

基礎からの周波数分析 (31) – 「ケプストラム」

今回は、音声処理分野でよく利用される「ケプストラム解析」についてお話します。

ケプストラム解析では、時間波形からパワースペクトルを求め、そのスペクトルデータを波形とみなしてそれを FFT して、そのスペクトル波形の周期性などを求めることができます。

また、ケプストラムを使うことにより、フィルタなどの系を通した時間信号から、そのフィルタ系の伝達特性を推定することができます。

一般に、実測された時間信号 $x(t)$ が、2 種類の時間信号 $f(t)$ と $g(t)$ から成り立っていると
して、 $x(t)$ から、 $f(t)$ あるいは $g(t)$ の信号情報を取り出すことを考えます。

まず簡単な例として、 $x(t)$ が、2 つの信号の和である場合とします。すなわち、

$$x(t) = f(t) + g(t) \quad \dots\dots\dots (1)$$

この例では、線形の演算なので、式(1)の両辺をフーリエ変換すると

$$X(f) = F(f) + G(f) \quad \dots\dots\dots (2)$$

すなわち、周波数領域でも、その周波数成分 $X(f)$ は、2 つの信号の周波数成分、 $F(f)$ と $G(f)$ の和となります。

ここで、図 1 のように、それぞれの周波数成分の帯域が違っていれば、フィルタ処理により(図 2)、簡単に分離して取り出すことができます。

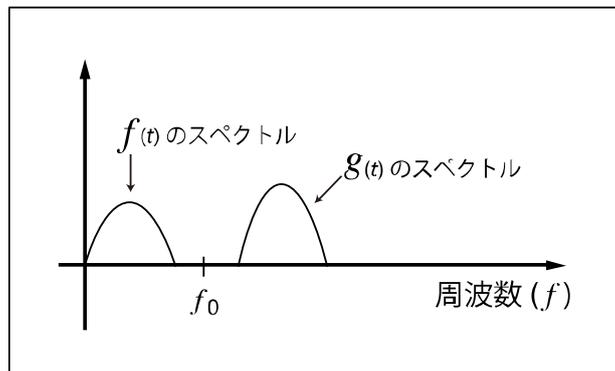


図 1 和信号 $x(t)$ のスペクトル

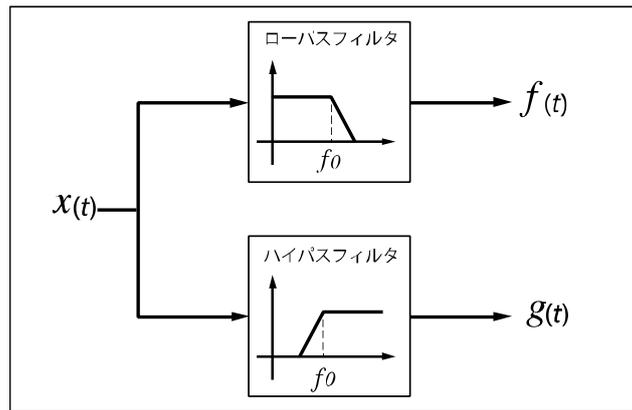


図2 フィルタリング処理による信号の分離

現実に得られる信号はこんな簡単な例は少なく、多くは、フィルタなどある線形系などの出力として得られる場合が一般的です。

いま、ある**インパルス応答** $g(t)$ の線形系の入力信号が $f(t)$ でその出力信号を $x(t)$ とすると、

$$x(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \quad \dots\dots\dots (3)$$

で、出力時間信号 $x(t)$ は、 $f(t)$ と $g(t)$ との**畳み込み積分**となります。式(3)の両辺をフーリエ変換すると

$$X(f) = F(f)G(f) \quad \dots\dots\dots (4)$$

すなわち、その周波数成分 $X(f)$ は、2つの信号の周波数成分、 $F(f)$ と $G(f)$ の積となります。式(4)から、元の入力信号 $f(t)$ を取り出すためには、 $G(f)$ の逆フィルタを求める必要があります、 $g(t)$ も未知の場合はなかなか困難です。そこで、式(4)において、積を和に変換できれば、帯域毎に分離するという上記の手法を使うことができます。具体的には、式(4)から、パワースペクトルを求めてその対数を取ることにより、2つの和成分として表現できます。

