

FFTケプストラムを用いた母音認識

68番 山崎 昌美

1. 目的

様々な音声認識方法を調べ、学習し、理解を深めることを目的とする。その学習した方法を用いて母音認識を実際に行う。

音声認識法について

音声認識法の中には、

- () 周波数領域での特徴を利用した方法
- () 時間領域での特徴を利用した方法
- () 上記の両領域の特徴を利用した方法

などがある。ここでは()の方法のひとつであるケプストラム法での音声認識法について調べた。

ケプストラム法の原理

音声信号を $x(n)$ とすると、そのフーリエスペクトル $X(j\omega)$ は

$$X(j\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-jn\omega}$$

$$X(j\omega) = Xr(\omega) + jXi(\omega)$$

$Xr(\omega)$:実数部 $Xi(\omega)$:虚数部

これを用いて、対数振幅スペクトルは

$$\ln|X(j\omega)| = 10\log_{10}|X(j\omega)|$$

対数スペクトルのフーリエ逆変換をケプストラムと呼ぶ。

ケプストラム $c(\tau)$ は (:ケフレンシー)

$$c(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln|X(j\omega)| e^{j\omega\tau} d\omega$$

これより、対数振幅スペクトルは次のように表現できる。

$$\ln|X(j\omega)| = \sum_{\tau=-\infty}^{\infty} c(\tau)e^{-j\omega\tau}$$

低ケフレンシー部分はホルマント成分を表す。ホルマント成分は発声時の口腔に由来する。ホルマントとは、口腔の形に対応する共振周波数である。このホルマント成分に音素の特徴が出る。

また、高ケフレンシー部分はピッチ成分を表す。これは音波形、つまり音の高さに由来している。

これより母音判定を行う際は低ケフレンシー成分のみを取り出せば良いことがわかる。この低ケフレンシーのケプストラムについて、母音の標準パターンとのユークリッド距離を求める。ユークリッド距離の最も小さい標準パターンを判定し、音素を決定する。

$$d = \sqrt{\sum_{m=1}^{15} \{c'(\tau) - c(\tau)\}^2}$$

窓関数について

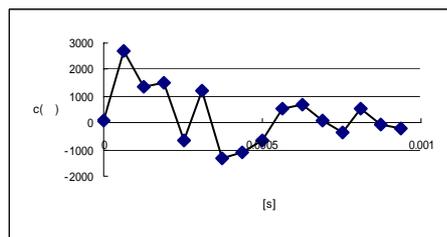
$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{M}\right)$$

音声信号を正しく高速フーリエ変換(FFT)する為に音声信号に窓関数を掛ける。ここでは次式のハミング窓を使用した。

3. 音声認識の実行

2節の方法で音声認識を実際に行う。

母音/a/のケプストラムの低ケフレンシー成分を取り出すと、このようになる。



このように標準パターンを求め、ユークリッド距離を求めると下記のような結果が得られる。

	入力 a	入力 i	入力 u	入力 e	入力 o
標準パターン a	2600	3249	3318	3876	4222
標準パターン i	3996	926	3282	2685	5263
標準パターン u	6304	4773	2163	5202	2385
標準パターン e	3975	2475	2825	1982	4655
標準パターン o	5469	4598	2199	5039	1137

女性の音声については、距離がやや大きくなるものの、同様に判定することが出来た。

4. まとめと今後の課題

FFTケプストラムを用いて、母音を認識するという目標は達成することが出来た。

課題としては、作成したプログラムが汎用的でないところがあげられる。音声信号の定常部分を取り出す手順をプログラムで処理できる方が望ましい。

定常部分では信号パワーが安定するという性質を利用し、1フレームのパワーが一定になったフレームについてケプストラムを求めれば良い。

5. 参考文献

臼井支朗, 信号解析, オーム社出版局, 1998

谷萩隆嗣, 音声と画像のデジタル信号処理, コロナ社, 1996

大類重範, デジタル信号処理, 日本理工出版会, 2001