Speech Intelligence Resolver (SIR)

User Manual

User Manual for Speech Intelligence Resolver (SIR)
Produced by: Phonexia, s.r.o.

Product Version:	1.2.14
User Manual Version:	1.2
Modified:	1 June 2015
Revision:	43

Table of Contents

1.Introduction.....2
Typical Use-Cases.....3
Technical Requirements.....3
2.Quick Start Guide.....3
3.Using the Application - Features.....4
a) First step – Storage Setting.....5
b) List of Speaker Models and Speaker Groups.....6
d) List of Language Models.....7
e) List of Keywords.....7
f) Processing Audios.....8
g) Results.....8
h) Settings.....11
i) Waveform Editor.....13
j) Layout.....16
4.Using the Application -Typical Work-flows.....17
a) Speaker Registration (speaker model enrollment).....17
b) Speaker Update (speaker model enhancement).....17
c) Speaker Search (Speaker Identification).....18
5.Speech Technologies Available in SIR.....19
6.Troubleshooting.....20
7.Contact.....20
8.Annexes.....21
Annex 8.1. - Abbreviations.....21
Annex 8.2. - Disclaimer.....22

1. Introduction

Phonexia Speech Intelligence Resolver (SIR) combines the power of speech technologies to single application. The application automatically performs visualization of record as well as effective filtration of speech metadata uncovered from your records.

Speech technologies implemented:

- Phonexia Speaker Identification (SID2)
- Phonexia Language Identification (LID2)
- Phonexia Gender identification (GID)
- Phonexia Voice Activity Detection (VAD)
- Phonexia Speaker Diarization (DIAR)
- Phonexia Keyword Spotting (KWS)
- Phonexia Speech Quality Estimator (SQE)
- Phonexia Speech Transcription (STT)

SIR is a client application cooperating with REST servers. It's possible to use it as a standalone application due to integrated local REST server. It was developed to use unique speech technologies so you can get valuable information from the content of your calls.

Typical Use-Cases

- Searching a target speaker in a large number of audio recordings
- Audio preparation for high quality voiceprint for future speaker searches
- Quick and advanced analysis of audio archive by metadata extraction from speech
- New monitored channels analysis for proper setting of speech technologies

Technical Requirements

Recommended hardware:

- for desktop installation of SIR with SID2_L, LID2_L, GID, VAD, SQE, DIAR_L:
  2,6GHz CPU with 4 physical cores available, 8GB RAM (or better)

- for client-server installation:
  - client (SIR) recommended:
    2GHz CPU single core or better, 1GB RAM
  - server (REST) with with SID2_L, LID2_L, GID, VAD, DIAR:
    2,6GHz CPU with 4 cores available, 8GB RAM (or better)
    (multi-core processing multiplies processing capacity)

Available platforms:

- for desktop installation of SIR: Windows 64bit (x86_64) and Linux 64bit (x86_64)

- for client-server installation:
  - client: Windows 32bit (x86) or 64bit (x86_64) or Linux 32bit (x86) or 64bit (x86_64)
  - server: Windows 64bit (x86_64) and Linux 64bit (x86_64)

2. Quick Start Guide

Installation (first start)

To install and run 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 SIR.exe for Windows)
3. Run SIR / SIR.exe

Update

To update the application, follow these three easy steps:

1. Backup your current installation of SIR (e.g. pack the whole SIR directory and storage to ZIP)
2. Unpack the archive (you should use the latest version of SIR X.Y.Z, where the X should be the same version as you have already installed)
3. Rewrite all the files in folder where SIR application already installed
4. Run SIR / SIR.exe
5. All results will be updated while you click “Start” button.

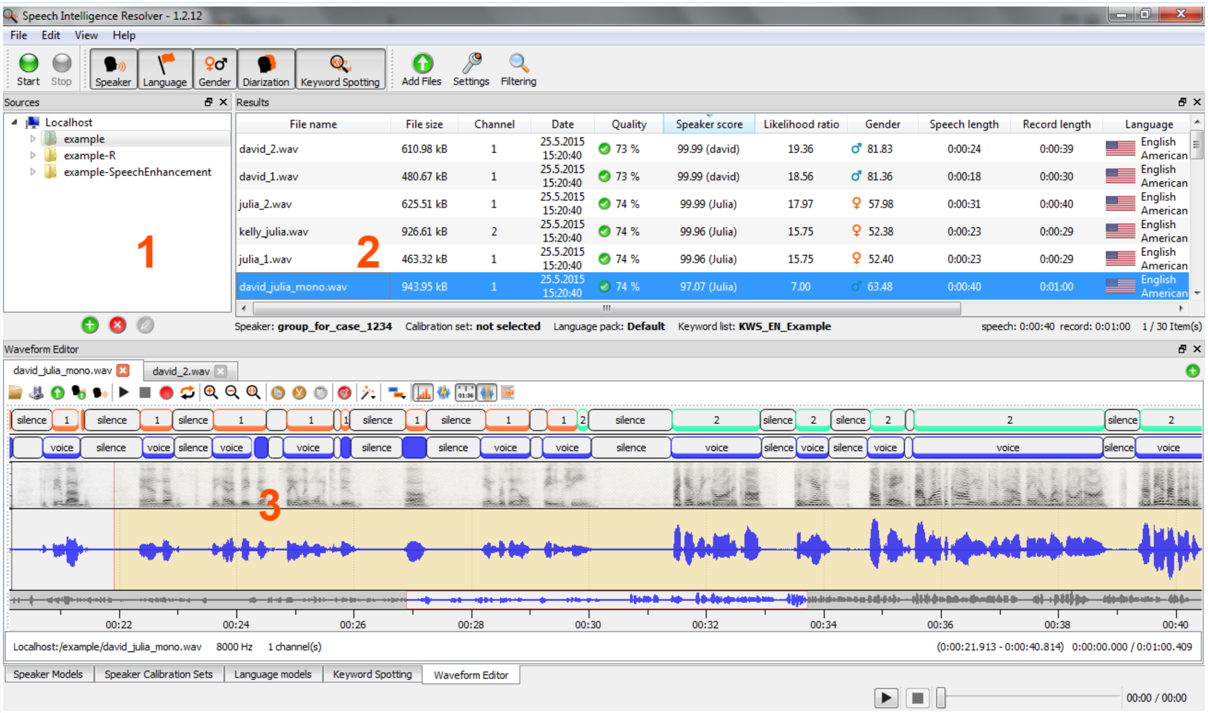
3. Using the Application - Features

Running SIR / SIR.exe, the application will be initialized. In case of desktop installation of SIR, the initialization time depends on HW of your computer.

