

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

正員 徳田 恵[†] 正員 小林 隆夫^{††} 正員 深田 俊明^{††}
 正員 斎藤 博徳^{†††} 正員 今井 聖^{††}

Spectral Estimation of Speech Based on Mel-Cepstral Representation

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

あらまし 本論文では、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ことが知られている。ここでは、メルケプストラムによるスペクトルモデルに対数スペクトルの不偏推定法におけるスペクトル評価を適用したスペクトル推定法について述べる。このときの最小化問題は、非線形最小化問題となるが、評価関数の凸性を示すことにより、一意解が存在することを明らかにする。更に、ニュートン・ラフソン法を利用した解法を示し、各繰返しにおいてFFTと再帰式を用いた効率的な計算法について述べる。本推定法により、低周波数域に帯域幅の狭い共振点、反共振点が集まっている信号に対して、低次数での分析が可能となる。最後に、合成信号、および自然音声の分析例により、本推定法の有効性を示す。

1. ま え が き

人間の音の高さに対する聴覚特性は、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ことが知られている。メルケプストラム⁽¹⁾は、このような聴覚特性を表す非直線周波数軸上で定義されたケプストラムであり、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⁽²⁾、音声認識のパラメータとして利用した場合、高い認識率が得られる⁽³⁾⁽⁴⁾などの利点をもつ。

従来、メルケプストラムの抽出法としては、線形予測法によって得られた対数振幅スペクトルを非直線周波数軸上で再サンプリングし、逆フーリエ変換することにより得る方法⁽⁵⁾、あるいはケプストラム法、線形予測法などによって得られたパラメータから再帰式⁽⁶⁾を用いて得る方法などが多く用いられている。しかし、

これらの手法によって得られるメルケプストラムは無限長の数列となり、これを有限次数で打ち切った場合には、もとのスペクトル分析によって得られた対数振幅スペクトルを非直線周波数軸上で線形平滑化することになる。従って、得られたメルケプストラムにより表されたスペクトルは、もとのスペクトル推定法によるスペクトルとは異なったものとなる。また、信号のピリオドグラムを非直線周波数軸上に周波数変換し、適当なスペクトル推定法を適用する⁽⁷⁾ことにより、メルケプストラムを得る手法もある。しかし、この手法によれば、非直線周波数軸上で適用した評価関数は、直線周波数軸では周波数重み付き評価となり、逆フィルタ出力の2乗誤差最小化、最ゆる性など、評価関数もつ元来の意味がなくなる。従って、メルケプストラムをパラメータとした音声のスペクトル分析を行うためには、以上で述べたような、打ち切り誤差や評価関数の変形が生じないスペクトル推定法を開発する必要があると考えられる。

以上の観点より、直線周波数軸上で対数スペクトルの不偏推定法を行った上で、打ち切り誤差の影響を避けるため、非直線周波数軸上でもう一度対数スペクトルの不偏推定法を行う方法が用いられており、音声認識などで良好な結果が得られている⁽⁸⁾。本論文では、より

† 東京工業大学工学部電気電子工学科, 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

††† 国際電信電話株式会社研究所, 上福岡町 Research and Development Laboratories, Kokusai Denshin Denwa Co., Ltd., Kamifukuoka-shi, 356 Japan

厳密に、メルケプストラムで表現されたスペクトルモデルに直接、対数スペクトルの不偏推定法⁽⁹⁾におけるスペクトル評価を適用する形のスペクトル推定法を提案し、その計算アルゴリズムを示す。

この際、評価式をメルケプストラム係数に関して直接最小化する方法⁽¹⁰⁾もあるが、ここではメルケプストラムを線形変換した係数に関して評価式を最小化している。このパラメータ変換を行うことによって、スペクトル評価関数の最小化問題が逆フィルタ問題に置き換えられること、変換されたパラメータを直接合成フィルタ⁽¹⁾の係数とできることなどの利点が生ずる。最小化問題は非線形問題であるが、評価関数の凸性から解の一意性を示し、ニュートン・ラフソン法に基づき、FFTと再帰式を利用した効率的な解法を与える。

更に、合成信号、および自然音声の分析を行い、本推定法によって得られるスペクトルは、人間の聴覚特性と同様、低周波数域で高い分解能をもつこと、従って低周波数域に共振点、反共振点をもつ信号に対しては、比較的低次数での分析が可能であることなどの有効性を示す。

2. メルケプストラムをパラメータとするスペクトル推定

2.1 メルケプストラムの定義

メルケプストラムの定義はいくつかあるが、ここでは、ある適当な位相特性 $\tilde{\omega} = \beta(\omega)$ をもつ因果的なオールパス関数を

