

汎用声質変換ソフト「こえへん！」取扱説明書

for Ver.Beta3

許可なき転載・再配布を禁ず

No unauthorized distribution

画面の概要

The screenshot shows the main interface of the 'こえへん!' software. The window title is 'こえへん! powered by UNVOCS'. The interface includes a top bar with 'ダークモード' (A) and 'サンプル周波数' (B) set to 16000. Below this is a 'リアルタイム声質変換' (C) section with a '変換開始' button (D) and '入出力チャンネル数' set to 1. To the right is a '音声ファイルを変換' (E) section with an '開く' button and a '変換' button. The next section contains 'API' (Default API), '入力デバイス' (Default input device), and '出力デバイス' (Default output device) (F), with an '更新' button. Below are '高音域の変換係数' (0.01) and '出力音量' (1.0). The '基本周波数の変換方法' (G) is set to '比例変換', with '(変換前基本周波数) × 1.0 + 0' (H) and (I). The interface is split into two columns: '変換元話者' (J) and '変換先話者'. Each column has a list of characters (あ, い, う, え, お) with '開く' buttons, and buttons for 'ファイル名読み込', 'ファイル名保存', '分割位置推定', and '手動補正'. Below each column are '現在の交換データ: (データなし)' and '交換データ読み込' / '交換データ保存' buttons (K). At the bottom, '変換先話者データの作成方法' (L) is set to '音声ファイルから'. A log window (N) at the bottom shows version information and a warning message.

This is a close-up of the '変換先話者' section. It features three buttons: '編集ウィンドウを開く', '音声パラメータ保存', and '音声パラメータ読み込' (M). Below these buttons, it displays '現在の交換データ: 人工的に作成' and '交換データ読み込' / '交換データ保存' buttons.

(A)ダークモード・ライトモード切替ボタン

ダークモード・ライトモードを切り替えます。(→ダークモードを利用する)

(B)サンプリング周波数の設定

読み込むファイル、および音声の入出力のサンプリング周波数を設定します。デフォルト値は 16000Hz です。他の周波数への変更も可能ですが、16000Hz 未満の周波数の設定はできません(PyWorld の仕様)。

(C)リアルタイム声質変換ボタン

リアルタイムで声質変換を行います。

(D)入出力チャンネル数

リアルタイムで声質変換を行う際の、入出力音声のチャンネル数を設定します。

「1」か「2」が選択できます。

(E)音声ファイルの変換

指定した音声ファイルを変換できます。(→音声ファイルを変換する)

(F)入出力デバイスの設定

リアルタイムで声質変換を行う際の入出力デバイスを設定します。(→入出力デバイスを設定する)

(G)高音域の変換係数の設定

高音域に掛ける係数を指定します。0～1.0 の範囲で指定できます。

小さな値にすると出力音声の音質が向上しますが、あまり小さくすぎると音声は不明瞭になります。

デフォルト値は 0.2 です。

(H)出力音量の設定

出力音量を設定します。0.01～2.00 の範囲で指定できます。

(I)基本周波数の変換の設定

声の高さ・抑揚の変換についての設定です。(→声の高さ・抑揚を変換する)

(J)変換データ作成[音声ファイルから]

音声ファイルから変換データを作成します。(→変換データを用意する(ア)音声ファイルから作る)

(K)変換データの読み込み・保存

変換データを JSON 形式で保存したり、それを読み込んだりする時に使用します。

(→変換データを保存する)

(L)変換データ作成モードの切替

「本ソフトで作成」を選択すると、変換先話者の側の画面が(M)に切り替わります。

(M)変換データ作成[音声ファイルから]

変換データを直接パラメータを与えて作成します。(→変換データを用意する(イ)本ソフトを用いて作る)