Main features of SIR application:

- Phonexia speech technologies
  (see. "Introduction" above, for details see technical papers of each technology).
- Filter for easy search in the results list
- Possibility to adjust dynamic range of result scores
- Speech files in MS Wave format
- The minimal length of voice records for detection is 10 s (default)

The graphical user interface consist of several main parts. For a better orientation we named particular sections of application as follows:



1. Sources View

The "**Sources view**" displays folder structure of particular servers (only names of folders). There is a "**Localhost**" item (only in case you obtained SIR with REST server). Through the localhost server you have an access to the local folder (it is possible to change it in "**Settings**" in the "**REST server**" tab). Using the buttons below or by context menu you may:

- Create a new server / folder
- Delete selected server / folder
- Edit selected server

2. Results View

The table includes files from the folder chosen in the "**Sources view**" and the results of all applied technologies. Its possible to show or hide certain columns by right clicking on the heading of whichever column. You can also shift the columns or sort the files by particular column. After right clicking on the selected record(s) you can:

- Play the record (CTRL+P)
- Delete file(s) – or press the „Del" key
- Open in wave editor (F2)
- Open with VAD segmentation
- Download file(s) (CTRL+D)
- Download all files
- Upload new file(s)
- Export selection
- Export all
- Show file information (CTRL+I)

3. Technologies View

- **Waveform Editor** - The table allows user to work with audio (e.g. open/save/edit/paste/delete audio, record, play (all / selection / in a loop), zoom in/out) and work with panels (show/hide). Details are described below.

- **Speaker Models View (List of Speakers)** - The table includes speaker groups and models of the speakers on the server and determines which speaker model (or group) is actually used for testing. Details are described below.

- **Speaker Calibration Set -** The table allows user to enter/import calibration set. This can be used for a calibrate speaker identification results for a specific value of maximum false alarms.

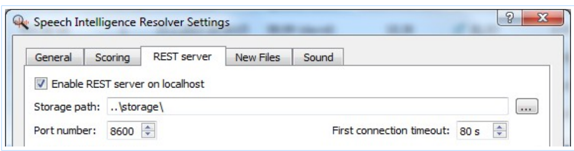
- **Language Models View (List of Languages)** - The table includes the language packs from the server and allows to select a certain pack intended for testing using the checkbox. Details are described below.

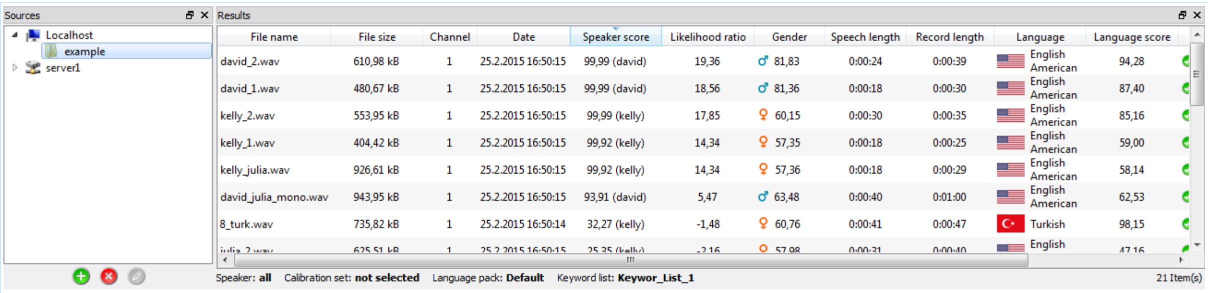
- **Keyword Spotting View (List(s) of Keywords)** - The table includes the lists of keywords and pronunciations. packs from the server and allows to select a certain pack intended for testing using the checkbox. Details are described below.

a) First step – Storage Setting

You should define the storage path first in a setting (Settings) > "REST server . Default storage path is "REST/storage" located in the folder with SIR. In the folder server expects to find the user records. To change storage path for REST server use the "**Settings**" button and the "**REST server**" tab. Click on the three dots button and choose a folder which contains the records. Details are described in section Setting below.

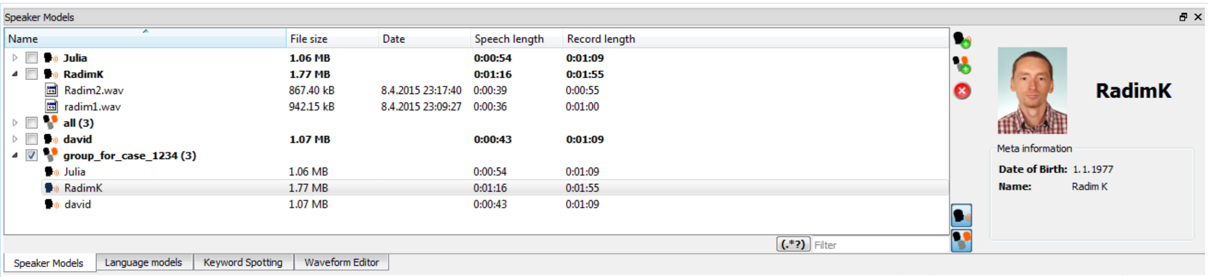
This folder will be accessible via "Localhost" item in a "**Sources view**" and its content is displayed in a "**Results view**":





b) List of Speaker Models and Speaker Groups

The list of speaker models is shown in Speaker models tab. Each speaker model includes audio files and information about their size, length and speech length. In the row with the speaker´s name there are visible total values about all included records. Operations on more items are available only if items of the same type are selected (models OR files). Details for selected speaker model will appear on right side.

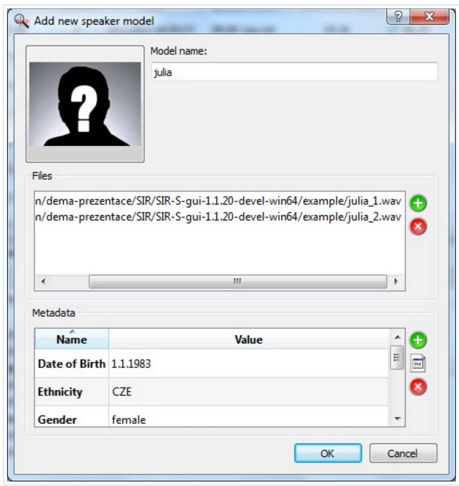


User can:

- Add new speaker model / group
- Show speaker models / groups
- Add more files to an existing model
- Delete a speaker model / file
- Download a model´s record
- Play a model´s record
- Filter speakers by Name

Hint: You can edit metadata of the speaker by double-click on speaker name.

Hint: You can add / delete speaker to/from speaker group by double-click on speaker group name.



You can import model(s) (right click > “import model(s)), if you created already voice-prints by Phonexia Speaker Identification technology. The data need to be in specific folder structure, where folder name will be used for speaker model name. The folder should contain at least voice-print file.

If you want to use speaker identification, it is necessary to create a speaker model(s) first.

Go to the “**Speaker Models view”** than click on the green “**Add new speaker**“ button.

You will get an “**Add new speaker**“ dialog box.

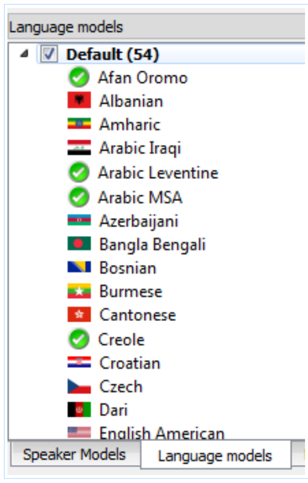
You must specify the model name and add one or more records of his voice using the green button. You can also fill in some metadata about the speaker. The table is editable, so it is possible to add and delete the rows.

