SVTOOLS software user's manual

Basic functions and offline processing

By Carlos Ishi, last revision 2020.06.10

General remarks:

The program "carlos-svtools.exe" includes several functions for recording, processing, playing, sending/receiving, labeling and visualizing single and multi-channel audio signals. In this manual, basic functions for opening, playing, visualizing and processing audio signals are described.

All configuration files have to be in a "config" directory placed in the same directory of the .exe file. For all items in the configuration file, the line right below a specified item (e.g., [paramdir]) is the valid one. All other lines are ignored.

Basic functions

--- Before running ---
In the config-svtools.txt file, set in the line below [paramdir] an existing directory path for saving parameters, and in the line below [tempdir] an existing directory path for saving temporary files. If the directories do not exist, “param” and “temp” directories will be created in the same directory of svtools.exe.

--- 起動する前に ---
パラメータ保存用のフォルダを作成し、config-svtools.txt 内の[paramdir]の下の行にそのパスを指定してください。
同様に、一時保存用のフォルダを作成し、config-svtools.txt 内の[tempdir]の下の行にそのパスを指定してください。

--- Opening a wav file ---
Menu bar Wave – Open wave or button <WAV> in the toolbar
Wav files are recommended to be 16 kHz/16 bits, mono or stereo. Other sampling rates can be opened, but appropriate processing of the signals is not guaranteed.
Another dialogue box appears asking for opening a label file. Press <cancel> if a label file does not exist.

--- wav ファイルを開く ---
ツールバーの<WAV>ボタン、もしくは、メニューの Wave – Open Wave で wav ファイルを開く。
16 kHz / 16 bits の形式がデフォルトである。モノ・ステレオのいずれも可能。
ラベルファイルを開くためのダイアログボックスも自動的に起動される。ラベルファイルが存在しない・もしくは開かない場合は、<Cancel>を押す。

--- Zooming ---
Click the button <z> in the toolbar for zoom out the time axis.
Click the button <Z> in the toolbar for zoom in the time axis.
Spin buttons at the right side of the panels can be used to zoom in/out the signals in time or amplitude.

--- ズーム機能 ---
ツールバーの<z>ボタンでズームアウト
ツールバーの<Z>ボタンでズームイン
それぞれのパネルの右側にあるスピンボタンで時間方向または振幅のズームが可能。

--- Selecting and playing ---
Use left and right buttons of the mouse to select an interval on the display.
Press <space> key or <enter> key to play the selected channel for the selected interval. (The <space> key cannot be used if "text input mode" is activated. (see "Transcription and labeling tools" below.)

--- 区間選択および再生 ---
波形パネルやスペクトログラムパネル上で、マウスの左ボタンおよび右ボタンをクリックすることにより、区間が選択される。
<space>キーもしくは<enter>キーを押すと、選択区間において選択されたチャンネルが再生される。(ただし <space>キーは "text input mode" が有効の場合には使用できない。以下の "Transcription and labeling tools"を参照。)

--- Opening a second wav file ---
Menu Wave – Open second wave
The first channel is shown in black, while the second channel is shown in blue, in the waveform and spectrogram panels. The first channel remains selected.
Press <F4> to switch the selected channel.

--- 2 つ目の wav ファイルを開く ---
Menu Wave – Open second wave
チャンネル 1 は黒で手前に、チャンネル 2 は青色で後ろに、波形およびスペクトログラムパネルで表示される。チャンネル 1 が選択された状態が保たれる。
<F4> キーを押すと、選択されたチャンネルが入れ替わる。

--- Spectrogram view option for two audio signals ---
Menu View – Spectrogram – Masked spectrogram (stereo)
Shows in black the predominant frequency components of channel 1 (relative to channel 2), and shows in blue the masked frequency components (predominant in channel 2), in the spectrogram panel.
When a second wav file is opened, this option is automatically switched on.

--- 2 つの信号を開いた場合のスペクトログラムの表示 ---
メニュー View – Spectrogram - Masked spectrogram (stereo)
スペクトログラムパネルで、左チャンネルの（右チャンネルに対する）強い周波数成分を黒で、右チャンネルの強い成分によりマスキングされた周波数成分を青色で表示する。

--- Stereo player ---

In the case two wave files are open:
Press <F1> to play channel 1.
Press <F2> to play channel 2.
Press <F3> to play channels 1 and 2 (in stereo).

