



HOKKAIDO UNIVERSITY

Title	デジタルフィルタによる高速母音特徴抽出アルゴリズム
Author(s)	三上, 直樹; Mikami, Naoki; 大場, 良次 他
Citation	北海道大學工学部研究報告, 125, 53-60
Issue Date	1985-03-29
Doc URL	https://hdl.handle.net/2115/41918
Type	departmental bulletin paper
File Information	125_53-60.pdf



デジタルフィルタによる 高速母音特徴抽出アルゴリズム

三 上 直 樹 大 場 良 次
(昭和59年11月30日受理)

Fast Algorithms to Extract Feature Parameters of Vowels by Digital Filters

Naoki MIKAMI and Ryoji OHBA
(Received November 30, 1984)

Abstract

This paper compares performances of two fast algorithms which were proposed to extract the feature parameters corresponding to the first and the second formant frequencies of vowels. Their principles and recursive forms for practical procedures to calculate the feature parameters are briefly reviewed. A natural Japanese continuous vowel phrase is analyzed by the two algorithms and by a linear prediction. The results of the analyses are compared with each other and it was confirmed that the first and the second formant frequencies obtained by the two algorithms are in good agreement with those by the linear prediction. Furthermore, it was also confirmed that both formant frequencies obtained by them are almost identical.

1. 緒 言

音声認識の為の特徴パラメータの抽出法は現在多数発表されており、それらは音声の生成モデルを用いる方法と用いない方法との2つに大別できる。モデルを用いる方法の代表的なものには合成による分析法(A-b-S法)¹⁾、線形予測分析法²⁾などがある。一方、モデルを用いない方法には零交差波分析法³⁾、短時間スペクトル分析法⁴⁾、準同形分析法⁵⁾などがある。モデルを用いる方法はモデルを用いない方法より精度の高い特徴抽出が可能であるが、一般にデータ処理が複雑である。このため、現在市販されている音声認識装置の特徴抽出部は大部分が帯域フィルター群による短時間スペクトル分析に基づいている。

著者等は、実時間音声認識の為の母音の特徴パラメータ抽出法を幾つか提案している⁶⁻¹⁰⁾。文献〔6〕では実時間ピッチ同期分析法を提案した。それ以前にすでにオフラインのピッチ同期分析法¹¹⁾が提案されているが、これはピッチ周期を視察で求める方法であり、実時間音声認識には使えない。著者等の方法は、ピッチを自動的に抽出して行う方法で、原理的には実時間特徴抽出が可能であるが、スペクトル分析の為にFFTを使用するので、マイクロプロセッサのような簡単な情

報処理装置のみで実時間の特徴抽出を行うのは無理で、専用の FFT 装置付加する必要があった。

音声の音響的特徴の中で最も重要なものはホルマント周波数で、これは通常平滑化されたスペクトルのピークから求められる。著者等は、この平滑化スペクトルのピーク周波数を高速に求めるため、ウォルシュ・アダマール変換を利用した方法を提案している⁷⁾。しかし、ホルマント周波数とこの方法で求められるスペクトルのピークの周波数とが必ずしも一対一には対応しないため、この方法で特徴抽出して音声認識を行っても高い認識率は望めない¹²⁾。

著者等は、母音の特徴を精度よく、非常に高速に、しかも非常に簡単に得る方法を提案している^{8~10)}。本論文では、著者等が提案した2つの高速特徴抽出法の比較を行うため、実際の連続して発声された母音のみからなるフレーズをこれらの方法で分析し、更に現在音声分析の主流となっている線形予測法により同じフレーズを分析し、比較検討を行った結果について述べる。

2. デジタルフィルタによる特徴抽出

著者等は、デジタルフィルタを用いた2つの高速母音特徴抽出アルゴリズムを提案している。次に2つの方法の概略を述べる。

2.1 方法 I^{8,9)}

音声信号のサンプル値系列 $\{x_n\}$ に対して次のパラメータ $P(x)$ を定義する。

$$P(x) = \frac{\sum_{n=0}^{\infty} (x_{n+1} - x_n)^2}{4 \sum_{n=0}^{\infty} x_n^2} \quad (1)$$

声道を1個の共振器で近似し、そのインパルス応答 $\{v_n\}$ に対する $P(v)$ を求める。ここでインパルス応答は式(2)で与えられる。

$$v_n = e^{-\pi B T n} \sin 2\pi F T n \quad (2)$$

ここで、 F : ホルマン周波数, B : ホルマント帯域幅, T : 標本化周期である。 $P(v)$ は

$$P(v) = \frac{e^{-2\pi B T} - 2e^{-\pi B T} \cos 2\pi F T + 1}{2(e^{-2\pi B T} + 1)} \quad (3)$$