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

但し、

$$\Psi(e^{j\omega}) = \exp(-j\tilde{\omega}) \quad (2)$$

として、

$$\tilde{c}(m) = \frac{1}{2\pi j} \oint_C \log \tilde{X}(\tilde{z}) \tilde{z}^{m-1} d\tilde{z} \quad (3)$$

$$\log \tilde{X}(\tilde{z}) = \sum_{m=-\infty}^{\infty} \tilde{c}(m) \tilde{z}^{-m} \quad (4)$$

でメルケプストラム $\tilde{c}(m)$ を定義する。ここで、 $X(z) = \tilde{X}(\tilde{z})$ は安定な実数係数 $x(n)$ の z 変換、 C は単位円を含む $\log X(z)$ の収束領域内で、原点を左回りに1周する閉路とする。式(3)、(4)は、単位円周上で、

$$\tilde{c}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \tilde{X}(e^{j\tilde{\omega}}) e^{j\tilde{\omega}m} d\tilde{\omega} \quad (5)$$

$$\log \tilde{X}(e^{j\tilde{\omega}}) = \sum_{m=-\infty}^{\infty} \tilde{c}(m) e^{-j\tilde{\omega}m} \quad (6)$$

但し、

$$\tilde{X}(e^{j\tilde{\omega}}) = \tilde{X}(e^{j\beta(\omega)}) = X(e^{j\omega}) \quad (7)$$

と書くことができ、 $\tilde{c}(m)$ は対数スペクトル $\log X(e^{j\omega})$ を非直線周波数軸 $\tilde{\omega} = \beta(\omega)$ に周波数変換したときのフーリエ係数となる。

ここで、オールパス関数を

$$\tilde{z}^{-1} = \Psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}, \quad |\alpha| < 1 \quad (8)$$

とすれば、 $\tilde{z}^{-1} = e^{-j\tilde{\omega}}$ の位相特性は、

$$\tilde{\omega} = \beta(\omega) = \tan^{-1} \frac{(1 - \alpha^2) \sin \omega}{(1 + \alpha^2) \cos \omega - 2\alpha} \quad (9)$$

で与えられる。式(9)は、単位遅延が $100\mu\text{s}$ (標準化周波数 10 kHz) の場合、 α を 0.35 と選べば、人間の音の高さに対する聴覚特性を表すメル尺度⁽¹¹⁾をよく近似し⁽¹⁾、また、同様に $\alpha = 0.47$ と選べば、バーク尺度をよく近似する⁽⁷⁾ことから、本論文では、式(8)で表される1次のオールパス関数を用いることにする。このような非直線周波数軸 $\tilde{\omega}$ 上でのケプストラム $\tilde{c}(m)$ を、メルなどの心理尺度を含む周波数目盛り上でのパラメータという意味で、メルケプストラムと呼んでいる。

メルケプストラムは、 $\alpha = 0$ とすれば、通常のケプストラムと等価となることから、ケプストラムの自然な拡張となっている。ケプストラムと同様、メルケプストラムは多くの興味深い性質をもっている⁽¹⁰⁾が、紙面の都合上、ここでは述べない。

2.2 スペクトルのモデルと評価関数

対数スペクトルの不偏推定法⁽⁹⁾は、ピリオドグラム の対数変換と非線形平滑化によって、相対値の意味で不偏なパワースペクトルの推定値を得る方法であり、 M 次までのケプストラム係数 $c(m)$ によって表現されたスペクトル

$$H(z) = \exp \sum_{m=0}^M c(m) z^{-m} \quad (10)$$

に、評価関数

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ \exp R(\omega) - R(\omega) - 1 \} d\omega \quad (11)$$

を適用する手法と言える。但し、

$$R(\omega) = \log I_N(\omega) - \log |H(e^{j\omega})|^2 \quad (12)$$

であり、 $I_N(\omega)$ は、弱定常過程 $x(n)$ の変形ピリオドグラムとする。

本論文では、信号のスペクトルを M 次までのメルケプストラム $\tilde{c}(m)$ によって、

$$H(z) = \exp \sum_{m=0}^M \tilde{c}(m) \tilde{z}^{-m} \quad (13)$$

とモデル化し、式(11)を適用することを考える。 $\alpha = 0$ とした場合には、対数スペクトルの不偏推定法と等価となることから、提案する方法は、対数スペクトルの

不偏推定法の拡張とみなすことができる。また、式(13)の伝達関数は、式(10)と同様、常に最小位相である。

式(13)に式(11)を適用して、 $\bar{c}(m)$ に関して直接、最小化問題を解くこともできる⁽¹⁰⁾が、ここでは、式(13)の右辺を変形してゲイン K をくくり出し、式(11)を適用することを考える。変形の仕方はいくつかあるが、ここでは、 $H(z)$ を文献(1)の方法IIと同様に、

$$H(z) = \exp \sum_{m=0}^M b(m) \Phi_m(z) = K \cdot D(z) \quad (14)$$

但し、

$$\Phi_m(z) = \begin{cases} 1, & m=0 \\ \frac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}} \bar{z}^{-(m-1)}, & m \geq 1 \end{cases} \quad (15)$$

