

メル一般化ケプストラム分析による音声のスペクトル推定

正員 徳田 恵一[†] 正員 小林 隆夫^{††}
 正員 千葉 健司^{†††} 正員 今井 聖^{††}

Spectral Estimation of Speech by Mel-Generalized Cepstral Analysis

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

あらまし メル一般化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。ケプストラム法と線形予測法とを統一的に扱う手法に、一般化ケプストラムをパラメータとするスペクトル推定法があるが、この推定法によれば、スペクトルモデルを全極形から指数形まで連続的に変化させることができる。本論文では、一般化ケプストラムに代えメル一般化ケプストラムを用いたスペクトル推定法について述べ、このときの最小化問題の解の一意性、解法、解の収束性を示す。提案する推定法は、一般化ケプストラムを用いる場合と同様、スペクトルのモデルを連続的に変化させることができること、得られる伝達関数の安定性が保証されることなどの特長をもつ。更に、メル周波数軸上で定義された一般化ケプストラムであるメル一般化ケプストラムをパラメータとして用いているため、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ったスペクトルを得ることができる。最後に、合成信号および自然音声の分析例により、本推定法の特徴および有効性を示している。

キーワード メル一般化ケプストラム、メル一般化ケプストラム分析、音声のスペクトル推定、スペクトルモデル

1. ま え が き

音声のスペクトル推定法として、全極形のスペクトルモデルを用いる線形予測法⁽¹⁾、得られたスペクトルが指数形となるケプストラム法⁽²⁾が多く用いられている。しかし、全極形のモデルでは零を表現することができず、逆に指数形ではスペクトルの鋭いピークを実現しにくくなる。極零モデルを用いる分析法も多く提案されているが、解の一意性、収束性、伝達関数の安定性などが問題となることが多い。一般化ケプストラム⁽³⁾をパラメータとするスペクトル推定法⁽⁴⁾は、このような問題を解決するための手法で、ケプストラム法と線形予測法とを統一的に扱い、べきパラメータ γ の値によって、スペクトルモデルを全極形から指数形

まで連続的に変化させるものである。

ところで、人間の聴覚特性は低い周波数域で細かい分解能を、高い周波数で粗い分解能をもつ⁽⁵⁾。従って、音声スペクトルを表現する際にも、同様の性質を取り入れることにより、少ないパラメータで効率的に音声スペクトルを表現することが可能となる。このような例として、メルケプストラム⁽⁶⁾によるスペクトル表現があり、音声の認識あるいは合成で有用であることが知られている。また、同様に聴覚特性を音声分析に導入した手法がいくつか提案されている^{(7),(8)}。

本論文では、一般化ケプストラムにこのような考え方を導入した、メル一般化ケプストラム⁽⁹⁾をパラメータとするスペクトル推定法について述べる。ここでは、メル一般化ケプストラムによるスペクトルモデルに対して、対数スペクトルの不偏推定法⁽¹⁰⁾におけるスペクトル評価を適用し、このときの最小化問題の解の一意性、解法、解の収束性を示している。提案する推定法は、一般化ケプストラムによる場合と同様、スペクトルのモデルを連続的に変化させることができること、得られる伝達関数の安定性が保証されることなどの特長をもちながら更に、周波数圧縮パラメータ α の値に

[†] 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

^{†††} 富士ゼロックスシステム技術研究所, 海老名市
 Fuji Xerox System Technology Research Laboratory, Ebina-
 shi, 243-03 Japan

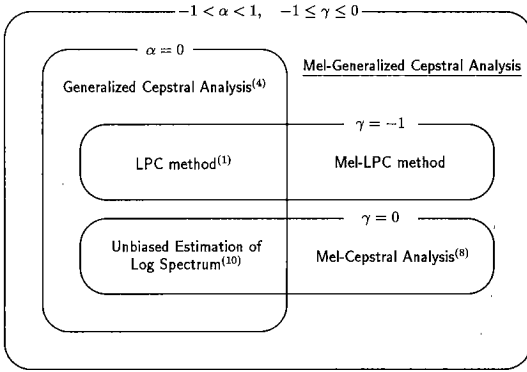


図1 他の手法との関係
Fig.1 Several methods included by the proposed method.

よって、低周波数域あるいは高周波数域に高い分解能をもたせることができる。例えば、 α を適当な正の値とした場合には、人間の音の高さに関する聴覚特性と同様、低周波数域で高い周波数分解能をもつスペクトル推定法となり、一般化ケプストラム分析では推定できないような低周波数域の鋭い共振点、反共振点も、分析次数を上げることなく推定が可能となる。また、本推定法は、既に提案されている線形予測法、対数スペクトルの不偏推定法、メルケプストラム分析法などを特別な場合として含む統一的手法となっており、これらの自然な拡張となっている。これらの関係をまとめると図1のように表すことができる (4,参照)。

2. メル一般化ケプストラムをパラメータとするスペクトル推定

2.1 スペクトルのモデルと評価関数

本論文では、 M 次までのメル一般化ケプストラム⁽⁹⁾ $\tilde{c}_\gamma(m)$ 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⁽¹⁰⁾ における評価関数

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

を適用することを考える。ここで、

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

であり、 $I_N(\omega)$ は長さ N の時間窓によって得られた弱定常過程 $x(n)$ の変形ピリオドグラムとする。なお、式(1)は、信号が正規定常過程の最尤推定法における評価関数と形式的に同じになっている。また、 $H(e^{j\omega})$ は、一般化対数関数 $s_\gamma(w)$ およびその逆関数 $s_\gamma^{-1}(w)$ を

$$s_\gamma(w) = \begin{cases} (w^\gamma - 1)/\gamma, & 0 < |\gamma| \leq 1 \\ \log w, & \gamma = 0 \end{cases} \quad (3)$$

$$s_\gamma^{-1}(w) = \begin{cases} (1 + \gamma w)^{1/\gamma}, & 0 < |\gamma| \leq 1 \\ \exp w, & \gamma = 0 \end{cases} \quad (4)$$

として、

$$H(z) = s_\gamma^{-1} \left(\sum_{m=0}^M \tilde{c}_\gamma(m) \tilde{z}^{-m} \right) \quad (5)$$

と M 次までのメル一般化ケプストラム $\tilde{c}_\gamma(m)$ によってモデル化される。ここで、 \tilde{z}^{-1} は 1 次オールパス関数

$$\tilde{z}^{-1} = \Psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (6)$$

但し、

$$\Psi(e^{j\omega}) = e^{-j\beta(\omega)} \quad (7)$$

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

$$\tilde{\omega} = \beta(\omega) = \tan^{-1} \frac{(1 - \alpha^2) \sin \omega}{(1 + \alpha^2) \cos \omega - 2\alpha} \quad (8)$$