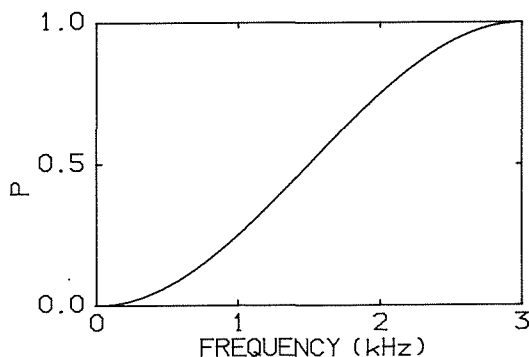


図1 パラメータ P の周波数特性。

となる。

図1に、 $B=0$ のときの F と P の関係を示す。ここで $T=1/6000$ 秒である。ホルマント周波数 F は P を用いて次のように求められる。

$$F = \frac{1}{2\pi T} \cos^{-1}\{\cosh \pi B T (1-2P(v))\} \quad (4)$$

ホルマント帯域幅 B をあらかじめ知ることはできないから、式(4)のみでは厳密な意味ではホルマント周波数を推定できない。しかし、 B は比較的小さいから、十分小さな T に対して $BT=0$ と見做せるとき式(4)は式(5)のように書き換えられる。このとき式(5)で求められる周波数を \hat{F}_p とすると、 \hat{F}_p をホルマント周波数の推定値とすることが可能である。

$$\hat{F}_p = \frac{1}{2\pi T} \cos^{-1}\{1-2P(v)\} \quad (5)$$

実際の声道には、複数のホルマントが存在する。母音の認識で重要な第1, 2ホルマントに対応するパラメータを求めたいときは、次の差分方程式で表される二つの簡単な前置フィルタで各ホルマント成分を強調した音声信号 $\{y_n^{(1)}\}$, $\{y_n^{(2)}\}$ から P を求めればよい。

$$y_n^{(1)} = x_{n+1} + x_n \quad (6)$$

$$y_n^{(2)} = x_{n+2} - 2x_{n+1} + x_n \quad (7)$$

又、実際の処理ではデータに有限長の時間窓を掛けて式(1)の計算を行う。 $\{x_n\}$ が周期信号の場合、時間窓を方形窓とし、その長さを周期の整数倍に選べば、パラメータ P は時間シフトに対して不変量となる。

2.2 方法 II¹⁰⁾

式(1)で定義される値は、パワの比である。従って、パワを近似するような値を用いても同じようなパラメータが求められることになる。最も簡単でしかも演算時間を短くするには、2乗演算の代わりに絶対値演算を使えばよい。そこで新しく、パラメータ Q を式(8)で定義する。

$$Q(x) = \frac{\sum_{n=0}^{\infty} |x_{n+1} - x_n|}{2 \sum_{n=0}^{\infty} |x_n|} \quad (8)$$

しかしこの場合 P とは異なり、方形窓の長さを信号の周期の整数倍に選んでも Q は時間シフトに対して不変量とはならない。

そこでインパルスの代わりに白色雑音を共振器に入力して狭帯域信号を作り、共振器の共振周波数を変えながらこの信号から Q の周波数特性 \tilde{Q} を求める。その結果を図2に示す。ここで、共振器の帯域幅は100Hzである。 \tilde{Q} は式(9)で近似できる。

$$\tilde{Q} = \sin \pi F T \quad (9)$$

この関係を使えば、ある狭帯域信号の中心周波数 \hat{F}_Q を

$$\hat{F}_Q = \frac{\sin^{-1} Q}{\pi T} \quad (10)$$

により推定できる。第1, 2ホルマントに対応するパラメータの求めかたは方法Iに準ずる。

方法IIでは2乗の代りに絶対値の計算を行うため、方法Iより次の2つの点で優れている。第1に処理速度が更に高速化されること、第2に必要な精度を得るために要求される計算機の語長が方法Iより短くてよいということである。

