

Phonexia Voice Inspector

User Manual



User Manual for Phonexia Voice Inspector - gui version
Produced by: Phonexia, s.r.o.

Product Version:	Phonexia Voice Inspector 1.3
User Manual Version:	1.3
Modified:	28 November 2015
Revision:	201

Table of Contents

1. Introduction.....	2
Typical Use-Cases.....	2
Technical Requirements.....	2
2. Quick Start Guide.....	3
3. Using the Application.....	3
New case.....	4
Open case.....	4
Case view.....	5
Wave editor.....	6
Graphs.....	16
Score table.....	16
Settings.....	17
Speaker identification evaluator.....	20
4. Annexes.....	22
Annex 1. - Disclaimer.....	23

1. Introduction

Voice Inspector (VIN) by Phonexia provides police forces and forensic experts with highly accurate speaker identification tool during investigation of criminal matters. It uses the power of voice biometry to automatically recognize a speaker by voice.

Typical Use-Cases

- Investigation support for police forces and Law Enforcement Agencies
- Quick analysis of 2 or more recordings (1:1 or 1:n) for identical voice/speaker recognition
- Advanced analysis realized by forensic experts
- Voiceprint extraction for future speaker searches

Technical Requirements

Recommended hardware:

- 2.6GHz CPU single core or better, 4GB RAM or better , 64bit platform

Available platforms:

- Windows 64bit (x86_64)
- Linux 64bit (x86_64)
- Other platforms on request

Note: VIN is a standalone application with complete Graphical User Interface (GUI)

2. Quick Start Guide

To use the application, follow these three easy steps:

1. Unpack the archive
2. Copy a license file (obtained during registration and download) to the program's root directory (e.g. next to VoicelInspector.exe for Windows)
3. Run VoicelInspector / VoicelInspector.exe

3. Using the Application

Main features of VIN application:

- Phonexia speech technologies implemented
- Scoring to likelihood ratio, log-likelihood ratio, percentage metrics (0-100%) and verbal ratio
- Detailed report output (expert opinion documentation automatically generated, including scoring tables and graphs)
- Speech files in MS Wave format supported
- Compatibility with the widest range of audio sources possible (applies channel compensation techniques): GSM/CDMA, 3G, VoIP, landlines, etc.
- Applies state-of-the-art channel compensation techniques, verified by NIST evaluation

To start using the application for the analysis purposes, a particular case needs to be selected. You can choose a case by clicking on "File" button at the top left corner, and then on "Open Case". Then you just choose a specific case out of the list, or you can also easily create a new case by clicking on "New Case" button.

New case

1. New case

Fill a name of the case and you can also add some further description.

2. Questioned recording

Add one recording or more of speaker whose identity should be examined. The maximum number of recordings isn't limited, every single recording will be evaluated separately. Note that only speech files in MS Wave format (Lin16, Lin8, Float, Lin8Offset, ALaw, MuLaw) are supported or voiceprints .

3. Suspected reference speaker

Create new speaker whose identity is known add ing some recordings or voiceprints – at least 3 (the maximum number isn't limited) . As in the previous step, only recordings in MS Wave format are supported (Lin16, Lin8, Float, Lin8Offset, ALaw, MuLaw). Questioned recordings will be compared against this speaker. You can of course use an existing speaker if you have already created the profile.

4. Potential population database

In this step the population database should be chosen or created. The population database is a set of recordings of various speakers. The absolute minimum of recordings in a population set is 10, but the recommended minimum is 30 – 40 recordings. The speakers should have as similar voice features as the suspected speaker, so it is necessary to choose speakers of the same gender, language etc. The more similar to the Suspected speaker they are the more reliable results you will get. On the contrary, the less similar to the Suspected speaker the speakers in a population set would be the more files the population set should contain in order to get reliable results.

Make sure that population set does not contain more recordings of the same speaker or any recording of suspected speaker!

Supported file format is still the same - MS Wave format (Lin16, Lin8, Float, Lin8Offset, ALaw, MuLaw) or voiceprints .

The graphical user interface consist of Case view (list of recordings/speakers) located on the left side of the application, 3 main tabs: Wave editor, Graphs, Score table and the main toolbar .

Open case

If you have already created some cases you can open them using the button “Open case” in the main toolbar. It opens a dialog with list of all cases and case details. The case details will be displayed after selecting a case in the list of cases. You can use filter tools for searching in the list of cases and filter by case name (RegExp = regular expression) or by associated speaker. If you want to filter cases by associated speaker you must click on the “selected speaker” and choose a speaker from “Speaker database”.

Case view

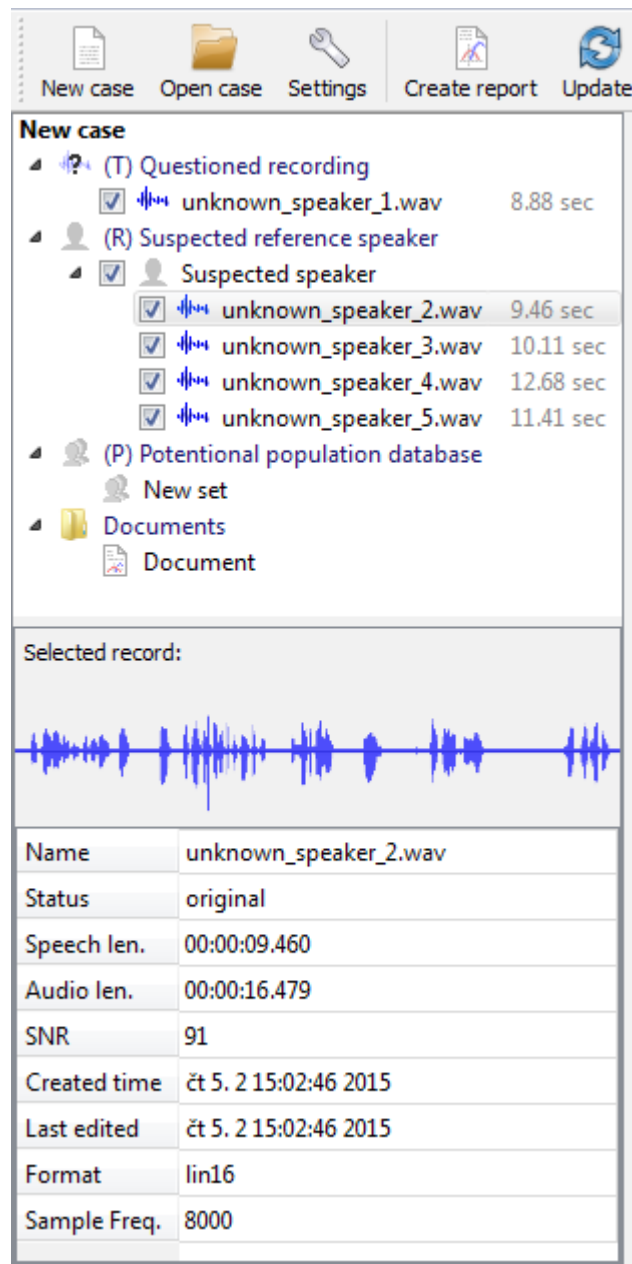
The opened cases are displayed in Case view. It shows a list of recordings/speakers in a particular cases.

In a section labeled as **Unknown recordings**, there are recordings where identities of speakers are not known.

Recordings with a speaker, whose identity is known, are displayed in a section called **Suspected Speakers**.

In a section called **Population set**, there is stored a default set with variety of speakers. This set is used for the purposes of identification analysis, which is based on Bayesian theory.

By right-clicking on a particular section name, the user can add/edit/delete recordings and speakers, or change a population set.