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⁽⁶⁾ ことが知られており、 $\tilde{c}_\gamma(m)$ はこのような非直線周波数軸上における一般化対数スペクトルのフーリエ係数となっている。

ところで、ケプストラムおよびメルケプストラムは定義より、第 1 次以降の係数は信号のゲインには依存しない。しかし、一般化ケプストラム同様、メル一般化ケプストラムはそれぞれの値がゲインの変化により全体的に変化してしまい、取扱いが簡単ではない。そこで、ゲインに依存しないパラメータによって $H(z)$ を表すことを考える。方法はいくつかあるが、ここでは文献(4)、(8)にならい、 K をゲインとして $H(z)$ を

$$H(z) = s_\gamma^{-1} \left(\sum_{m=0}^M b_\gamma(m) \Phi_m(z) \right) = K \cdot D(z) \quad (9)$$

但し、

$$\Phi_m(z) = \begin{cases} 1, & m=0 \\ \frac{(1 - \alpha^2) z^{-1}}{1 - \alpha z^{-1}} \tilde{z}^{-(m-1)}, & m \geq 1 \end{cases} \quad (10)$$

また、

$$K = s_\gamma^{-1}(b_\gamma(0)) \quad (11)$$

$$D(z) = \begin{cases} \left(1 + \gamma \sum_{m=1}^M b'_\gamma(m) \Phi_m(z) \right)^{1/\gamma}, & 0 < |\gamma| \leq 1 \\ \exp \sum_{m=1}^M b'_\gamma(m) \Phi_m(z), & \gamma = 0 \end{cases} \quad (12)$$

と変形することにする。ここで、

$$b'_\gamma(m) = b_\gamma(m) / (1 + \gamma b_\gamma(0)), \quad m = 1, 2, \dots, M \quad (13)$$

である。

係数 $b_\gamma(m)$ と $\tilde{c}_\gamma(m)$ は、

$$b_\gamma(m) = \begin{cases} \tilde{c}_\gamma(m), & m=M \\ \tilde{c}_\gamma(m) - ab_\gamma(m+1), & 0 \leq m < M \end{cases} \quad (14)$$

あるいは、

$$\tilde{c}_\gamma(m) = \begin{cases} b_\gamma(m), & m=M \\ b_\gamma(m) + ab_\gamma(m+1), & 0 \leq m < M \end{cases} \quad (15)$$

で互いに交換することができ、行列表現した場合には、

$$\mathbf{b}_\gamma = \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{c}}_\gamma \quad (16)$$

$$\tilde{\mathbf{c}}_\gamma = \tilde{\mathbf{A}} \mathbf{b}_\gamma \quad (17)$$

但し、

$$\tilde{\mathbf{A}} = \begin{bmatrix} 1 & a & 0 & \cdots & 0 \\ 0 & 1 & a & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & & \ddots & \ddots & a \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad (18)$$

$$\tilde{\mathbf{A}}^{-1} = \begin{bmatrix} 1 & (-a) & (-a)^2 & \cdots & (-a)^M \\ 0 & 1 & (-a) & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & (-a)^2 \\ \vdots & & \ddots & \ddots & (-a) \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad (19)$$

$$\tilde{\mathbf{c}}_\gamma = [\tilde{c}_\gamma(0), \tilde{c}_\gamma(1), \dots, \tilde{c}_\gamma(M)]^T \quad (20)$$

$$\mathbf{b}_\gamma = [b_\gamma(0), b_\gamma(1), \dots, b_\gamma(M)]^T \quad (21)$$

と書くことができる。従って、 $\tilde{\mathbf{c}}_\gamma$ と \mathbf{b}_γ は線形変換の関係をもつことから、式(1)の $\tilde{\mathbf{c}}_\gamma$ に関する最小化は、 \mathbf{b}_γ に関する最小化に等価となる。更に、式(11)、(13)の関係より、これは K および

$$\mathbf{b}'_\gamma = [b'_\gamma(1), b'_\gamma(2), \dots, b'_\gamma(M)]^T \quad (22)$$

の最小化に等価である。

ここで、 $H(z)$ を音声の合成フィルタと考えれば、 $H(z)$ は安定でなければならないことから、 $H(z)$ が最小位相であると仮定すれば、文献(4)と同様にして、

$$\tilde{\mathbf{c}}_\gamma = [\tilde{c}_\gamma(0), \tilde{c}_\gamma(1), \dots, \tilde{c}_\gamma(M)]^T \quad (23)$$

に関する式(1)の最小化問題は、 \mathbf{b}'_γ に関する

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \quad (24)$$

の最小化、および K に関する E の最小化と等価になることが示される。式(1)を最小化する K は、 $\partial E / \partial K = 0$ とおけば、 \mathbf{b}'_γ に関して最小化された ε の値 ε_{\min} によって、

$$K = \sqrt{\varepsilon_{\min}} \quad (25)$$

で与えられる。

2.2 解の一意性と安定性

ε は、 $-1 \leq \gamma \leq 0$ において、 \mathbf{b}'_γ に関して凸であることが示される(付録1.)ので、 ε の極小点は唯一存在し、

$\nabla \varepsilon = \partial \varepsilon / \partial \mathbf{b}'_\gamma = \mathbf{0}$ を満たす \mathbf{b}'_γ によって与えられる。 $\nabla \varepsilon$ は、

$$\tilde{\mathbf{r}} = [\tilde{r}(1), \tilde{r}(2), \dots, \tilde{r}(M)]^T \quad (26)$$

$$\tilde{r}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \Phi_k^*(e^{j\omega}) d\omega \quad (27)$$

とおけば、

$$\nabla \varepsilon = -2 \tilde{\mathbf{r}} \quad (28)$$

と書ける。

音声の分析では、得られた伝達関数は音声の合成フィルタとして用いられるため、分析によって得られる伝達関数ははじめから安定であることが望ましい。本推定法では、付録2.に示すように、 $-1 \leq \gamma \leq 0$ のもとで、 $\nabla \varepsilon = \mathbf{0}$ を満たす \mathbf{b}'_γ によって定められる $H(z)$ は、必ず安定となる。但し、 $0 < \gamma \leq 1$ では、 ε は必ずしも凸とはならないので、数学的取扱いは容易ではない。従って、本論文では、 $-1 \leq \gamma \leq 0$ の場合についてのみ考えることにする。

3. 最小化問題の解法

3.1 ニュートン・ラフソン法

ニュートン・ラフソン法を用いて、 ε を最小にする \mathbf{b}'_γ を得ることを考える。つまり、文献(4)、(8)と同様、 i 番目の近似値 $\mathbf{b}'_\gamma^{(i)}$ に対し、線形方程式

$$(\tilde{\mathbf{P}} + (1+\gamma)\tilde{\mathbf{Q}}) \Delta \mathbf{b}'_\gamma^{(i)} = \tilde{\mathbf{r}} |_{\mathbf{b}'_\gamma = \mathbf{b}'_\gamma^{(i)}} \quad (29)$$