式(4)から、

$$P_{xx}(f) = P_{ff}(f)P_{gg}(f) \quad \dots\dots\dots (5)$$

ここで、 $P_{xx}(f)$, $P_{ff}(f)$, $P_{gg}(f)$ は、各々 $x(t)$, $f(t)$, $g(t)$ のパワースペクトルを求め、両辺の対数を取ると

$$\log P_{xx}(f) = \log P_{ff}(f) + \log P_{gg}(f) \dots\dots\dots (6)$$

となり、和の形になります。ここで、上記の式(1)と同様な考えで、式(6)の左辺を時間データとみなしてフーリエ変換をすると、帯域分離することが可能になります。

このように、元の時間信号をフーリエ変換してパワースペクトルを求め、その対数データを再度フーリエ変換する手法をケプストラム解析と呼びます。ケプストラム (cepstrum) の横軸は、時間になり、通常的时间波形の時間と違うので、周波数 (frequency) をもじって、**ケフレンシ** (quefrequency) と呼びます。また、帯域制限することを、周波数軸上でのフィルタ (filter) をもじって、**リフタ** (lifter) と呼びます。周波数スペクトルとケプストラムの用語の対比をまとめると、表1となります。

表1 スペクトルとケプストラムの用語の比較

名称	スペクトル (spectrum)	ケプストラム (cepstrum)
横軸	周波数 (frequency)	ケフレンシ (quefrequency)
帯域分離	フィルタ (filter)	リフタ (lifter)

具体的な例として、音声の分析を考えます。人間の声は、声帯のパルス的な振動を音源として、喉や口腔などの声道でフィルタリング処理された音として発生されます。

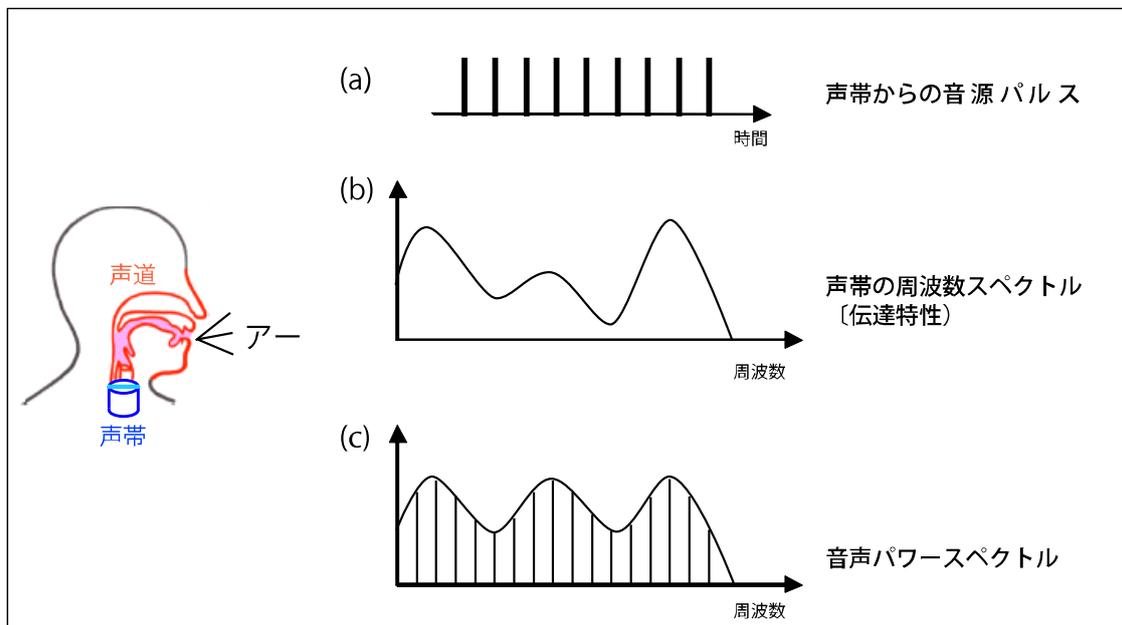


図3 音声の発生モデル

図3において、音源(a)が $f(t)$ 、声道のスペクトル(b)に対応するのが $g(t)$ 、発生した音声(c)が、観測される時間信号 $x(t)$ に対応します。

