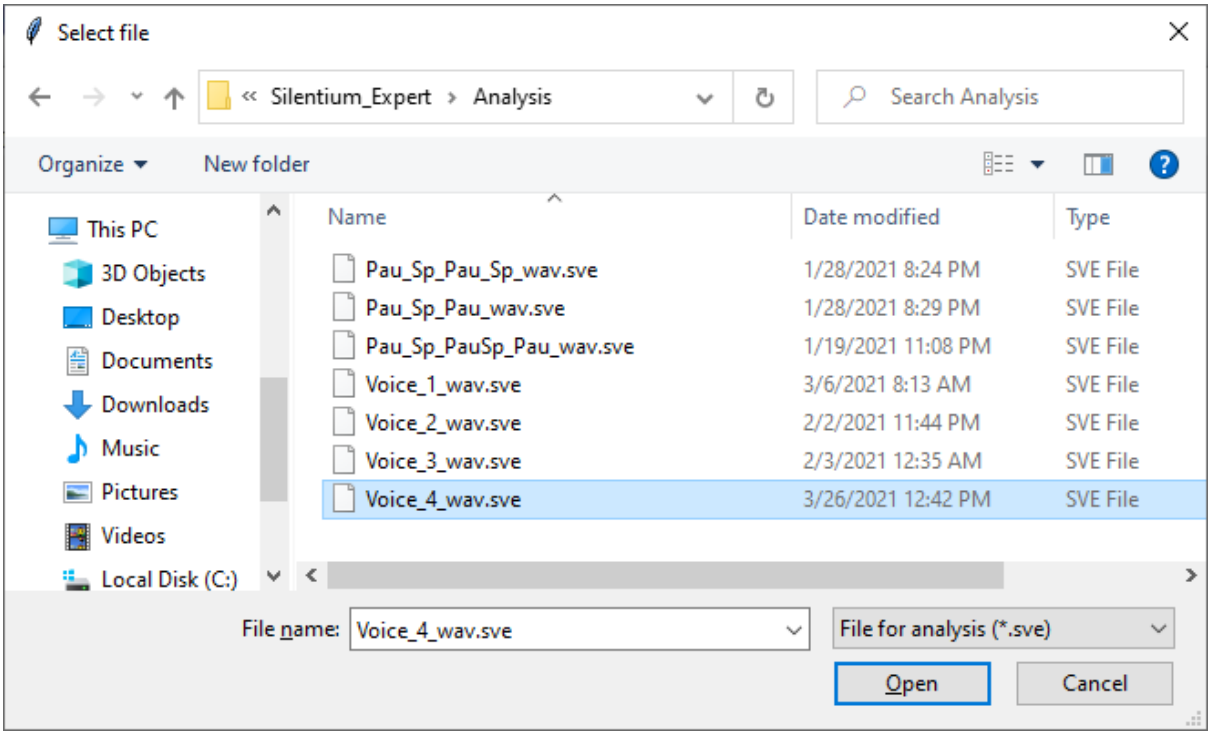

Fig. 6 Select a file with analysis parameters

By default, the directory used is Analysis. If a file with analysis parameters was saved in another folder, open that directory to select the file.

3. Select the file.

Note

When analyzing the entire pause, the System decides if there is editing based on the maximum probability of editing in one of the fragments of the pause.

There may be other fragments of 20 ms in this pause fragment with the probability of editing more than Ppor = 0.9. In view of this, as the tests on pauses with editing show, the localization accuracy of a fragment with editing can be of the order of 30-100 ms in the particular pause.

4. A window of the operating control panel will appear for viewing the analysis results of a fragment of speech of a phonogram without editing:

