

音声の適応メルケプストラム分析

正員 徳田 恵一[†] 正員 小林 隆夫^{††}
 正員 深田 俊明^{††} 正員 今井 聖^{††}

Adaptive Mel-Cepstral Analysis of Speech

Keiichi TOKUDA[†], Takao KOBAYASHI^{††}, Toshiaki FUKADA^{††}
 and Satoshi IMAI^{††}, *Members*

あらし メルケプストラムをパラメータとする音声の適応分析法を提案する。メルケプストラムは、低い周波数域で細かい分解能を、高い周波数域で粗い分解能をもつという人間の聴覚特性を考慮したパラメータであり、音声スペクトルの表現に適している。本論文では、メルケプストラムをパラメータとするスペクトル推定法における最小化問題を解く際に、LMS アルゴリズムと同様の瞬時的なこう配の推定値を用いることにより、適応メルケプストラム分析アルゴリズムを導いている。提案する適応分析法は、MLSA (メル対数スペクトル近似) フィルタによって実現される指数形伝達関数をもつ IIR 形適応フィルタである。分析次数を M として $O(M)$ の演算量で標本点ごとのメルケプストラム係数を得ることができ、出力のみから係数が更新されること、フィルタの安定性が保証されることなどの特徴をもつ。合成信号の分析によって、本アルゴリズムの収束性を示すと共に、自然音声分析への適用例によって本方法の有効性を示す。

1. まえがき

音声信号は時間的に変化するシステムの出力と考えることができ、また、その分析は汎用 DSP などでの実時間処理が要求されることが多い。このような場合の信号処理には適応信号処理法の応用が期待できるが、適応ケプストラム分析法^{(1),(2)} は音声スペクトルの表現に適したケプストラムをパラメータとした適応分析法であり、次数を M として $O(M)$ の演算で標本点ごとのケプストラム係数が得られること、線形予測形の LMS アルゴリズムに比べ速い収束特性をもつ⁽³⁾ ことなどの特長から、音声の適応分析などに有用である。

ところで、メルケプストラム⁽⁴⁾ は、人間の音の高さに対する聴覚特性⁽⁵⁾ を表すメル周波数軸上で定義されたケプストラムであり、メルケプストラムによって表現されるスペクトルは、人間の聴覚特性と同様、低い周波数域で細かい分解能を、高い周波数域で粗い分解能をもつ。メルケプストラムは、通常のケプストラム

の半分程度の次数で音声スペクトルを表現することができること⁽⁶⁾、また、音声認識のパラメータとして用いた場合、高い認識率が得られること^{(7),(8)} などの利点をもつ。従って、適応ケプストラム分析法を拡張し、メルケプストラムをパラメータとする適応分析アルゴリズムを開発することができれば、音声の適応分析法として更に有効なものとなることが予想される。

本論文では、メルケプストラムをパラメータとするスペクトル推定法⁽⁹⁾ の最小化問題において、適応ケプストラム分析アルゴリズムと同様、瞬時的なこう配の推定値を用いることにより、適応メルケプストラム分析アルゴリズムを導いている。本適応分析系は、MLSA フィルタによって実現される指数形伝達関数をもつ IIR 形適応フィルタとなり、フィルタの安定性が保証される。

アルゴリズム導出の際には、パラメータをメルケプストラムから線形変換された係数に置き換えている。これは、スペクトル評価関数の最小化問題を逆フィルタ問題に置き換えるためであり、同時に、フィルタ実現の際のディレーフリーループを取り除くという意味をもつ。更に、この変換によって、未知系と推定系の特性が一致したときの Hesse 行列が対角行列となるため、適応ケプストラム分析法と同様の速い収束特性

[†] 東京工業大学電気電子工学科, 東京都
 Faculty of Engineering, Tokyo Institute of Technology,
 Tokyo, 152 Japan

^{††} 東京工業大学精密工学研究所, 横浜市
 Precision and Intelligence Laboratory, Tokyo Institute of
 Technology, Yokohama-shi, 227 Japan

をもつことが示される。

本アルゴリズムは、係数の更新のために逆フィルタ出力の値のみを必要とするという特徴をもち、係数更新および MLSA フィルタの演算量は、分析次数を M として $O(M)$ である。

合成信号の分析例により本アルゴリズムの収束性を示した上で、汎用 32bit 浮動小数点 DSP を用いて実現した適応メルケプストラム分析系によって自然音声进行分析し、本適応分析法の有効性を示すと共に収束性、特徴についての検討を行う。

2. 適応メルケプストラム分析^{(10),(11)}

2.1 メルケプストラムをパラメータとするスペクトル推定⁽⁹⁾

メルケプストラムをパラメータとするスペクトル推定法⁽⁹⁾は、メルケプストラムによって表されたスペクトルモデル $H(e^{j\omega})$ に、対数スペクトルの不偏推定法⁽¹²⁾におけるスペクトル評価

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\exp R(\omega) - R(\omega) - 1\} d\omega \quad (1)$$