但し、

$$\tilde{\mathbf{P}} = \begin{bmatrix} \tilde{p}(0) & \tilde{p}(1) & \cdots & \tilde{p}(M-1) \\ \tilde{p}(1) & \tilde{p}(0) & \cdots & \tilde{p}(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}(M-1) & \tilde{p}(M-2) & \cdots & \tilde{p}(0) \end{bmatrix} \quad (30)$$

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \tilde{q}(2) & \tilde{q}(3) & \cdots & \tilde{q}(M+1) \\ \tilde{q}(3) & \tilde{q}(4) & \cdots & \tilde{q}(M+2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{q}(M+1) & \tilde{q}(M+2) & \cdots & \tilde{q}(2M) \end{bmatrix} \quad (31)$$

但し、

$$\tilde{p}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \frac{(1-a^2)^2 e^{j\omega}}{(1-ae^{j\omega})(e^{j\omega}-a)} \cdot \Psi^{*k}(e^{j\omega}) d\omega \quad (32)$$

$$\tilde{q}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) \frac{(1-a^2)^2 e^{j2\omega}}{(1-ae^{j\omega})^2} \cdot \Psi^{*k-2}(e^{j\omega}) d\omega \quad (33)$$

を

$$\Delta \mathbf{b}'_\gamma^{(i)} = [\Delta b'_\gamma^{(i)}(1), \Delta b'_\gamma^{(i)}(2), \dots, \Delta b'_\gamma^{(i)}(M)]^T \quad (34)$$

について解き、次回の近似値

$$b_r^{(i+1)} = b_r^{(i)} + \Delta b_r^{(i)} \quad (35)$$

を得る。

式(29)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(例えば文献(11), (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ための計算が支配的となるが、これも次節で述べるように、FFT と再帰式を利用した効率的な算法によって計算することができる。

付録 1. に示すように、式(24)の ε は b_r に関して凸であることから、 ε の Hesse 行列 H 、つまり式(29)の係数行列は正定である。従って、式(29)および式(35)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(29)も、 H が正定であることから、必ず一意解をもつ。

i 番目の繰返しにおける ε の値は、式(28)より

$$\varepsilon^{(i)} = \tilde{r}(0) + \gamma r^T \cdot b_r |_{b_r = b_r^{(i)}} \quad (36)$$

で計算することができ、 $\varepsilon^{(i)}$ の変化率を見ることにより、繰返しを終了することができる。変化率が 0.1% 程度以下で終了すれば、十分な収束が得られ、このときの平均繰返し回数は、音声信号の分析において数回程度となることが確かめられている。また、 K の i 番目の近似値 $K^{(i)}$ は、

$$K^{(i)} = \sqrt{\varepsilon^{(i)}} \quad (37)$$

で求めることができる。

3.2 方程式の係数計算

式(29)の方程式の係数は、文献(8)と同様の方法によって計算することもできるが、ここでは以下に示す計算法⁽¹⁴⁾を用いる。 $D(e^{j\omega})$ が与えられれば、

$$p(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} e^{j\omega k} d\omega \quad (38)$$

$$q(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+2\gamma}(e^{j\omega})|^2} D^{2\gamma}(e^{j\omega}) e^{j\omega k} d\omega \quad (39)$$

$$r(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} D^{\gamma}(e^{j\omega}) e^{j\omega k} d\omega \quad (40)$$

は FFT によって計算することができる。従って、付録

3. に示すように、 $\tilde{p}(k)$ は $p(k)$ から

$$\tilde{p}'(k) = \begin{cases} \frac{1}{2} p(0), & k=0 \\ p(k), & k=1, 2, \dots, N \end{cases}$$

$$\tilde{p}^{(j)}(k) = \begin{cases} (1-\alpha^2) \tilde{p}'(-j) + \alpha \tilde{p}^{(j-1)}(0), & k=0 \\ (1-\alpha^2) \tilde{p}^{(j-1)}(0) + \alpha \tilde{p}^{(j-1)}(1), & k=1 \\ \tilde{p}^{(j-1)}(k-1) + \alpha(\tilde{p}^{(j-1)}(k) - \tilde{p}^{(j)}(k-1)), & k=2, 3, \dots, M-1, \end{cases}$$

$$j = -N, \dots, -1, 0$$

$$\tilde{p}(k) = \begin{cases} 2\tilde{p}^{(0)}(0), & k=0 \\ \tilde{p}^{(0)}(k), & k=1, 2, \dots, M-1 \end{cases} \quad (41)$$

で、また $\tilde{q}(k)$ から再帰式

$$\tilde{q}^{(j)}(k) = \begin{cases} (1-\alpha^2)q(-j+1) + \alpha \tilde{q}^{(j-1)}(1), & k=1 \\ (1-\alpha^2) \tilde{q}^{(j-1)}(1) + \alpha \tilde{q}^{(j-1)}(2), & k=2 \\ \tilde{q}^{(j-1)}(k-1) + \alpha(\tilde{q}^{(j-1)}(k) - \tilde{q}^{(j)}(k-1)), & k=3, 4, \dots, 2M, \end{cases}$$

$$j = -N+1, \dots, -1, 0$$

$$\tilde{q}(k) = \tilde{q}^{(0)}(k), \quad k=2, 3, \dots, 2M \quad (42)$$

で、また $\tilde{r}(k)$ は $r(k)$ から再帰式

$$\tilde{r}^{(j)}(k) = \begin{cases} (1-\alpha^2)r(-j+1) + \alpha \tilde{r}^{(j-1)}(1), & k=1 \\ \tilde{r}^{(j-1)}(k-1) + \alpha(\tilde{r}^{(j-1)}(k) - \tilde{r}^{(j)}(k-1)), & k=2, 3, \dots, M, \end{cases}$$

$$j = -N+1, \dots, -1, 0$$

$$\tilde{r}(k) = \tilde{r}^{(0)}(k), \quad k=1, 2, \dots, M \quad (43)$$

によって計算することができる。但し、 N は理論上 $N \rightarrow \infty$ であるが、実際には打切り誤差の問題のならない 100 ないし数百程度の値を用いればよい。

