

音声認識システム「おんたま」の使い方
(ONTAMA: ONsei Total Analysis system by MATsumura)

1. 「おんたま」とは

「おんたま」は、オフラインで音声認識(文字起こし)をするためのソフトです。Google ドキュメントの音声入力や Word for Web (Microsoft Office 365)のディクテーションでも音声認識は可能ですが、オンラインでないと使えません。通常の音声データであれば上記を使えば良いですが、何らかの事情でネットが使えないときや、極秘情報のため情報漏洩を防ぎたいときには、オフラインでの音声認識が必要です。そのような時、おんたまでのオフラインでの音声認識が可能です。

おんたまは、Vosk というオフラインの音声認識ソフトと Python を利用して作成しました。

<https://alphacephei.com/vosk/>

音声認識の精度は、オンラインのツールとそれほど異ならないと思われそうですが、興味があるときはご自身で比較してみてください。以下のナレーション音声のように、きれいな音声だと非常に認識精度は高いです。しかし、開発者がふつうに話している音声だと、精度は非常に低くなります。

<http://pro-video.jp/voice/announce/>

2. 免責事項

本ソフト「おんたま」の使用による不利益への責任は負えませんので、自己責任でご利用ください。不具合がありましたら、松村(matutosi@konan-wu.ac.jp)にご連絡いただけると助かります。

3. 導入方法

(1) USB メモリ等で ontama.exe と vosk-model (フォルダ) がまとめて配布されている場合

- ① ontama.exe と vosk-model を任意のフォルダに保存(ここでは「ontama」とする)。
- ② コピーに時間がかかるので、コーヒーを飲みながら(任意)しばらく待機。
- ③ 全てのファイルがコピーされると完了。

(2) ontama.exe と vosk-model をインターネットから取得する場合

ファイルのダウンロードと解凍・移動には時間がかかるので、コーヒーと本(あるいは他の仕事)などの準備がオススメです。

- ① ontama.exe を任意のフォルダに保存(ここでは「ontama」とする)。
- ② 以下の URL から日本語モデルのダウンロード

<https://alphacephei.com/vosk/models> から vosk-model-ja-0.xx.zip
(xx はバージョン番号)をダウンロード。

2023 年 7 月現在の最新版

<https://alphacephei.com/vosk/models/vosk-model-ja-0.22.zip>

ダウンロードに結構な時間がかかるので、コーヒーを飲みながら(任意)しばらく待機。

- ③ ダウンロードした zip ファイルを解凍(右クリックして「全て展開」)(ここでもしばらく待機)。
解凍してできたフォルダ内の「vosk-model-ja-0.xx」の名前を「model-ja」に変更。
- ④ ontama 内に vosk-model というフォルダを作成して、3 の model-ja を vosk-model の中に全て移動。ここでもしばらく待機。
- ⑤ 全てのファイルがコピーされると完了。



4. フォルダ・ファイル構成の概要

フォルダ・ファイル構成が正しいか確認するには、ontama.exe と vosk-model を選択肢して、右クリックで「プロパティ」を選択してください。プロパティが以下のとおりであれば、おそらく大丈夫です。

・ファイル数:30(README.txtと README.pdfを含む),フォルダ数:8

・ファイルサイズ 1.56GB (環境によって多少の違いの可能性あり)

名前の後ろに「/」があるものはフォルダです。model-ja の下位フォルダの内容は省略しました。

ontama/

- └ README.txt (文章のみの説明)
- └ ontama.exe (実行ファイル)
- └ manual.pdf (本ファイル)
- └ vosk-model/
 - └ model-ja/
 - └ am/
 - └ conf/
 - └ graph/
 - └ ivector/
 - └ rescore/
 - └ README