を適用する手法である。但し、

$$R(\omega) = \log I_N(\omega) - \log |H(e^{j\omega})|^2 \quad (2)$$

であり、 $I_N(\omega)$ は長さ N の時間窓によって得られる弱定常過程 $x(n)$ の変形ピリオドグラムとする。また、 $H(e^{j\omega})$ は M 次までのメルケプストラム $\tilde{c}(m)$ によって、

$$H(z) = \exp \sum_{m=0}^M \tilde{c}(m) \tilde{z}^{-m} \quad (3)$$

とモデル化している。ここで、 \tilde{z}^{-1} は 1 次オールパス関数

$$\tilde{z}^{-1} = \Psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (4)$$

但し、

$$\Psi(e^{j\omega}) = e^{-j\beta(\omega)} \quad (5)$$

であり、位相特性 $\tilde{\omega} = \beta(\omega)$ は

$$\tilde{\omega} = \beta(\omega) = \tan^{-1} \frac{(1 - \alpha^2) \sin \omega}{(1 + \alpha^2) \cos \omega - 2\alpha} \quad (6)$$

で与えられる。単位遅延が $100 \mu\text{s}$ (標準化周波数 10 kHz) のとき、 α を 0.35 に選べば、 $\tilde{\omega} = \beta(\omega)$ は人間の音の高さに対する聴覚特性を表すメル尺度⁽⁵⁾をよく近似する⁽⁴⁾ことから、以下では $\alpha = 0.35$ とする。

ここで、式(1)の評価式を適応ケプストラム分析^{(1),(2)}と同様、逆フィルタ問題に置き換えるため、 $H(z)$ からゲイン K を除くことを考える。方法はいくつかあるが、ここでは、 $H(z)$ を

$$H(z) = \exp \sum_{m=0}^M b(m) \Phi_m(z) = K \cdot D(z) \quad (7)$$

但し、

$$\Phi_m(z) = \begin{cases} 1, & m=0 \\ \frac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}} \tilde{z}^{-(m-1)}, & m \geq 1 \end{cases} \quad (8)$$

また、

$$K = \exp b(0) \quad (9)$$

$$D(z) = \exp \sum_{m=1}^M b(m) \Phi_m(z) \quad (10)$$

と変形することにする。この変形は、MLSA フィルタにおいてディレーフリーループを取り除く操作 (文献(4)の方法 II) と同じである。

ここで、 $b(m)$ と $\tilde{c}(m)$ は、

$$b(m) = \begin{cases} \tilde{c}(m), & m=M \\ \tilde{c}(m) - ab(m+1), & 0 \leq m < M \end{cases} \quad (11)$$

あるいは、

$$\tilde{c}(m) = \begin{cases} b(m), & m=M \\ b(m) + ab(m+1), & 0 \leq m < M \end{cases} \quad (12)$$

で互いに交換することができる。 $\tilde{c}(m)$ と $b(m)$ は線形変換の関係にあるため、式(1)の $\tilde{c}(m)$ に関する最小化と $b(m)$ に関する最小化は等価である。このとき

$$\tilde{\mathbf{c}} = [\tilde{c}(0), \tilde{c}(1), \dots, \tilde{c}(M)]^T \quad (13)$$

に関する式(1)の最小化問題は、

$$\mathbf{b} = [b(1), b(2), \dots, b(M)]^T \quad (14)$$

に関する

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \quad (15)$$

の最小化、および K に関する式(1)の最小化と等価になることが示される⁽⁹⁾。ここで、時間窓長 N が十分大きいとすれば、 $I_N(\omega)$ を信号 $x(n)$ のパワースペクトル $|X(e^{j\omega})|^2$ と考えることができ、このとき問題は、逆フィルタ $1/D(z)$ に $x(n)$ を入力したときの出力を $e(n)$ として、

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|X(e^{j\omega})|^2}{|D(e^{j\omega})|^2} d\omega = E[e^2(n)] \quad (16)$$

の最小化、つまり $e(n)$ の最小 2 乗平均問題となる。

また、式(1)を最小化する K は、 \mathbf{b} に関して最小化された ε の値 ε_{\min} によって、

$$K = \sqrt{\varepsilon_{\min}} \quad (17)$$

と与えられる⁽⁹⁾。

2.2 最急降下法

ε は \mathbf{b} に関して凸であることが示される⁽⁹⁾ ので、 ε の最小点は一意に存在し、 $\nabla \varepsilon = \partial \varepsilon / \partial \mathbf{b} = 0$ を満たす \mathbf{b} によって与えられる。評価関数が凸関数である場合に

は、適当なステップ幅 μ を用いれば、最急降下法によって大域的な最小点を見つけることが可能である。つまり、 i 番目の近似値を $\mathbf{b}^{(i)}$ として、次回の近似値 $\mathbf{b}^{(i+1)}$ を

$$\mathbf{b}^{(i+1)} = \mathbf{b}^{(i)} - \mu \nabla \varepsilon |_{\mathbf{b}=\mathbf{b}^{(i)}} \quad (18)$$

とする操作を繰り返すことにより解を得る。

$\nabla \varepsilon$ は、式(16)を \mathbf{b} に関して微分し、

$$\bar{\mathbf{r}} = [\bar{r}(1), \bar{r}(2), \dots, \bar{r}(M)]^T \quad (19)$$

$$\bar{r}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|X(e^{j\omega})|^2}{|D(e^{j\omega})|^2} \Phi_m^*(e^{j\omega}) d\omega \quad (20)$$