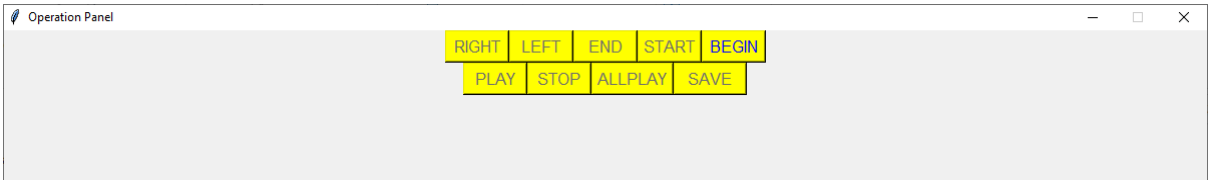

Fig. 7 Operational Panel window for viewing of analysis results

5. Press START to view fragments of the phonogram with no editing

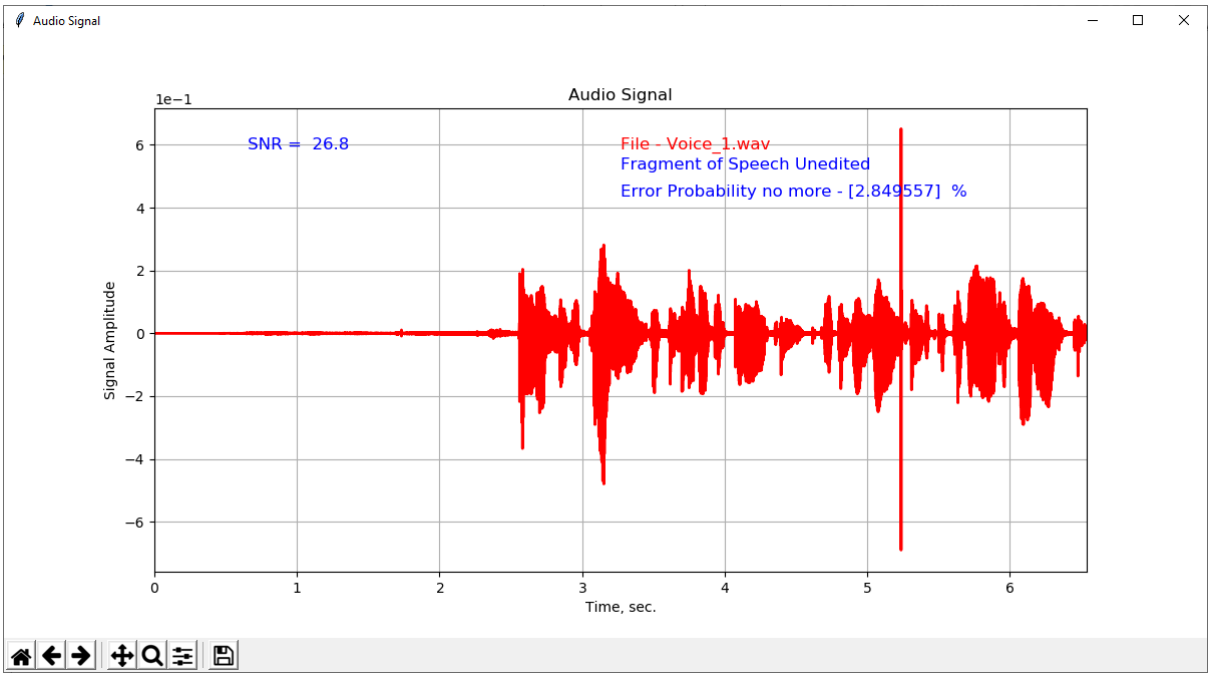

Fig. 8 Viewing fragments of a phonogram with no editing

In the upper left corner of the Audio Signal graph there is SNR (Signal to Noise Ratio) for the entire track.

Operational buttons for View Analysis

RIGHT provides navigation through fragments of speech of a phonogram without editing until the end of the phonogram.

LEFT provides navigation through fragments of speech of a phonogram without editing until the beginning of the phonogram.

END shows the last speech fragment in the phonogram without editing.

START shows the first speech fragment in the phonogram without editing.

BEGIN opens the analysis window.

PLAY plays a speech fragment of the phonogram without editing.

STOP stops playing the phonogram speech fragment without editing.

AllPLAY plays the entire phonogram.

SAVE saves a speech fragment without editing in a file. wav format:

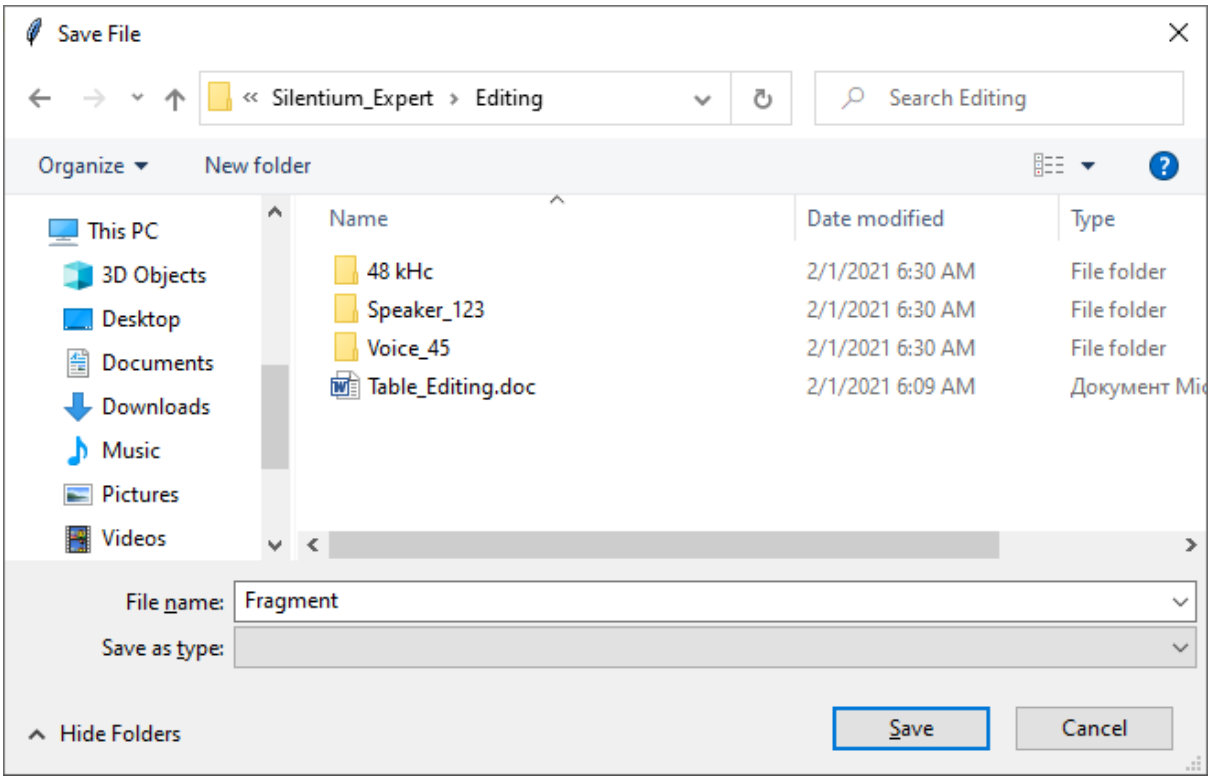

Fig. 9 Saving a speech fragment

Refer to Fig. 10 and Fig. 11 for more information on the navigation options:

(604) 710-2407 Canada
(650) 319-8866 US

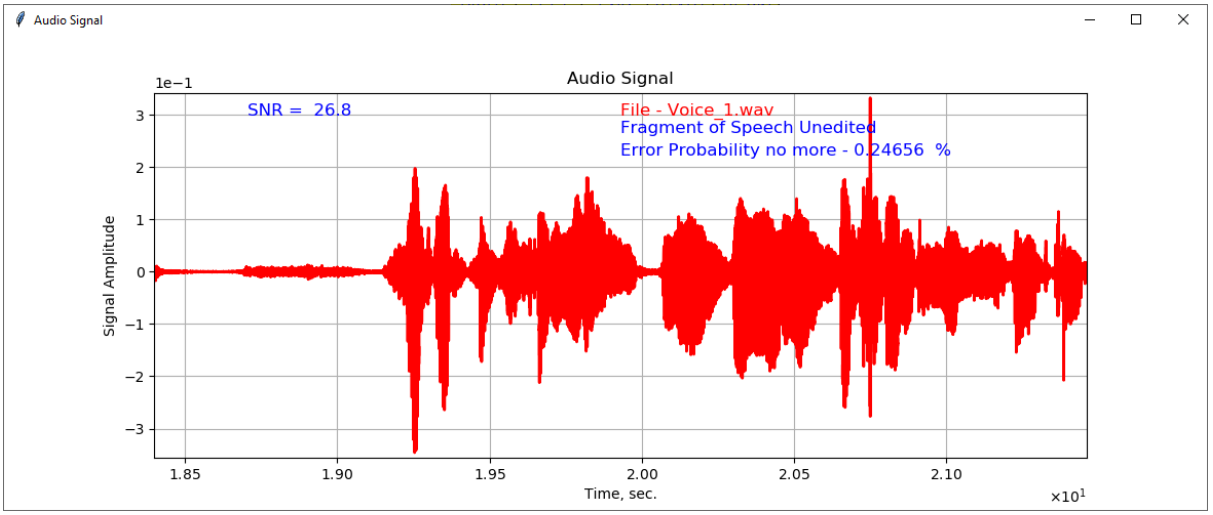

Fig. 10 Viewing speech fragments without editing

Note

Fragments containing speech begin and end with pauses. Although the fragments themselves are editing free, the pauses at the beginning and at the end may contain signs of editing.

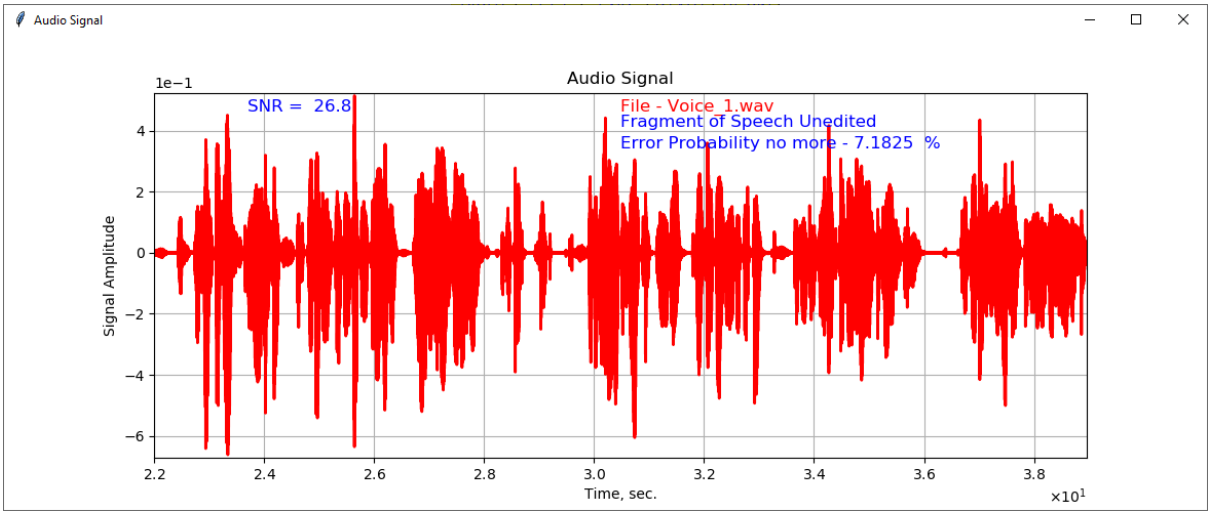

Fig. 11 Viewing fragments of speech without editing

The probability in the graphs is determined on the basis of the plots of the errors of the first and second kind. These graphs are based on a large amount of data thus the error probabilities are determined based on a priori information.

The value of the probability is given for one of the pauses of the phonogram speech fragment with the maximum error probability.

4.2.2 Problematic Pauses

This section analyzes pauses that contain signs of editing.

Steps:

1. Click on Click on View Analysis → Problematic Pauses in the main menu.
2. A window for selecting a file with phonogram parameters opens
3. Select the file.
4. A window of the operating control panel opens for viewing the pause analysis results with signs of tampering.