および、

$$K = \exp b(0) \quad (16)$$

$$D(z) = \exp \sum_{m=1}^M b(m) \Phi_m(z) \quad (17)$$

と変形することにする。

ここで、 $\bar{c}(m)$ と $b(m)$ は、

$$\bar{c} = [\bar{c}(0), \bar{c}(1), \dots, \bar{c}(M)]^T \quad (18)$$

$$\mathbf{b}_0 = [b(0), b(1), \dots, b(M)]^T \quad (19)$$

とおけば、

$$\bar{c} = \mathbf{A} \mathbf{b}_0 \quad (20)$$

但し、

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha & 0 & \dots & 0 \\ 0 & 1 & \alpha & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \alpha \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \quad (21)$$

の関係をもつ。ここで、 \mathbf{b}_0 は、 $H(z)$ を直接実現する MLSA フィルタ⁽¹⁾の係数となる。 \mathbf{b}_0 と \bar{c} は線形変換の関係にあるため、式(11)の \bar{c} に関する最小化と、 \mathbf{b}_0 に関する最小化は等価である。これは、更に、式(16)で定義されるように K と $b(0)$ が単調な関係にあるため、 E の (K, \mathbf{b}) に関する最小化と等価となる。このとき、

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log |D(e^{j\omega})|^2 d\omega = 0 \quad (22)$$

の関係が成り立つので、式(11)は、

$$E(K, \mathbf{b}) = \epsilon(\mathbf{b})/K^2 + \log K^2 + \text{Const.}$$

但し、

$$\epsilon(\mathbf{b}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \quad (23)$$

$$\mathbf{b} = [b(1), b(2), \dots, b(M)]^T \quad (24)$$

$$\text{Const.} = -\frac{1}{2\pi} \int_{-\pi}^{\pi} \log I_N(\omega) d\omega - 1 \quad (25)$$

と書くことができる。この式より、 $E(K, \mathbf{b})$ を最小化する \mathbf{b} は、 K の値とは独立に $\epsilon(\mathbf{b})$ を最小化する \mathbf{b} として与えられる。ここで、 $I_N(\omega)$ を信号 $x(n)$ のパワースペクトルと考えれば、式(23)は、逆フィルタ出力の2乗平均最小化問題となることがわかる。

また、式(11)を最小化する K は、式(23)において、 $\partial E / \partial K^2 = 0$ とおけば、 \mathbf{b} に関して最小化された ϵ の値 ϵ_{\min} によって、

$$K = \sqrt{\epsilon_{\min}} \quad (26)$$

と与えられる。

3. 最小化問題の解法

3.1 ニュートン・ラフソン法

ϵ は、 \mathbf{b} に関して凸であることが示される(付録1.)ので、 ϵ の最小点は一意に存在し、 $\nabla \epsilon = \partial \epsilon / \partial \mathbf{b} = 0$ を満たす \mathbf{b} によって与えられる。このような \mathbf{b} を得るために、ニュートン・ラフソン法を用いることを考える。

つまり、 i 番目の近似値 $\mathbf{b}^{(i)}$ に対し、線形方程式

$$\mathbf{H} \Delta \mathbf{b} = -\nabla \epsilon|_{\mathbf{b}=\mathbf{b}^{(i)}} \quad (27)$$

を

$$\Delta \mathbf{b} = [\Delta b(1), \Delta b(2), \dots, \Delta b(M)]^T \quad (28)$$

について解き、次回の近似値

$$\mathbf{b}^{(i+1)} = \mathbf{b}^{(i)} + \Delta \mathbf{b} \quad (29)$$

を得る。ここで、 \mathbf{H} は、 ϵ の Hesse 行列 $\mathbf{H} = \partial^2 \epsilon / \partial \mathbf{b} \partial \mathbf{b}^T$ とする。

式(23)、(27)より次の方程式が得られる。

$$(\tilde{\mathbf{P}} + \tilde{\mathbf{Q}}) \Delta \mathbf{b} = \tilde{\mathbf{r}}|_{\mathbf{b}=\mathbf{b}^{(i)}} \quad (30)$$

但し、

$$\tilde{\mathbf{P}} = \begin{bmatrix} \tilde{p}(0) & \tilde{p}(1) & \dots & \tilde{p}(M-1) \\ \tilde{p}(1) & \tilde{p}(0) & \dots & \tilde{p}(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}(M-1) & \tilde{p}(M-2) & \dots & \tilde{p}(0) \end{bmatrix} \quad (31)$$

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \tilde{q}(2) & \tilde{q}(3) & \dots & \tilde{q}(M+1) \\ \tilde{q}(3) & \tilde{q}(4) & \dots & \tilde{q}(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}(M+1) & \tilde{q}(M+2) & \dots & \tilde{q}(2M) \end{bmatrix} \quad (32)$$

$$\tilde{\mathbf{r}} = [\tilde{r}(1), \tilde{r}(2), \dots, \tilde{r}(M)]^T \quad (33)$$