とおけば、

$$\nabla \varepsilon = -2\bar{\mathbf{r}} \quad (21)$$

と書くことができる。

一般に、収束のためのステップ幅 μ の理論的な条件を導くことは困難であるが、ここでは評価関数が滑らかな凸関数であることから、最小点の周りの比較的広い範囲にわたって、Hesse 行列 $\mathbf{H} = \partial^2 \varepsilon / \partial \mathbf{b} \partial \mathbf{b}^T$ をもつ 2 次関数で近似できると考え、ステップ幅のおおまかな収束条件を導いておくことにする。但し、 \mathbf{H} の要素は、

$$h_{k,l} = 2 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|X(e^{j\omega})|^2}{|D(e^{j\omega})|^2} \Phi_k^*(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega + 2 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|X(e^{j\omega})|^2}{|D(e^{j\omega})|^2} \Phi_k^*(e^{j\omega}) \Phi_l^*(e^{j\omega}) d\omega \quad (22)$$

で与えられる。 ε が \mathbf{b} に関して凸であることから、 \mathbf{H} は \mathbf{b} の値によらず正定となり、 \mathbf{H} の固有値はすべて正となる。このとき、 \mathbf{H} の最大固有値を λ_{\max} とすれば、最急降下法が収束する μ の範囲は、

$$0 < \mu < 2/\lambda_{\max} \quad (23)$$

である⁽¹³⁾。 λ_{\max} の値を知ることは簡単ではないが、

$$\lambda_{\max} \leq \text{tr} \mathbf{H} \quad (24)$$

の関係式(23)に適用すれば、

$$0 < \mu < 2/\text{tr} \mathbf{H} \quad (25)$$

で最急降下法の収束条件を見積もることができる。

2.3 適応メルケプストラム分析アルゴリズム

式(21)のこう配 $\nabla \varepsilon$ を時間領域で表すと、

$$\nabla \varepsilon = -2E[e(n)\mathbf{e}_0^*(n)] \quad (26)$$

となる。但し、

$$\mathbf{e}_0^*(n) = [e_1(n), e_2(n), \dots, e_M(n)]^T \quad (27)$$

であり、 $e_m(n)$ は逆フィルタの出力 $e(n)$ を図 1 に示すフィルタ $\Phi_m(z)$ に通した出力である。ここで、LMS アルゴリズム⁽¹⁴⁾、適応ケプストラム分析法^{(1),(2)} などと同様に、こう配 $\nabla \varepsilon$ の時刻 n における瞬時的な推定値

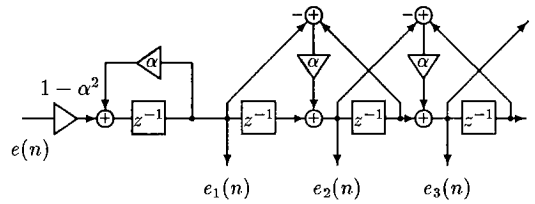


図 1 フィルタ $\Phi_m(z)$
Fig. 1 Filter $\Phi_m(z)$.

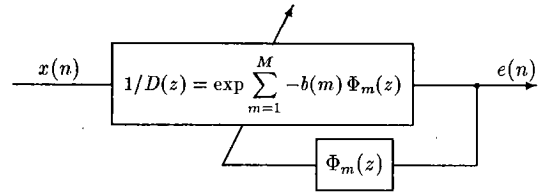


図 2 適応メルケプストラム分析のブロック図
Fig. 2 Block diagram of the adaptive mel-cepstral analysis.

$\bar{\nabla} \varepsilon^{(n)}$

$$\bar{\nabla} \varepsilon^{(n)} = -2e(n)\mathbf{e}_0^*(n) \quad (28)$$

を考える。 \mathbf{b} の値が一定のとき、 $\bar{\nabla} \varepsilon^{(n)}$ の期待値は $\nabla \varepsilon$ となる。式(28)の推定値を用いたとき、 \mathbf{b} の時刻 n から $n+1$ への係数更新アルゴリズムは、

$$\mathbf{b}^{(n+1)} = \mathbf{b}^{(n)} + 2\mu e(n)\mathbf{e}_0^*(n) \quad (29)$$

で与えられる。

以上で導かれた適応分析系の構成を図 2 に示す。 $1/D(z)$ は、非有理式の伝達特性であるが、MLSA フィルタ⁽⁴⁾によって高い精度で実現することができる。1 回の係数更新に必要な演算量は、分析次数を M とすれば、MLSA フィルタの演算を含めて $O(M)$ である。

2.4 ゲインの推定

音声信号などのように、パワーが時間的に変化する信号を分析する際には、 $b(0)$ あるいは K を推定する必要があるが、 K は逆フィルタ出力 $e(n)$ の最小平均 2 乗誤差 ε_{\min} 、つまり最小化された ε の値で得られるので、時刻 n の ε_{\min} を

$$\varepsilon^{(n)} = (1-\lambda) \sum_{i=-\infty}^n \lambda^{n-i} e^2(i) \quad (30)$$