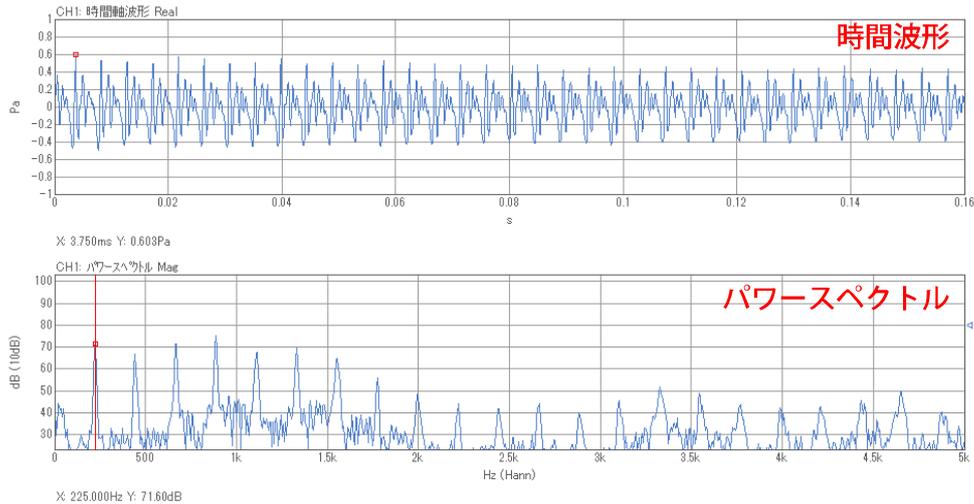


図4 女性の声/あ/の時間波形とパワースペクトル（縦軸は対数）

図4は、成人女性の声（あ）の時間時間信号とその分析結果です。これから、声（あ）の**基本周波数（ピッチ）**は、約 225 Hz であることがわかります。図4下図のパワースペクトル（対数化）を更にFFTをして、ケプストラムを求めると、図5となります。

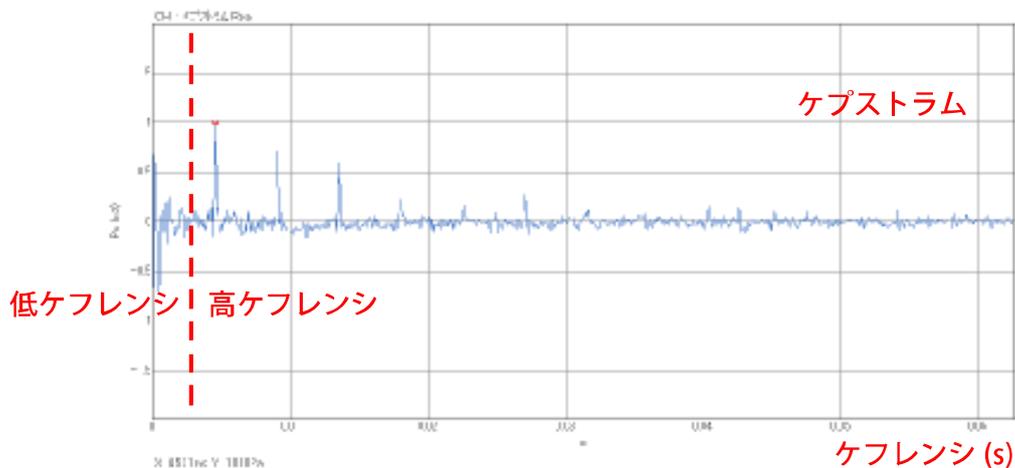


図5 女声/あ/のケプストラム

図5の横軸は、ケフレンシで単位は時間（s）です。

この図から、以下のことがわかります。図の赤点線（約 3.5 ms）より高いケフレンシ帯域は、声帯振動による音源に対応したケプストラム成分で、元のスペクトルの周期性（基本周波数とその高調波成分の繰り返し）を表しています。ピーク値が約 450 ms でスペクトルの間隔周期が約 450 ms となりますから、その逆数は約 220 Hz で、基本周波数 220 Hz とその高調波成分の情報であることを示しています。

また、図の赤点線（約 3.5 ms）より低いケフレンシ帯域は、元の周波数軸上ではゆっくりした変化の情報ですので、声道のスペクトルを表しています。

次に、赤点線（約 3.5ms）付近を境に高ケフレンシを 0 にしてリフタリング処理（ローパスリフタ）して、再度 FFT して周波数軸上に戻すと、図 6 の上図となります。これを**リフターードスペクトル**と呼び、元のパワースペクトルの**包絡線（エンベロープ）**に相当します。元のパワースペクトルとリフターードスペクトルを重ね書きしたものが、図 7 です。

このように、ケプストラム解析の用途として、スペクトルの周期情報の抽出、スペクトルの包絡線（エンベロープ）をもとめる、ことなどです。その他の応用としては、音の反射波などの分離などがあります。