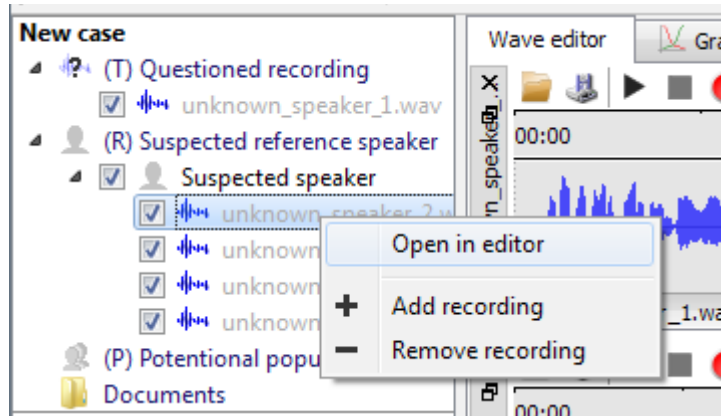
Name	unknown_speaker_2.wav
Status	original
Speech len.	00:00:09.460
Audio len.	00:00:16.479
SNR	91
Created time	čt 5. 2 15:02:46 2015
Last edited	čt 5. 2 15:02:46 2015
Format	lin16
Sample Freq.	8000

Wave editor

With a waveform editor you can easily manage your recordings using several features. You can display several recordings next to each other and easily compare them, play, cut or copy selected parts and save the results

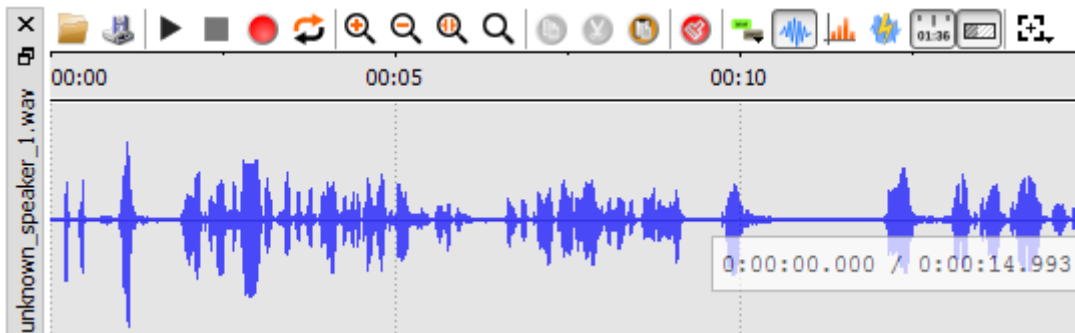
Open a recording

A recording can be opened in Wave editor from the Case view by double clicking or using a context menu (right-click on the item).

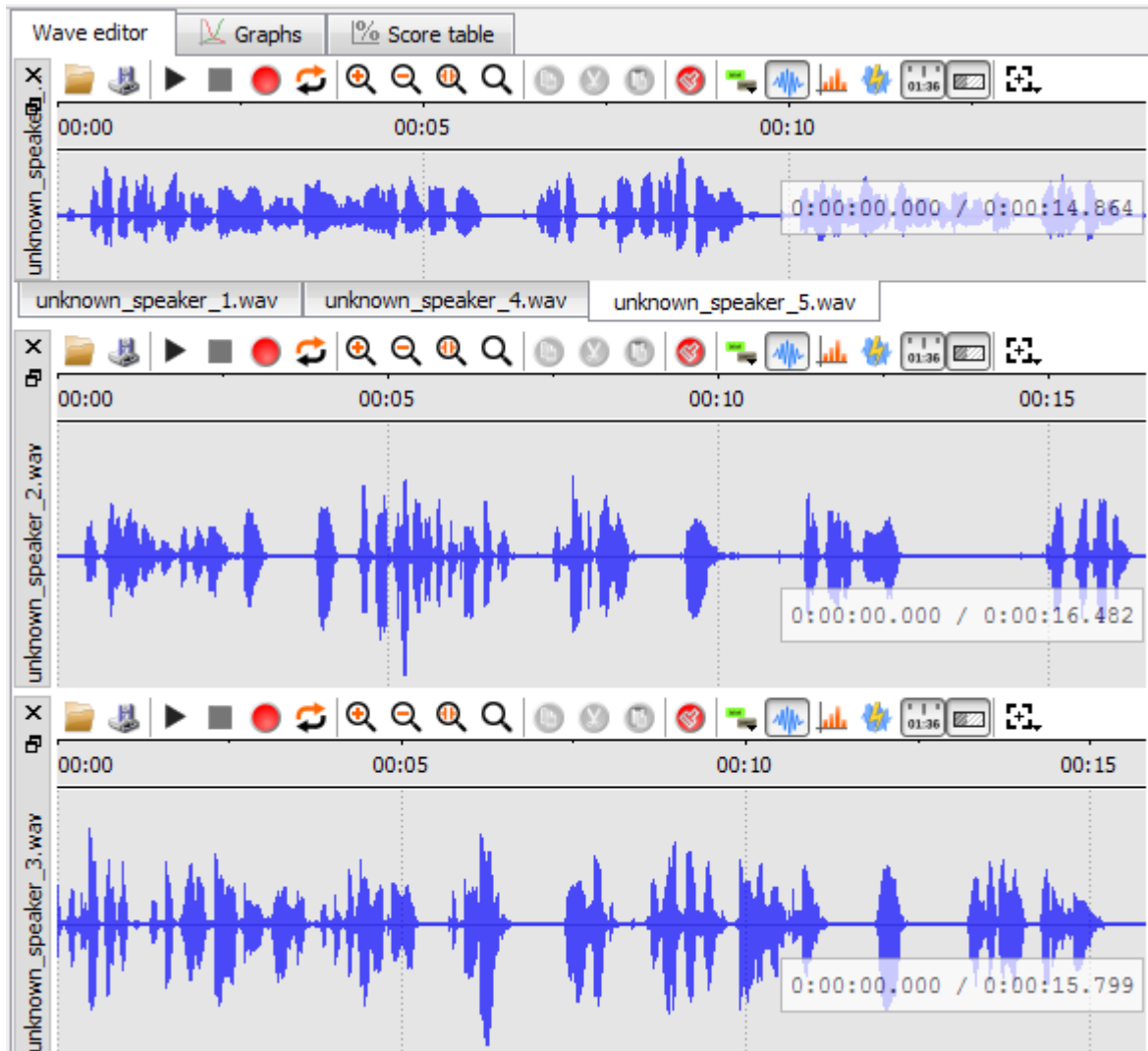


Close a recording

Pay attention that the closing button is situated in unusual upper left hand corner!





Notice that two or three recordings will be opened below each other, but if you want to open four or more recordings, they will be opened as tab in the first window.











Tools

Open record file, Save waveform to file local storage



You easily open the recordings from actual Case just by double clicking on them in the Case view. For opening a file which is not included in opened cases you can open by Open button  or use CTRL+O as a shortcut. To save the waveform to file on local storage you can do likewise, either use the Save button  or CTRL+S as a shortcut.



Play/Pause, Stop playing, Record new waveform, Enable/disable loop playback





Click on the waveform to choose a place to start, then click the Play button . Click the Pause button  or Stop button  to stop playback. After using the Pause button  the next playback will start at that position where it stopped. After using the Stop button  the next playback will start from the beginning of recording or selection.

Click and drag to create a selection, and then when you click Play button  only the selection will play. You can use the SPACE key on the keyboard as a shortcut for Play or Pause. Enable or disable playing the entire selection repeatedly (or the entire track if no selection) until stopped using the Loop playback button . Using the Record new waveform button  you can create a new recording.



    **Zoom in, Zoom out, Zoom to selection, Zoom default**

The image above shows the Zoom buttons. This  is the Zoom In tool, and this  is the Zoom Out tool. Keep clicking the Zoom In button until you see the detail you need. For zooming in or out you can use also CTRL+scroll as a shortcut.

To get a closer look at particular selection of waveform click and drag to create a selection you're interested in, then click the Zoom to selection button . To get back to the default view use Zoom to default button .