また、式(38)~(40)における $D^{\gamma}(e^{j\omega})$ は、 $D(e^{j\omega})$ の一般化ケプストラム $c_r(m)$ を、 $b_r(m)$ から式(17)と再帰式⁽⁸⁾

$$c_r^{(i)}(m) = \begin{cases} (1-\alpha^2)b_r'(-j+1) - \alpha b_r^{(j-1)}(1), & m=1 \\ c^{(j-1)}(m-1) - \alpha(c^{(j-1)}(m) - c^{(j)}(m-1)), & m=2, 3, \dots, N, \end{cases}$$

$$j = -M, \dots, -1, 0$$

$$c_r(m) = c_r^{(0)}(m), \quad m=1, 2, \dots, N \quad (44)$$

によって求め、

$$D^{\gamma}(e^{j\omega}) = 1 + \gamma \sum_{m=1}^N c_r(m) e^{-j\omega m} \quad (45)$$

を FFT で計算することにより求めることができる。 N の値については、式(41)~(43)と同様である。

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、このような初期値として、メル一般化対数スペクトルの線形平滑化⁽¹⁵⁾によって得られるメル一般化ケプストラムを用いることができる。つまり、従来のケプストラム法によって得られる N 次のケプストラム $c(m)$ から 1~ M 次のメル一般化ケプストラム $\tilde{c}_r(m)$ を求める場合には、再帰式

$$\tilde{c}^{(i)} = \begin{cases} c(-i) + \alpha \tilde{c}^{(i-1)}(0), & m=0 \\ (1-\alpha^2) \tilde{c}^{(i-1)}(0) + \alpha \tilde{c}^{(i-1)}(1), & m=1 \\ \tilde{c}^{(i-1)}(m-1) + \alpha(\tilde{c}^{(i-1)}(m) - \tilde{c}^{(i)}(m-1)), & m>1 \end{cases}$$

$$i = -N, \dots, -1, 0$$

$$\tilde{c}(m) = \tilde{c}^{(0)}(m), \quad 1 \leq m \leq M \quad (46)$$

よって、 $c(m)$ からメルケプストラム $\tilde{c}(m)$ へ周波数変換を行い、その後、

$$\tilde{c}_\gamma(m) = \tilde{c}(m) + \sum_{k=1}^{m-1} \frac{k}{m} \gamma \tilde{c}(k) \tilde{c}_\gamma(m-k), \quad 1 \leq m \leq M \quad (47)$$

よって $\gamma=0$ から任意の γ への一般化対数変換を行ってメル一般化ケプストラム $\tilde{c}_\gamma(m)$ を求める。この後、式(14)、(13)を用いることにより、初期値 $\mathbf{b}_\gamma^{(0)}$ を得る。

4. 他の手法との関係

本推定法は、特別な場合として、 $\alpha=0$ とした場合には一般化ケプストラムをパラメータとするスペクトル推定法⁽⁴⁾に等価であり、更に、 $\gamma=-1$ とすれば線形予測法⁽¹⁾、 $\gamma=0$ とした場合にはケプストラム法⁽²⁾ をスペクトル推定の観点から厳密化した対数スペクトルの不偏推定法⁽¹⁰⁾ と等価となる。また、 α を $-1 < \alpha < 1$ の任意の値、 γ を $\gamma=0$ とした場合に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⁽⁶⁾ に等価である。このように、本推定法は線形予測法、ケプストラム法、メルケプストラム分析法などを統一化した上で、これらを特別な場合として含むスペクトル推定法となっており、これらの関係は図1に示したとおりである。なお、図中では、 $-1 < \alpha < 1$ 、 $\gamma=-1$ の場合をメル線形予測法(Mel-LPC Method)と呼んでいる。

解の一意性、および得られる伝達特性の安定性は2.2で述べたとおり、図中いずれの場合にも保証されているが、計算量は α, γ の値によって異なるため、以下にまとめる。

提案法において $\gamma=-1$ とした場合、評価関数 ε は \mathbf{b}_γ に関する2次関数となる。従って、 $\nabla \varepsilon = 0$ は、式(28)からもわかるとおり、対称Toeplitz形の係数行列をもつ線形方程式となるため、直接解くことができる。また、 $\gamma=-1$ の場合、ニュートン・ラフソン法の繰返しは初期値によらず1回で収束し、特に初期値を $\mathbf{b}_\gamma^{(0)} = \mathbf{0}$ とした場合には、1回目の繰返しにおける線形方程式(29)は、 $\nabla \varepsilon = 0$ に等しくなる。

提案法において $\alpha=0$ とした場合には、式(41)~(44)および(46)の再帰式の計算を省略することができる。

更に、 $\alpha=0$ かつ $\gamma=-1$ とした場合には、繰返し計算、再帰式の計算がともに不要となり、 $\nabla \varepsilon = 0$ で与え

られる線形方程式は従来の線形予測法(自己相関法)における正規方程式に一致する。

以上より、定量的な評価は行っていないが、線形予測法を基準とした提案法の計算量は、 α 、および γ の値により、1倍から数十倍程度となると見込まれる。

5. シミュレーション

本章では、いくつかの合成信号の分析例により、本推定法の特徴を示す。標準化周波数は10 kHzとし、時間窓は102.4 msのブラックマン窓を用いた。分析次数は $M=10$ である。また、FFT長は1,024点とし、再帰式は512次まで計算した。いずれの合成信号もデジタルフィルタを周期10.8 msのパルス列で励振することにより生成した。未知系のスペクトルは細線で、推定されたスペクトルは太線で示されている。スペクトル推定値の評価として、ADPCMなどの音声符号化において予測器の評価尺度として用いられる予測利得(Prediction Gain) $G = (\text{入力信号のエネルギー}) / (\text{予測誤差のエネルギー})$ を用いている。式(24)は利得が1の逆フィルタ出力のエネルギー、つまり予測残差のエネルギーを表しているため、式(24)の評価関数を小さくする推定値ほど、予測利得 G を大きくすることになる。

5.1 α の値に関する検討

提案法によれば、 α の値によって、低周波数域あるいは高周波数域で高い周波数分解能をもったスペクトル推定値が得られる。このことを示すため、未知系スペクトルの山と谷が、低周波数域に集まっている場合、周波数軸にそって均一に分布する場合、高周波数域に集まっている場合の三つの例について、シミュレーション結果を図2に示す。いずれも、 γ の値は $\gamma=0$ とした。

図2(a)からわかるとおり、スペクトルの低周波数域に鋭い山と谷をもった未知系の場合には、 α を正の値にすることにより、 α が負の値のときには推定できなかったスペクトルの山と谷を推定することができている。また、このとき予測利得も高い値が得られている。逆に高周波数域に鋭い山と谷をもった未知系の場合には、図2(c)に示すように、 α を負の値にすることにより山と谷を推定することができている。予測利得も高い値となっている。更に、図2(b)のようにスペクトルの山と谷が周波数軸上に均一に分布する場合には、 $\alpha=0$ とすればよいことがわかる。