(N)ログエリア

エラーメッセージなどが必要に応じて出力されます。

変換データを用意する

(ア)音声ファイルから作る

①目的の話者が「あ」「い」「う」「え」「お」と発音している音声ファイルを用意します。

音声ファイルは wav 形式で、チャンネル数は1か2である必要があります。

②「あ」「い」「う」「え」「お」の各段の「開く」ボタンをクリックし、①の音声ファイルを読み込みます。

ファイルを読み込む際には、右上の「サンプル周波数」の値と読み込むファイルのサンプル周波数が一致している必要があります。

③「分割位置推定」のボタンをクリックします。

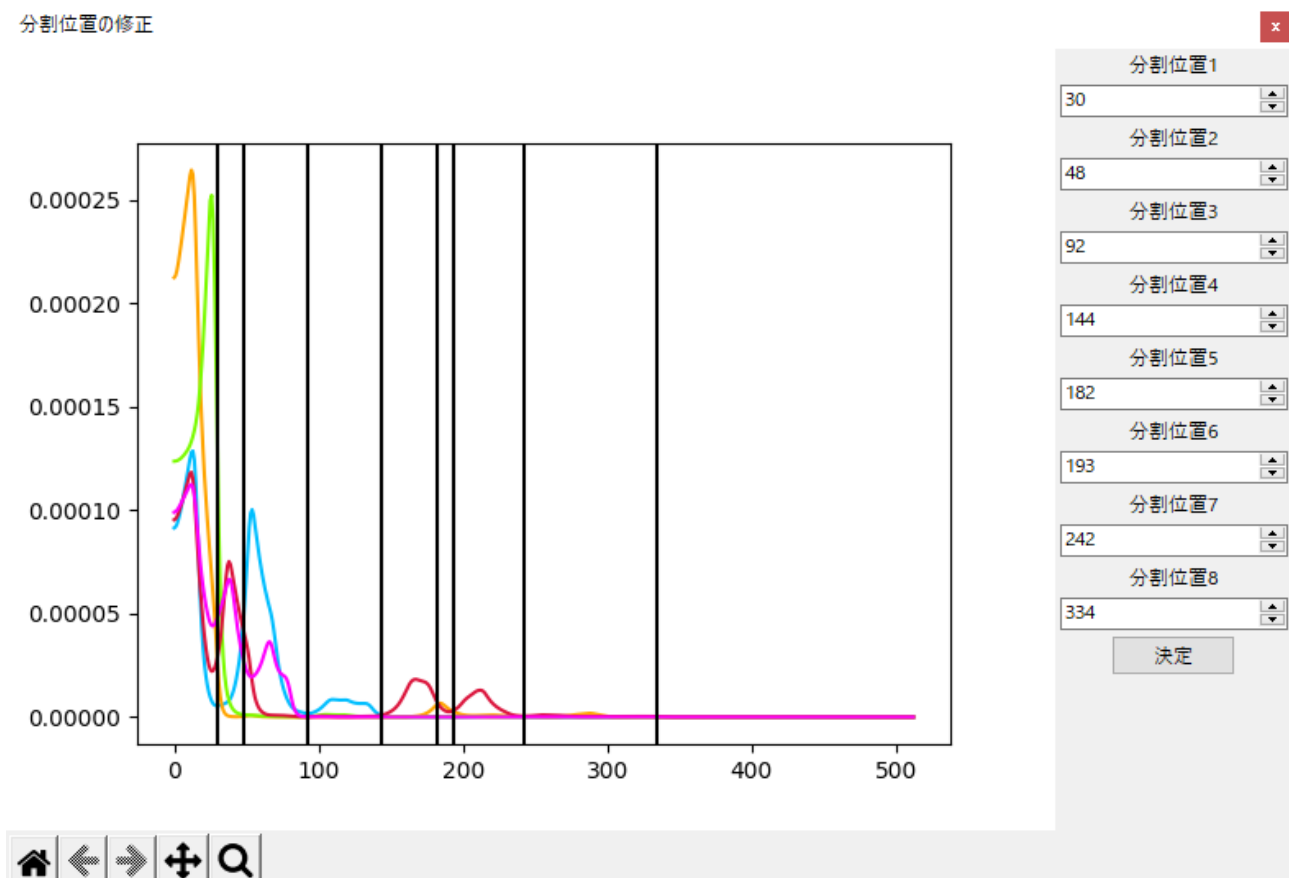
全てのファイルのサンプル周波数が同一でない場合、エラーとなります。

④「手動補正」のボタンをクリックすると、各母音のスペクトル包絡と分割位置が表示されます(下図)。

必要なら、右の入力欄を用いて修正します。修正が完了したら、「決定」ボタンをクリックします。

ウィンドウ右上の×ボタンを押すと修正をキャンセルできます(誤って閉じないようご注意ください)。

分割位置の修正



分割位置について

分割後の領域の特徴は、理論上以下ようになります：

一番左：橙(い)・黄緑(う)がピーク

左から2番目：赤(え)・桃色(お)がピーク

左から3番目：水色(あ)がピーク

左から4番目：水色(あ)がピーク

左から5番目：赤(え)がピーク

左から6番目：橙(い)がピーク

左から7番目：いずれかの母音がピーク

左から8番目：いずれかの母音がピーク

一番右：目立ったピークなし

本ソフトウェアは、この特徴に基づいて自動的に分割を試みますが、音声の個人差により分割が成功しない場合があります。必ず「手動補正」のボタンをクリックして、分割が適切に行われているかを確認してください。