--- ステレオ再生 ---
音声信号が２つ開かれている場合：
<F1> キーを押すと、チャンネル１が再生される。
<F2> キーを押すと、チャンネル２（２番目に開いたファイル）が再生される。
<F3> キーを押すと、チャンネル１とチャンネル２が左右（ステレオ）で再生される。

--- Signal Processing Tools for the selected channel (Menu SigProc) ---
Frame feature extraction: Extract acoustic features (F0, power, LPC, SACF, formant, and other voice quality related parameters), each 10ms.
Gain (x2): Amplifies the whole signal by a factor of 2.
Attenuate (/2): Attenuates the whole signal by a factor of 1/2.
Cut DC and low freq. or toolbar button <FILT LOW>: Filters DC and low frequency components below 70 Hz.
Left by Right WF: Attenuates the frequency components of the left channel, which are smaller than that of the right channel, based on Wiener filtering.

--- Signal Processing Tools for both channels (Menu SigProc) ---
Manual WF: Background noise suppression, based on Wiener filtering. A segment where only stationary noise is present in both channels has to be set for estimating the noise components.
Dual channel WF: Inter-channel leakage suppression, based on Wiener filtering. Frequency components predominant in the counterpart channel are attenuated. The scroll bar at the left side of the spectrogram panel controls the level of suppression.

--- 選択されているチャンネルに対する信号処理ツール （メニュー SigProc）---
Frame feature extraction: 10ms 毎の音響特徴を抽出（基本周波数、パワー、LPC、SACF、フォルマント、その他の声質に関連したパラメータ）。
Gain (x2): 波形全体の振幅を 2 倍にする。
Attenuate (/2): 波形全体の振幅を半分にする。
Cut DC and Low Freq. または <FILT LOW>ボタン: 直流や 70 Hz 以下の低い周波数成分をフィルタリングする。
Left by Right WF: 左チャンネルの周波数成分に対して、右チャンネルよりも弱い周波数成分を Wiener フィルタにより抑圧する。

--- 両チャンネルに対する信号処理ツール （メニュー SigProc）---
Manual WF: Wiener フィルタによる背景雑音の抑圧。雑音成分の推定に、定常的雑音のみが存在する区間を選択する必要がある。
Dual channel WF: Wiener フィルタによるチャンネル間音漏れの抑圧。 相手チャンネルの方が強い周波数成分を抑圧する。スペクトログラムパネルの左側のスクロールバーで抑圧レベルを調整できる。

--- Transcription and labeling tools ----
Click the button <NEW LAB> in the toolbar for opening a new label file.
An edit panel appears above the time bar, and a label bar appears below the waveform panel.
Select an interval by clicking the left and right mouse buttons.
Press <F8> to insert a new segment in the selected label tier.
A segment is selected by clicking the left mouse button on a segment in the label bar.
A segment can be split (i.e., a segment boundary is created) by clicking the left mouse button on the label bar, while pressing the <Shift> key.
Press <F10> to delete a selected segment in the selected label tier.
Subsequent segments can be merged (i.e., a segment boundary can be deleted) by clicking the left mouse button on a segment boundary, while pressing the <Shift> key.
A segment boundary can be moved, by pressing the left mouse button on a segment boundary, and dragging the mouse to the desired position, on the label bar.

A segment can be selected either by clicking the left mouse button on a segment in the label bar, or by clicking an edit box in the edit panel.
A text or label can then be input in the selected label segment.
The two columns at the left side can be used to text transcription (at most 512 bytes per item), while the five columns at the right side can be used to label annotation (at most 32 bytes per item).

If the "Auto Hiragana" option is activated in the menu bar Label - Auto Hiragana, Hiragana characters are automatically input in the second column of the text panel, when something is typed in the first column of the text panel. (However, mistyping in the first column has to be manually corrected in the second column.)

Click the button <SAVE LAB> for saving the selected label file (in text format).
The extension .ctr is default for label files.

