

9 母音の音色の分析

9.1 母音の生成過程

母音が生成される過程を簡単にまとめると次のようになる。まず、声帯を周期的にふるわせて音を出す。次に、舌の位置や口の形で/a/ や /i/ など生成したい音韻にあわせて音色を決定する。最後に、口から空気中に放射される。それぞれの段階の周波数特性を $G(f), V(f), L(f)$ とすると、観測された音の周波数特性 $X(f)$ は次の式でかける。

$$X(f) = G(f) \cdot V(f) \cdot L(f)$$

演習 9-1 母音の特徴だけを明確に捉えるためには、上記の $V(f)$ にだけ着目するのがよい。なぜか、その理由を述べよ。

演習 9-2 $L(f)$ の部分の影響を軽減するためには、演習 8-6 で適用したフィルタを利用するのがよい。このようなフィルタをかける処理は、プリエンファシスと呼ばれることが多い。適当な母音のデータを用いてプリエンファシスが対数振幅スペクトルにあたえる影響を観察せよ。

9.2 ケプストラム

母音が /a/ であるか /i/ であるかは、スペクトルの細かい形状とは余り関係がない。むしろ、スペクトルのおおまかな形状 (概形と呼ぶことにする) で特徴づけられることが知られている。このスペクトルの概形が上記の $V(f)$ である。

プリエンファシスした音声を $V(f)$ と $G(f)$ に分解する方法としてここではケプストラムを用いた方法を紹介する。

周波数特性 $V(f)$ のフィルタと $G(f)$ のフィルタをかけあわせると $X(f)$ という特性になるとする。

$$X(f) = G(f) \cdot V(f)$$

対数をとる。

$$\log X(f) = \log G(f) + \log V(f)$$

掛け算だった関係が足し算となる。ここで $G(f)$ と $V(f)$ が例えば、 $G(f)$ の変動は細かいが $V(f)$ の変動は細かくない、などの特徴があれば、分解することができる。

上式の両辺に対して IDFT (逆離散フーリエ変換) を適用する。この変換を \mathcal{F}^{-1} とあらわすことにする。

$$\mathcal{F}^{-1}(\log X(f)) = \mathcal{F}^{-1}(\log G(f)) + \mathcal{F}^{-1}(\log V(f))$$

これをケプストラムとよぶ (FFT を用いて求めたので、FFT ケプストラムとよぶこともある)。ケプストラムは、周波数領域を IDFT したものであるため、時間の次元を持つ。この次元の単位は、周波数領域の単位を周波数 (frequency) とよぶのにちなんで、ケフレンシー (quefreny) と呼ぶ。

今説明のため、ケプストラム係数 C の長さを 16 とする。(つまり $C_0 \dots C_{15}$ の値があるとする) とすると、このケプストラム係数では、次の関係が成り立つ。

$$C_1 = C_{15}, C_2 = C_{14}, \dots, C_7 = C_9$$

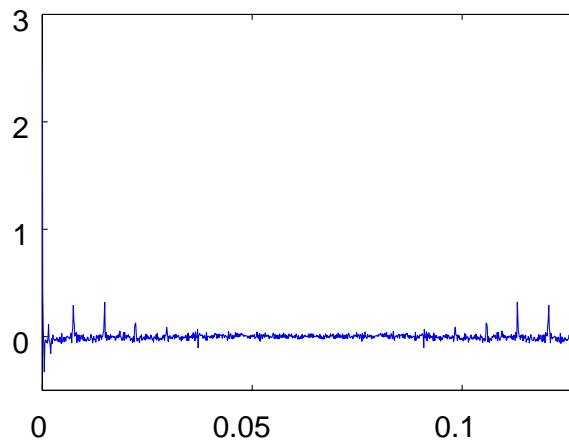
C_0 を 0 次の係数、 C_1 を 1 次の係数などと呼ぶ。

このケプストラム係数の 0 次から 12 次程度以外の要素を 0 にすると、スペクトルの概形 (スペクトル包絡とよぶ) を求めることができる。

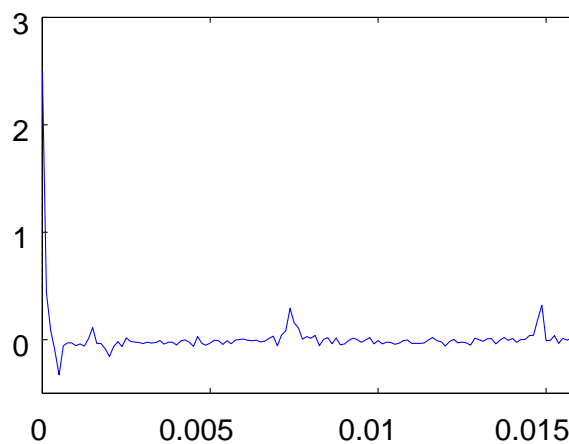
ケプストラムを求める手順は次の通りである。

1. 振幅スペクトルを求める。
2. 対数をとる。
3. 逆離散フーリエ変換をおこなう。(MATLAB では `ifft` で計算できる。)