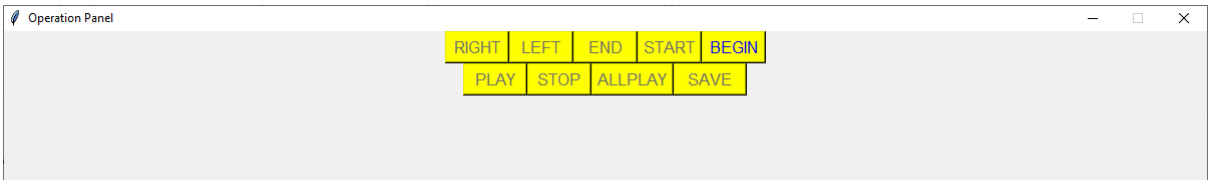

Fig. 12 Operation panel window for managing the results view

Refer to Viewing fragments without editing - Operational buttons section for the functional description of the buttons.

5. Click BEGIN to open the analysis window:

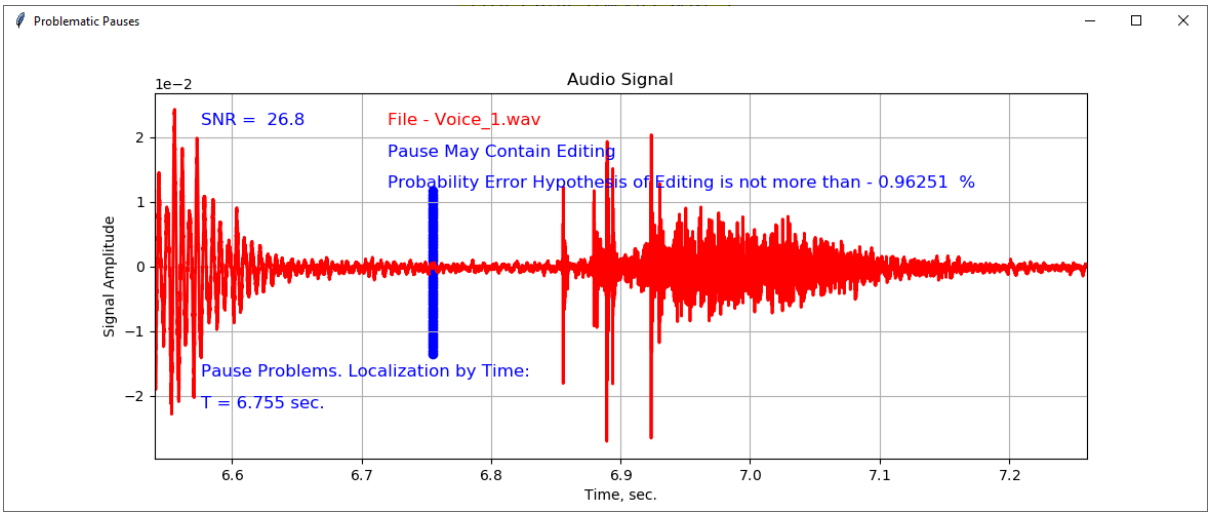

Fig. 13 Viewing a pause with signs of tampering

Note

The pause can contain several fragments with evidence of editing, each 20 ms long. Data is displayed only for the fragment with maximum probability of editing.

To analyze and view other fragments of the pause with potential signs of editing the pause should be divided into parts, and then those parts are analyzed.

4.2.3 Full soundtrack

This option is for viewing all the parts of the soundtrack, cutting fragments and pauses, and for analysis of spectra in a different frequency range of different fragments of the phonogram.

Steps:

1. Click on Click on View Analysis → Full soundtrack in the main menu.
2. A window for selecting a file with phonogram parameters opens:

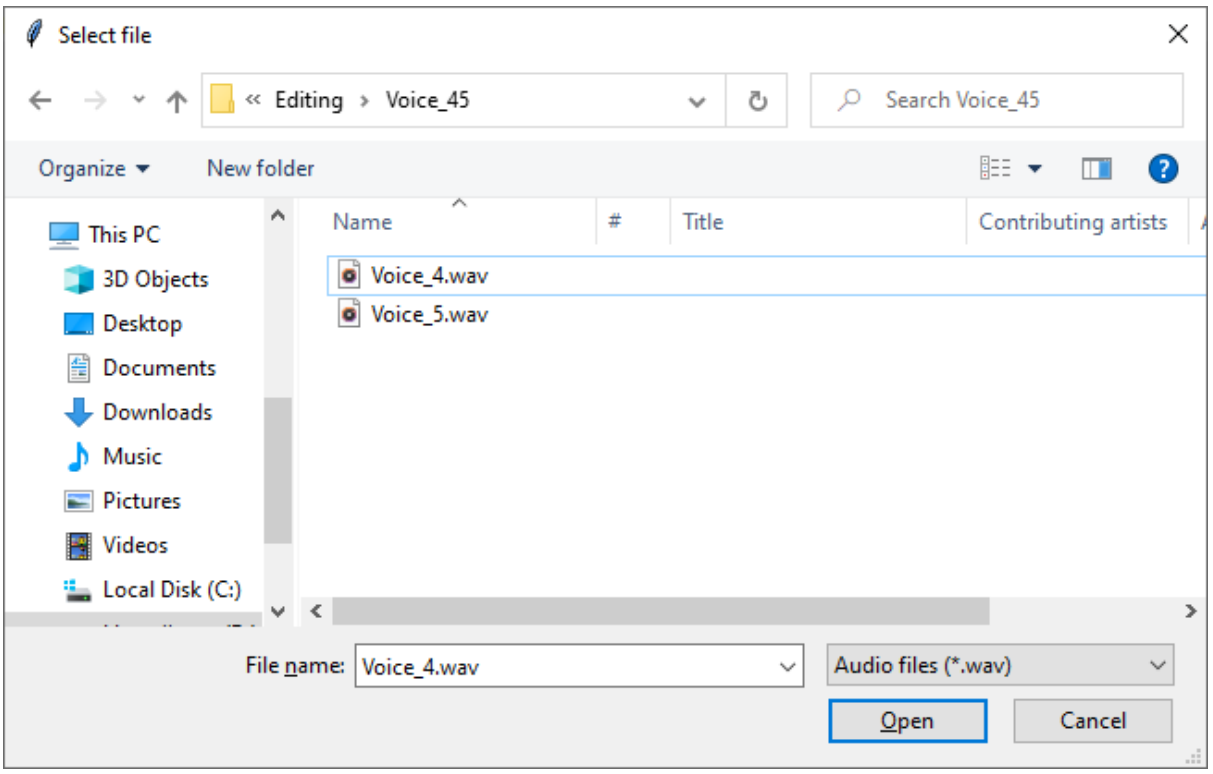

Fig. 14 Selecting a file for analysis

3. Select the file.
4. Operating panel window for working with the phonogram opens:

Fig. 15 Operation panel for working with phonogram

Operational buttons for Full Soundtrack

Begin opens the phonogram analysis window.
Right provides navigation through the track until the end of the file. Initial view window is 20 ms. The navigation shift is equal to half the size of the viewport (see Fig. 16)

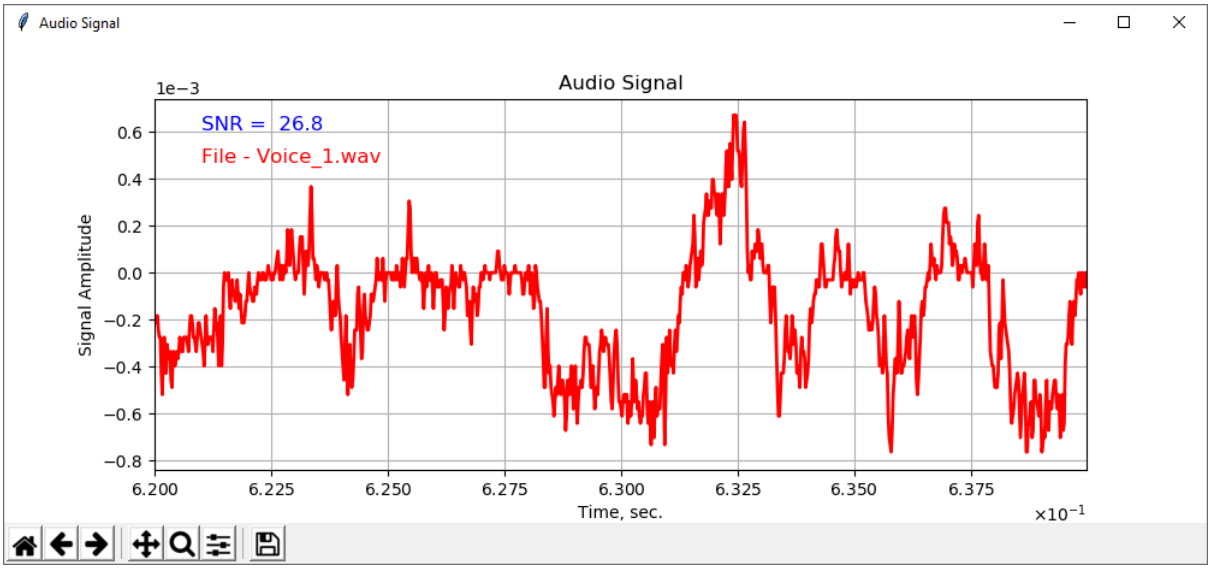

Fig. 16 Audio signal window of 20 ms

Left provides navigation through the track until the beginning of the file.
End opens the audio signal window for the last fragment.
Start opens the audio signal window for the first fragment
Dright makes the viewport twice longer
Left / 2 makes the viewport shorter by half
Play starts playing a fragment of the phonogram
Stop stops playing of the fragment
Spectr toggles the view between the following options: Audio signal only, Audio signal and spectrum, Fragment spectrum only (see Fig. 17 and Fig. 18).