で推定することにする。但し、 λ は $0 < \lambda < 1$ の定数である。このような $\varepsilon^{(n)}$ は、時刻 $n-1$ における ε の推定値 $\varepsilon^{(n-1)}$ から

$$\varepsilon^{(n)} = \lambda \varepsilon^{(n-1)} + (1-\lambda)e^2(n) \quad (31)$$

で得ることができる。従って、時刻 n における K は、

$$K^{(n)} = \sqrt{\varepsilon^{(n)}} \quad (32)$$

で、 $b(0)$ は、

$$b^{(n)}(0) = \frac{1}{2} \log \varepsilon^{(n)} \quad (33)$$

で得られる。

また、5.1 で後述するように、収束性を考慮し、時刻 n の μ の値を $\varepsilon^{(n)}$ 、つまり K^2 の推定値によって、

$$\mu^{(n)} = \frac{a}{M\varepsilon^{(n)}} \quad (34)$$

と正規化することにする。但し、 a は $0 < a < 1$ の定数とする。

3. 適応メルケプストラム分析系の実現

3.1 指数形伝達関数の実現と安定性

実際に適応メルケプストラム分析を行うには、図 2 に示すように、指数形の伝達関数をもつ逆フィルタ $1/D(z)$ が必要となる。ここでは、この逆フィルタを MLSA フィルタ⁽⁴⁾ により実現する。式(10)の指数形伝達関数の複素指数関数部分は、

$$\exp w \cong R_L(w) = \frac{1 + \sum_{l=1}^L A_{L,l} w^l}{1 + \sum_{l=1}^L A_{L,l} (-w)^l} \quad (35)$$

の形の L 次有理式により近似することができる。従って、次式の IIR フィルタ

$$F(z) = \sum_{m=1}^M -b(m) \Phi_m(z) \quad (36)$$

を基礎フィルタとすれば、

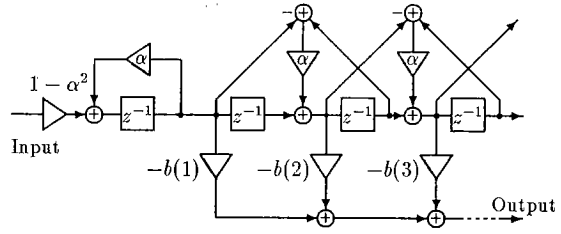
$$R_L(F(z)) \cong \exp F(z) = 1/D(z) \quad (37)$$

により指数形の逆フィルタ $1/D(z)$ が近似的に実現できる。ここで、 $F(z)$ の構成は図 3 (a) のとおりである。 $F(z)$ は入力から出力へのディレーフリーパスを含まないので、図 3 (b) のように構成された $R_L(F(z))$ はディレーフリーループをつくらず、そのまま実現することができる。

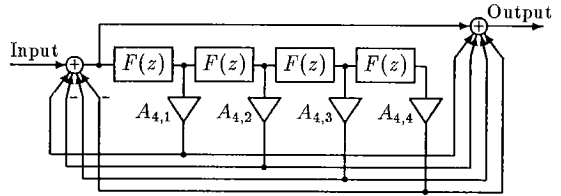
ところで、 $F(z)$ の極はすべて $z=a$ にあり、 $a=0.35 < 1$ としているので、 $F(z)$ は安定である。従って、 $b(m)$ が有限の値をとるならば、 $|F(z)|, |z| \geq 1$ の値も有限であり、

$$|F(z)| \leq r, |z| \geq 1 \quad (38)$$

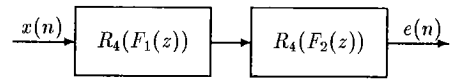
となるような定数 r を選ぶことができる。例えば、 $r=4.5$ として、有理近似式の $R_4(w)$ の係数 $A_{4,l}$ に文献(1)と同様のものを用いれば、式(38)のもとで、指数形伝達特性との近似誤差 $|\log(1/D(e^{j\omega})) - \log R_4(F(e^{j\omega}))|$ は、0.24 dB 以下となり⁽⁵⁾、また



(a) Basic filter $F(z)$



(b) $R_L(F(z)) \cong 1/D(z) \quad L=4$



(c) Two-stage cascade structure
 $R_4(F_1(z)) \cdot R_4(F_2(z)) \cong 1/D(z)$

図 3 指数形伝達関数 $1/D(z)$ の実現

Fig. 3 Implementation of the exponential transfer function $1/D(z)$.

$$|F(z)| \leq r_{\max} = 6.2, |z| \geq 1 \quad (39)$$

のもとで、 $R_4(F(z))$ は安定、かつ最小位相となる⁽⁶⁾。

3.2 基礎フィルタの分割

基礎フィルタの伝達関数 $F(z)$ が、

$$F(z) = F_1(z) + F_2(z) \quad (40)$$

と和の形に分解できる場合には、指数形伝達関数を

$$\begin{aligned} 1/D(z) &= \exp(F(z)) \\ &= \exp F_1(z) \cdot \exp F_2(z) \\ &\cong R_L(F_1(z)) \cdot R_L(F_2(z)) \end{aligned} \quad (41)$$