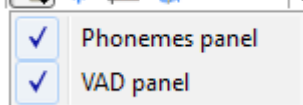
    **Copy selected waveform, Cut selected waveform, Paste waveform from clipboard, Clear editor**




The copy  and cut  buttons are enabled only after creating a selection. It is also possible to use the shortcuts CTRL+C for copying or CTRL+X for cutting a selection. After using one of these functions the Paste button are enabled and you

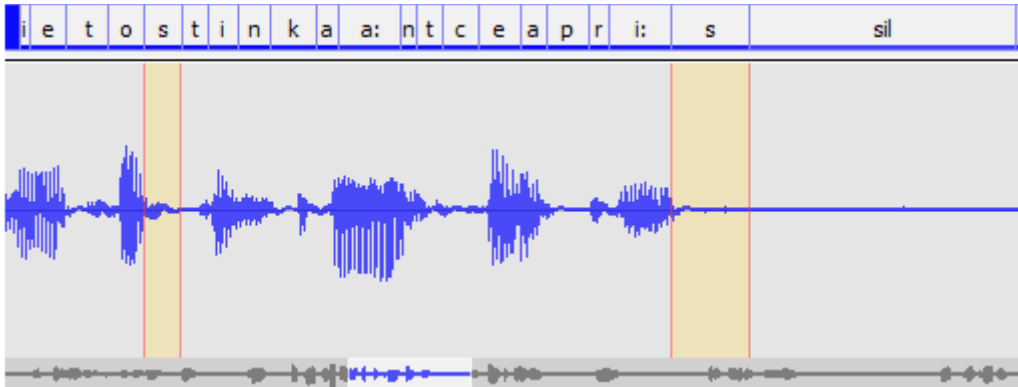
can paste waveform from clipboard (using the button  or CTRL+V as a shortcut) to the position of cursor. To discard all changes use the Clear button .

  **Phonemes panel & VAD panel**



Using this button  you can choose if you want to display Phonemes panel or VAD (Voice Activity Detection) panel.

Phonemes panel breaks words down into phonemes and show them above corresponding sections of recording. You can select only phonemes you are interested in and play them separately from the rest of audio.



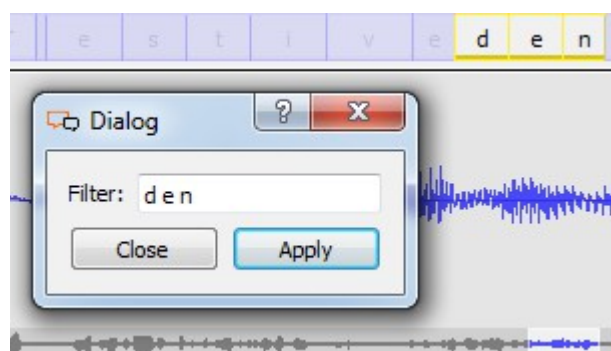
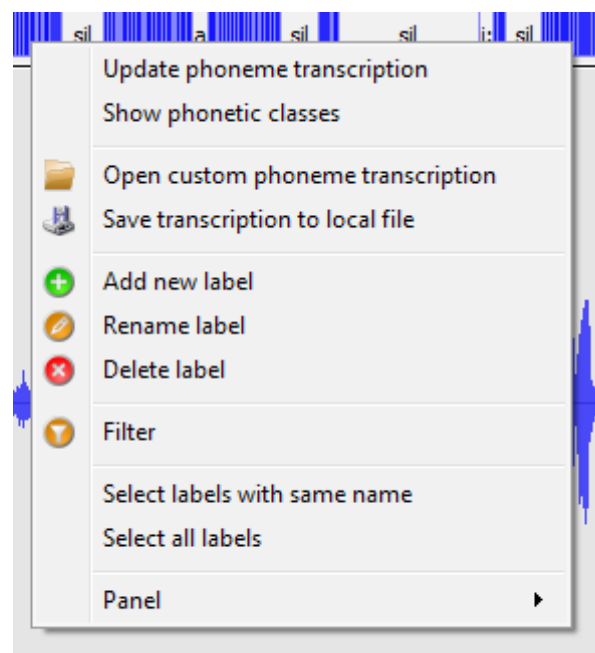
There is a context menu available after right click on the panel.

The first option **Update phoneme transcription** is useful especially after making some changes in the waveform or in the Phonemes panel itself.

Using **Show phonetic classes** option you can display a phonetic class (vowel, fricative, nasal, trill, plosive, approximant, none) for each label.

You can also open your own phoneme transcription using **Open custom phoneme transcription** option which could be useful especially if you you've already saved a transcription using **Save transcription** to local file option and you want to apply it. You might as well **add new label**, **rename a label** or **delete a label**.

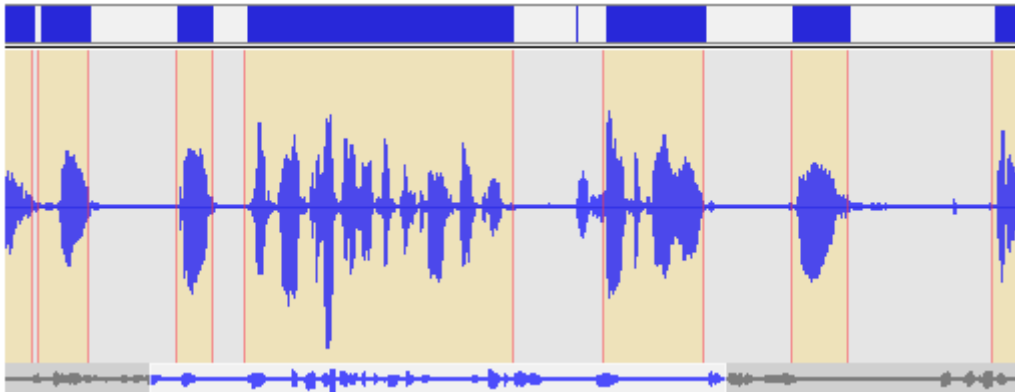
There is a **Filter** option which opens a simple tool which helps you with searching for a particular phoneme or phoneme sequence. Note that it is necessary to write phonemes separated by space.



The context menu also enables **selecting labels with same name**. It depends on what label you right clicked to show the context menu. There is also a possibility to **select all labels**.

To **hide** the phonemes panel or **move** it **up** or **down** use the **Panel** option.

The **VAD panel** divides the speech stream into the silent sections (white parts of the panel) and sections with some voice activity (blue parts of the panel).



There is a context menu available after right click on the panel.

Similarly to the Phonemes panel you may use the **Process VAD transcription** option especially after making some changes in the waveform or in the VAD panel itself. You can **Load custom VAD transcription** as well what could be useful especially if you you've already saved VAD using **Save VAD to custom file** option.

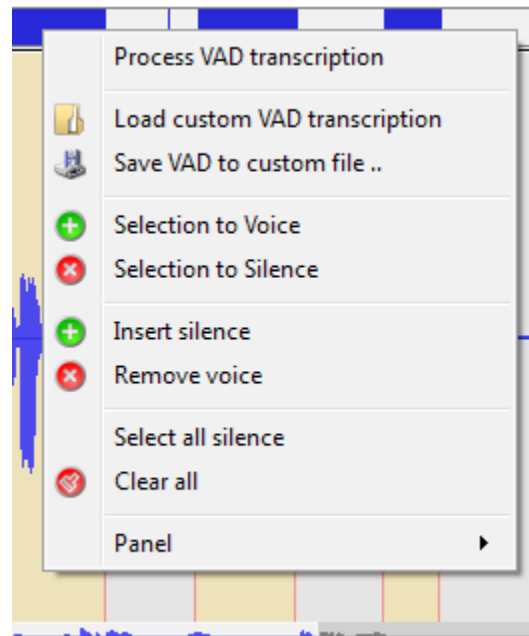
You can change the markings of labels using options **Selection to Voice** or **Selection to Silence**. Note that you can create the selection either by clicking on a section in the VAD panel or by select it independently in the Waveform panel.

It is possible also to **Insert silence/voice** or **Remove voice** in VAD panel.

Select all silence/voice (it depends where you click – on segment with silence/voice detected).

The option **Clear all** erases all detected sections in VAD panel.

To **hide** the panel or **move it up** or **down** use the **Panel** option.



Waveform panel

The default view displays a recording in a waveform panel. You can hide it or show it again using this

 button.