For faster processing you can group created speaker models into speaker groups using “**Add new speaker group**“ button.

Using the buttons “**Show speaker models**“ and “**Show speaker group**“ you can affect which models will be visible.

Hint: **One or several voiceprint(s) can be automatically imported** by right-click at “Speaker view” and option “**Import model(s)**”.

d) List of Language Models



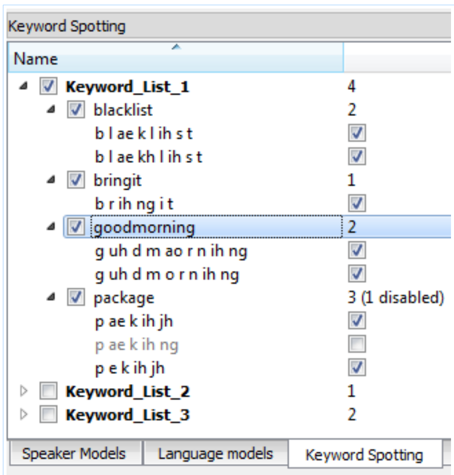
The list of language models is shown in Language models tab. The default Language pack contains 54 languages. The language identification is done on selected language pack.

The REST server administrator can add more language packs and/or add languages for language identification. This requires technical knowledge – please contact Phonexia representative for additional information.

e) List of Keywords

It is required to prepare a list of keywords or phrases first.

The keyword list can be created by right click in the “**Keyword Spotting View**” and using “**Add new keyword list**”.



The new keyword/phrase can be added by right click on the name of keyword list and using “**Add new keyword into list**”.

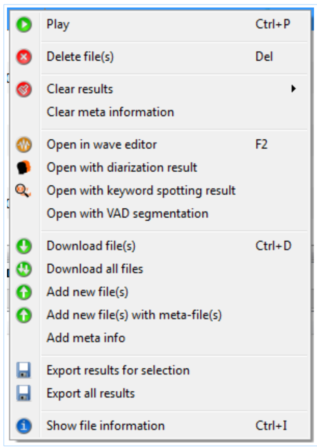
When keyword added (phrase should be added without space character(s)), the pronunciation is generated automatically. User can even add several pronunciations for one keyword/phrase or change pronunciation by doubleclick. Only the active pronunciation variants are searched.

The new keyword/phrase can be added by right click on the name of “. keyword list and using “**Add new keyword into list**

f) Processing Audios

1. Select a folder of records in "**Sources view**". You can see the list of included files in "**Results view**". You can use filtering to test only some of your files.
2. You can add files to Results view using or **Drag&Drop** functionality or button "**Upload Files**" in the main toolbar.
3. Select a technology/technologies in the main toolbar – Speaker / Language / Gender identification / Diarization / Keyword Spotting.
4. For Speaker identification, select a speaker in "**Speaker Models view**"
5. For Language identification, select a language pack in "**Language Models view**"
6. For Keyword Spotting, select a list of keywords in "**Keyword Spotting view**"
7. Start testing using the green **"Start"** button in the toolbar (you can see progress in status bar at the bottom of the window).
8. The results will appear in "**Results view**" for each active technology.

g) Results

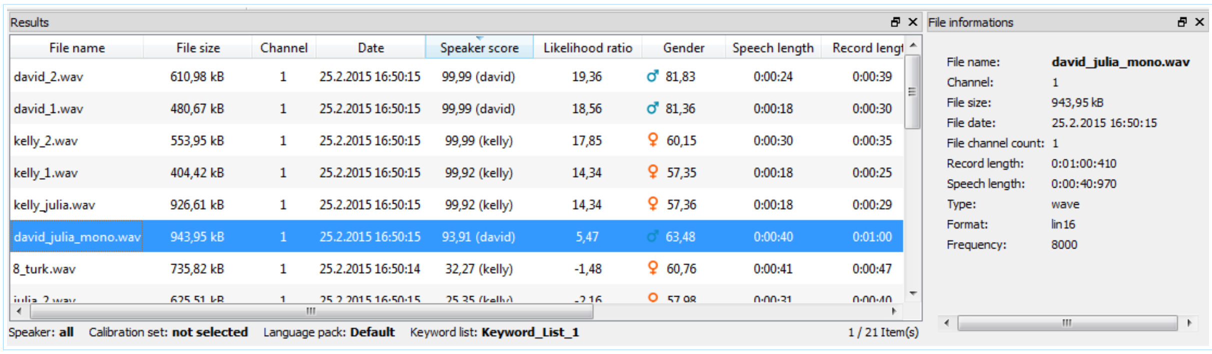


After right clicking on a selected item (audio file) in the "**Results view**" you can see several options of content menu.

"**File informations**" (Ctrl+I) can appear on right side.

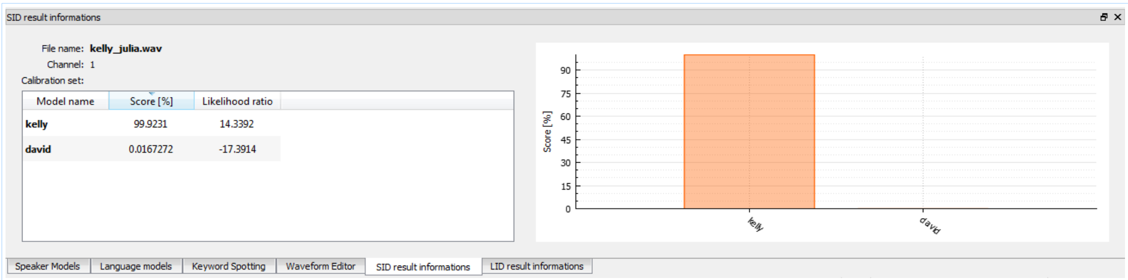
The following information can appear after left double-click on specific value of requested item (audio file) in particular tabs:

- Waveform (in Waveform View) after left double click on File name / File size / Channel / Speech length / Record length.
- GID results after left double click on Gender.
- SID results after double-click on Speaker score / Likelihood ratio.
- SQE results after left double click on Quality score.
- LID results after left double click on Language / Language score.
- DIAR results after left double click on Diarization. The results are shown in Waveform Editor as through Diarization label panel (see Waveform section below).
- KWS results after left double click on Keyword Spotting. The results are shown in Waveform Editor as through Keyword Spotting label panel (see Waveform section below).
- Quality information after left double click on Quality. The quality in Results view is shown in percentage. If the number is higher than 70% the quality is good, if it is lower than 30% it means very bad quality. It is important to be aware of the fact that the technological quality is not the same like audible quality.



The graphs in the SID or LID results information tabs you can zoom in or out with mouse scroll or drag and move them. Don´t forget that it shows the results only for the models which have been already tested.