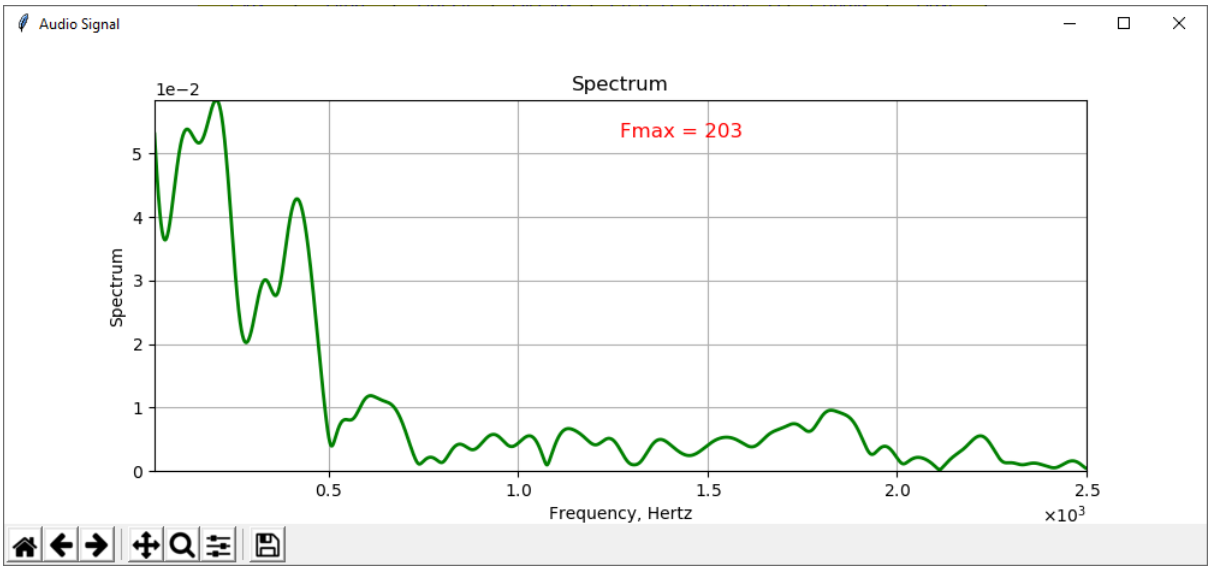

Fig. 17 Spectrum of a phonogram fragment

(604) 710-2407 Canada
(650) 319-8866 US

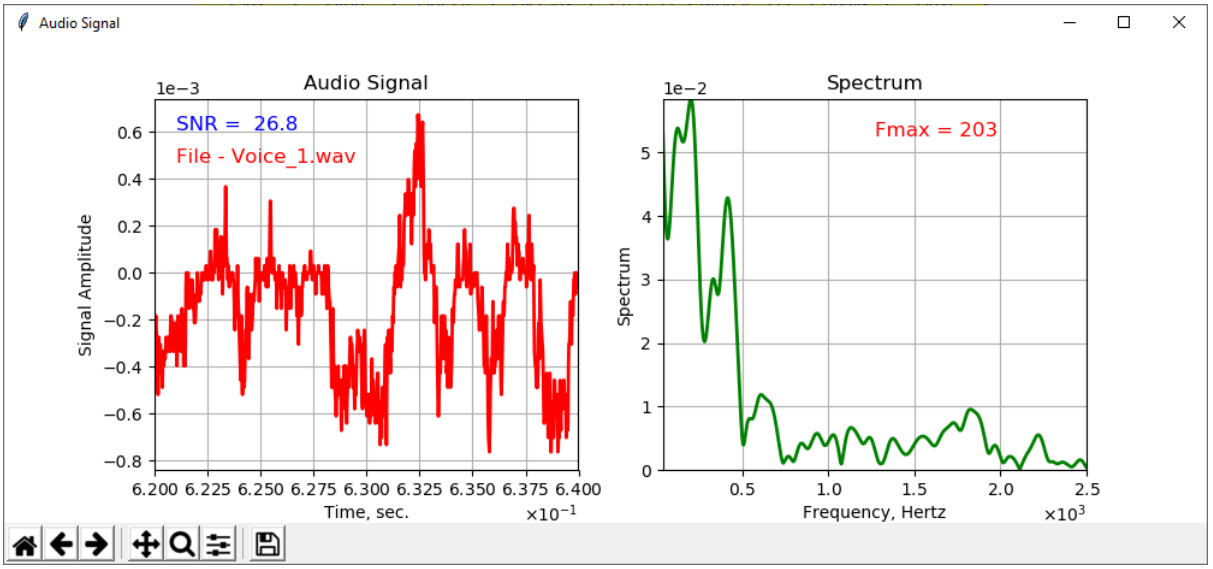

Fig. 18 Audio signal and spectrum of a phonogram fragment

AllPlay starts playing the entire soundtrack

Freq_C selects spectrum viewing range (see Fig 19). Click SAVE to apply your selection.

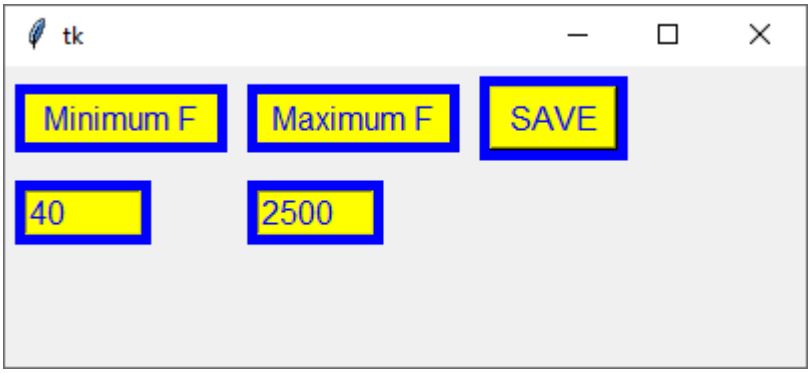

Fig. 19 Selecting the spectrum viewing range (Hz)

Mouse_C enables cutting the selected fragment- with two mouse clicks on the audio signal graph.

Equally sets the “ear” filter for the spectrum according to the characteristics of human auditory perception.

Save saves the selected fragment of the audio track in a .wav file. When saving, you only need to specify the name of the saved file.