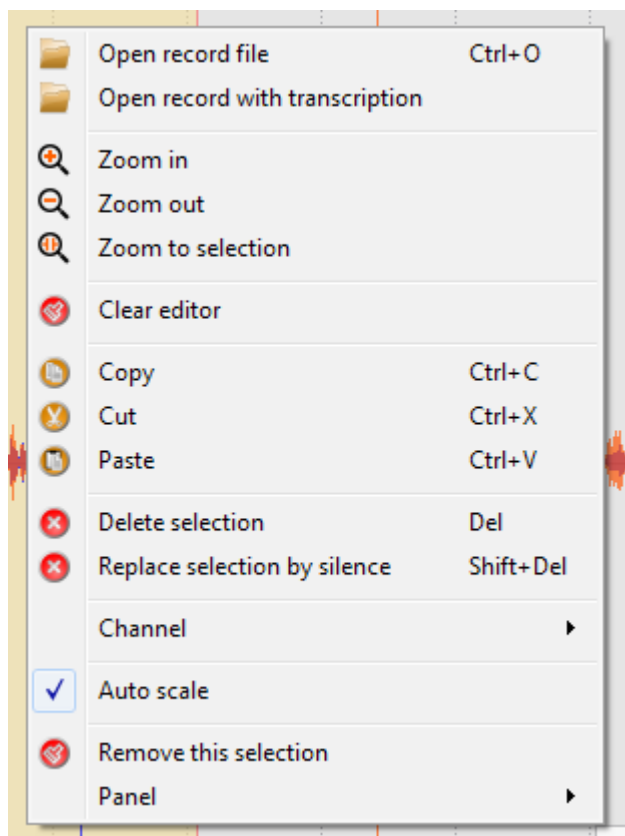
There is a context menu available after right click on the panel.

Besides **opening a record file** you can **open a recording with transcription**. Besides some ordinary functions like **zooming tools**, **copy/cut/paste** function and **clear editor** option there is also an option of **deleting selection**, **replacing selection by silence** or **removing a selection**, which means that for example one of multiple selection can be canceled. Some options are of course enabled only if you have created a selection before.


If you have the stereo recording opened you may use the **Channel** option. It is available only if it is a stereo record. You can adjust what channel you want to see. (all channels, channel 1, channel 2).

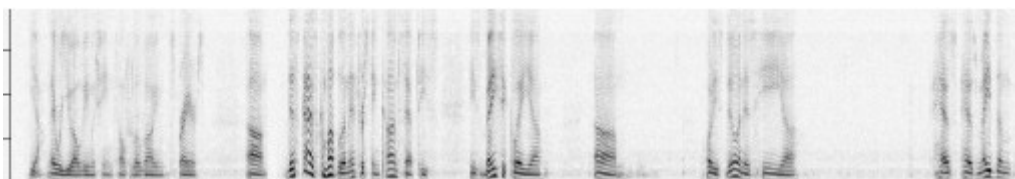
There is an **Auto scale** option checked as a default, which sets the optimal scaling of a waveform.

To hide the panel or move it up or down use the **Panel** option.



Spectrum panel

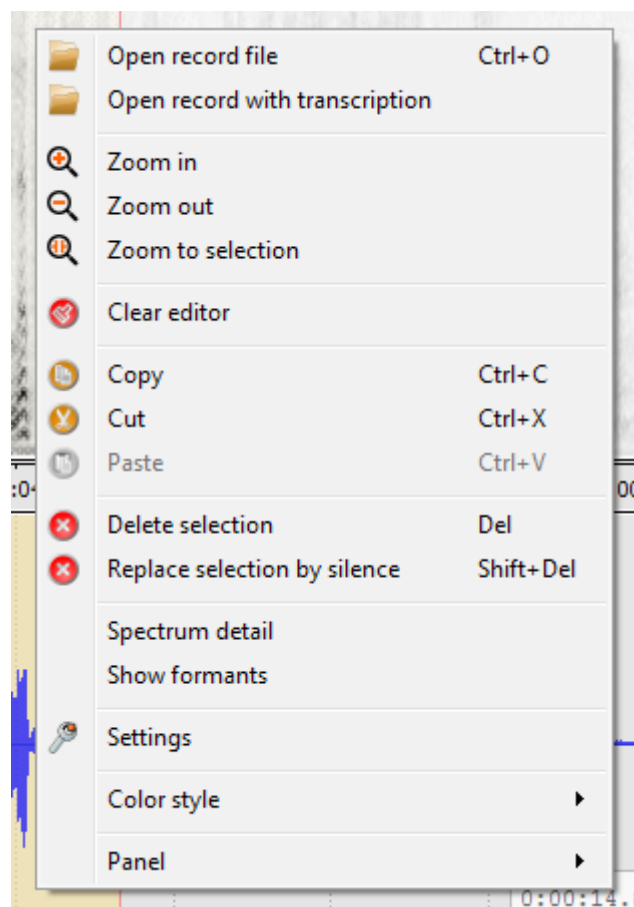
You can also view an audio file visualized in Spectrum panel using this  button. Spectrum panel gives you a way to find areas of significant frequency change that aren't visible in a waveform. It shows the frequency spectrum on scale from 0 to x kHz (x- the maximum is equal to the frequency of the recording).



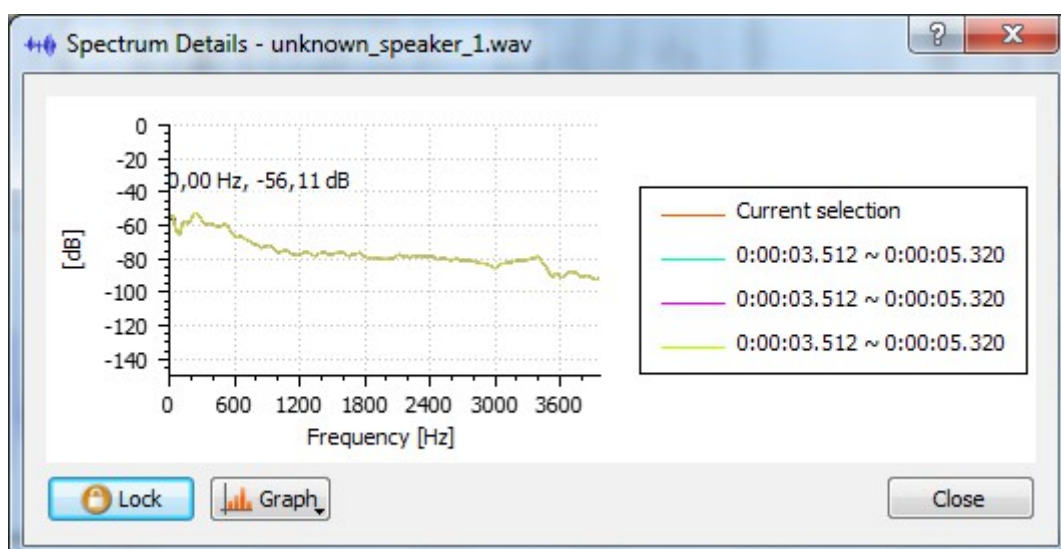
Using CTRL+SHIFT+scroll you can zoom the spectrum and then with CTRL+SHIFT you can drag and move in the panel to focus on what you are interested in.

There is a context menu available after right click on the panel.

You can use several common functions this way like **opening a record file**, **zooming tools** as well as **cut**, **copy** and **paste** functions for chosen selections. As in waveform panel options you can use **Delete selection** or **Replace selection** by silence.



There are several specific functions too. First of them is a Spectrum detail. It shows you a detailed and interactive graph which you can print, save/copy as image or just save/copy the values.

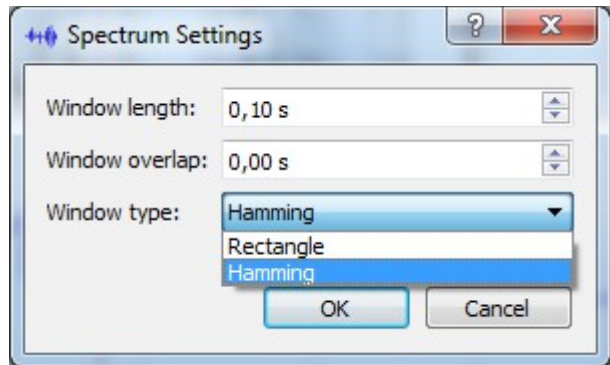


Another specific option in the context menu on Spectrum panel is Show formants. It displays formants directly in the Spectrum panel.

In the Settings dialog you can adjust window length, window overlap and window type – rectangle or hamming.


In a separate dialog you can also adjust a color style of the spectrum.

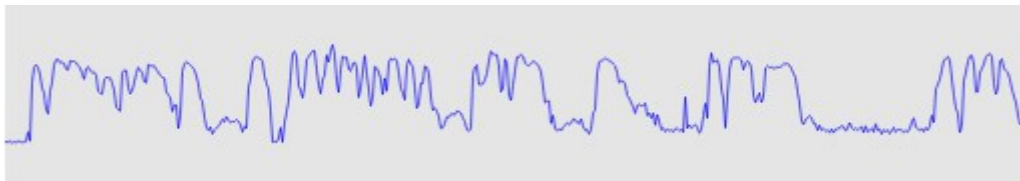
To hide the panel or move it up or down use the Panel option.