Speaker Identification Results



The speaker score can be found in SID results tab. There are two metrics:

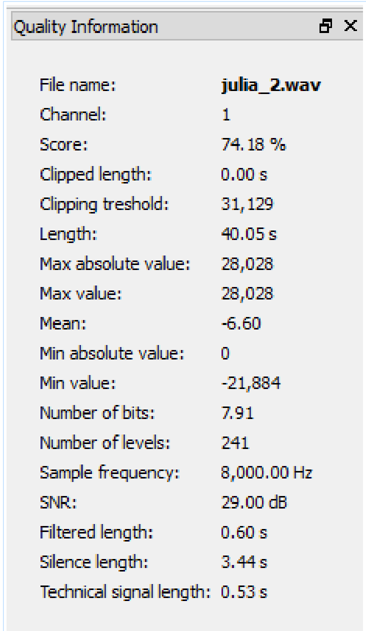
Percentage score (probability of speaker in percentage)
Even the likelihood ratio is the primary metrics, our system recalculate it to percentage metric for easy understanding. The log-likelihood score is converted to the interval between 0 and 100% through “Speaker score sharpness”. It can be adjusted in setting tab.

- 100% – the system is sure that it is the speaker
- 50% – the system is not sure
- **0% – the system is sure that it is not the speaker**

Likelihood ratio (raw score)
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. If the likelihood ratio is not visible in the list of tested records, it can be added by right click to the headers of columns.

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

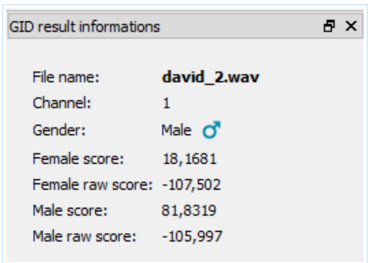
- a high negative number means that the system is sure that it is not the speaker



Speech Quality Estimator Results

Quality of audio is summarized to final score in percentage metric. The results lower than 30% are marked with red icon for warning. The results better than 70% are marked with green icon. The details of speech quality can be achieved with double-left-click on item in Quality column.

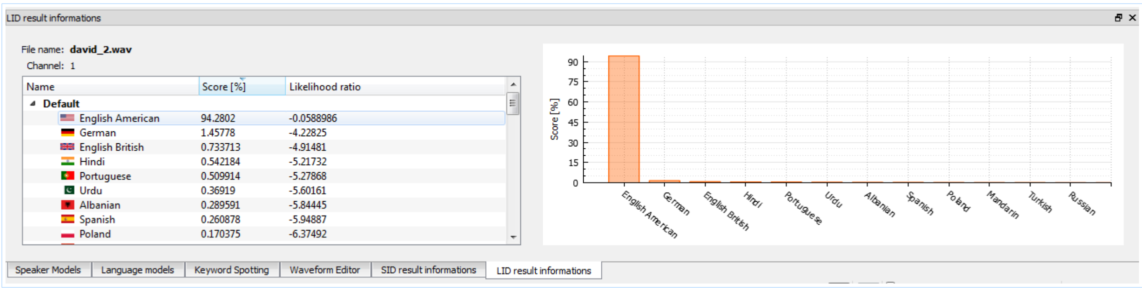
Gender Identification Results



Female / Male percentage score and Female / Male raw score can be found on right side. The raw score is converted to the interval between 0 and 100% (percentage score) through "Gender score sharpness". It can be adjusted in setting tab.

Language Identification Results

The Language percentage score and Likelihood ratio score can be found in LID results tab. The sum of percentage score for all active language models gives 100%.

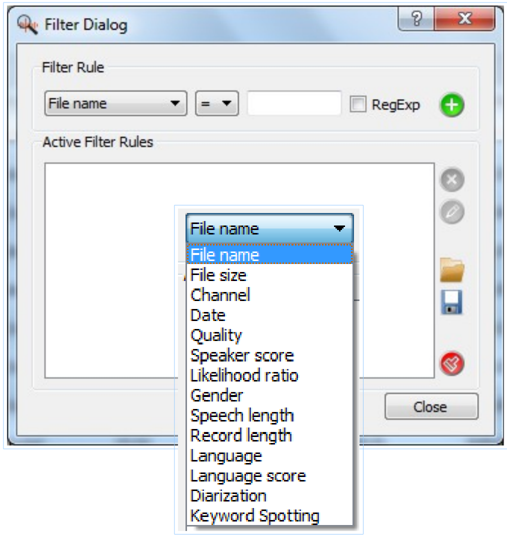


Filtering Results

To search in the files in "**Results view**", click on the "**Filtering**" button in the toolbar.

You will get the "**Filter Dialog**" box appears so you can set the filter rules.

If any filter rule is set, you can see only items corresponding to all of active rules in the Results view. An active filter is notified by green borderline on edges of Results View.



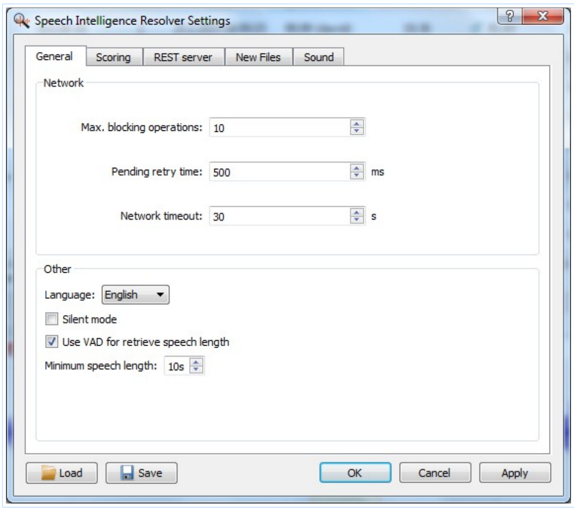
You may:

- Create a rule – use the green button to add it to the “**Active Filter Rules**“
- Save current filter rules to file
- Load previously saved file with filter rules
- Delete all filter rules
- Set filter rule as regular expression

After selecting particular rule:

- Delete selected filter rule (by red button)
- Edit selected filter rule (by orange button or double-click on rule)

h) Settings



The system can be configured by clicking on the “**Settings**“ button.

In the “**General**“ tab you can adjust:

- **Max. blocking operations** - Maximum of simultaneously sent blocking operations. Blocking operations are testing and remove operations. High value may cause server overload, recommended is 10.

- **Pending retry time** - Defines how often the application tries to send requests for pending operations (e.g. request for testing results). Higher value decreases server load but may cause slower response.

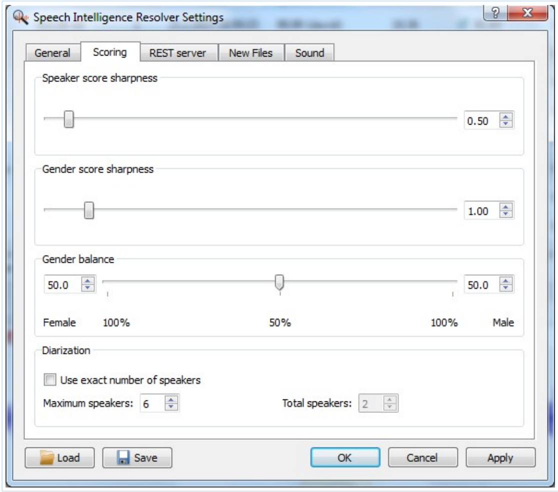
- **Network timeout** - Time when the connection is deemed to be lost.