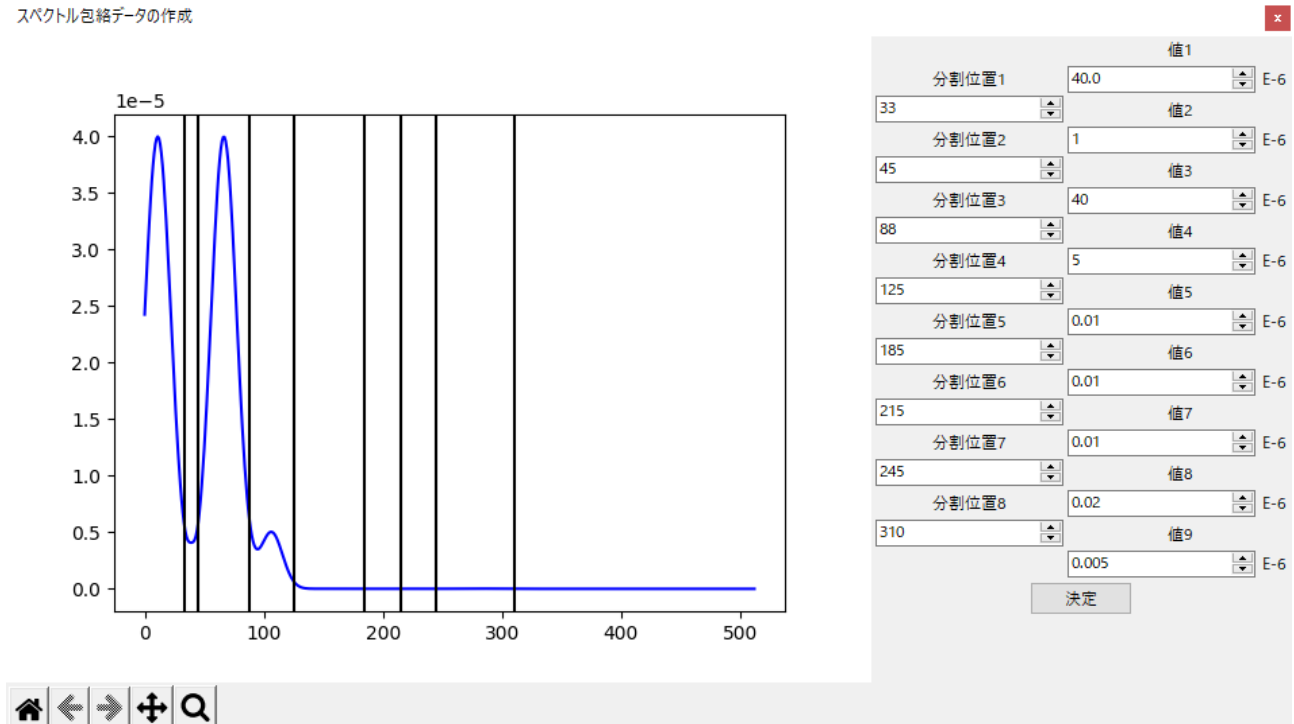
(イ)本ソフトを用いて作る(実験的)

- ①まず、変換元話者の変換データを(ア)音声ファイルから作るに従って準備してください。
- ②「変換データ作成モードの切替」をクリックし、「本ソフトで作成」を選択します。
- ③「編集ウィンドウを開く」をクリックします。下のような画面が表示されます。
- ④右の入力欄を用いて、パラメータを調整します。
- ⑤「決定」ボタンをクリックします。