(604) 710-2407 Canada
(650) 319-8866 US

4.3 System Parameters

System parameters menu configures the following parameters: Decision Threshold, Parameter Pauses, Pause scan interval, Dynamic Filter, and Pause Minimum.

Additionally, the System Parameters tab contains references to Error Graphs the System uses.

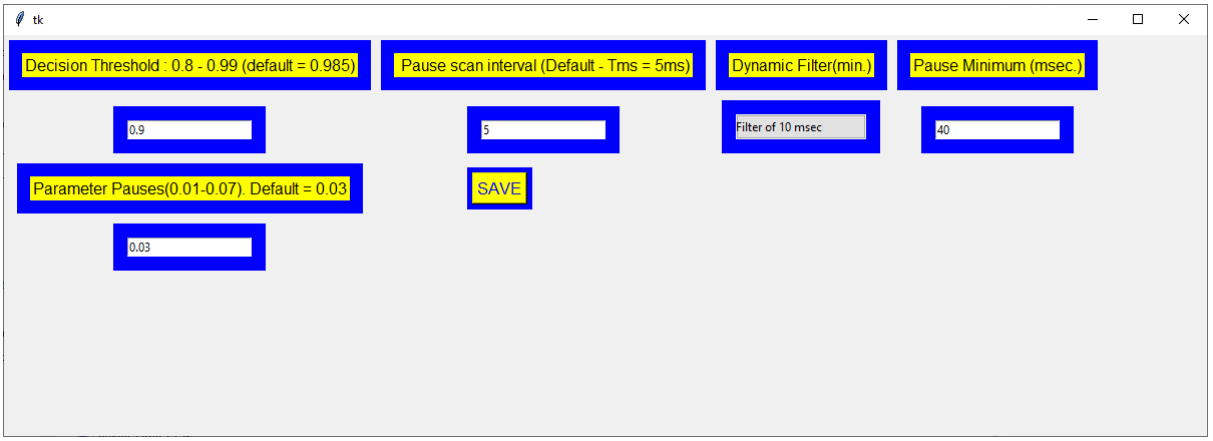

Fig. 20 System parameters window

Decision Threshold

The default value is Ppor = 0.9.

When the threshold is changed, both the efficiency of selecting fragments with editing signs and the number of fragments with false editing signs change.

It is not recommended to change this parameter.

Scan interval

The default value is Tms = 5ms.

It has a significant impact on computing speed and decision making efficiency.

It is not recommended to change this parameter.

Dynamic filter

The default value is 10 ms.

Studies have shown that for a 20 ms fragment, at least two or three consecutive fragments with signs of editing must be present.

When this parameter is set to 10 ms the System uses a filter by fragments with signs of editing from two consecutive fragments.

With a value of 15 ms the System uses a filter by fragments with signs of editing from three consecutive fragments.

(604) 710-2407 Canada
(650) 319-8866 US

There is also a no filter option.

Pause Minimum

By default, the minimum duration is set to 40 ms.

This parameter sets the minimum duration of a pause, which is taken into account when detecting a cut. With pauses 20-40 ms long, editing is unlikely, since the gap between the fragments of speech is very small. Moreover, a series of pauses less than 40 ms long may be pauses within a word or phrase, which also lessens the probability of audio tampering within such pauses.

Parameter Pauses

Default value is 0.03 for an audio signal format in the range -1 to +1.

This parameter defines the threshold as an absolute value of the audio signal amplitude for pause detection. Changing this parameter can affect the efficiency of highlighting and skipping pauses and it is recommended that you use the default value for this parameter.

Research shows that a significant proportion of fragments with false signs of editing have editing fragments at the beginning of a pause or at the end. Increasing the value of this parameter can increase the number of pauses with false evidence of editing.

Note

To apply the changes made to any of the parameters, click on the SAVE button.

Error Graphs

This option opens the graph of errors of the first and second kind:

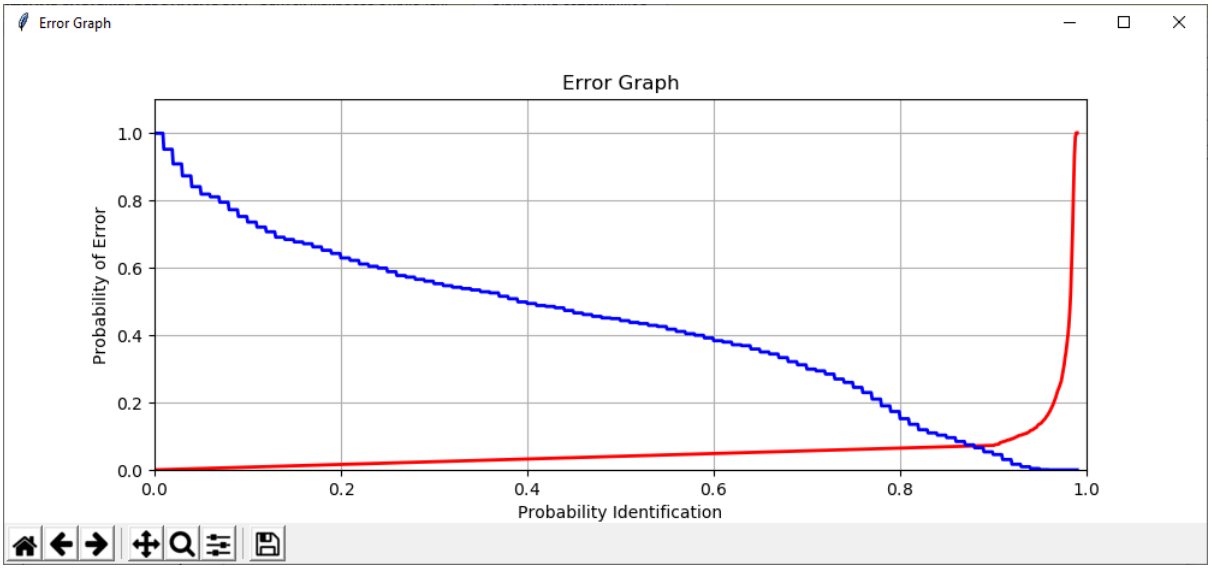

Fig. 21 Graphs of errors of the first and second kind

(604) 710-2407 Canada
(650) 319-8866 US

Error Graphs (complete)

This option opens the graph of errors of the first and second kind with an illustration of the probabilities at the point of intersection of the graphs.

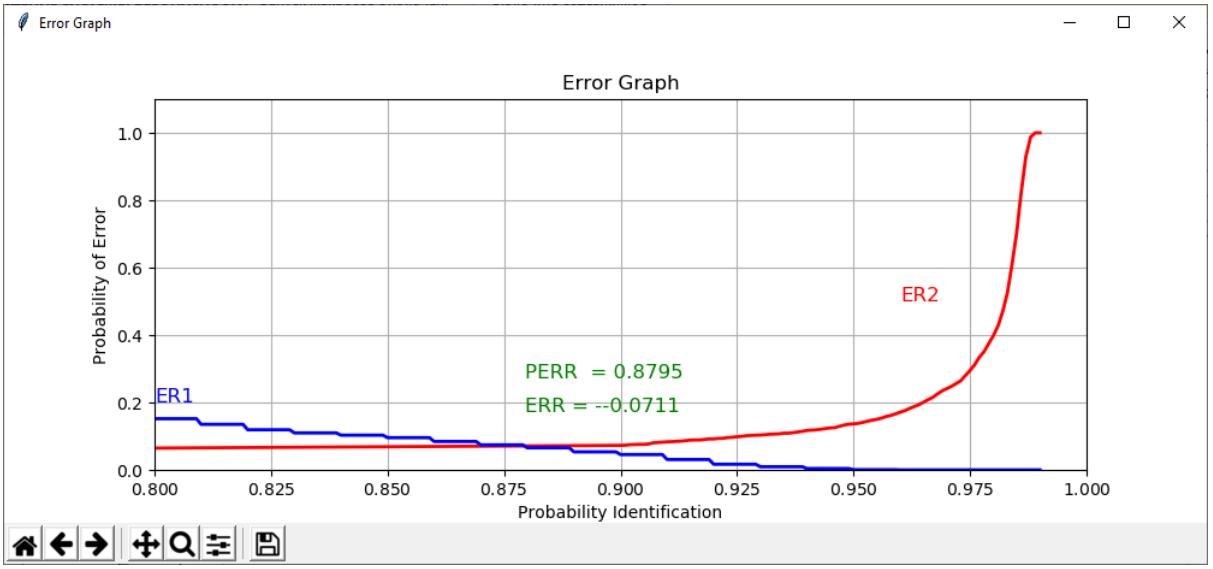

Fig. 22 Graphs of errors of the first and second kind

Error graphs are built based on hundreds of thousands of fragments of pauses 20 ms long that did not participate in the training of the neural network.

Main principles of the Error Graphs

1. The input for the neural network is the spectra of fragments of pauses with and without editing 20 ms long. The deep learning neural network model depending on the spectral characteristics of a particular pause fragment outputs two probabilities: the probability that the pause fragment is a pause without editing and the probability that the fragment was edited.

2. The spectrum of the pause fragments covers the range from 1Hz to the maximum frequency equal to half the sampling frequency of the audio file. As a detailed analysis of the operation of the neural network model shows, the output results are affected significantly by the high-frequency components of the spectra. The visual appearance of changes in the amplitude of the audio signal in the pause fragment rarely has direct correlation with potential editing.

3. The horizontal axis (abscissa) of the error graphs shows the probability of editing in a 20 ms pause fragment. The vertical axis (ordinate) shows the probability of error under the hypothesis (H0) that the pause fragment is a pause with editing for the blue graph and the probability of error under the hypothesis (H1) that the pause fragment is a pause without editing for the red graph.

The probabilistic characteristics of the presence of editing in a pause for this version of the System refer to small fragments of a pause of 20 ms. All pauses in the system are scanned

(604) 710-2407 Canada
(650) 319-8866 US

with a time shift of 5 ms. At each time offset, a 20 ms fragment is analyzed based on the model.

4. Analysis of numerous pauses shows that a good threshold for making a decision to classify a fragment of a pause as a fragment with editing is the probability of identification (abscissa axis) more than Ppor = 0.9. This parameter is set by default in the System. If necessary, it can be changed in the System Parameters.

5. If potential edited fragments are indicated in a long pause fragment with error probability less than 1% (0.01), then results reliability is not sufficiently high. For example, for a phonogram about 1 min long which contains around 10% pauses (6 seconds) scanning all pauses with an interval of 5 ms will produce about 1200 fragments of 20 ms each. With an error probability of 1% up to 12 false negative fragments could be present.

However, the error probability of the System's decisions can be very small: less than 0.01% - (0.0001) which indicates a high level of reliability and precision.

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