- **Language** – Language of application. If you select "System", the default language is used if available, otherwise English is selected. The language change will appear after the application restart.

- **Silent mode** – Do not show any information, warning or error dialogs (all messages are visible in Error Console tab).

- **Use VAD for retrieve speech length** – If the speech length is not´t interesting for you, don´t use it, because it takes some time.

- **Minimum speech length** – this option gives minimum limit for net speech for processing. If the net speech is lower, the results are not shown. We strongly recommend to use 9 sec of net speech as minimum.



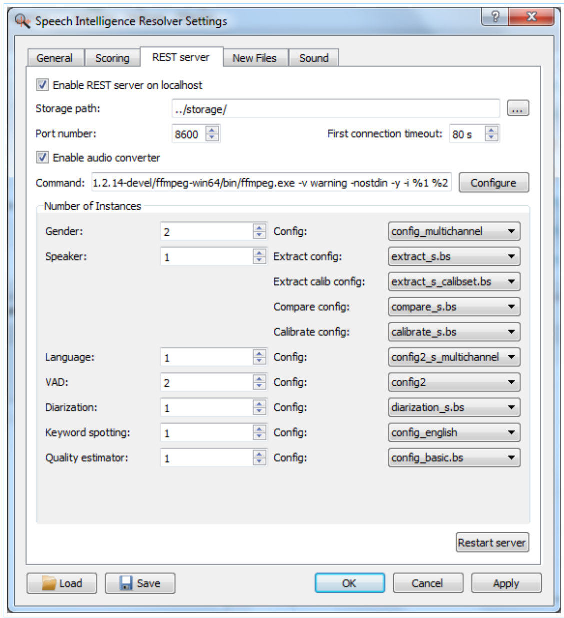
The “**Scoring**“ tab adjusts the score to a particular task.

For example if you feel like having too many scores with 100%, it is possible to decrease the “**Speaker score sharpness**“ and non-linearly map the scores closer to 50%.

A similar option “**Gender score sharpness**“ is available for the gender identification. In some cases, the technology favors one gender over the other. For example, the female voices have higher fundamental frequency that can be more affected by the transmission channel. This can be corrected by the “**Gender balance**“.

The “**Diarization**” adjusts the number of speakers in diarization. You can set maximum number of speakers (technology index audio for up to this maximum) or total number (technology use this amount of speakers, used for older version of diarization). We recommend to set "Maximum speakers" = 2 for phone calls.

The “**REST server**” tab adjusts:



- **Enable / disable running REST server** on localhost. If localhost REST server is disabled, local directories will not be accessible.
- Storage path for REST server. This folder will be accessible via “Localhost” item in “**Sources view**”.
- Port on which server will run.

- **Enable / disable audio converter** You can use some of the free availbale audio converters such as SOX, FFMPEG etc. Installation is easy – example for FFMPEG:
  - Download tool from some of the sourcepages (eg. http://ffmpeg.zeranoe.com/builds/ (use static package . eg. FFmpeg git-* 64-bit Static)
  - Unpack it on your HDD (eg. To SIR folder)
  - Go to SIR > menu Setting > REST server tab
  - Click “Configure” button
  - Select where the FFMPEG (EXE file) is installed. When SIR test the FFMPEG by reading parameters automatically, confirm OK. SIR will add command automatically.

- Number of gender / speaker / language identification instances which run on REST server. More instances speed up testing, but need more resources.
- Number of voice activity detector (VAD) and Speech Quality Estimator (SQE) instances which run on REST server.

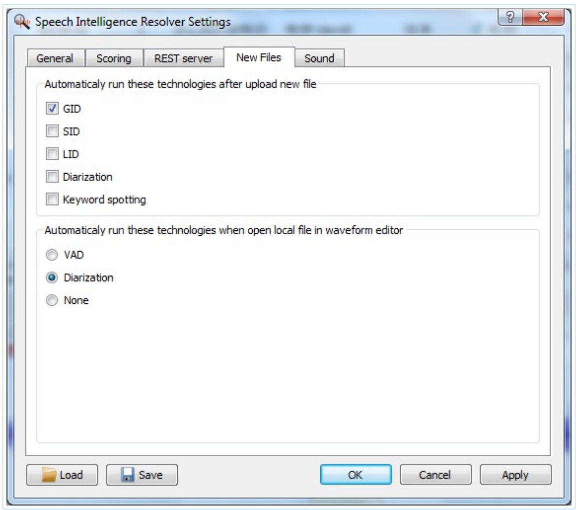
- Configuration files for technologies. Version (l / s / o) of SID extractor configuration must be the same as comparator configuration.

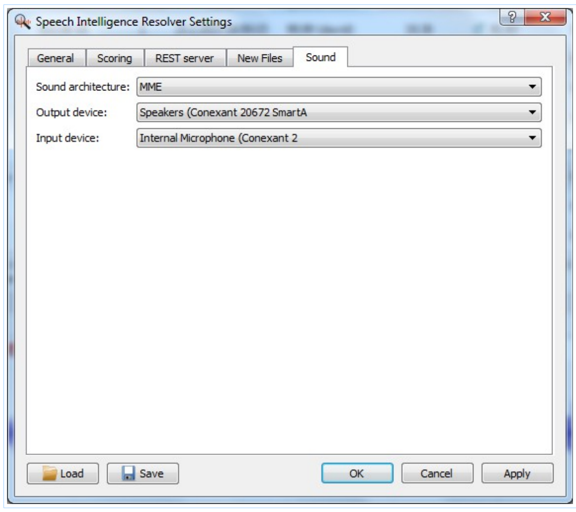
To apply changes user can restart server by "Restart server" button. Otherwise the changes will be applied after the restart of the application.