The created label file can be opened when opening a wav file, from the next time.
The label file can also be opened by clicking the button <LAB> in the toolbar.

Warning: This tool was originally created for text transcriptions in Japanese, so that the insertion of spaces is not necessary, and <space> key is used to play segments. If you need the insertion of spaces, e.g. in English text transcription, activate "Text input mode" option in the menu bar Label – Text input mode. In this case, <space> key cannot be used to play segments. Use <enter> key instead.

--- ラベリング ツール ----
ツールバーの<NEW LAB>ボタンを押すと、新しいラベル・バーが作成される。
時間バーの上にテキストパネルが表示され、波形パネルの下にラベル・バーが表示される。
ラベル・バー上で、<shift>キーを押しながら、マウスの左ボタンをクリックすると、その位置にラベルの切り出しが挿入される。
切り出しを削除したい場合は、<shift>キーを押しながら、削除したい切り出しをラベル・バー上で、マウスの左ボタンでクリックする。
切り出し位置を移動したい場合は、移動したい切り出しをラベル・バー上で、マウスの左ボタンを押したまま、移動したい位置までマウスを動かす。

ラベル・バー上で、マウスの左ボタンをクリックすると、その区間が選択される。
テキストパネルで、選択されたラベルの書き起こしが可能である。
メニューの Label - Auto Hiragana で、"Auto Hiragana"が有効化されている場合、テキストパネルの 1 番左の列に漢字かな混じりの日本語文字を入力した場合、自動的に 2 番目の列にひらがな表記で入力される。(ただし、打ち間違いは反映されないので、2 列目は手動で訂正する必要がある。)

ツールバーの<SAVE LAB>ボタンを押すと、ラベルファイルが保存される。
ラベルファイルの拡張子は　.ctr となる。

作成されたラベルファイルは、次回 wav ファイルを開いた際に、開くことが出来る。
ツールバーの<LAB>ボタンを押して、既存のラベルファイルを開くこともできる。

本ツールは当初、日本語の書置きしを行うために作成されており、<space>キーは、音声の再生用に使用している。英語の書き起こしで文字列に空白を記入する必要がある場合は、メニューバーの Label – Text input mode を有効にしてください。この場合、音声再生には<enter>キーを使用してください。

Format of the .ctr file:
TAB separated text file; can be opened with Excel to allow better visualization.
The first line may contain the global time in Unix time, corresponding to the file name

(YYYYMMDD-HHMMSS-MSC-XX.ctr)
The next line contains the header.
The subsequent lines contain the following segment information, one line per segment.
A: filebase (file name of the wav file without extension)
B: ti (initial time of the segment, in seconds)
C: tf (final time of the segment, in seconds)
D: text1 (text data): We use this slot to input text transcription （漢字かな）
E: orig_filebase (file base of the original file: this is used when utterance segments from different wav files are joined into a new file; empty, if not created by an utterance list.)
F: orig_ti (initial time of the segment in the original wav file, in seconds; -1 if not created by an utterance list)
G: orig_tf (final time of the segment in the original wav file, in seconds; -1 if not created by an utterance list)
H: lsti (initial time of the last syllable in the segment): The last syllable conveys important information on intentions and attitudes, through its prosodic information. In the parameter extractions these are automatically extracted, but they are not robust, so that manual corrections are necessary before analysis!
I: lstf (final time of the last syllable)
J: text2 (text data): We use this slot to input hiragana transcription corresponding to text1
K: dur (segment duration, in seconds)
L: ls_dur (last syllable duration, in seconds)
M~Q: label1~label5 (label data): We use these slots to input different types of labels corresponding to the segment (e.g., dialogue acts, intonation, voice qualities, etc.)
R: prev_pause (pause duration relative to the previous segment, in seconds; -1 for the first segment)
S: next_pause (pause duration relative to the next segment, in seconds; -1 for the last segment)
T: cleantext1 (cleaned text1; remove symbols and keep only text information; automatically created)
U: cleantext2 (cleaned text2; automatically created)

The subsequent columns are created after Sigproc – Segment Feature Extraction. The label file can be overwritten or saved in a new parameter file. (See svtools-features-202006.pdf)

Utterance segmentation

1. Open the wav file to be segmented
Menu bar Wave – Open wave or toolbar button <WAV> or drag the wav file to the svtools icon

2. Extract acoustic features
Menu bar SigProc – Frame Feature Extraction
Wait until the message "Parameter extraction finished" appears in the status bar.
The pitch and power panels appear between the waveform and spectrogram panels.
Once the feature extraction is conducted, the parameters will be loaded when opening the wav file from the next time.