と縦続接続構成とすることができ、 $|F_1(z)|, |F_2(z)|$ の最大値がそれぞれ $|F(z)|$ の最大値より小さければ、近似の精度を上げることができる。分解の仕方はいくつかあるが、以下で述べる分析例では、

$$F_1(z) = -b(1) \Phi_1(z) \quad (42)$$

$$F_2(z) = \sum_{m=2}^M -b(m) \Phi_m(z) \quad (43)$$

と分解し、それぞれ $L=4$ とした図 3 (c) のような構成で逆フィルタを実現する。

実際の音声信号では、 $|F_1(z)|, |F_2(z)|$ の最大値は 4 程

度であり、 $r=4.5$ を超えないことを実験的に確かめることができる⁽¹⁵⁾。従って、図3(c)のフィルタは、 $1/D(z)$ に対し誤差0.24 dB以下の近似精度をもち、かつ安定である。

但し、以上は、 \mathbf{b} が一定値をとる場合の議論であり、 \mathbf{b} が変化している場合の議論は、 $1/D(z)$ の動的システムとしての取扱いが必要となる。

4. 分析例

4.1 合成信号の分析

合成信号の分析により、本方法の収束性を示す。合成信号は、5次までのメルケプストラムで与えられる図4の特性をもったMLSAフィルタを分散1の白色ガウス雑音、およびパルス列で励振した出力を用いた。標本化周波数は10 kHzであり、パルス列の周期は2.0 ms、分析次数は $M=5$ とし、初期値は $\mathbf{b}^{(0)}=0$ としている。ここでは、合成信号のゲインは一定であるため、式(34)の正規化は行っていない。

結果は図5に示すとおりである。図より100ないし200回程度の繰返しで収束している様子がわかり、未知系を変えた実験でも数百回程度で収束することが確かめられている。分析条件や、得られるパラメータ(適応ケプストラム分析ではケプストラム、提案法ではメルケプストラム)の差異等から、適応ケプストラム分析と提案法を単純に比較することはできないが、適応ケプストラム分析系においても数百回程度の繰返しで収束することが報告されており⁽¹⁾、提案法と適応ケプストラム分析法はほぼ同程度の収束特性をもつと考えられる。また、収束時の係数の分散も小さいことがわ

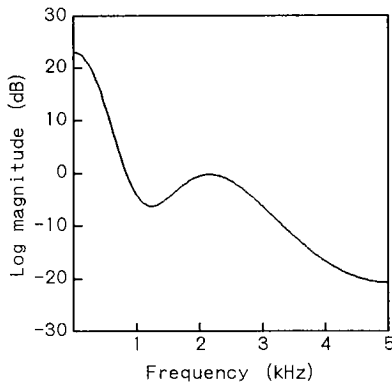


図4 未知系の対数振幅特性
Fig. 4 Log magnitude response of the unknown system.

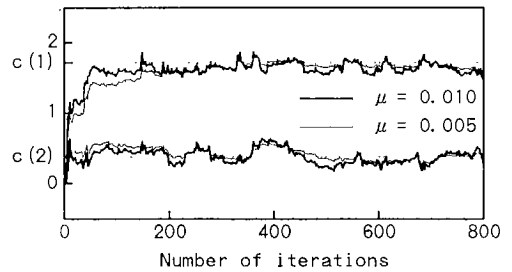
かるが、特に、未知系の励振信号がパルス列のときには、係数はほぼ真値に収束し、ほとんど変動していない。これらの理由については、5.1および5.2で検討する。

4.2 自然音声の実時間分析

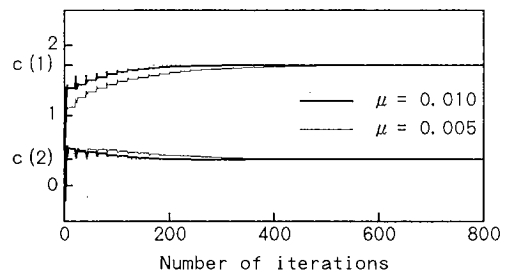
ここでは、DSPにより実現した適応分析系によって自然音声の分析を行った結果を図6に示す。

信号は男性話者が発声した「南部では東の風」で、標本化周波数は10 kHzである(図6(a))。分析次数 $M=12$ 、 $\alpha=0.10$ 、 $\lambda=0.97$ とし、推定されたメルケプストラム係数から得られる対数振幅スペクトルを100サンプルごとにプロットしたのが図6(c)である。比較のため、図6(b)にメルケプストラムをパラメータとするスペクトル推定法⁽¹⁶⁾(メルケプストラム分析法)によって得られた対数振幅スペクトルを示してある。メルケプストラム分析法は、25.6 msのブラックマン窓を用いて行った。図6(d)は、1次と2次のメルケプストラム係数 $\bar{c}(1)$ 、 $\bar{c}(2)$ の時間変化をメルケプストラム分析と適応分析法で比較したものである。

図6(d)から、適応メルケプストラム分析系は音声信号の変化に十分追従できる収束特性をもつこと、また、図6(c)から、得られるスペクトルは、 $M=12$ と



(a) White noise



(b) Pulse train

図5 メルケプストラム係数 $\bar{c}^{(n)}(1)$ 、 $\bar{c}^{(n)}(2)$ の収束の様子

Fig. 5 Convergence of mel-cepstral coefficients $\bar{c}^{(n)}(1)$ and $\bar{c}^{(n)}(2)$.