演習 9-3 母音のケプストラムを求めプロットしてみよ。上手く求まると次のようにプロットされる。



このプロットの 0.01 あたりを拡大してみる。



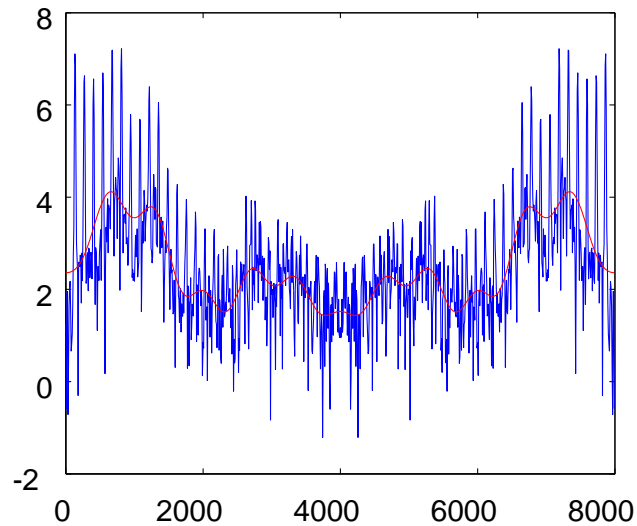
このピークは基本周波数をあらわす。(この図の場合、0.0074 あたりの方が基本周波数をあらわしている)

演習 9-4 ピークが表示されたら、その位置 (単位に注意せよ) から、基本周波数を求めよ。ただし、単位は Hz で求めること。

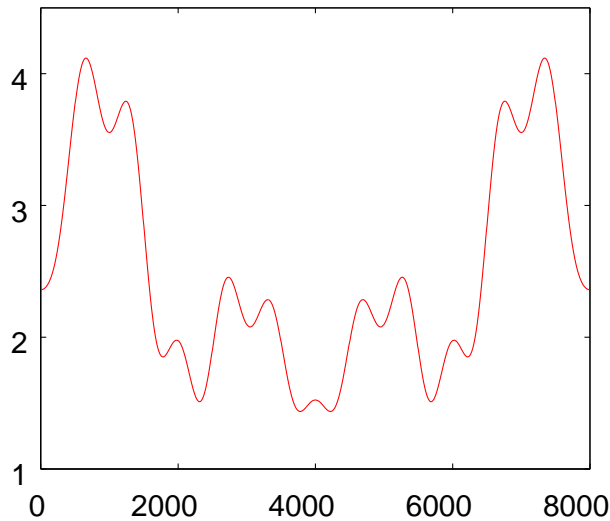
ケプストラムを使ってスペクトル包絡を求める手順は次の通りである。

1. 高次の要素の値を 0 にする。
2. 離散フーリエ変換をおこなう。
3. (もともと実数だったので) 実数部を取る。

演習 9-5 0 次から 12 次以外の係数を 0 にすることで、スペクトル包絡を求めよ。(注意: 1 次から 12 次の係数については対応する係数があり、そこは 0 にしてはいけない。例えば、ケプストラム c の長さが 256 だとすると、 $c(14)$ から $c(244)$ までを 0 とする。) 上手く求められると、スペクトルとあわせてプロットすると、次のようになる。



スペクトル包絡だけプロットすると次のようになる。



母音によって形は異なるが、どの母音でもいくつかピークがある形状になる。このピークのことをホルマントとよぶ。低い方から第1ホルマント、第2ホルマントというような呼び方をする。上記のプロットでは、第1ホルマントは約660Hz、第2ホルマントは約1250Hzである。

演習 9-6 最低でも2名の話者で5母音を収録し、それぞれ第1ホルマント、第2ホルマントを求めよ。人によってどのように異なるか?

演習 9-7 いろいろな声の高さで母音を収録し、それぞれ第1ホルマント、第2ホルマントを求めよ。声の高さによってどのように異なるか?