3. Adjust the power threshold
The spin button at the left side of the power panel changes the power threshold.
The red straight line shows the power threshold. Adjust the power threshold so that the red line becomes lower than the black and green lines (representing the low and high frequency-band power) for the intervals the target speech is active, and higher for the background noise (non-speech) intervals.
If the green lines are much higher than the black lines in the non-speech intervals due to presence of high frequency noise, it is recommended to conduct background noise suppression (step 5 in inter-channel suppression) and repeat from step 2.

4. Utterance segmentation
Menu bar SigProc – Utterance segmentation
A label tier appears below the waveform panel showing the utterance segmentation results. You can check how the segmentation worked by selecting a segment (left mouse button click in the label tier, and playing (by <space> bar). Cursor up and down keys can be used to change segments.
The segmentation results are saved in the same directory of the wav file "XXXX.wav", with a file name "XXXX-autoseg.ctr".

If you are not satisfied with the segmentation results, repeat steps 3 and 4, by adjusting a new threshold.
The new segmentation results are overwritten in the "XXXX-autoseg.ctr" file, and appear as new label tier in the label panel.
You can close the unnecessary label tiers by selecting a segment in the label tier to be closed and execute

Menu bar Label – Close label or toolbar button <CLOSE LAB>

5. Save the final version of segmentation
Menu bar Label – Save label or toolbar button <SAVE LAB>

Creating small wav files for each utterance

1. Prepare a list of utterances according to the following format (.txt):

==========

C:¥inputdir¥

C:¥outputdir¥

test1 0.32 1.45 こんにちは

test1 3.12 3.652

test2 0.2 1.10 バイバイ

==========

In the first line, set the input directory where the input wav files are stored.
In the second line set the output directory where the utterance wav files are to be saved.
From the third line, set the wav file names (with or without extension; default extension is .wav), utterance start time (in seconds), utterance end time (in seconds), utterance text transcription (optional), separated by tab, one line per utterance.
Be sure to save the file in ANSI text format! (Unicode will not work!)

2. For creating the utterance wav files, open the file list by:
Menu bar Batch – Create mini wav from utterance list

The output file names are the same of the first input wav file in the list. For the example above, the following wav files will be created.

C:¥outputdir¥test1-0.320-1.450-こんにちは.wav
C:¥outputdir¥test1-3.120-3.652.wav
C:¥outputdir¥test2-0.200-1.100-バイバイ.wav

Joining multiple small wav files in big wav files

Multiple wav files recorded in small cut files (e.g. every minute) can be joined to a single wav file, in order to allow suitable post-processing.

1. Prepare a list of wav files in the following format (.txt):

==========

C:¥inputdir¥
C:¥outputdir¥
121510.wav
121600.wav
121700.wav

122015.wav
122100.wav
122200.wav

==========

In the first line, set the input directory where the input wav files are stored.
In the second line set the output directory where the joined wav files are to be saved.
From the third line, set the wav file names to be joined, one per line. For creating multiple joined files, set an empty line between the file names to be separated.
Be sure to save the file in ANSI text format! (Unicode will not work!)

2. For joining the wav files, open the file list by:
Menu bar Batch – Join Wav files

The output file names are the same of the first input wav file in the list. In the example above, the following two joined wav files will be created.
C:¥outputdir¥121510.wav
C:¥outputdir¥122015.wav

Warnings:
Too large files might cause lack of memory for processing. If two wav files are open, it is recommended that their lengths do not exceed 30 minutes (for 16kHz sampling rate).

Offline inter-channel suppression (2 channel case)

Requirements: two audio files with enough synchronization, and 16kHz sampling rates. Other conditions might work, but performance is not guaranteed.