とし、 $\tilde{\mathbf{P}}$ 、 $\tilde{\mathbf{Q}}$ 、 $\tilde{\mathbf{r}}$ の各要素は次式で与えられる。

$$\tilde{p}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \cdot \frac{(1-\alpha^2)^2 e^{j\omega}}{(1-\alpha e^{j\omega})(e^{j\omega}-\alpha)} \Psi^{*m}(e^{j\omega}) d\omega \quad (34)$$

$$\tilde{q}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \cdot \frac{(1-\alpha^2)^2 e^{2j\omega}}{(1-\alpha e^{j\omega})^2} \Psi^{*m-2}(e^{j\omega}) d\omega \quad (35)$$

$$\tilde{r}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} \Phi_m^*(e^{j\omega}) d\omega \quad (36)$$

式(30)の線形方程式は、係数行列が対称 Toeplitz 行列と Hankel 行列の和になっているため、 $O(M^2)$ の高速算法⁽¹²⁾によって解くことができる。従って、演算量は、方程式の係数を求める計算が支配的となるが、これも次節で述べるように、FFT と再帰式によって効率的に計算することができる。

式(23)の ϵ は、 \mathbf{b} に関して凸、つまり ϵ の Hesse 行列 \mathbf{H} が正定であることから、式(30)および式(29)の繰返しにより 2 次収束解を得ることができる⁽¹³⁾。また、式(30)も、 \mathbf{H} が正定であることから、必ず一意解をもつ。

3.2 方程式の係数計算

i 番目の近似解 $\mathbf{b}^{(i)}$ に対する $D(e^{j\omega})$ が与えられれば、

$$\mathbf{r}_N = [r(0), r(1), \dots, r(N)]^T \quad (37)$$

$$r(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} e^{j\omega m} d\omega \quad (38)$$

を FFT によって計算することができるので、付録 2 に示す再帰式による変換 T を用いると、

$$\tilde{\mathbf{r}}_{2M} = [\tilde{r}(0), \tilde{r}(1), \dots, \tilde{r}(2M)]^T \quad (39)$$

は \mathbf{r}_N から

$$\tilde{\mathbf{r}}_{2M} = T(\mathbf{r}_N, N, 2M, \alpha) \quad (40)$$

で求めることができる。ここで N は理論上 $N = \infty$ であるが、実際には打ち切り誤差が問題とならない 100 程度の値を用いれば十分である。 $\tilde{\mathbf{r}}_{2M}$ が求まれば、

$$\tilde{\mathbf{p}} = [\tilde{p}(0), \tilde{p}(1), \dots, \tilde{p}(M-1)]^T \quad (41)$$

$$\tilde{\mathbf{q}} = [\tilde{q}(2), \tilde{q}(3), \dots, \tilde{q}(2M)]^T \quad (42)$$

は、

$$\tilde{p}(m) = \begin{cases} (1-\alpha^2)\tilde{r}(0) + 2\alpha\tilde{r}(1), & m=0 \\ \tilde{r}(m) + \alpha\tilde{r}(m+1), & 1 \leq m \leq M-1 \end{cases} \quad (43)$$

$$\tilde{q}(m) = \tilde{r}(m) + \alpha\tilde{r}(m-1), \quad 2 \leq m \leq 2M \quad (44)$$

によって計算できる。

また、 i 番目の近似解 $\mathbf{b}^{(i)}$ に対する $|D(e^{j\omega})|^2$ は、 $\mathbf{b}^{(i)}$ に対応するケプストラム