The “**New Files**” tab allows user to set:

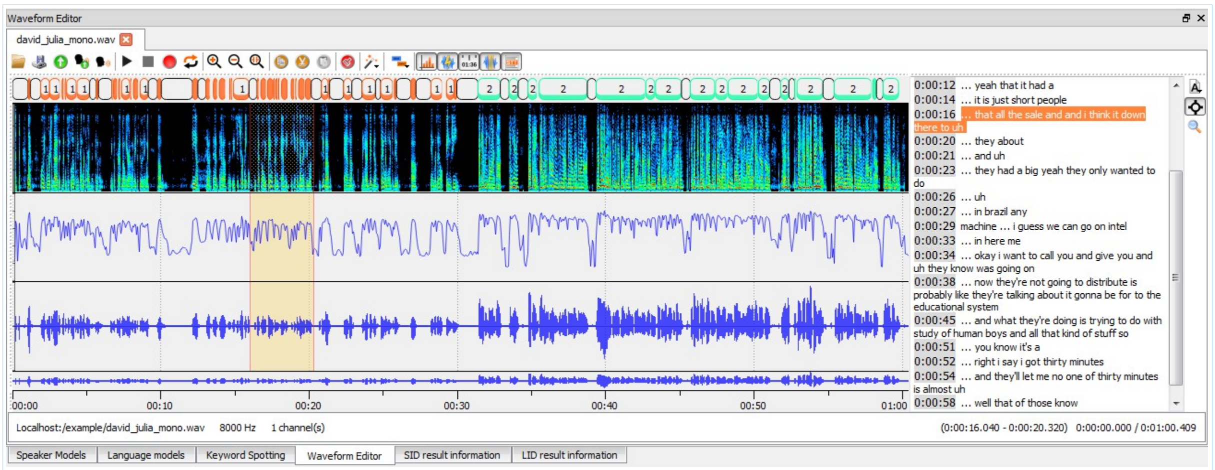
- What technologies should be processed automatically after the new file upload
- What technologies should be processed automatically when local file is opened in Waveform Editor.

The “**Sound**” tab adjusts the output device.





i) Waveform Editor

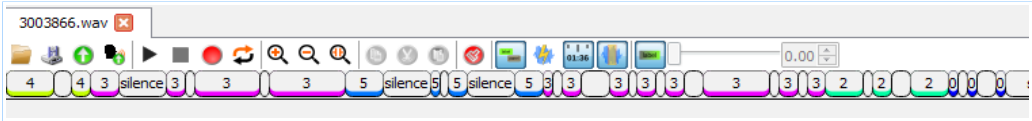


With a waveform editor you can easily manage your records using several features.

You can open a local file with „Open record file” button . You can open records from:

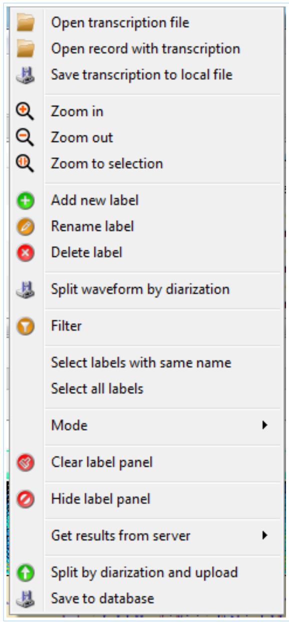
- a local file (CTRL+O)
- from the REST server (in the Test view, by right clicking on particular file)

- with diarization results (in the Test view, by right clicking on particular file), which mark one speaker off another and also voice off silence



- with VAD segmentation (in the Test view, by right clicking on particular file) which marks voice off silence

The context menu appears after right click on selected label. Example for right click on diarisation label on right.

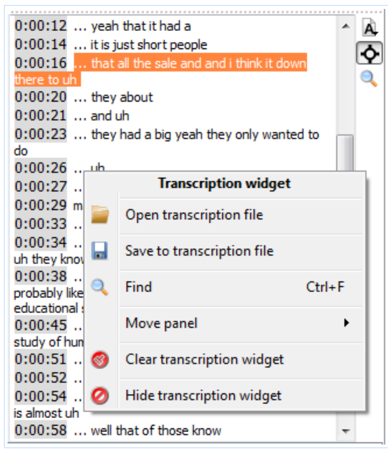


Several label panels can be shown:

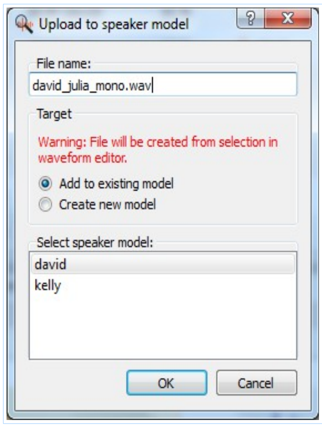
You can use/hide also

- Add label panel" allows you to add some label panel type (Generic, VAD, Diarization, KWS)
- Minimap panel
- Time panel
- Power panel
- Spectrum panel
  - here you can use several color schemes
- Label panel
  With a Keyword Spotting label feature you can play around with:
  - showing a label score
  - colors by threshold colors by label score
  - edit
  - move
- colors by threshold colors by label score

- STT label panel allows you to show transcript output
  With a STT (transcription) label feature you can play around with:
  - Open/Save transcript of audio
  - showing a transcript (synchronized with waveform through time-lines)
  - show transcript as Plain text / Sentences / Blocks
  - Snap to mouse (if active, a particular words are highlighted while user move above waveform)
  - Request automatic transcription of audio (license required)



 You can save a record (CTRL+S) to the file on local storage or to the REST server.



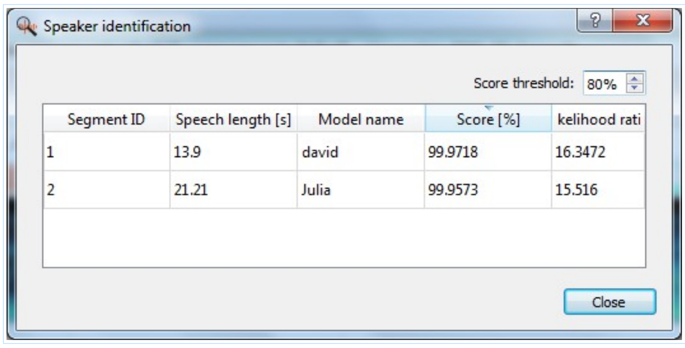
You can upload your waveform to the server to an actual folder in the Test view (CTRL+U).

You can use the current your waveform (or selection to :

- Create new speaker model (create new voice print) or
- Add to existing model (enhance voice print)

  The small window will appear for confirmation. If user select some part(s) of waveform only that part(s) will be used. If user does not select any part the whole audio is used.

  This is useful feature for new speaker registration or voice-print enhancement!



You can run Speaker Identification task directly from Waveform Editor on the waveform (or selected part). The table results will appear for comparison against speaker models / group selected in Speaker Models tab.

Play your record with a „Play / Pause“ (SPACE) button:

- play from the cursor position
- play only elected parts
- play only visible channel (set which channel should be visible by right clicking on the waveform and choose channel 1 / channel 2 / all channels)

With the „Loop playback button” you can:

- play from the cursor position
- play only selected parts (if several parts selected cursor skip a non-selected part)
- play in loop (de)activated by icon

You can record new voice using your microphone (before recording starts, you will be asked to specify input device (external closed microphone recommended), frequency (8kHz recommended) and number of channels (1 recommended)).

Hint: There are several possibilities you can do with recorded audio:

- Create a speaker model
- Add the audio to existing speaker model.
- Upload audio to server for processing (as WAV file).

Cut / Copy / Paste only selected parts of a waveform. You can use this function for:

- a single selected part
- multiple selected parts (using CTRL)
- if you have label panel displayed, it adapts to changes, but if you save the waveform, the changes in label panel aren´t saved

There are also several typical instruments for a better customer usage such as „Zoom in“, „Zoom out“, „Zoom to selection“. You can move waveform position by mouse dragging with middle button pressed. With a red button you can clear your editor.