5. 使い方

(1) ontama.exe をクリック。

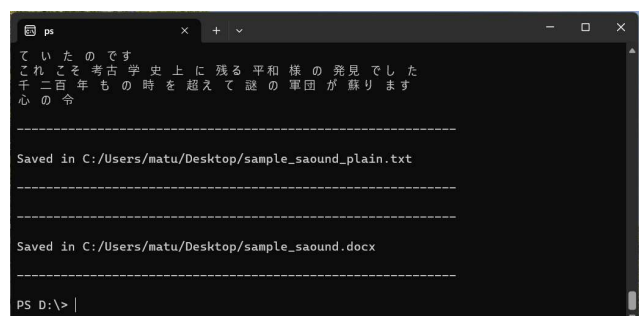
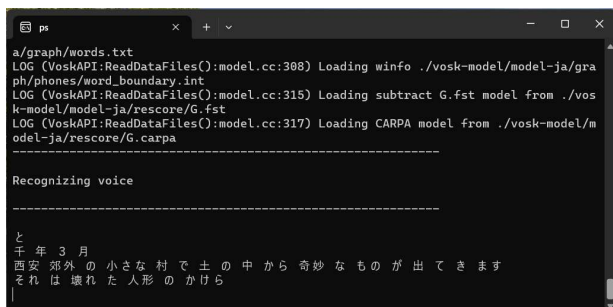
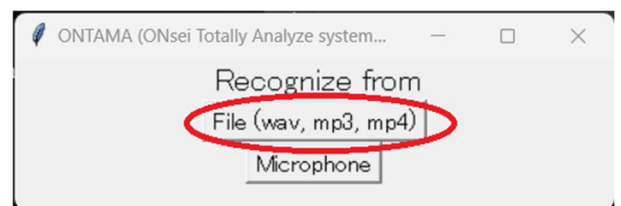
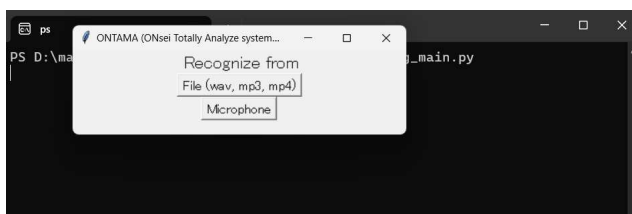
(2) 黒い画面が現れ、少し待っているとメニューが現れる。

(3) 音声ファイルや動画ファイル内の音声を認識させる場合は、「File(wav, mp3, mp4)」を選択。

(3)-1. ファイルを選択する画面があるので、音声ファイルか動画ファイルを選択して、「開く」。

(3)-2. 2 の黒い画面に、色々と実行経過が表示される。

(3)-3. 3-1 で選択したファイルと同じフォルダに、「***.docx」「***_plani.txt」というファイルが作成される(***は入力したファイルと同じ名前)。mp3 と mp4 の入力時は、wav 形式の音声データ「***.wav」が生成される(不要な場合は削除する)。



(4)パソコンのマイクから入力する音声を認識させる場合は、「Microphone」を選択。

(4)-1. 2 の黒い画面に色々并表示されるのでしばらく待つ。

(4)-2. 黒い画面に以下が表示されたら、マイクから音声を入力する。

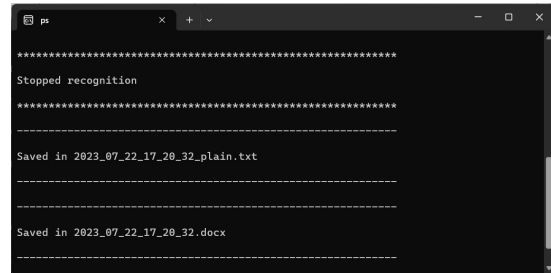
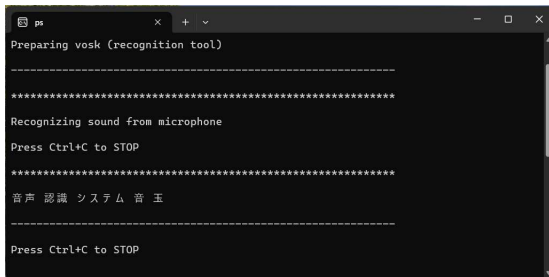
Recognizing sound from microphone

Press Ctrl+C to STOP

(4)-3. 認識結果が黒い画面に表示される。

(4)-4. 終了するときは、[Ctrl] を押しながら [c] を押す。

(4)-5. ontama.exe と同じフォルダに「yyyy_mm_dd_hh_mm_ss.docx」(年_月_日_時_分_秒)と「yyyy_mm_dd_hh_mm_ss_plain.txt」というファイルが作成される。



6. 出力ファイルの内容

Word ファイル(.docx)とテキストファイル(.txt)の文字データ自体には違いはありません。Word ファイルの場合は、認識の信頼度によってフォントが異なります。ただし、あくまでもプログラムが判定した信頼度であり、実際の音声との一致度ではありません。

高:通常

中:太字

低:太字・下線

と。千年3月。西安郊外の小さな村で土の中から奇妙なものが出てきます。それは壊れた人形のかげら。首や胴手足など精巧に作られた人形の破片が土の中に大量に埋まっていたのです。これこそ考古学史上に残る平和様の発見でした。千二百年もの時を超えて謎の軍団が蘇ります。心の金。

7. 名前の由来

「おんたま」に大した意味はありません。温泉玉子は美味しいのと、なんとなく可愛らしい名前にしたかっただけです。英語(ONTAMA: ONsei Totally Analyze system by MATsumura)は無理やりです。あえて漢字をあてるなら、「音魂」あるいは「温玉」でしょうか。