$$\mathbf{c}_N = [c(0), c(1), \dots, c(N)]^T \quad (45)$$

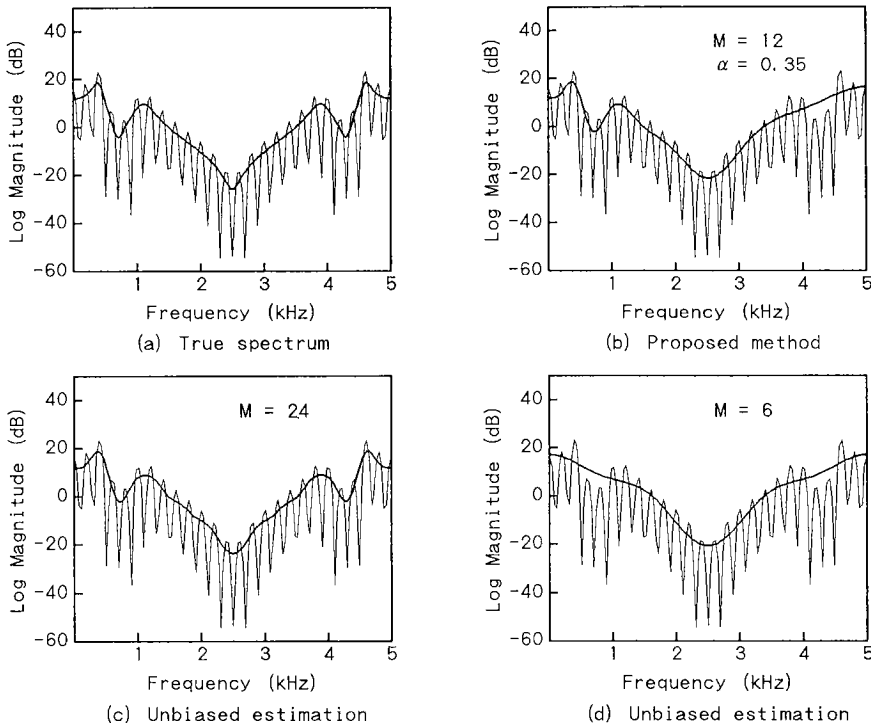


図 1 合成信号のスペクトル推定例
Fig.1 Spectral estimates of a synthetic signal.

を再帰式

$$c_N = T([0, \mathbf{b}^T]^T, M, N, -\alpha) \quad (46)$$

によって求めれば、 c_N から FFT を利用して計算することができる。但し、 N については式(40)の場合と同様、100 程度の値を用いれば十分である。

3.3 初期値

初期値は、できるだけ収束値に近いことが望ましいが、ここでは、従来のケプストラム法⁽⁴⁾により得られるケプストラムを $c^{(0)}$ として再帰式

$$[b(0), \mathbf{b}^{(0)T}]^T = \frac{1}{1-\alpha^2} T(c^{(0)}, N, M, \alpha) \quad (47)$$

によって得られる $\mathbf{b}^{(0)}$ を用いる。但し、 N については式(40)の場合と同様である。

4. 分析例

4.1 本推定法の特徴

低周波数域で細かい分解能をもつという、本推定法の特徴を示すために行った合成信号の分析結果を示す。標準化周波数は 10 kHz、時間窓は 25.6 ms のブラックマン窓とし、すべてのフーリエ変換は 256 点 FFT により計算した。 $\alpha=0.35$ とし、分析次数は $M=$

12 である。また、式(40)、(46)、(47)における N は、 $N=128$ とした。信号は図 1 (a) に示す周波数特性をもつデジタルフィルタを周期 5 ms のパルス列で励振した出力信号が定常状態に達した後の任意の 1 フレームを用いる。また、ニュートン・ラフソン法における繰返しは、 ϵ の変化率が 0.1% 以下となったときに終了する。

図 1 (b) に本推定法の分析結果から得られるスペクトルを示す。また、同じ信号を対数スペクトルの不偏推定法により分析次数 $M=24$ 、および $M=6$ で分析した結果をそれぞれ図 1 (c)、(d) に示す。本推定法により得られたスペクトルは、低周波数域では、対数スペクトルの不偏推定法における $M=24$ の分析結果と、また、高周波数域では、 $M=6$ の分析結果と同等なスペクトルが得られている。つまり、低周波数域に帯域幅の狭い共振点、反共振点が集まっている信号の分析を行う際に、低次数での分析が可能である。このような周波数特性として、図 2 (a) の特性を与えたとき、本推定法により $M=12$ で分析した結果を図 2 (b) に示す。また、比較のため、対数スペクトルの不偏推定法における $M=24$ の分析結果と、線形予測法におけ