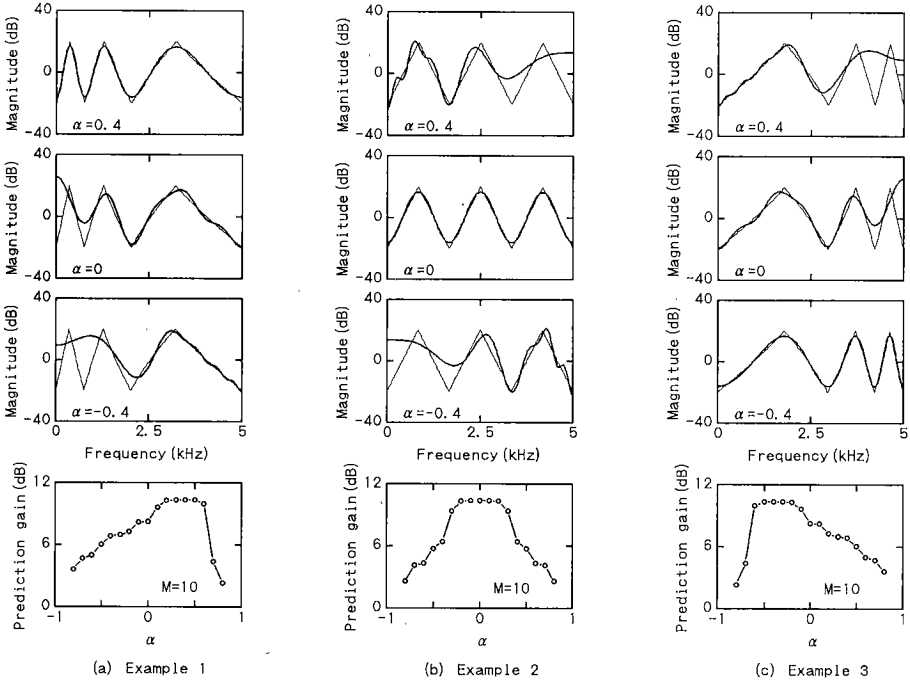


図2 α の値に関する検討 ($\gamma=0$)
 Fig. 2 Spectral estimates for several values of α . ($\gamma=0$)

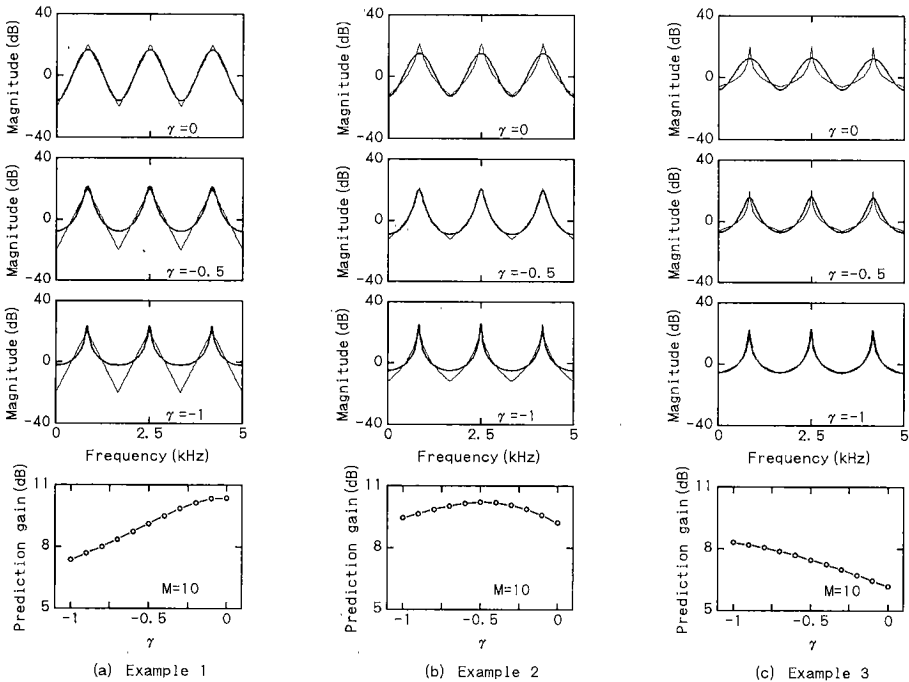


図3 γ の値に関する検討 ($\alpha=0$)
 Fig. 3 Spectral estimates for several values of γ . ($\alpha=0$)

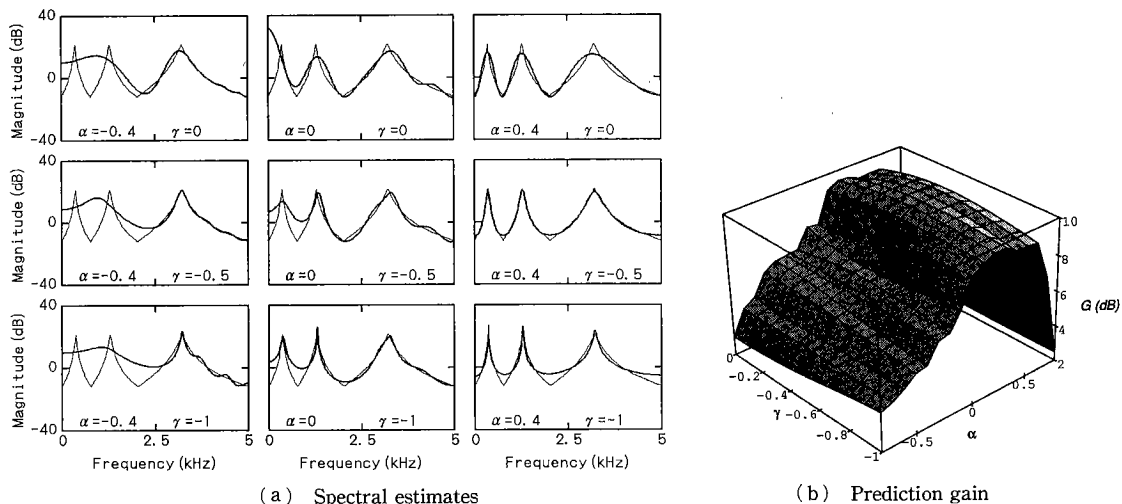


図4 α および γ の値に関する検討
Fig. 4 Spectral estimates for several values of α and γ .