Power panel




Another available visualization is the Power panel which can be displayed using this button  on the toolbar..

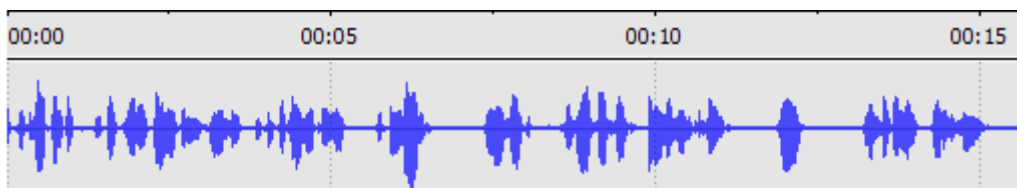


There is a context menu available after right click on the panel. It contains just the ordinary options which you already know from Waveform or Spectrum panel.

Time panel



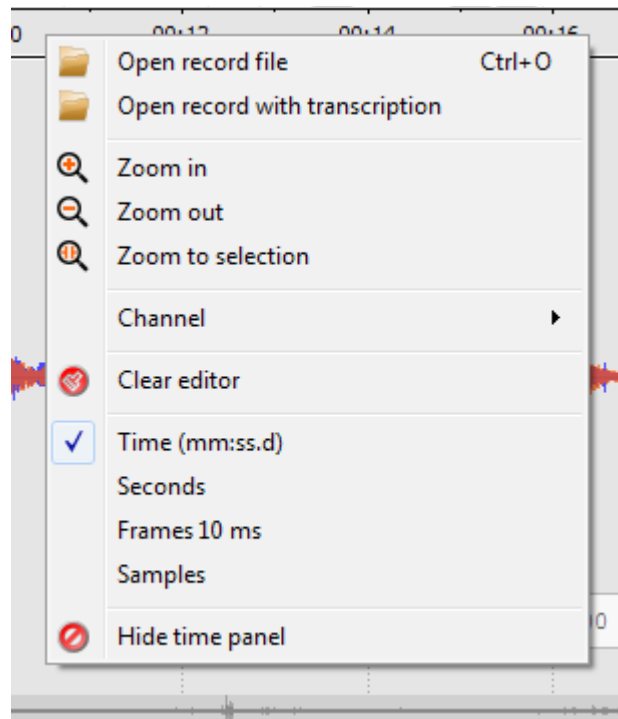
The time panel is another part of default view (like Waveform panel and Status panel). You can hide it or show it again using this button .




There is a context menu available after right click on the panel.

Besides few common tools you can use a **Chanel** option (like in the Waveform panel), which is available only if it is a stereo record, and adjust what channel you want to see. (all channels, channel 1, channel 2).

The specific functions here are the options of time axes. You can choose from the four different format in which the time will be displayed in the time axis – **time** (mm:ss.d), **seconds**, **frames 10 ms**, and **samples**.



Status panel 

This panel is also a part of the default view (like Waveform panel and Time panel). You can hide it or show it again using this button .

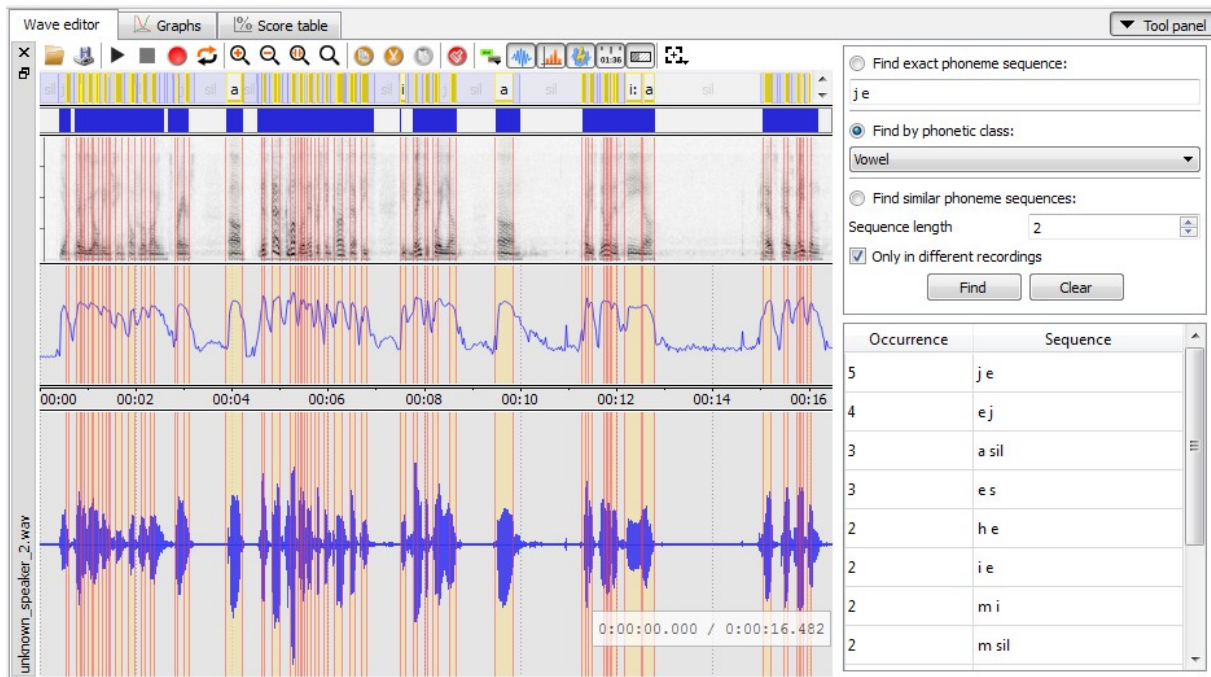


Print / Copy / Save current view

Using this button you can preview and print actual Waveform editor view. You can also copy it as a picture and paste to another document or save it as a picture wherever you want.

Useful shortcuts:

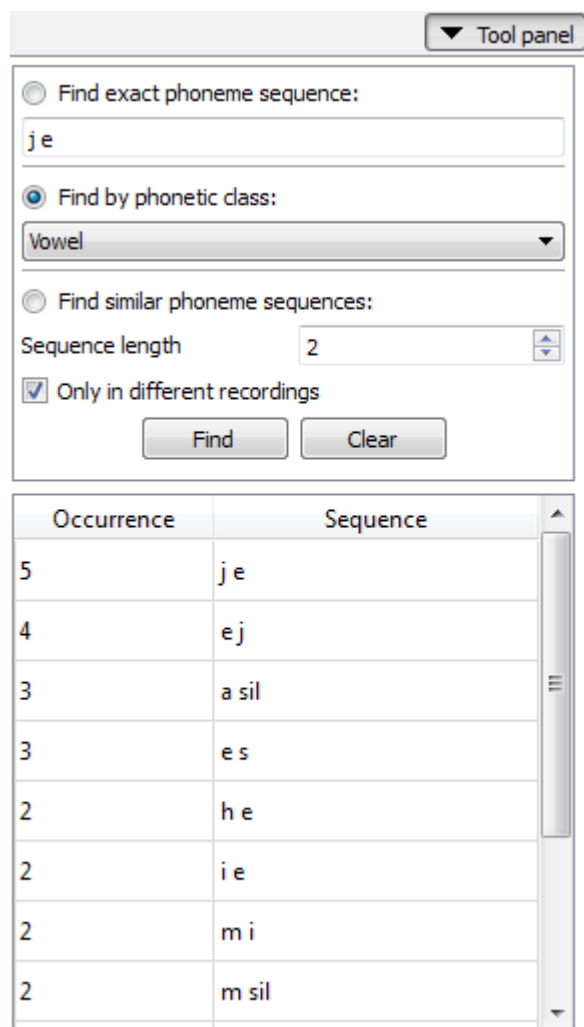
- CTRL+C copy
- CTRL+X cut
- CTRL+V paste
- CTRL+O Open record file
- CTRL+S Save waveform to local file
- CTRL+U Upload waveform to server into actual folder
- SPACE Play/Pause
- CTRL+A select all
- DELETE delete selection