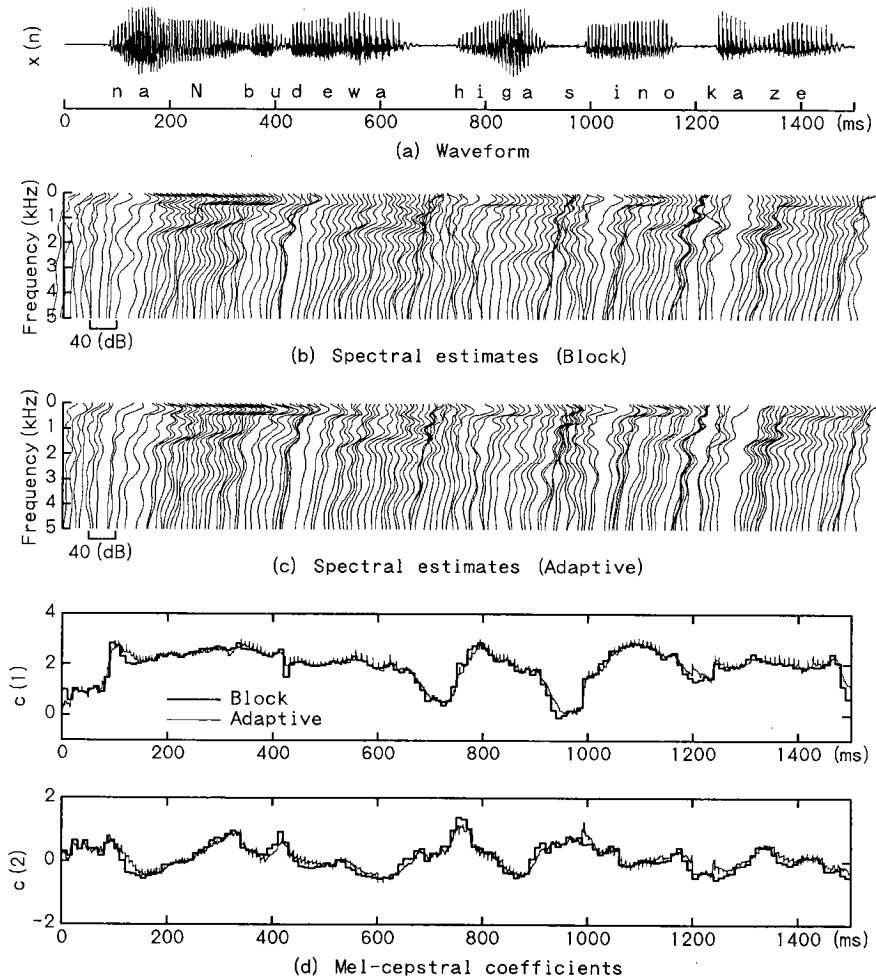


図 6 「南部では東の風」の分析結果 ($M=12$)
 Fig. 6 An example of the adaptive mel-cepstral analysis for natural speech ($M=12$).

いう低い次数にもかかわらず、1~2 kHz 以下の周波数域で高い分解能をもっていることがわかる。このようなメルケプストラムをパラメータとしたことによる利点については、文献(9)、(16)で詳しく議論されているため、ここでは詳しく述べない。

ここで用いた適応メルケプストラム分析系は、32 bit 浮動小数点 DSP (NEC μ PD77230)⁽¹⁷⁾ によって実現されている。1 サンプル当りの処理時間は、 $b(0)$ の推定、 \mathbf{b} の更新およびメルケプストラム係数 \mathbf{c} への変換に約 40 μ s、MLSA フィルタの演算に約 54 μ s、合計約 94 μ s であり、サンプリング周期 100 μ s (標準化周波数 10 kHz) に対しても実時間処理が可能である⁽¹⁸⁾。

5. 検 討

5.1 収束性に関する検討

本分析法における Hesse 行列は式(22)で与えられるが、 ϵ の最小点では $\mathbf{r} = \mathbf{0}$ であり、未知系の特性と推定された特性が一致したときには、

$$h_{\text{diag}} = 2 \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|X(e^{j\omega})|^2}{|D(e^{j\omega})|^2} \frac{(1-a^2)^2}{(1-ae^{-j\omega})(1-ae^{j\omega})} d\omega = 2\epsilon_{\text{min}} \quad (44)$$

を要素とする対角行列となり、Hesse 行列の固有値はすべて同じ値 h_{diag} となる。従って、 \mathbf{b} が最小点付近に追従し続ければ、本アルゴリズムは、適応ケプストラム分析アルゴリズムと同様、LMS アルゴリズムが理