5.2 γ の値に関する検討

図3に、 γ の値によってスペクトル推定値がどのように変化するかを示すために、いくつかの推定例を示す。図3(a)は、スペクトルの山と谷が対数振幅スペクトル上で同じ鋭さをもつ場合の例で、 $\gamma=0$ とすることで山と谷をともに同等の正確さで推定できている。予測利得もこの場合が最も高くなっている。 $\gamma=-1$ の場合、つまり線形予測法による分析では、スペクトルの谷が推定されておらず、得られたスペクトルの山も未知系のそれより鋭くなってしまっている。逆に、図3(c)は、スペクトルが鋭い山をもつ場合の例で、 $\gamma=-1$ とすることによりスペクトルの山がうまく推定されているが、 $\gamma=0$ とした場合にはスペクトルの山がうまく推定されていない。図3(b)は、図3(a)と図3(c)との中間的な例であり、 γ を-1と0の間の値にすることにより、スペクトルの山、谷ともに同等の正確さで推定されている様子が図よりわかる。図3(a)~(c)を通して、適切なスペクトル推定値が得られている γ の値で、予測利得も高い値を示していることがわかる。

5.3 α および γ の値に関する検討

図4(a)は、低周波数域にスペクトルの山と谷が集まり、更にスペクトルの山が谷に比べやや鋭い場合の分析例である。5.1および5.2での検討より、 α が正の値、 γ が-1と0の中間の値でスペクトルの山、谷が同じ正確さで推定できることが予想できる。ここに示した分析例の中では、 $\alpha=0.4$ 、 $\gamma=-0.5$ の場合にこのよ

うなスペクトル推定値が得られていることがわかる。図4(b)には、 α と γ の値を変化させた場合の予測利得を示してある。この図でも、 $\alpha=0.4$ 、 $\gamma=-0.5$ 付近で最も高い予測利得が得られている様子がわかる。

このように、

(1) 低周波数域、高周波数域のどちらに鋭いスペクトルの山谷があるか。

(2) スペクトルの山と谷のどちらが鋭いか。

のような未知系のスペクトルの性質がある程度わかっている場合には、提案法を α と γ の値を適切に設定して用いることにより、同じ分析次数で推定できないようなスペクトルの山や谷を推定することができ、また高い予測利得を得ることができる。

6. 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」を、いくつかの γ の値に対して、分析した結果を図5に示す。いずれも $\alpha=0.35$ としている。標準化周波数は10 kHz、時間窓は25.6 msのブラックマン窓を用いた。分析次数は $M=12$ である。またFFT長は256点とし、再帰式は128次まで計算した。フレーム周期は5 msとしている。

ケプストラムをパラメータとする場合には、分析次数を $M=20\sim 25$ とする必要があるが、図より、 $M=12$ 程度の小さな次数で、特に1~2 kHz以下の低周波数域でのスペクトルの形状をよくとらえている様子がわかる。また、 γ の値を-1に近づけることによりスペク

トルの山がはっきりと抽出され、 γ の値を 0 に近づけることによりスペクトルの山と谷が同等に抽出されていることがわかる。

MGLSA フィルタ⁽⁹⁾ あるいは文献(16)のフィルタ

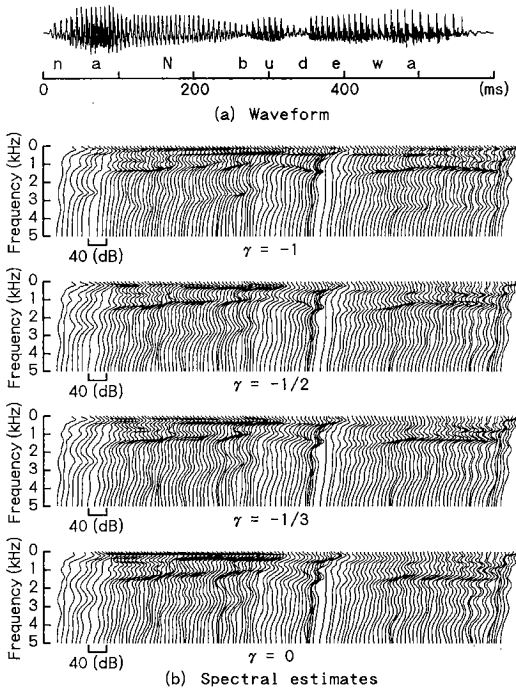


図5 自然音声の分析例 ($M=12$)
Fig. 5 Spectral estimates of a natural speech. ($M=12$)

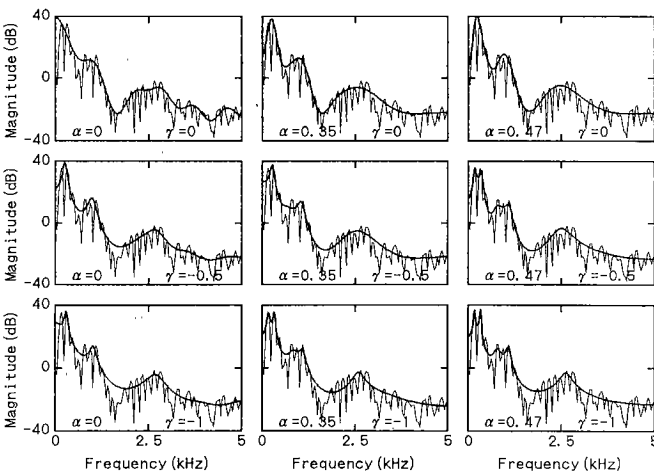
を用いることにより、提案法によって得られた分析結果から直接、音声を作成することができる。分析合成音の非公式な聴取によれば、

(1) $\alpha=0.35$ とした場合の合成音は、 $\alpha=0$ の場合の半分から 2/3 程度の次数で、 $\alpha=0$ の場合とほぼ同等の品質が得られる。

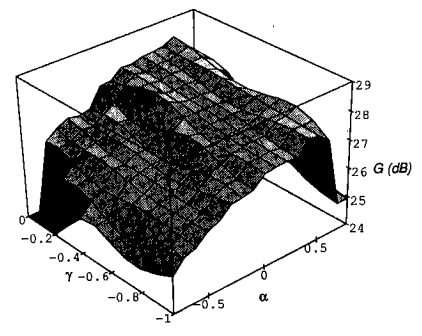
(2) γ の値により音質が変化し、 γ を -1 に近づけることにより歯切れのよい感じ、0 に近づけることにより滑らかな感じの音になる。

(3) 次数 M を十分大きくした場合には、 γ の値による合成音の音質の違いはさほど大きくはない。などがわかっているが、音声の分析・合成における α , γ の最適値を検討するためには、次数 M とも関連づけて、詳しい聴取試験を行う必要がある。

以上は、 α , γ を分析フレームによらず、固定とした場合であるが、自然音声では、スペクトルの性質は分析フレームごとに变化するため、 α , γ を可変とすることも考えられる。図 6, 7 は、図 5 に示した自然音声の鼻音/N/および母音/e/部分の分析結果である。図 6 の鼻音/N/の場合、低い周波数に共振点、反共振点が集まっていることから、 α が正の値で予測利得が大きくなっており、また、1.6 kHz 付近の反共振の影響のため、 $\gamma=0$ で予測利得は最も大きくなっている。図 7 の母音/e/の場合には、共振点が周波数軸に沿ってほぼ一様に分布しており、これは $\alpha=0$, $\gamma=-1$ 付近で予測利得が最大となっていることと合致している。図 6, 7 より、



(a) Spectral estimates



(b) Prediction gain

図6 自然音声の分析例/N/
Fig. 6 Spectral estimates of natural speech /N/.