Tool panel

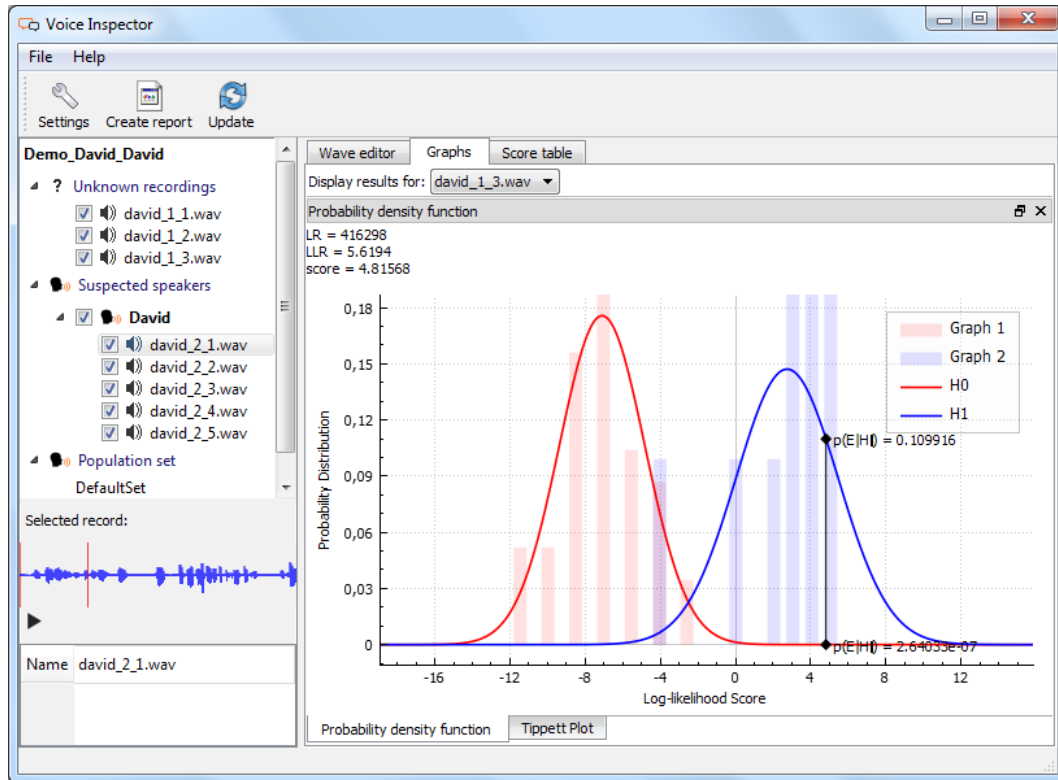
You can search in a recording for a phoneme using filter in phoneme panel. But the tool panel is useful for searching in two or more recordings simultaneously. There are two main options:

- Find exact phoneme sequence
 - make sure that you write spaces between each phoneme otherwise the searching can't be successful
- Find by phonetic class – vowel, fricative, nasal, trill, plosive, approximant, none
- Find similar phoneme sequences - it searches for sequences,
 - sequence length – specify number of phonemes which the sequence should contain
 - only in different recordings – decide if you are interested only in similar sequences between two or more recordings or if you are interested also in similar sequences occurring only in the same recording
 - see in the table below particular sequences – by clicking on a row, the sequence (all occurrence) will be highlighted as a selection in the waveform



Graphs

The tab called “Graphs” shows a graphical expression of identification analysis with respect to a particular unknown recording, suspected speaker and population set. You can display results for each unknown recording individually. This analysis method is based on Bayesian theory.



Score table

The last tab called “Score table” shows several statistical results of the match for each unknown recording. To display all the scores, you also need to click the “Update” button first.

Recording name	Speech length	Score	LR	Log10 LR	Verbal ratio
unknown_speaker_1.wav	00:00:08.880	-3.53	2.19	0.341	Limited evidence to support prosecution hypothesis

Verbal ratio

So called “Verbal Ratio” uses scale of the probability, which is expressed as:

- Strong evidence against
- Moderately strong evidence against
- Moderate evidence against
- Limited evidence against
- Limited evidence to support
- Moderate evidence to support
- Moderately strong evidence to support
- Strong evidence to support

Likelihood ratio

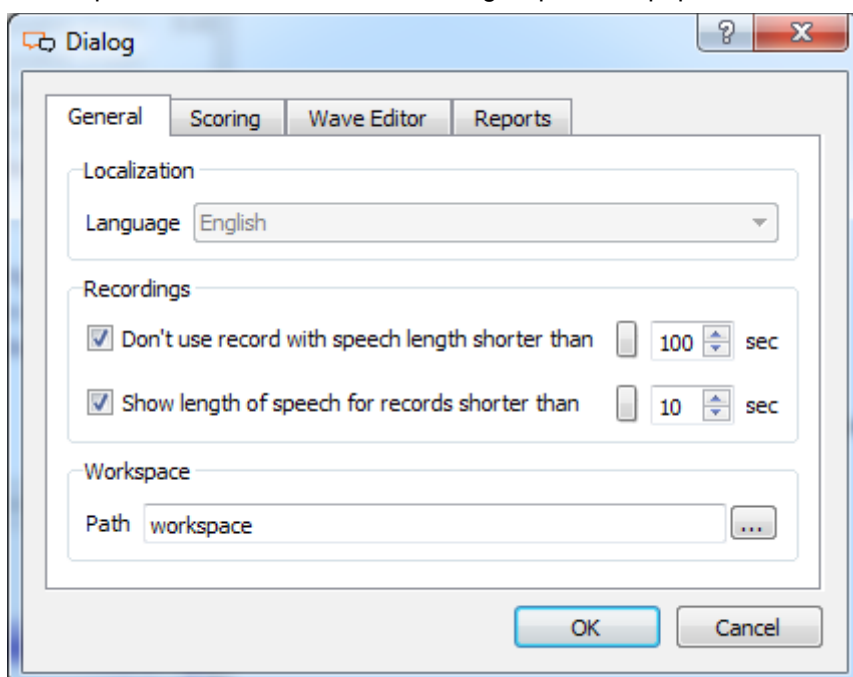
A likelihood ratio is metrics used by forensic experts. The system trains a model of the speaker's voice. Internally it also uses a world model (trained on the voices of several thousand speakers). During scoring, the system first evaluates how close the record is to the model of the speaker (its "proximity"), and how close it is to the model of the world. The likelihood ratio for the speaker and world is then evaluated. This ratio gives the record's proximity to the speaker model divided by its proximity to the world model. The logarithm of this score is taken before saving it.

- a high number means that the system is sure that it is the speaker
- zero means that the system is not sure
- a high negative number means that the system is sure that it is not the speaker

Settings

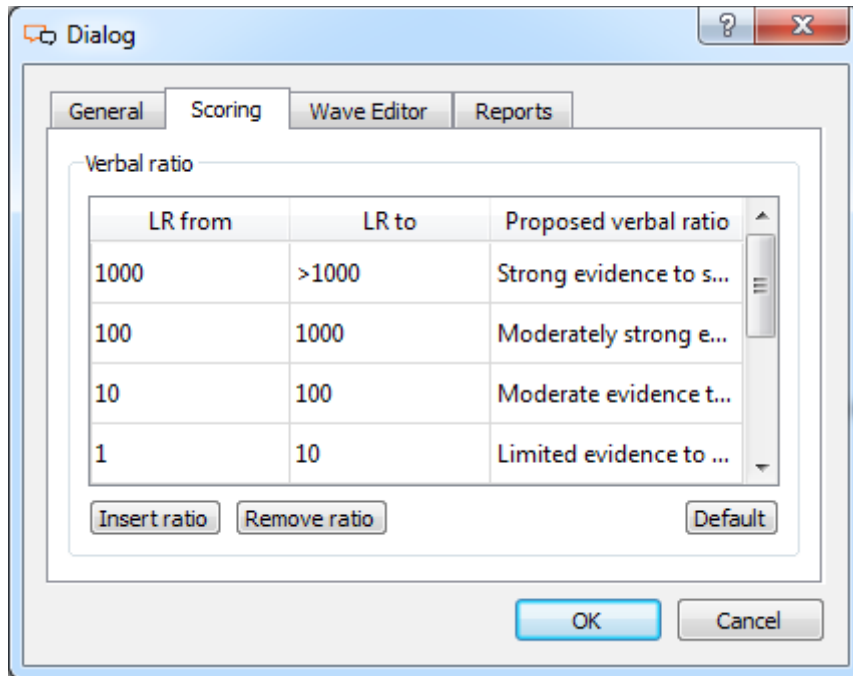
General

- Localization – set a language of the application
- Recordings
 - Reduce an amount of processed recordings to the recordings with sufficient speech length. Recordings with short speech length can cause irrelevant results.
 - Set that the speech length will be displayed for recordings whose speech length is shorter than the required.
- Workspace – choose a folder in which will be stored
 - all wav files in folders
 - cases
 - speakers
 - population s
 - in the file database.db (SQLite database)
 - voiceprints and all data about recordings, speakers, populations etc.