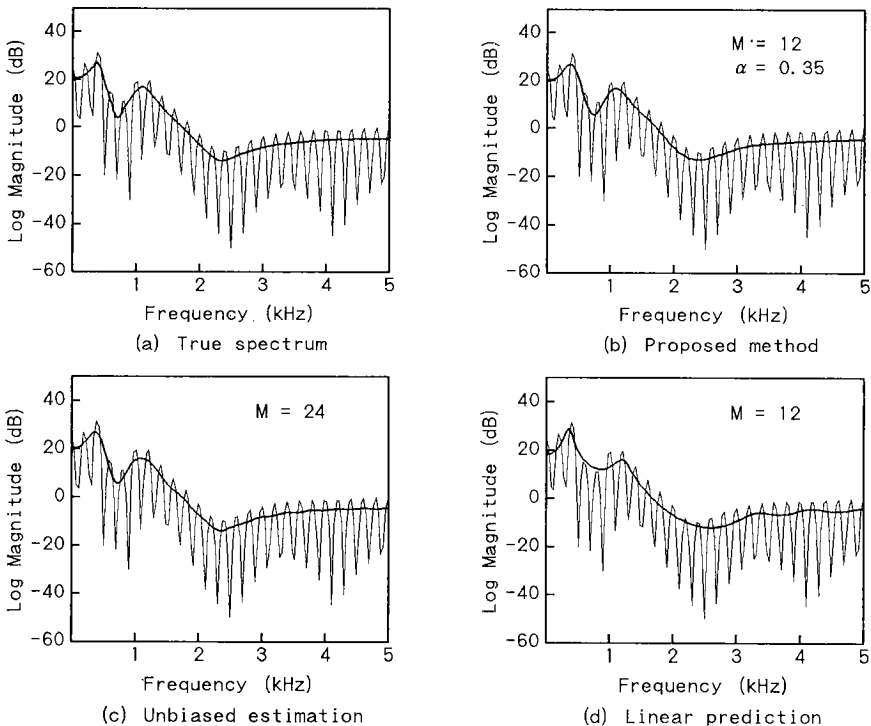


図 2 他のスペクトル推定法との比較

Fig. 2 Comparison of several methods with a synthetic signal.

る $M=12$ の分析結果をそれぞれ図 2 (c), (d) に示す。この図より、対数スペクトルの不偏推定法と同様のスペクトルが半分の分析回数で得られていること、線形予測法で抽出できない零点が抽出されていることがわかる。

4.2 他のメルケプストラム抽出法との比較

線形予測法などの他のスペクトル推定法から、再帰式によりメルケプストラムを求める際に生じる打ち切り誤差の影響を見るために、図 3 (a) に示す周波数特性をもつデジタルフィルタを周期 5 ms のパルス列によって励振した信号の分析を行った。分析条件は、4.1 と同様である。従来のメルケプストラムの抽出法として、対数スペクトルの不偏推定法により得られるケプストラム、線形予測法により得られる予測係数から、それぞれ再帰式⁽⁶⁾ によって、メルケプストラムを得る方法を選び、本推定法との計 3 種類で比較を行った。

図 3 (b) に本推定法の分析結果を示す。図 3 (c) の実線は、分析回数 $M=24$ の対数スペクトルの不偏推定法で得られたケプストラムを 12 次のメルケプストラムに変換したときのスペクトルである。図 3 (d) の実線は、分析回数 $M=12$ の線形予測法で得られた予

測係数を 12 次のメルケプストラムに変換したときのスペクトルである。また、もとの推定法によって得られたスペクトルをそれぞれ図 3 (c), (d) 中に破線で示す。

図 3 (c), (d) の結果から、メルケプストラムから得られるスペクトルは、破線で表されるもとのスペクトルと、形が大きく異なっていることがわかる。これは、再帰式によって求めたメルケプストラムは、無限長の数列となるが、これを有限回数で打ち切ったためである。しかし、本推定法によれば、メルケプストラムをパラメータとしてスペクトル評価を行っているため、このような打ち切り誤差の影響が出ず、式 (11) の評価関数 E をメルケプストラムに関して最小化するという意味で、最適なスペクトルが得られる。

4.3 自然音声分析への適用

男性話者による音声資料「南部では東の風」を、連続的に分析した結果を図 4 に示す。分析条件は、 $M=15$ とした他は、前節と同様であり、フレーム周期は 10 ms としている。このとき、ニュートン・ラフソン法における繰返し回数の平均値は、4.1 の終了条件において 3.5 回となった。

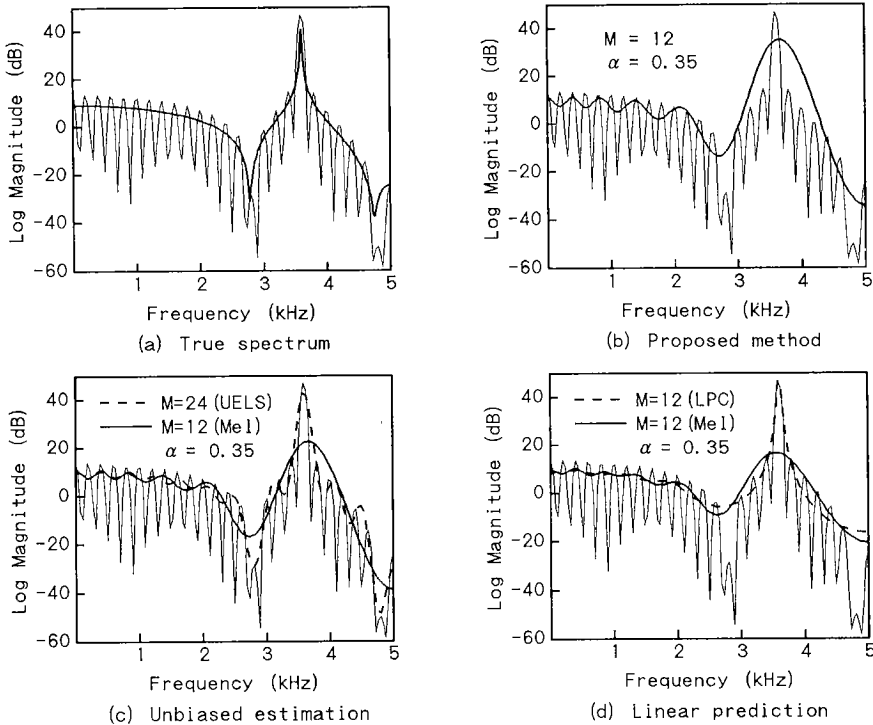


図 3 打ち切り誤差の影響
Fig. 3 Effect of truncation of mel-cepstral coefficients.

