

1 概要

`python main.py` として実行すると図 1 のような画面が現れる。

wav ファイルを '参照' して、'開く' と画面が変化して、参照された wav ファイルに関する以下の情報が現れる。

- スペクトログラム
スペクトル成分の大きさと色は図 2 に表されるように対応している。
- 基本周波数
基本周波数を自己相関によって推定した結果をスペクトログラムに重ねた折れ線グラフにより示している。ただし、ゼロ交差数が基本周波数と比較して著しく大きい区間に対しては、無声音と判定して基本周波数のグラフは表示しないようにしている。
- 母音の識別結果
train/ 以下の訓練用 wav ファイルを読み込み、正規分布モデルのパラメータを与える。このモデルによって、参照された wav ファイルの母音識別を行い、結果をグラフ表示する。無声音に対しては 'NA' と表示する。

図 3 はサンプルとして用意した `data/aiueo.wav` を開いた結果である。

以下の環境で動作確認を行った。

- macOS Sierra version10.12
- python 3.6.7
- numpy 1.14.2
- matplotlib 3.0.1
- scipy 1.1.0

2 詳細

2.1 main.py

tkinter による GUI の制御を行う。'ファイル:' Label, wav ファイルへのパスが入力される Entry, '参照' ボタン, '開く' ボタン, スペクトログラムが表示される Frame と母音の識別結果が表示される Frame と順に構成している。

'参照' ボタンが押されたときには `refbutton_clicked` 関数が呼ばれ、wav ファイルの選択ウィンドウが開いて、選択されたファイルへのパスが Entry に入力される。

'開く' ボタンが押されると、コールバック関数 `openbutton_clicked` が呼ばれ、現在 Entry にあるパスにあたる音響信号の解析を行い、スペクトログラム + 基本周波数と母音の識別結果を描画する。

2.2 specgram.py

自己相関による基本周波数の推定を行う。

自己相関は $\tau = 0, 1, \dots$ に対して

$$AC_\tau = \sum_t x_t x_{t+\tau}$$

と求められる.

はじめのピークは $\tau = 0$ にあり、2 番目のピークが基本周波数にあたる. 2 番目のピークを近似的に求める方法として、まず $\tau = i - 1$ と $\tau = i$ を比較して、 $\tau = i$ の方が自己相関の値が大きくなったときにはじめのピークを抜けたと判断する. その後、自己相関が最大となった τ に対応する周波数を基本周波数とした.

- `predict_fundamentals`
main.py から呼ばれる関数. 信号を一定時間のフレームに分割して、各フレームの波形に対して `fundamental_by_frame` 関数と `zero_crossing` 関数を呼び出す. ゼロ交差数が基本周波数の 10 倍以上または 500Hz 以上ならば無声音と判定して基本周波数は表示しない.
- `fundamental_by_frame`
フレームに分割された波形に対して、自己相関を用いた基本周波数の推定を行う.
- `zero_crossing`
信号波形が 1 秒あたり振幅 0 の軸を何回交差するかを求める.

2.3 recog.py および recog_helper.py

recog.py に母音識別のモデルを表すクラス、recog_helper.py にそのクラスに必要な関数を実装した.
多次元正規分布の確率密度関数は

$$f(x; \mu, \sigma) = \frac{1}{(2\pi)^{D/2} \prod_d \sigma_d} \exp\left(-\sum_d \frac{(x_d - \mu_d)^2}{2\sigma_d^2}\right)$$

ある母音のケプストラムの集合を $X = x_1, x_2, \dots, x_N$ とするとケプストラム集合に対する対数尤度 L は以下.

$$L = \log p(X|\mu, \sigma) = \log \prod_n f(x_n; \mu, \sigma)$$

各母音 (a, i, u, e, o) のモデルに対して、その対数尤度を最大化するパラメータ μ と σ を学習する.

- `VowelClassifier::init` train/a.wav, i.wav, u.wav, e.wav, o.wav に対して `get_cepstrum` 関数を用いてケプストラム係数を求めて、モデルのパラメータを学習する.
- `VowelClassifier::predict` 信号を一定時間のフレームに分割して、各フレームの波形に対してケプストラム係数を求める. `calc_likelihood` 関数により各母音のモデルに対する対数尤度を求めそれが最大となる母音を予測結果とする.

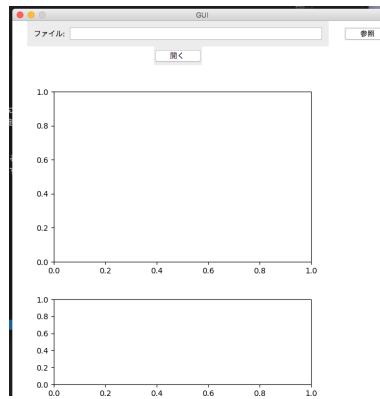


図1 はじめの画面



図2 colormap(右にいくほど値が大きいことを表している)

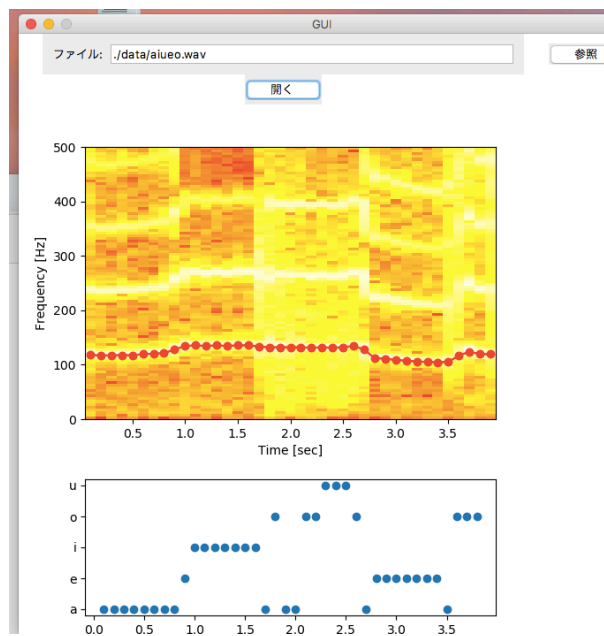


図3 wav ファイルを開いた後