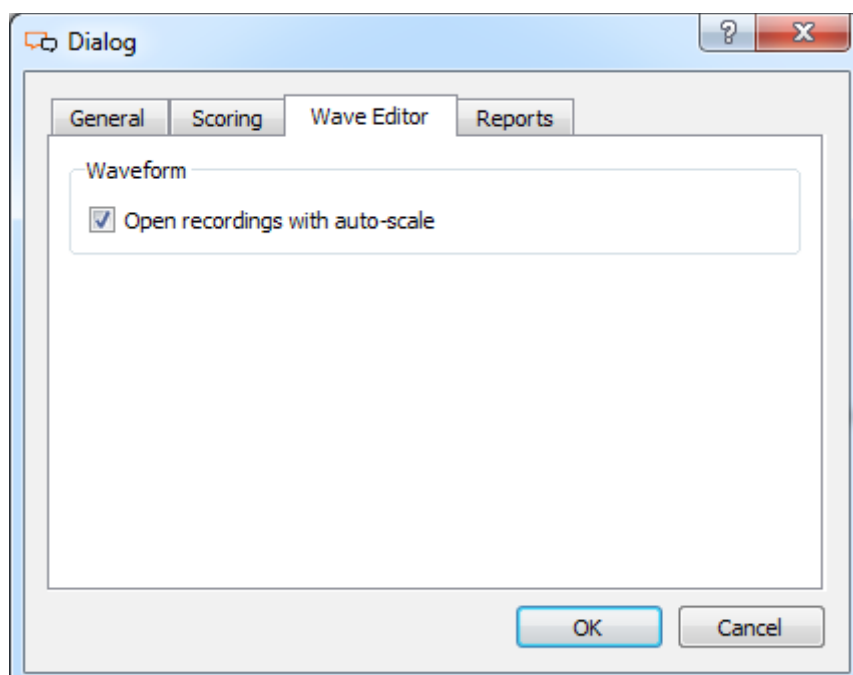
Scoring

In the Scoring tab you can create your custom scale for likelihood ratio and verbal ratio. Anyway it is easy to store the default scoring.



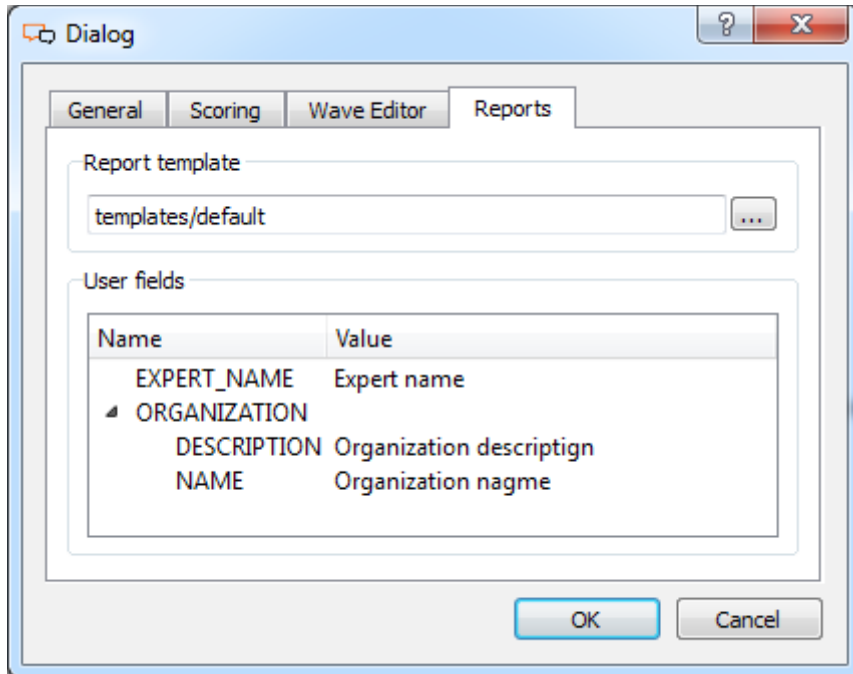
Wave editor

In the Wave editor tab you can decide if the recordings should be (while opening them in Wave editor) automatically opened with auto-scale in Waveform panel.



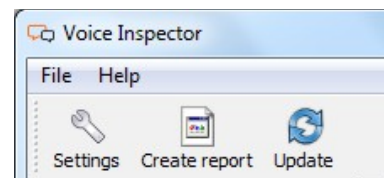
Reports

In the Reports tab you can set the path to the Report template and customize the layout and structure of the document. You can change the names of particular sections and characterization of their values.



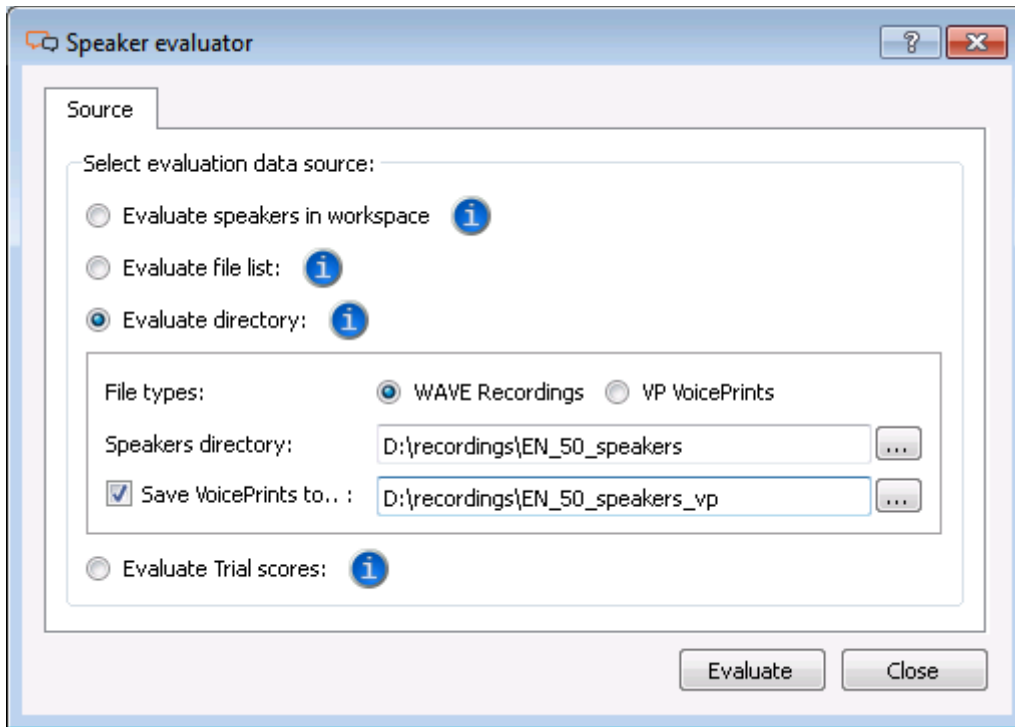
Create Report

Very important feature of the application is a possibility to generate Expert opinion report. The report is easily editable, and automatically inserts particular graphs, scoring results, and all the information needed. The report significantly saves time of forensic experts during investigation.



Speaker identification evaluator

This tool allows evaluate speaker identification accuracy and show statistics such as Detection Error Tradeoff (DET) plot, Probability Density Function (PDF) plot, Equal Error Rate (EER).