想的に動作した場合の収束特性を示すと考えられる⁽¹⁾⁽²⁾。また、未知系と推定系の特性が一致した場合、式(34)の正規化は式(25)の収束条件にあたることがわかる。

このように Hesse 行列が対角行列となるのは、パラメータを $\bar{c}(m)$ から $b(m)$ へ変換した際、 $\Phi_m(z)$ が単位円上の直交関数系となる(付録参照)ことに起因しており、 $\bar{c}(m)$ をパラメータとした場合にはこのような性質を導くことはできない。

5.2 未知系励振源がパルス列の場合の検討

式(29)より、ある時刻 n における m 次の係数 $b^{(n)}(m)$ は、逆フィルタ出力 $e(n)$ と、 $e(n)$ をフィルタ $\Phi_m(z)$ に通した出力 $e_m(n)$ の積で更新される。この際、 $\Phi_m(z)$ のインパルス応答は時刻 $n=0$ でゼロであり、 $n \geq 1$ の減衰も速い。また、係数 b が収束し、未知系と推定系の特性が一致したなら、 $e(n)$ は未知系の励振信号であるパルス列となる。従って、パルスの周期が $\Phi_m(z)$ のインパルス応答の減衰に比較して長ければ、 $e(n)$ と $e_m(n)$ のいずれかはゼロと考えてよく、係数はほとんど変動しないことになる。

5.3 パラメータの変換に関する検討

適応メルケプストラム分析アルゴリズムを導くためには、

(1) 式(1)の評価式 E を式(16)の形の逆フィルタリング問題に置き換えなければならない。

(2) MLSA フィルタの構成の際、ディレーフリー

ループがあってはならない。などの条件が満たされる必要があるが、パラメータを $\bar{c}(m)$ から $b(m)$ に変換することにより、これらが満たされ、更に、

(1) 評価式 ε の最小点における Hesse 行列が対角行列となるため、速い収束性が得られる。

(2) 5.2 で述べたように未知系励振源がパルス列の場合、収束した係数がほとんど変動しない。

などの利点が生ずる。これらのことから、 $\bar{c}(m)$ から $b(m)$ への変換法が妥当なものであることがわかる。最終的に、得られた $b(m)$ から $\bar{c}(m)$ を得るには、式(12)の線形変換を行えばよい。

6. む す び

適応メルケプストラム分析アルゴリズムを提案し、その特徴、性質について述べた。提案する適応分析系は、指数形の伝達関数をもつ IIR 形の適応フィルタとなり、適応ケプストラム分析法と同様、フィルタの安

定性が保証されること、適応ケプストラム分析と同様の速い収束が得られることなどの特長をもつことを示した。少ない演算量 ($O(M)$) で、標本点ごとのメルケプストラム係数を得ることができ、推定されたスペクトルは人間の聴覚特性と同様、低い周波数域で高い分解能をもつため、音声の符号化、認識などでの有用性が期待される。また、新しい形の適応フィルタとして、適応等化器、エコーキャンセラなどにも応用が可能と考えられる。

謝辞 日ごろ、御助言、御討論頂く小野田真穂樹教授、國枝博昭助教授、金子峰雄講師に感謝致します。

文 献

- (1) 徳田恵一, 小林隆夫, 塩本祥司, 今井 聖: “適応ケプストラム分析—ケプストラムを係数とする適応フィルタ—”, 信学論(A), **J73-A**, 7, pp. 1207-1215(1990-07).
- (2) 徳田恵一, 小林隆夫, 塩本祥司, 今井 聖: “適応ケプストラム分析の検討”, 信学技報, **SP88-53**, **DSP88-23**(1988-09).
- (3) 深田俊明, 徳田恵一, 小林隆夫, 今井 聖: “適応一般化ケプストラム分析の検討”, 1990 信学春季全大, A-150, 1-150.
- (4) 今井 聖, 住田一男, 古市千枝子: “音声合成のためのメル対数スペクトル近似 MLSA フィルタ”, 信学論(A), **J66-A**, 2, pp. 122-129(1983-02).
- (5) G. Fant: “Speech sound and features”, MIT Press, Cambridge(1973).
- (6) 北村 正, 今井 聖, 古市千枝子, 小林隆夫: “メルケプストラムを用いる音声の情報圧縮”, 信学論(A), **J67-A**, 11, pp. 1092-1093(1984-11).
- (7) 小林隆夫, 近藤俊一, 今井 聖: “単語音声認識における一般化ケプストラム距離尺度の評価”, 信学技報, **SP87-18**(1987-06).
- (8) 新田恒夫, 上原堅助, 渡辺貞一: “不特定話者認識のための複合 LPC 距離尺度”, 信学論(D), **J70-D**, 10, pp. 1902-1910(1987-10).
- (9) 徳田恵一, 小林隆夫, 深田俊明, 齋藤博徳, 今井 聖: “メルケプストラムをパラメータとする音声のスペクトル推定”, 信学論(A), **J74-A** (1991).
- (10) 徳田恵一, 小林隆夫, 深田俊明, 今井 聖: “適応メルケプストラム分析の検討”, 1989 信学秋季全大, A-123, A-1-126.
- (11) 徳田恵一, 小林隆夫, 深田俊明, 今井 聖: “適応メルケプストラム分析による音声信号処理”, 信学技報, **DSP89, SP89**(1989-10).
- (12) 今井 聖, 古市千枝子: “対数スペクトルの不偏推定”, 信学論(A), **J70-A**, 3, pp. 471-480(1987-03).
- (13) Jacoby S. L. S., Kowalik J. S. and Pizzo J. T. 著, 関根智明訳: “非線形最適化問題の反復解法”, 培風館(1976).
- (14) Widrow B. and Stearns S. D.: “Adaptive Signal Processing”, Prentice-Hall, Englewood Cliffs, N. J. (1985).
- (15) 深田俊明, 小林隆夫, 徳田恵一, 今井 聖: “音声の適応メルケプストラム分析とその応用”, 第4回デジタル信号