1. Open the first wav file
Menu bar Wave – Open wave or toolbar button <WAV> or drag the wav file to the svtools icon

2. Open the second wav file
Menu bar Wave – Open second wave

3. Check and fix synchronization between the two signals
It is important that the signals in the different audio channels are synchronized (within 10ms), otherwise the suppression processing might not be satisfactory. If you use signals recorded by different devices and/or different machines, their synchronization might not be good enough even if the clocks are synchronized, due to imperfect clock synchronization and/or different latencies by different audio capture devices.

3.1. Look at the spectrogram panel, and press <F4> key a few times to change the selected channel. Zoom in the time bar to allow better visualization. If there is leakage of one signal in the other channel, you will be able to observe similar spectral shapes changing from black to blue and vice-versa. If the signals are not well synchronized, you will observe that the similar spectral shapes (in black and blue) are shifted in time. (in black and blue)
3.2 If the signals are not well synchronized, conduct manual synchronization. Click the left mouse button to set the target position (on the blue signal), and the right mouse button to set the reference position (on the black signal) to be shifted.
3.3. Shift the signal
Menu bar SigProc – Fix – Shift signal
Note that the beginning or the end of the signal is filled with zeros for the shifted amount.
3.4. Steps 3.1 ~ 3.3 can be repeated until the signals are well synchronized.
3.5. Save shifted signal (you can save now or later after the inter-channel suppression)
Menu bar Wave – Save Wave

4. Optional gain adjustment
The levels of the recorded signals can be optionally adjusted, to make the adjustment of

suppression levels in step 6 easier.

4.1. Adjust the gain of the first channel
Menu bar SigProc – Gain (x2)
or
Menu bar SigProc – Attenuate (/2)
4.2. Adjust the gain of the second channel
Press <F4> to switch the selected channel, and repeat step 4.1.

5. Optional background noise suppression
Background noise can be optionally conducted before inter-channel suppression.
5.1. Select an interval (of about 0.5 ~ 1 second) where none of the speakers are active in both audio channels.
5.2. Conduct background noise suppression
Menu bar SigProc – Manual WF

6. Adjust the suppression levels
The threshold for inter-channel suppression can be adjusted in the scroll bar at the left side of the spectrogram panel. Negative thresholds will suppress less, while positive thresholds will suppress more the interfering signal. You can check that by setting positive thresholds, the spectrogram of the interfering signal becomes blue. However, too positive thresholds will also degrade (suppress) the target signal. Thresholds should be set in a way to have spectral components of the interfering signal in blue, while keeping the spectral components of the target signal in black.

7. Conduct inter-channel suppression
Menu bar SigProc – Dual channel WF

8. Save the processed signals
8.1. Save one of the channels.
Menu bar Wave – Save Wave
8.2. Press <F4> and save the other channel.
Menu bar Wave – Save Wave

Warnings:
<u>If you are working with small cut files (e.g. every minute), and the utterances are cut in the boundary between files, it is recommended to join them before processing</u>.

Batch processing for frame-level parameter extraction

1. Prepare a list of utterances according to the following format (.txt):

==========

C:¥inputdir¥

test1.wav

test2.wav

testN.wav

==========

In the first line, set the input directory where the input wav files are stored.
From the second line, set the wav file names, one line per wav file.
Be sure to save the file in ANSI text format! (Unicode will not work!)

2. For batch-processing the wav files, open the file list by:
Menu bar Batch – Frm. Par. Extr. of multiple Wav files

The output files will be saved in the [paramdir] specified in the config-svtools.txt file.
For saving the parameter files in text format, set [parameter format] to "1" in the config-svtools.txt file.

Batch processing for segment-level parameter extraction

1. Prepare a list of utterances according to the following format (.txt):

==========

C:¥inputdir¥

test1.ctr

test2.ctr

testN.ctr

==========

In the first line, set the input directory where the input ctr files are stored.
From the second line, set the ctr file names, one line per label file.
Be sure to save the file in ANSI text format! (Unicode will not work!)

2. For batch-processing the wav files, open the file list by:
Menu bar Batch – Utt. Par. Extr. of multiple Utt lists

The output files will be saved in the same folder of the ctr files with the format XXX-param.txt.