Select evaluation data source

Before evaluation is necessary to select data which will be used for evaluation. If wave recording is used as source it will be needed to extract VoicePrints from recordings. This process may take long time depends on count and lengths of recording.

Evaluate speakers in workspace

All speakers from Speaker Database are used for evaluation.

Evaluate file list

Process three-column list file. Paths to compared recordings are in first and second columns. Third column contains character T (for Target score) or N (for Non-target score). Columns are separated by space. Instead of WAVE recordings can VoicePrints be used.

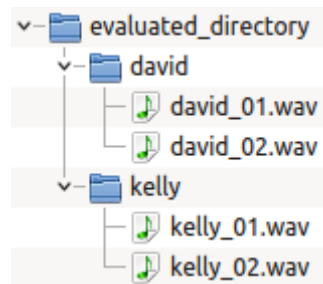
Example of list file structure:

```
./david_1.wav ./david_2.wav T
./david_1.wav ./kelly_1.wav N
```

Evaluate directory

Select directory which contains for every speaker sub-directory. Every sub-directory has to contain at least two recordings of speaker. Instead of Wave recordings can VoicePrints be used. For future evaluations extracted VoicePrints can be saved.

Example of evaluated directory structure:



Evaluate Trial scores

Select two files. One with target scores and second with non-target score.

Example of target scores file:

```

3.77935630561994
8.10854848100454
4.69067347134316
2.20380387631228
    
```

Example of non-target scores file:

```

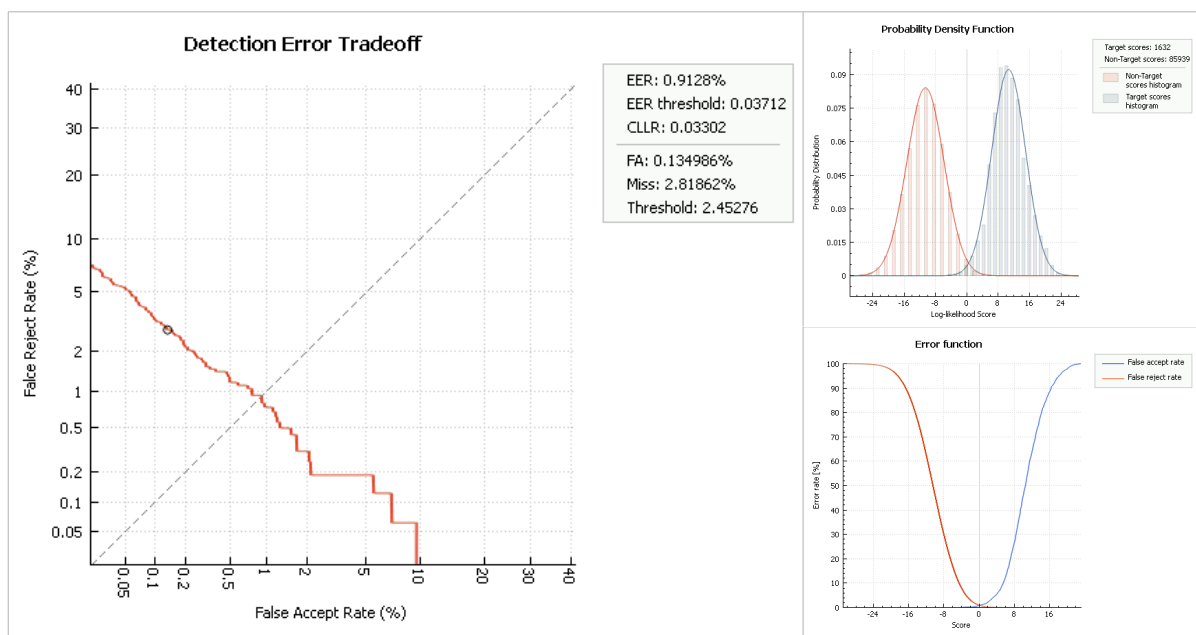
-0.56013115401013
-1.16668882627714
0.211325357197232
-3.53186749100894
    
```

Numbers can be in scientific format (e.g. 3.9265e+2).

If evaluation process is complete scores can be saved from Scores tab.

Evaluation results

After evaluation process is complete following statistics and plots will appear.



Contact

Phonexia helps clients to automatically extract the maximum amount of valuable information from spoken speech. We develop technologies for data mining from speech, speech analytics, and voice biometry. These technologies are used by call centers, telecommunication companies, banks, government agencies, media servers, and broadcast service providers.

You might be interested to see the full list of our:

- speech technologies: www.phonexia.com/technologies
- solutions and services: www.phonexia.com/solutions

Phonexia also provides research & development services such as: speech technology optimization for target channels, development of new language versions, etc.

Contact us for additional information

Phonexia s.r.o.
info@phonexia.com | +420 511 205 265
Chaloupkova 1a, 61200 Brno, Czech Republic, Europe

4. Annexes

List of Annexes

- Annex 1 . - Disclaimer

Annex 1. - Disclaimer

The following text provide only certain chapters from Phonexia Software License Agreement describing warranty disclaimer and limitation of liability.

- For complete Software License Agreement for Phonexia demoversions please visit www.phonexia.com/docs/license_agreement.txt
- For complete Software License Agreement for production versions please contact Phonexia.

WARRANTY DISCLAIMER

THE SOFTWARE IS PROVIDED ON AN "AS IS" BASIS. Phonexia AND ITS SUPPLIERS MAKE AND YOU RECEIVE NO OTHER WARRANTIES OR CONDITIONS, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE OR IN ANY COMMUNICATION WITH YOU, AND Phonexia AND ITS SUPPLIERS SPECIFICALLY DISCLAIM ANY IMPLIED WARRANTY OF MERCHANTABILITY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT AND THEIR EQUIVALENTS.

Phonexia does not warrant that the software (i) be error free or (ii) meet Your specific requirements. SOME STATES OR OTHER JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO THE ABOVE EXCLUSIONS MAY NOT APPLY TO YOU. YOU MAY ALSO HAVE OTHER RIGHTS THAT VARY FROM STATE TO STATE AND JURISDICTION TO JURISDICTION.

LIMITATION OF LIABILITY

IN NO EVENT WILL Phonexia OR ITS SUPPLIERS BE LIABLE FOR ANY DAMAGES, INCLUDING BUT NOT LIMITED TO LOSS OF OR CORRUPTION TO DATA, LOST PROFITS, LOSS OF CONTRACTS, COST OF PROCUREMENT OF SUBSTITUTE PRODUCTS OR ANY OTHER DIRECT, SPECIAL, INCIDENTAL, PUNITIVE, CONSEQUENTIAL OR INDIRECT DAMAGES, LOSSES, COSTS OR EXPENSES OF ANY KIND ARISING FROM THE SUPPLY OR USE OF THE demo, HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY (INCLUDING WITHOUT LIMITATION NEGLIGENCE). THIS LIMITATION WILL APPLY EVEN IF Phonexia OR AN AUTHORIZED DISTRIBUTOR HAS BEEN ADVISED OF THEIR POSSIBILITY. THIS LIMITATION WILL NOT APPLY TO THE EXTENT THAT LIABILITY MAY NOT BY LAW BE LIMITED OR EXCLUDED. SOME STATES OR OTHER JURISDICTIONS DO NOT ALLOW THE EXCLUSION OR LIMITATION OF LIABILITY FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES, SO THE ABOVE LIMITATIONS AND EXCLUSIONS MAY NOT APPLY TO YOU.

SAFE OPERATION

YOU ASSUME RESPONSIBILITY FOR, AND HEREBY AGREE TO USE YOUR BEST EFFORTS IN, USING THE SOFTWARE TO PROVIDE FOR SAFE OPERATION THEREOF, INCLUDING, BUT NOT LIMITED TO, COMPLIANCE OR QUALIFICATION WITH RESPECT TO ALL SAFETY LAWS, REGULATIONS AND AGENCY APPROVALS, AS APPLICABLE. THE SOFTWARE IS NOT DESIGNED OR INTENDED FOR USE AS COMPONENTS IN ANY APPLICATION IN WHICH THE FAILURE OF THE EXAMPLE CODE COULD CREATE A SITUATION IN WHICH PERSONAL INJURY OR DEATH MAY OCCUR, AND YOU SHALL HAVE NO RIGHTS HERE UNDER FOR ANY SUCH APPLICATIONS.