処理シンポジウム講論集, B-5-1, pp. 279-284(1989).

- (16) 徳田恵一, 小林隆夫, 斎藤博徳, 今井 聖: “メルケプストラムをパラメータとする音声のスペクトル推定”, 信学技報, **DSP89-17**(1989-06).
- (17) 日本電気: “μPD77230 ユーザーズ・マニュアル”, (1988).
- (18) 小林隆夫, 徳田恵一, 深田俊明, 今井 聖: “音声の適応メルケプストラム分析”, 音響学会講論集, 1-3-15(1989-10).
- (19) 小林昭治, 久保田一, 高原幹夫: “正規直交関数系の一般的導出法とその適応デジタルフィルタへの応用”, 信学論(A), **J72-A**, 7, pp. 1045-1051(1989-07).
- (20) Laugunas-Hernández M. A., Figueiras-Vidal A. R., Mariño-Acebal J. B. and Vilanova A. C.: “A linear transform for spectral estimation”, IEEE Trans. Acoust., Speech & Signal Process., **ASSP-29**, pp. 989-993(Oct. 1981).

付 録

$\Phi_m(z)$ が直交関数系であることの証明

ここでは, 関数系

$$\Phi_m(z) = \begin{cases} \frac{\sqrt{1-\alpha^2} z}{1-\alpha z} \tilde{z}^{m-1}, & m < 0 \\ 1, & m = 0 \\ \frac{\sqrt{1-\alpha^2} z^{-1}}{1-\alpha z^{-1}} \tilde{z}^{-(m-1)}, & m > 0 \end{cases} \quad (\text{A}\cdot 1)$$

が単位円上の正規直交系であること, つまり

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_k'(e^{j\omega}) \Phi_l'^*(e^{j\omega}) d\omega = \delta_{kl}, \quad k, l = \dots, -2, -1, 0, 1, 2, \dots \quad (\text{A}\cdot 2)$$

であることを示す。

k と l が異符号, あるいは一方がゼロのとき, 式(A・2)の左辺がゼロとなること, また, $k=l=0$ のとき 1 となることは自明であり, $k, l > 0$ または $k, l < 0$ のときには,

$$\frac{\partial \tilde{\omega}}{\partial \omega} = \frac{1-\alpha^2}{(1-\alpha e^{-j\omega})(1-\alpha e^{j\omega})} \quad (\text{A}\cdot 3)$$

より, 式(A・2)の左辺の積分を $\tilde{\omega}$ の積分に書き換えれば,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j\tilde{\omega}(k-l)} d\tilde{\omega} = \delta_{kl}, \quad k, l = 1, 2, \dots \quad (\text{A}\cdot 4)$$

となるので, $\Phi_m(z)$ は単位円上の正規直交関数系である。従って, $\Phi_m(z)$ は単位円上の直交関数系である。

同様の関数系を用いる適応フィルタが提案されているが⁽¹⁹⁾, パラメータの線形変換によって式(3)から式(7)が導かれることを考えれば, これは1次オールパス関数によって周波数変換された FIR 形適応フィルタ⁽²⁰⁾ に等価であることが明らかとなる。

(平成2年9月18日受付, 3年2月13日再受付)



徳田 恵一

昭59 名工大・工・電子卒。平1 東工大大学院博士課程了。同年東工大電気電子工学科助手。工博。デジタル信号処理, 音声情報処理の研究に従事。



小林 隆夫

昭52 東工大・工・電気卒。昭57 同大大学院博士課程了。工博。同年東工大精密工学研究所助手。現在同大助教授。デジタルフィルタ, 音声の分析・合成, 音声認識の研究に従事。日本音響学会, IEEE 各会員。



深田 俊明

昭63 東工大・工・電気電子卒。平2 同大大学院修士課程了。在学中, 適応信号処理, 音声情報処理の研究に従事。現在, キヤノン(株)情報システム研究所勤務。



今井 聖

昭34 東工大・工・電気卒。昭39 同大大学院博士課程了。同年東工大精密工学研究所助手。昭43 同大助教授。昭54 同大教授, 工博。デジタル信号処理, 音声の合成および認識の研究に従事。昭45 年度精機学会論文賞受賞。著書「デジタル信号処理」など。計測自動制御学会, 日本音響学会, IEEE, ASA 各会員。