パラメータの変更をキャンセルする場合は、ウィンドウ右上の×ボタンを押してください。
どんな声になっているかは、実際に何か音声を変換して確かめてください。

- ⑥気に入った音声ができたら、「音声パラメータ保存」のボタンをクリックして、パラメータを保存してください。**パラメータ変更の履歴は残りません。**
保存したパラメータは、「音声パラメータ読込」のボタンから読み込むことができます。

スペクトル包絡データの作成



変換データ作成用のファイル名を記録する

- ①「ファイル名記録」のボタンをクリックしてください。ファイル名がjson形式で記録されます。
- ②保存したjsonファイルは、「ファイル名読込」のボタンをクリックして読み込むことができます。これにより、毎回5つのファイルを開かなくてもよくなります。

変換データを保存する / 読み込む

- ①「変換データ保存」のボタンをクリックしてください。変換データがjson形式で保存されます。形式は、音声ファイルから作成した場合も本ソフトで合成した場合も同一です。
- ②保存したjsonファイルは、「変換データ読込」のボタンから読み込むことができます。

リアルタイムに声質変換をする

- ①まず、変換元・変換先の話者それぞれの変換データを用意してください。(→変換データを用意する)
- ②「変換開始」のボタンを左クリックすると、「変換開始」の文字が「変換停止」に変わり、声質変換が開始されます。
- ③変換を終了するには、「変換停止」のボタンを左クリックします。

※リアルタイム声質変換の際には処理が十分高速に実行される必要があるため、**環境によっては、処理が間に合わず音声途切れる場合があります**。特に入出力チャンネル数を2(ステレオ)にしている場合、チャンネル数が1(モノラル)の場合の2倍の計算コストがかかるため、音声途切れる可能性が高くなりますのでご注意ください。

声の高さ・抑揚を変換する

「基本周波数の変換方法」から、「比例変換」と「オクターブ変換」の2通りの変換方法を選択することができます。

「比例変換」:

左側の入力欄の数値を基本周波数に掛け、右側の入力欄の数値を足します。例えば、入力の基本周波数が100Hz、左側の入力欄の数値が1.5、右側の入力欄の数値が50だったとすると、変換後の周波数は $100 \times 1.5 + 50 = 200\text{Hz}$ となります。

左側の入力欄に1より大きな数を入力したり、右側の入力欄に正の数を入力したりすると元より高い声に、左側の入力欄に1より小さな数を入力したり、右側の入力欄に負の数を入力したりすると元より低い声になります。

なお、右側の入力欄に負の数を入力した場合、変換後の基本周波数がマイナスになるとエラーとなり、変換は直ちに中止されます。**元の声より低い声へ変換する場合でも、右側の入力欄に負の数は入力しないことをお勧めします。**

「オクターブ変換」:

入力欄の数値が1大きくなるごとに半音ずつ声が高くなります。12を入力するとちょうど1オクターブ上の声になります。歌声を変換するときにお使いください。

入出力デバイスを変更する

- ①「API」のプルダウンリストから、サウンド API を選択します。
- ②「入力デバイス」・「出力デバイス」のプルダウンリストから、使用したい入出力デバイスを選択します。

入出力に使用している PyAudio ライブラリの仕様上、入出力で異なるサウンド API を使用することはできません。必要に応じ仮想オーディオデバイスを利用してください。

※デバイスの変更を検出するには、「更新」のボタンを押してください(使用デバイスは、一旦デフォルトのデバイスにリセットされます)。

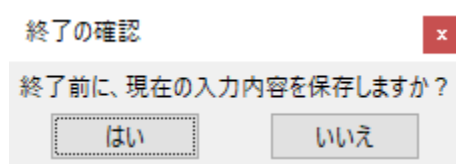
※APIとして Windows-WASAPI を使用している場合、**本ソフトのサンプル周波数と、使用するデバイスのデフォルトの周波数設定が一致している必要があります**(PyAudio ライブラリの仕様?)

ダークモードを利用する

画面左上の切替ボタンをクリックすると、画面がダークモードに切り替わります。OS の設定がダークモードの場合、最初からダークモードの状態です。

終了時の処理

メインウィンドウの×ボタンを押すと、以下の画面が表示されます。



「はい」を選択した場合、実行ファイルと同じディレクトリの中に Latest.json ファイルが生成され、終了時の入力内容が保存されます。

Latest.json ファイルは終了時に「はい」を選択するたびに上書きされるので、必要に応じてバックアップをとってください。