You can use some of the Effect included:

- Amplify selected segment of signal to higher or lower volume (dB)

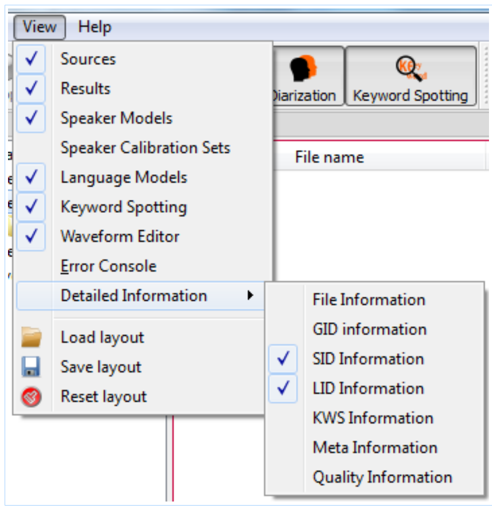
SIR system support standard shortcuts:

- 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

j) Layout

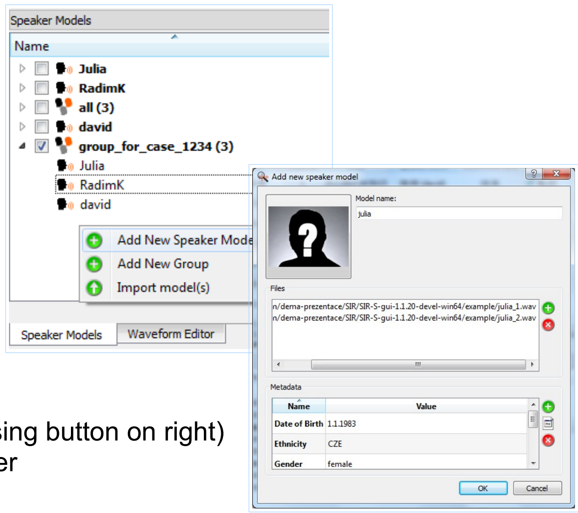
Dragging and moving particular windows you can customize a layout of the application. You can also merge several windows into one window with several tabs. In the View option in the main menu you can choose which windows should be displayed and which not There is also Save layout option. If you save it, you can use preferred layout anytime using the Load layout option. If you make some changes in the layout, you can refresh the original layout easily using Reset layout option.

Hint: The layout might be created by senior expert and send to junior analytic. The layout is also stored with closing SIR application.



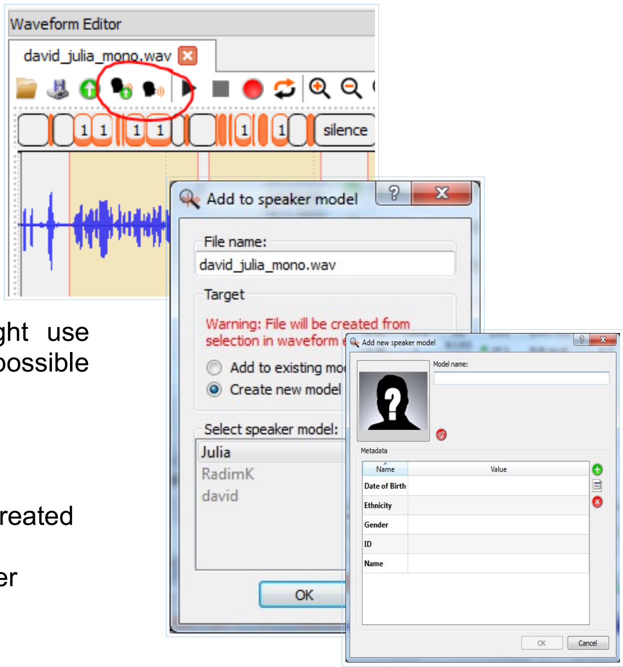
4. Using the Application -Typical Work-flows

a) Speaker Registration (speaker model enrollment)



There are two possibilities for speaker registration:

1. **using Speaker Models view**
   a) Right-click in the view
   b) Choose "Add New Speaker Model"
   c) Enter
   - Model name
   - Audio files (audio should contain only voice of target speaker, 40sec speech recommended)
   - Metadata (you can created any field using button on right)
   - you might add also photo of the speaker
   d) Confirm with OK



2. **using Waveform Editor**
   a) Open audio in Waveform Editor
   b) Run Diarization first if you expect several speakers in audio
   c) Select segment(s) with voice of target speaker only (you might use double-click on diarization label or select several segment holding Ctrl key.
   d) Click "Add waveform to server into speaker model" button (you might use "Speaker Identification button to check possible duplicities in speaker database first)
   e) Select "Create new model"
   f) Enter file name
   g) Enter
   - Model name, and metadata (you can created any field using button on right)
   - you might add also photo of the speaker
   h) Confirm with OK

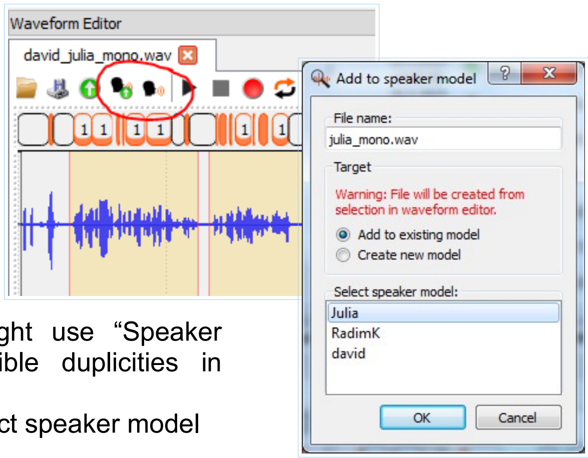
b) Speaker Update (speaker model enhancement)

There are two possibilities for speaker registration:

1. **using Speaker Models view**
   a) Right-click on the Speaker model name
   b) Choose "Add Files"
   c) Select audio files from you HDD (audio should contain only voice of target speaker, 40sec speech recommended)

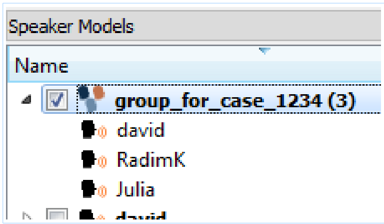
- Metadata (you can created any field using button on right)
- you might add also photo of the speaker

d) Confirm with OK



2. **using Waveform Editor**
   a) Open audio in Waveform Editor
   b) Run Diarization first if you expect several speakers in audio
   c) Select segment(s) with voice of target speaker only (you might use double-click on diarization label or select several segment holding Ctrl key.
   d) Click “Add waveform to server into speaker model” button (you might use “Speaker Identification” button to check possible duplicities in speaker database first)
   e) Select “Add to existing model” and select speaker model
   f) Enter file name
   g) Confirm with OK