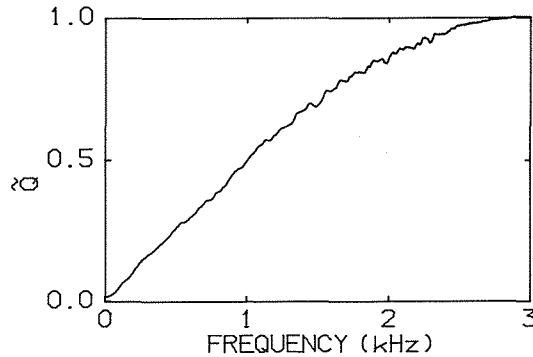


図2 狭帯域信号(帯域幅=100Hz)の中心周波数に対するパラメータ Q 。

3. パラメータの逐次計算法¹³⁾

パラメータを計算する際には、まず入力データからの分析フレームの切り出しが行われる。このときフレームの切り出しによる影響を抑えるために、方形窓ではなく、その両端で荷重が徐々に小さくなるハミング窓などを用いることが望ましい。又、分析結果の連続性を保つため、隣り合う分析フレーム間でその一部を重ね合わせることが望ましい。通常これらを実現するために、入力データを一度バッファに貯えて1フレームのデータがすべて揃ったところでパラメータを計算するという方法がとられるが、その場合演算回数が増え、更に余分なメモリが必要になってくる。

これを避ける方法として、著者等は1標本化データが入力されるたびに計算する逐次計算法を提案している¹³⁾。パラメータの計算で時間窓を掛けて和をとるという計算は、見方を変えれば重み付き移動平均の計算すなわち非巡回型の低域通過フィルタを作用させることである。従って、巡回型の低域通過フィルタによっても同様のことが可能である。

ここでは、計算をできるだけ高速に行うため、次の差分方程式で表される一次の低域通過フィルタを用いる。

$$v_n = cv_{n-1} + u_n \quad (11)$$

ここで、 $\{u_n\}$ 、 $\{v_n\}$ はそれぞれフィルタの入力、出力のサンプル値系列である。又、 c による掛け算を高速に行うため、 $c = 1 - 2^{-K}$ とし、シフトと減算で $c \cdot v_{n-1}$ の計算を行う。

次に分析フレームの等価的な長さを求めるために実効フレーム長 N_e を式(12)で定義する。

$$N_e = \frac{\left\{ \sum_{n=0}^{\infty} h_n \right\}^2}{\sum_{n=0}^{\infty} h_n^2} \quad (12)$$

ここで、 $\{h_n\}$ は式(11)のフィルタのインパルス応答であり、

$$h_n = c^n \quad (13)$$

である。従って、式(11)を用いるときの等価的なフレームの長さは

$$N_e = 2^{K+1} - 1$$

となる。

この逐次計算のブロック図を図3に示す。ここで関数 $G(\cdot)$ はそれぞれ P のときは2乗, Q のときは絶対値を計算する関数である。図中では1標準化データが入力されるたびに除算が行われてパラメータが求められるようになっているが、実際には適当な時間間隔毎にパラメータを求めればよいから、除算は適当なサンプル数毎に1度実行される。

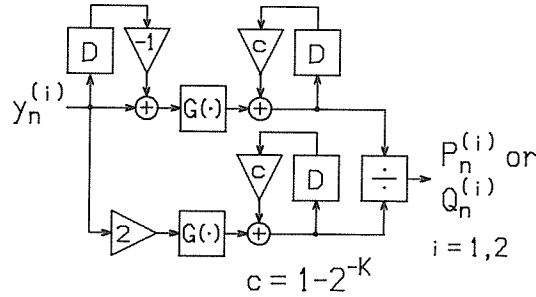


図3 パラメータ逐次計算法のブロック図。

4. 実音声の分析

2.で述べた著者等の2つの方法で求めたホルマント周波数の例を示す。連続して/*aiueo*/と発声された日本語母音を遮断周波数3 kHz, 減衰率-48dB/oct.の低域フィルタを通した後, 6 kHzで標準化, 12ビットで量子化してデータとした。パラメータの計算は $K=7$ ($N_e=255$)として行った。これは実時間に換算すると42.5msに相当する。パラメータの値は100サンプル毎に求め, 式(5)又は(10)を用いてホルマント周波数を求めた。図4に方法Iで, 図5に方法IIで求めた結果を示す。比較のため, 線形予測法で求めた結果も示す。線形予測法では1フレームの長さを200サンプル(33.3ms)とし, このフレームを100サンプル毎に移動してデータを切りだし, 1フレームのデータにハミング窓を掛け, 次数を10として分析を行った。×印は2.の方法で求めたもの, 縦棒は線形予測法で求めたもので, 棒の中心がホルマント周波数, その長さが帯域幅を示す。2つの方法で得られたホルマントは線形予測法で求められたものと比較して, ホルマントの遷移