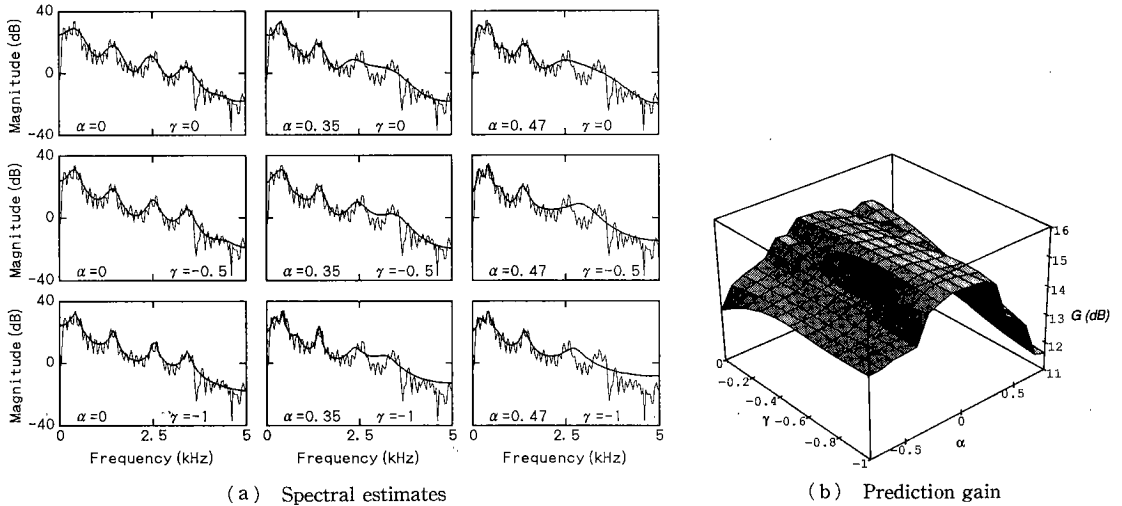


図7 自然音声の分析例/e/
Fig. 7 Spectral estimates of natural speech /e/.

α, γ を分析フレームごとに最適な値となるように変化させることによって、 α, γ を固定する場合に比べて大きな予測利得が得られることが期待できる。但し、先見的に α, γ の各フレームにおける最適値を知ることは難しいため、一つのフレームに対して何通りかの α, γ の値により分析を行い、この中から、予測利得などの基準により、最適な α, γ の値を選択する必要がある。従って、例えば α, γ の値をそれぞれ3通りの中から選ぶとしても、計算量は、 α, γ が固定の場合に比べ、 $3 \times 3 = 9$ 倍となり、計算量の大きさが問題となると考えられる。

7. むすび

メル一般化ケプストラムにより表現されたスペクトルモデルを用いるスペクトル推定法を提案した。また、このときの最小化問題の解の一意性を示した上でその解法を与え、繰返し計算の収束性、得られる伝達特性の安定性を示した。更に、実際の解法において、再帰式、FFT、 $O(M^2)$ の高速算法により、効率的に計算を行うことができることを示した。

提案した推定法では、 α および γ の値によって、直線周波数あるいはメルなどの非直線周波数軸上で、全極形、ケプストラムで表された形、またはそれらの中間的な形のスペクトルを得ることができ、信号に適したスペクトル表現法を選ぶことができる。

本論文では、計算法、安定性および収束性の証明について重点をおいて述べ、いくつかの分析例によりそ

の特徴と有効性を示したが、実際の音声合成、音声認識、その他のスペクトル推定への適用については今後の課題とする。

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) Markel J. D. and Gray Jr. A. H.: "Linear Prediction of Speech", Springer-Verlag, New York (1976).
- (2) Oppenheim A. V. and Schaffer R. W.: "Digital Signal Processing", Prentice-Hall, Englewood Cliffs, N. J. (1975).
- (3) Kobayashi T. and Imai S.: "Spectral analysis using generalized cepstrum", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-32**, pp. 1087-1089 (Oct. 1984).
- (4) 徳田恵一, 小林隆夫, 山本竜太郎, 今井 聖: "一般化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), **J72-A**, 3, pp. 457-465 (1989-03).
- (5) Fant G.: "Speech sound and features", MIT Press, Cambridge (1973).
- (6) 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).
- (7) 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).
- (8) 徳田恵一, 小林隆夫, 深田俊明, 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", 信学論(A), **J74-A**, 8, pp. 1240-1248 (1991-08).
- (9) 小林隆夫, 今井 聖, 福田 豊: "メル一般化対数スペクトル近似 (MGLSA) フィルタ", 信学論(A), **J68-A**, 6, pp. 610-611 (1985-06).
- (10) 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).

- (11) Gohberg I. and Koltracht I.: "Efficient Algorithm for Toeplitz Plus Hankel Matrices", *Integral Equations and Operator Theory*, **12**, pp. 136-142(1989).
- (12) Yagle A. E.: "New Analogs of Split Algorithms for Arbitrary Toeplitz-plus-Hankel Matrices", *IEEE Trans. Acoust., Speech & Signal Process.*, **ASSP-39**, **11**, pp. 2457-2463(Nov. 1991).
- (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) 徳田恵一, 小林隆夫, 千葉健司, 今井 聖: "メル一般化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", 第4回デジタル信号処理シンポジウム議論集, B-5-2, pp. 285-290(1989-12).
- (15) 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).
- (16) 千葉健司, 徳田恵一, 小林隆夫, 今井 聖: "メル一般化ケプストラムをパラメータとする音声合成", 1990 信学春季全大, A-243, 1-243.
- (17) 美多 勉: "デジタル制御理論", 昭晃堂 (1984).

付 録

1. ϵ が凸であることの証明

文献(4)における凸性の証明と同様, $w = x + jy$ の関数

$$f(w) = 1/|s_\gamma^{-1}(w)|^2 \tag{A.1}$$

が $-1 \leq \gamma \leq 0$ で凸であることを用いれば, 任意に選んだ二つの b_γ を b_a, b_b として,

$$\epsilon(s b_a + (1-s) b_b) \leq s \epsilon(b_a) + (1-s) \epsilon(b_b), 0 \leq s \leq 1 \tag{A.2}$$

を示すことができ, ϵ は $-1 \leq \gamma \leq 0$ で b_γ に関して凸である.