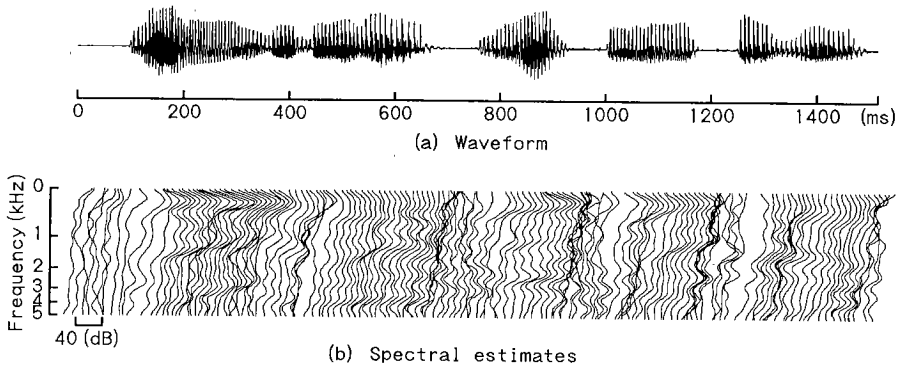


図 4 自然音声の分析例 ($M=15$)
Fig. 4 Spectral estimates of a natural speech ($M=15$).

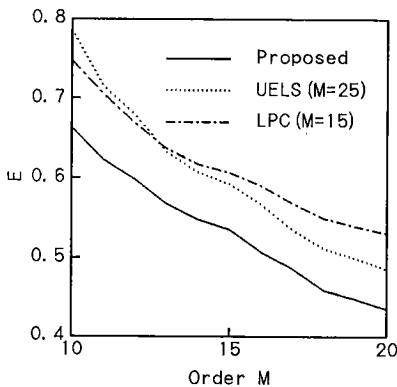


図 5 評価関数 E の平均値
Fig. 5 Average of the criterion E .

この音声 を 4.2 で述べた 3 通りの方法で分析し、得られたメルケプストラムから式(11)の評価関数 E の平均値を計算した結果を図5に示す。但し、対数スペクトルの不偏推定法の分析次数は $M=25$ 、線形予測法の分析次数は $M=15$ とし、これらから再帰式により 10~20 次までのメルケプストラムを求めている。本推定法は、メルケプストラムで表されたスペクトルを得る方法として、式(11)の評価関数 E を最小化するという意味において最適であり、 E の下限を与えている。それに対して、他の二つの手法によって得られるスペクトルは、メルケプストラムによって表されているが、他の推定法で得られたスペクトルから変換して得たものであるため、評価関数 E のもとで最適なスペクトルになっていないことが図5からも確かめられる。

5. む す び

メルケプストラムにより表現されたスペクトルモデ

ルに対数スペクトルの不偏推定法におけるスペクトル評価を適用したスペクトル推定法を提案し、その効率的な計算法を示した。

本推定法によれば、低周波数域に帯域幅の狭い共振点、反共振点が存在するような場合、対数スペクトルの不偏推定法と同様の推定結果を半分近い分析次数で得ることができる。また、このような低い次数でも線形予測法などで抽出することができない零点の抽出も可能である。更に、ケプストラム法、線形予測法などによって得られたスペクトルあるいはパラメータから変換してメルケプストラムを得る方法とは異なり、メルケプストラムをパラメータとするスペクトルモデルを用いているため、パラメータ変換による次数打切りの影響が生じない。実際に音声合成、音声認識に応用したときの有効性の確認については今後の課題である。