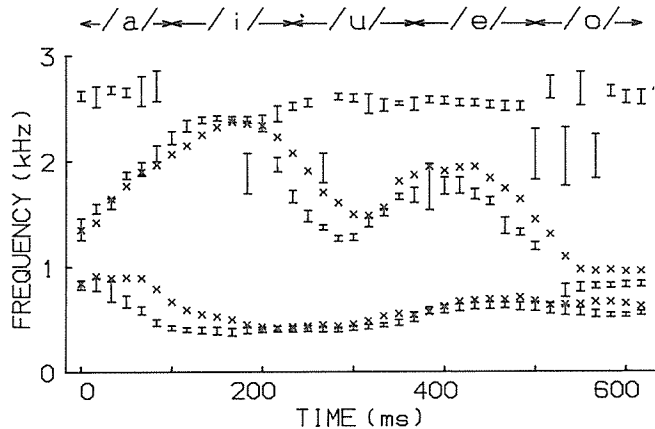


図4 方法I(×)と線形予測法(I)で得られたホルマント周波数。

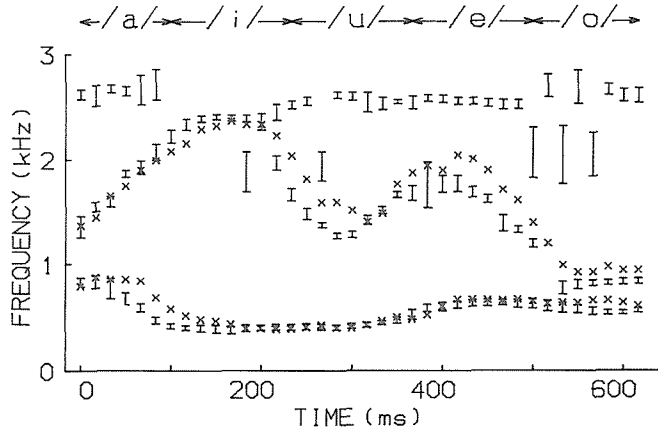


図5 方法II(x)と線形予測法(I)で得られたホルマント周波数。

の速い部分では幾分かの遅れが見られるが、その他では良く一致している。又、線形予測法では /e/から/o/に移行する部分で第2ホルマントの得られない部分があるが、本方法では2つの方法共に原理的にそのようなことはない。

2つの方法で得られたホルマント周波数と線形予測法で求められたものの一致の程度を見るためにホルマント周波数の対数をとってその散布図を求め、それらの間の回帰分析を行った。対数を用いたのは、我々の聴覚が2つの異なる基本周波数の音の隔たりすなわち音程を2つの周波数の差ではなく比により聞き分けるためである。図6, 7, 8はそれぞれ線形予測法と方法I ($\tilde{F}-\hat{F}_p$)、線形予測法と方法II ($\tilde{F}-\hat{F}_q$)、方法Iと方法II ($\hat{F}_p-\hat{F}_q$)で得られたホルマント周波数の間の散布図で、図中の直線は回帰直線を表している。表1に回帰分析の結果を示す。

これらの結果から、2つの方法で得られたホルマント周波数と線形予測法で得られたものとの相関が非常に高いことがわかる。又、方法Iと方法IIで得られたものはほぼ一致しているといえることができる。