2. 安定性の証明

$\gamma = 0$ のとき, $H(z)$ は b_γ の値によらず最小位相となるため安定である. $-1 \leq \gamma < 0$ のときには,

$$H(z) = K \left\{ \frac{1}{C(z)} \right\}^{-1/\gamma} \tag{A.3}$$

但し,

$$C(z) = 1 + \gamma \sum_{m=0}^M \tilde{c}_\gamma(m) z^{-m} \tag{A.4}$$

として, $1/C(z)$ が安定ならば, 式(A.3)で与えられる $H(z)$ も安定となる. 従って, $C(z)$ の零点が z 平面上の単位円内であれば, $C(z) = C(\Psi(z))$ の z 平面上の零も単位円内にあることから, 問題は $1/C(z)$ の安定性に帰着される. ここで,

$$\tilde{c}_\gamma = [\tilde{c}_\gamma(1), \tilde{c}_\gamma(2), \dots, \tilde{c}_\gamma(M)]^T \tag{A.5}$$

とおけば, \tilde{c}_γ と b_γ は,

$$\begin{bmatrix} 1 \\ \gamma b_\gamma \end{bmatrix} = \tilde{A}^{-1} \begin{bmatrix} 1 + \gamma \tilde{c}_\gamma(0) \\ \gamma \tilde{c}_\gamma \end{bmatrix} \tag{A.6}$$

の関係をもつ.

決定方程式(28)および式(25)より,

$$\begin{bmatrix} \tilde{p}(0) & r^T \\ r & \tilde{P} \end{bmatrix} \begin{bmatrix} 1 \\ \gamma b_\gamma \end{bmatrix} = \begin{bmatrix} K^2 \\ 0 \end{bmatrix} \tag{A.7}$$

が成り立つ. 但し, \tilde{P} は式(30)で与えられ,

$$\tilde{p}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D^{1+\gamma}(e^{j\omega})|^2} \Psi^{*k}(e^{j\omega}) d\omega \tag{A.8}$$

である. 式(A.6)を式(A.7)に代入し, 更に両辺に左から $(\tilde{A}^{-1})^T$ を掛けること,

$$\begin{bmatrix} \tilde{p}(0) & \tilde{P}^T \\ \tilde{P} & \tilde{P} \end{bmatrix} \begin{bmatrix} 1 + \gamma \tilde{c}_\gamma(0) \\ \gamma \tilde{c}_\gamma \end{bmatrix} = K^2 \begin{bmatrix} 1 \\ \alpha \end{bmatrix} \tag{A.9}$$

但し,

$$\tilde{P} = \begin{bmatrix} \tilde{p}(0) & \tilde{p}(1) & \dots & \tilde{p}(M-1) \\ \tilde{p}(1) & \tilde{p}(0) & \dots & \tilde{p}(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{p}(M-1) & \tilde{p}(M-2) & \dots & \tilde{p}(0) \end{bmatrix} \tag{A.10}$$

$$\tilde{p} = [\tilde{p}(1), \tilde{p}(2), \dots, \tilde{p}(M)]^T \tag{A.11}$$

$$\alpha = [(-a), (-a)^2, \dots, (-a)^M]^T \tag{A.12}$$

の関係が得られる.

ここで, $C(z)$ のコンパニオン行列は,

$$A = \left[\begin{array}{cccc|c} \tilde{c}_\gamma^T / (1 + \gamma \tilde{c}_\gamma(0)) & & & & \\ 1 & 0 & \dots & 0 & \\ 0 & 1 & \ddots & \vdots & \\ \vdots & \ddots & \ddots & 0 & \\ 0 & \dots & 0 & 1 & \end{array} \right] \mathbf{0} \tag{A.13}$$

と与えられるが, 次のような行列 P を考え,

$$P = \tilde{P} - K^2 \alpha \alpha^T \tag{A.14}$$

式(A.9)の関係をを用いれば,

$$A P A^T = \tilde{P} - K^2 \frac{(1 - a^2)}{a^2} \alpha \alpha^T \tag{A.15}$$

となることが示されるので,

$$G = K^2 \frac{(1 - a^2)}{a^2} \alpha \alpha^T \tag{A.16}$$

とおけば, リヤプノフ方程式⁽¹⁷⁾

$$A \tilde{P} A^T - \tilde{P} = -G \tag{A.17}$$

を得ることができる. \tilde{P} が正定, G が半正定, かつ

$$\text{rank} \begin{bmatrix} G^{1/2} \\ G^{1/2} A \\ \vdots \\ G^{1/2} A^{M-1} \end{bmatrix} = M \tag{A.18}$$

但し,

$$G^{1/2} = K \sqrt{\frac{1-a^2}{a^2}} \alpha \quad (\text{A}\cdot 19)$$

となることを示すことができるので、 $C(z)$ の零点は単位円内にあることが示される。よって、 $C(\bar{z}) = C(\Psi(z))$ の z に関する零点も単位円内にあり、式(A・3)で与えられる $H(z)$ は最小位相、かつ安定である。

3. 再帰式の導出

ここでは、一般的に

$$x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) d\omega \quad (\text{A}\cdot 20)$$

から

$$\bar{x}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) Y_k^*(e^{j\omega}) d\omega \quad (\text{A}\cdot 21)$$

を再帰的に求めることを考える。式(A・21)は、 $Y_k(z)$ の逆 z 変換を $y_k(n)$ とすれば、

$$\bar{x}(k) = \sum_{n=-\infty}^{\infty} x(n) y_k(n) \quad (\text{A}\cdot 22)$$

と書くことができる。ここで $x(n)$ 、 $y_k(n)$ のいずれかが因果的であれば、式(A・22)は、

$$\bar{x}(k) = \sum_{n=0}^{\infty} x(-(i-n)) y_k(n) \Big|_{i=0} \quad (\text{A}\cdot 23)$$

となるので、 $\bar{x}(k)$ はフィルタ $Y_k(z)$ 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。従って、 $\bar{x}(k)$ をそれぞれ $\tilde{p}(k)$ 、 $\tilde{q}(k)$ 、 $\tilde{r}(k)$ に置き換えて考えれば、式(41)～(43)の再帰式が得られる。但し、式(41)の場合は、 $x(n)$ 、 $y_1(n)$ がともに因果的でないため、 $x(n)$ を因果的な部分と反因果的な部分に分けて考えている。

(平成3年10月21日受付、4年1月31日再受付)



千葉 健司

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、音声分析・合成、音声情報処理の研究に従事。現在、富士ゼロックス(株)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。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。



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、音声処理の研究に従事。IEEE会員。



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。