また、本推定法から、ケプストラム法の適応化手法⁽¹⁵⁾にならない、適応メルケプストラム分析法⁽¹⁶⁾を導くことができるが、これについては稿を改めて報告したい。

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) 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).
- (2) 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).
- (3) 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).
- (4) 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).
- (5) 板橋秀一, 横山晶一: “変形ケプストラム法による音声の特徴抽出”, 音響学会講義集, 2-2-4(1974-05).
- (6) 徳田恵一, 小林隆夫, 今井 聖: “メル一般化ケプストラムの再帰的計算法”, 信学論(A), **J71-A**, 1, pp.128-131(1988-01).
- (7) Strube H. W.: “Linear prediction on a warped frequency scale”, J. Acoust. Soc. Am., **68**, 4, pp.1071-1076(Oct. 1980).
- (8) 古市千枝子, 今井 聖: “多様な音韻環境における音素的単位のセグメンテーション”, 信学論(D-II), **J72-D-II**, 8, pp.1221-1227(1989-08).
- (9) 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).
- (10) 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**(1980-06).
- (11) Fant G.: “Speech sound and features”, MIT Press, Cambridge (1973).
- (12) Merchant G. A. and Parks T. W.: “Efficient solution of a Toeplitz-plus-Hankel coefficient matrix system of equations”, IEEE Trans. Acoust., Speech & Signal Process., **ASSP-30**, pp.40-44(Feb. 1982).
- (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) Oppenheim A. V. and Johnson D. H.: “Discrete representation of signals”, Proc. IEEE, **60**, pp.681-691(June 1972).
- (15) 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ー”, 信学論(A), **J73-A**, 7, pp.1207-1215(1990-07).
- (16) 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).

付 録

1. ε が凸であることの証明

$w = x + jy$ に関する凸関数

$$f(w) = 1/|\exp w|^2 = \exp(-2x) \quad (\text{A}\cdot\text{1})$$

を考えたとき, 任意の w_a, w_b に対して不等式

$$f(sw_a + (1-s)w_b) \leq sf(w_a) + (1-s)f(w_b), \quad 0 \leq s \leq 1 \quad (\text{A}\cdot\text{2})$$

が成り立つ。この不等式を式(23)に適用すれば, 任意の $\mathbf{b}_a, \mathbf{b}_b$ に対して,

$$\varepsilon(s\mathbf{b}_a + (1-s)\mathbf{b}_b) \leq s\varepsilon(\mathbf{b}_a) + (1-s)\varepsilon(\mathbf{b}_b), \quad 0 \leq s \leq 1 \quad (\text{A}\cdot\text{3})$$

が成り立つため, ε は \mathbf{b} に関して凸であることが示される⁽¹⁶⁾。

2. 再帰式の導出

ここでは, 一般的に

$$x(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega m} d\omega \quad (\text{A}\cdot\text{4})$$

から

$$\bar{x}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) \Phi_m^*(e^{j\omega}) d\omega \quad (\text{A}\cdot\text{5})$$

を再帰的に求めることを考える。 $\Phi_m(z)$ の逆 z 変換を $\phi_m(n)$ とすれば, 式(A・5)は Parseval の関係から

$$\bar{x}(m) = \sum_{n=-\infty}^{\infty} x(n) \phi_m(n) \quad (\text{A}\cdot\text{6})$$

と書くことができる。ここで $x(n), \phi_m(n)$ のいずれかが因果的であれば, 式(A・6)は,

$$\bar{x}(m) = \sum_{n=0}^{\infty} x(-i-n) \phi_m(n) \quad (\text{A}\cdot\text{7})$$

となるので, $\bar{x}(m)$ はフィルタ $\phi_m(z)$ に数列 $x(-i)$ を入力して得られる出力の $i=0$ における値となる。これより, N 次までの $x(n)$ に対応する $\bar{x}(m)$ を M 次まで求める再帰式は,

$$\bar{x}^{(i)}(m) = \begin{cases} x(-i), & m=0 \\ (1-\alpha^2)\bar{x}^{(i-1)}(0) + \alpha\bar{x}^{(i-1)}(1), & m=1 \\ \bar{x}^{(i-1)}(m-1) + \alpha\bar{x}^{(i-1)}(m) - x^{(i)}(m-1), & m=2, 3, \dots, M \\ & i=-N, \dots, -1, 0 \end{cases}$$

$$\bar{x}(m) = \bar{x}^{(0)}(m), \quad m=0, 1, \dots, M \quad (\text{A}\cdot\text{8})$$

となる。但し, 初期値 $\bar{x}^{(-N-1)}(m), 0 \leq m \leq M$ は 0 とする。

ここでは, 簡単のため, $\mathbf{x} = [x(0), x(1), \dots, x(N)]^T$ から $\bar{\mathbf{x}} = [\bar{x}(0), \bar{x}(1), \dots, \bar{x}(M)]^T$ を求める式(A・8)の再帰式を記号 T を用いて,

$$\bar{\mathbf{x}} = T(\mathbf{x}, N, M, \alpha) \quad (\text{A}\cdot\text{9})$$

と表すことにする。また, 逆に M 次までの $\bar{x}(m)$ があったとき, これに対応する $x(n)$ を N 次まで求める再帰式は,

$$\mathbf{x} = T(\bar{\mathbf{x}}, M, N, -\alpha) \quad (\text{A}\cdot\text{10})$$

で与えられることが示される。

(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，キャノン(株)情報システム研究所勤務。



斎藤 博徳

昭62 東工大・工・電子物理卒，平1 同大大学院修士課程了，在学中，音声認識音声情報処理の研究に従事，現在，KDD 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。