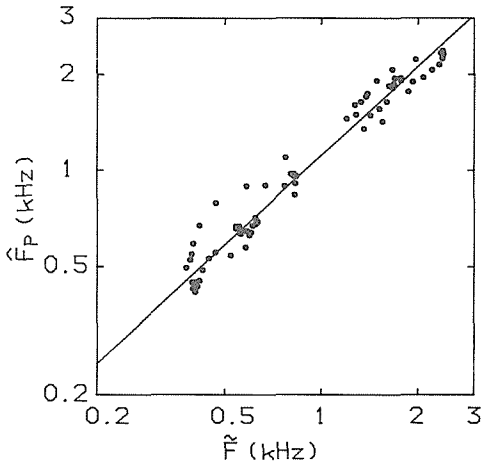


図6 方法I (\hat{F}_p) と線形予測法 (\tilde{F}) で求められたホルマント周波数の散布図。

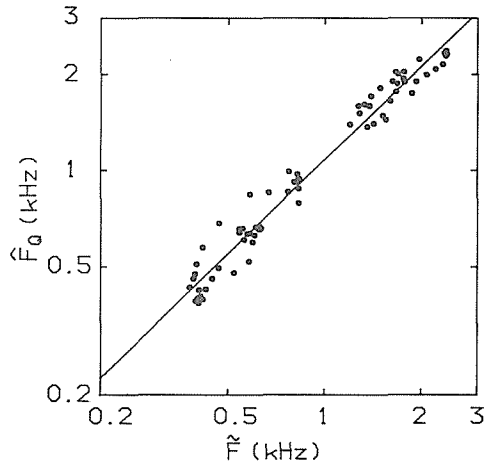


図7 方法II (\hat{F}_q) と線形予測法 (\tilde{F}) で求められたホルマント周波数の散布図。

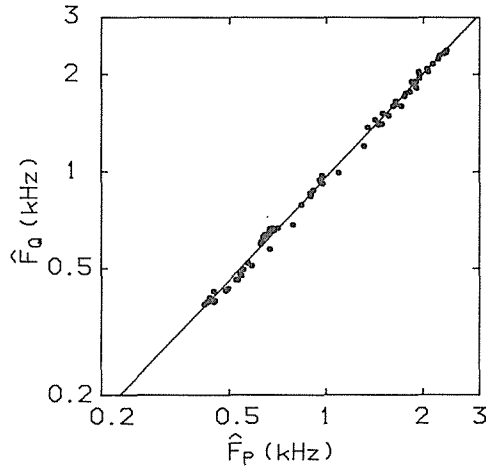


図8 方法I (\hat{F}_P) と方法II (\hat{F}_Q) で求められたホルマント周波数の散布図。

表1 方法I, II, 及び線形予測法で得られたホルマント周波数に対する回帰分析の結果。

	相関係数	回帰直線	回帰直線のまわりの縦軸方向の標準偏差
線形予測法 —方法I	0.981	$\log \hat{F}_P = 0.908 \log \hat{F} + 0.324$	0.049
線形予測法 —方法II	0.984	$\log \hat{F}_Q = 0.971 \log \hat{F} + 0.118$	0.045
方法I —方法II	0.998	$\log \hat{F}_Q = 1.063 \log \hat{F}_P - 0.209$	0.016

\hat{F}_P : 方法Iによるホルマント周波数.

\hat{F}_Q : 方法IIによるホルマント周波数.

\hat{F} : 線形予測法によるホルマント周波数.

5. 結 言

著者等が提案した2つの母音高速特徴抽出法の比較を行なった。その結果、2つの方法で得られたホルマント周波数は線形予測法で得られたものと非常に高い相関があることが示された。更に方法Iと方法IIで得られたホルマント周波数は実質的に同一と見做してよいことがわかった。従って、音声認識の前処理としてこの2つの方法を考えると、精度的にはどちらを用いても差し支えないと言える。しかし方法Iでは2乗の計算を行っているので、方法IIに比べて演算の際により長い語長の演算つまり広いダイナミックレンジが要求され、又、一般に2乗の計算の方が絶対値をとる計算よりも時間がかかるから、実用上方法IIの方が優れているといえることができる。

文 献

- 1) Bell, C. G., et al. : J. Acoust. Soc. Amer., **33** (1961), 12, pp.1725-1736.
- 2) Markel, J. D. & Gray, A. H., Jr. : Linear prediction of speech (1976), Springer-Verlag.
- 3) Peterson, E. : J. Acoust. Soc. Amer., **23**(1951), 6, pp.668-674.
- 4) Schafer, R. W. & Rabiner, L. R. : Proc. of IEEE, **63** (1975),4, pp.662-677.
- 5) Oppenheim, A. V. & Schafer, R. W. : IEEE Trans. Audio Electroacoust., **AU-16** (1968), 2, pp.221-226.
- 6) 三上, 大場 : 電子通信学会論文誌(A), **J62-A** (1979), 12, pp.849-855.
- 7) 三上, 大場 : *ibid.***J63-A** (1980), 10, pp.728-729.
- 8) 三上, 大場 : *ibid.***J64-A** (1981), 8, pp.692-693.
- 9) 三上, 大場 : *ibid.* **J65-A** (1982), 11, pp.1039-1045.
- 10) Mikami, N. & Ohba, R. : Proc. of IECON'84 (1984), pp.1148-1151.
- 11) Mathews, M. V. et al. : J. Acoust. Soc. Amer.,**33** (1961), 2, pp.179-186.
- 12) 三上, 大場 : 第19回 SICE 学術講演会予稿集 (1980), pp.335-336.
- 13) 三上, 大場 : 第22回 SICE 学術講演会予稿集 (1983), pp.695-696.