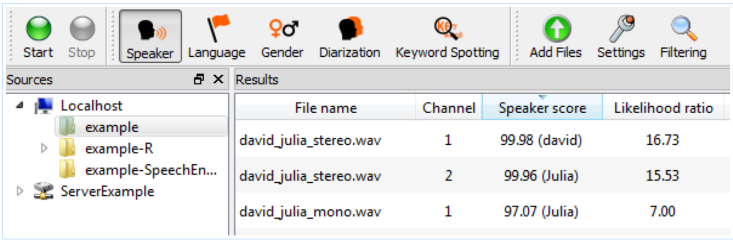
c) Speaker Search (Speaker Identification)

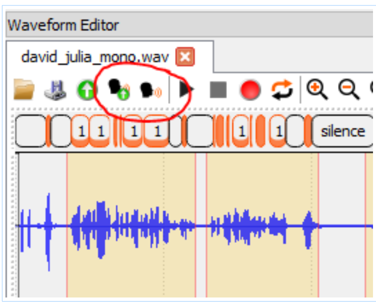


Prerequisites: If you want to use speaker identification, it is necessary to create a speaker model(s) first. Searched speaker (or group with target speaker(s)) should be selected in Speaker Models view.

There are two possibilities for speaker identification:

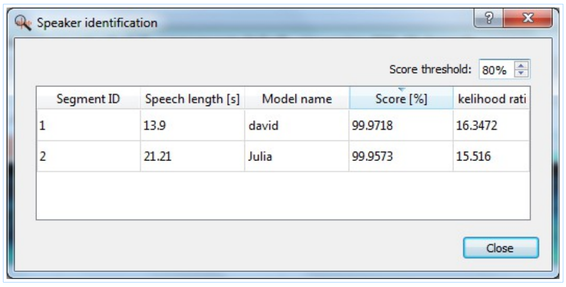
1. **using Results view**
   a) Add all investigated audio files to Results view (you might use Drag&Drop or use icon “Add Files” button in top menu
   b) Click “Start” button in top menu (be sure that the Speaker Identification technology is active
   c) Results for Speaker Identification appears in columns “Speaker score” and “Likelihood ration”
   d) If speaker group selected, the speaker with the highest score will appear in brackets. (For score interpretation, please see details for results in Results > Speaker Identification results above.)





2. **using Waveform Editor**
   a) Open audio in Waveform Editor
   b) Run Diarization first if you expect several speakers in audio
   c) Click “Speaker Identification” button

d) Speaker Identification will be processed for each speaker label (Segment ID produced by diarization).
e) The score is ordered by “Score [%]”
f) You might lower “Score threshold” if you wish to see more results.



5. Speech Technologies Available in SIR

Phonexia Speaker Identification (SID2)

This highly accurate technology uses the power of voice biometry to automatically recognize a speaker by voice and to search for a specific speaker in an archive of speech records. It allows you to verify the caller by his/her voice against thousands of voiceprints. It verify the speaker very quickly by his/her voice biometry using his/her unique voiceprint. Basic features: small voiceprint size (<2kB for standard and <15kB for); text-, language-, and channel independent.

Phonexia Language Identification (LID2)

Phonexia Language Identification will help you distinguish the spoken language or dialect. Our technology is very fast and it has been an integral part of many solutions in the security/defense sector. Pre-trained languages (in alphabetical order): Afan_Oromo, Albanian, Amharic, Arabic_Iraqi, Arabic_Levant, Arabic_MSA, Azerbaijani, Bangla_Bengali, Bosnian, Burmese, Chinese_Cantonese, Chinese_Mandarin Creole, Croatian, Czech, Dari, English_American, English_British, Farsi, French, Georgian, German, Greek, Hausa, Hebrew, Hindi, Indonesian, Italian, Japanese, Khmer, Kirundi_Kinyarwanda, Korean, Lao, Macedonian, Ndebele, Pashto, Polish, Portuguese, Russian, Serbian, Shona, Slovak, Somali, Spanish, Swahili, Tamil, Thai, Tibetan, Tigrigna, Turkish, Ukrainian, Urdu, Uzbek, Vietnamese.

Phonexia Speech Transcription (STT)

Converts natural spontaneous speech into text. Unlike other technologies, Phonexia Speech Transcription is able to handle noisy recordings. You can easily search and categorize recognized text with text-based data-mining tools. Available languages (in alphabetical order): Arabic (Levantine, prototype), Czech, English_US, Chinese (Mandarin), Russian, Spanish (N&S American), Slovak.

Phonexia Keyword Spotting (KWS)

Phonexia Keyword Spotting (KWS) helps you to find keywords or phrases of your interest. You need only specify the keywords while the pronunciation is generated automatically. The user can add several variants of pronunciation for each keyword or phrase. Available languages (in alphabetical order): Arabic (Levantine, prototype), Czech, English_US, German, Hungarian, Italian, Polish, Russian, Spanish (N&S American), Slovak.

Phonexia Gender Identification (GID)

Did you know that men are arrested more frequently than women? This technology identifies whether a speaker is male or female. This engine reacts extremely quickly and accurately and is robust under all sound conditions. Phonexia Gender Identification helps you to distinguish the gender of a speaker. With this very precise speech technology, you can halve the search area.

Phonexia Speaker Diarization (DIAR)

With speaker diarization you are able to distinguish different speakers in one monochannel audio recording and divide it into several recordings by each speaker.

Phonexia Speech Quality Estimator (SQE)

Helps you to estimate the quality of the speech in audio recording. The results may serve for e.g. discarding low quality recordings from processing by other speech recognition technologies.

Speech Preprocessor (incl. Voice Activity Detection)

Contains several techniques to preprocess the speech signal: voice activity detection, technical-noise removal (e.g., DTMF), speech enhancement techniques, noise reduction, etc.

6. Troubleshooting

In case of any problems, please search for particular keywords in this manual (availbale in SIR package as “SIR_manual.pdf” or in SIR menu > Help > User Guide (F1).

If problem remains, please describe intended action to support@phonexia.com . In case of technical problems, please send us also “Error console description” (available from SIR menu > View > Error Console (use “Save to file” icon on right)).

7. 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.

If any bug appears, please contact us at support@phonexia.com.

Contact us
for additional information

Phonexia s.r.o.
support@phonexia.com | +420 511 205 265
U Vodárny 2a, 616 00 Brno, Czech Republic, Europe

8. Annexes

List of Annexes

- Annex 8.1. - Abbreviations
- Annex 8.2. - Disclaimer

Annex 8.1. - Abbreviations

ASR	Automatic Speech Recognition (several technologies possible - see LVCSR, STT or KWS)
DIAR	Phonexia Speaker Diarization
GID	Phonexia Gender identification
KWS	Phonexia Keyword Spotting (acoustics based ASR, language dependent)
LID2	Phonexia Language Identification v.2
LVCSR	Large-Vocabulary Continuous Speech Recognition
SID2	Phonexia Speaker Identification v.2
SQE	Phonexia Speech Quality Estimator
STT	Phonexia Speech Transcription (LVCSR based ASR, language dependent)
VAD	Phonexia Voice Activity Detection

Annex 8.2. - 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.