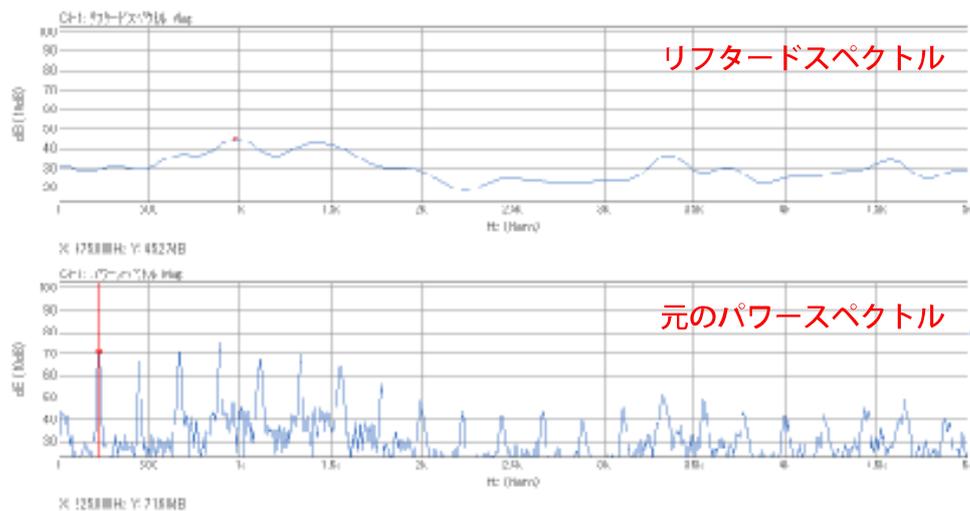


図 6 リフターードスペクトルと元のパワースペクトル

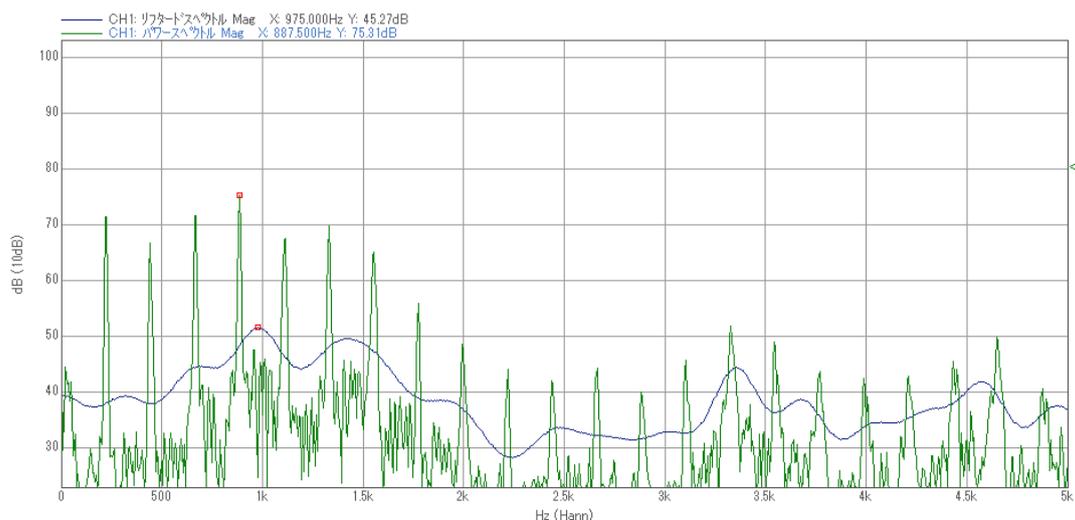


図 7 リフターードスペクトルと元のパワースペクトルの重ね書き表示

最後に、まとめです。

- (1) ケプストラムは、対数化されたパワースペクトルを時間信号とみなしてそれをフーリエ変換したものです。
- (2) ケプストラムは横軸が周波数であるスペクトルをフーリエ変換したものですから、その横軸は時間軸（単位：s）にもどり、ケフレンシと呼ばれます。
- (3) 音声の分析にケプストラムを利用すると、声帯の基本周波数と声道のスペクトル形状を検出することができます。
- (4) 低ケフレンシをリフタリングして再度フーリエ変換すると、元のスペクトルの包絡線が得られ、それをリフタードスペクトルと呼びます。
- (5) ケプストラムの用途としては、
 1. スペクトルの周期情報の抽出
 2. スペクトルの包絡線の検出
 3. 音の反射波などの分離
 などが、あります。

○ キーワード

ケプストラム、インパルス応答、畳み込み積分、ケフレンシ、リフタ、基本周波数、ピッチ、低ケフレンシ、高ケフレンシ、リフタードスペクトル、包絡線、エンベロープ

【キーワード】

ケプストラム、インパルス応答、畳み込み積分、ケフレンシ (quefrensy)、リフタ (lifter) 基本周波数 (ピッチ)、リフタードスペクトル、包絡線 (エンベロープ)

【参考】

森下巖、小畑秀文共著「信号処理」(社)計測自動制御学会 (1982年)
 城戸健一著「デジタルフーリエ解析(Ⅱ)ー上級編ー」コロナ社 (2007年)

以上
 (Hima)