

メル一般化ケプストラム係数に基づく音声のスペクトル表現と
その諸特性

小石田和人[†] 徳田 恵一^{††} 小林 隆夫[†] 今井 聖^{†*}

Spectral Representation of Speech Based on Mel-Generalized Cepstral
Coefficients and Its Properties

Kazuhito KOISHIDA[†], Keiichi TOKUDA^{††}, Takao KOBAYASHI[†], and Satoshi IMAI^{†*}

あらまし メル一般化ケプストラム分析法では、スペクトルモデルを全極形からケプストラム形まで連続的に変化させることができ、更に人間の聴覚特性を考慮することも可能となる。本論文では、メル一般化ケプストラム分析法の音声符号化や分析合成への利用を目的として、メル一般化ケプストラム係数によるスペクトル表現について述べる。提案するスペクトル表現パラメータにより合成フィルタの安定性条件が明確になり、量子化後の安定性を容易に保証することができる。まず、提案法の分布特性、スペクトル感度などを示し、量子化・補間性能をLSPと比較した。合成音による主観評価の結果、提案法はLSPよりも高い量子化・補間性能をもつことを確認した。

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、安定性、スペクトル符号化

1. まえがき

これまでに種々の音声符号化・分析合成方式が提案され、実用化されている。これらの多くは、何らかのスペクトル分析に基づいており、どのようなスペクトル分析を行うかがシステムの特徴を決める大きな要因の一つとなっている。

ところで、音声のスペクトル分析法として、線形予測法[1]、ケプストラム法[2]などが広く用いられてきた。しかし、線形予測法では、音声スペクトルを全極モデルで表現することから零を表現することができず、また、ケプストラム法によって得られる指数形の伝達関数ではスペクトルの鋭いピークを表現しにくくなる。この問題を解決する手法として、一般化ケプストラム分析法[3]があるが、これはケプストラム法と線形予測法を統一的に扱い、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。

一方、人間の聴覚特性は低周波数域で細かい分解能、高周波数域で粗い分解能をもつ[4]ことから、この性質を音声のスペクトルを表現する際に取り入れることにより、少ないパラメータ次数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、周波数変換された線形予測法[5]やメルケプストラム分析法[6]が提案され、音声符号化の分野での有効性が報告されている[7]~[9]。

更に、以上の分析法を統一的に扱うことのできる手法として、メル一般化ケプストラム分析法[10]がある。この分析法では、スペクトルモデルの極と零の表現精度をべきパラメータ γ により、周波数軸を周波数圧縮パラメータ α により、それぞれ連続的に変化させることができ、音声のスペクトルを少ない次数でより効率的に表現することが可能となる。従って、メル一般化ケプストラム分析法を用いて音声の符号化・分析合成を行うシステムを実現することができれば、特に低ビットレートへの応用において、その有効性が大きいと考えられる。

音声符号化や分析合成システムにおいて、スペクトル包絡情報を担うパラメータの高能率符号化は、システムの性能にかかわる重要な課題の一つとなる。スペ

[†] 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, 226 Japan

^{††} 名古屋工業大学知能情報システム学科, 名古屋市
Faculty of Engineering, Nagoya Institute of Technology, Nagoya-shi, 466 Japan

* 現在, 千葉工業大学電子工学科

クトルパラメータの符号化において特に問題となるのは、合成フィルタの安定性と量子化・補間特性である。従来広く用いられている線形予測分析法に基づくシステムでは、線形予測係数を PARCOR 係数 [11] や LSP 係数 [12] に変換することで、安定性の条件を明確にしている。また、量子化・補間性能を比較すると、LSP 係数は PARCOR 係数よりも優れていることが明らかにされている [12]。

本論文では、低ビットレート音声符号化や分析合成への利用を目的として、メル一般化ケプストラム係数による音声のスペクトル表現について述べる。まず、提案するスペクトル表現をシステムの安定性問題の立場から考察する。更に、提案するスペクトル表現の諸特性を明らかにし、評価実験により量子化・補間特性を LSP と比較する。また、提案するスペクトル表現は、特別な場合として線形予測係数の LSP 表現を含む統一的手法とみることもできる。

2. メル一般化ケプストラム分析

ここでは、本論文で用いるメル一般化ケプストラムによるスペクトルモデルとその分析手法 [10] について簡単に述べる。

音声のスペクトル $H(e^{j\omega})$ を M 次までのメル一般化ケプストラム $c(m)$ によって、

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

$$= \begin{cases} \left(1 + \gamma \sum_{m=0}^M c(m) \bar{z}^{-m} \right)^{1/\gamma}, & -1 \leq \gamma < 0 \\ \exp \sum_{m=0}^M c(m) \bar{z}^{-m}, & \gamma = 0 \end{cases} \quad (2)$$

とモデル化する。但し、 $s_\gamma^{-1}(w)$ は一般化対数関数 $s_\gamma(w)$ の逆関数であり、

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

で与えられる。また、 \bar{z}^{-1} は 1 次のオールパス関数

$$\bar{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}, \quad |\alpha| < 1 \quad (4)$$

で与えられ、 $z = e^{j\omega}$ としたときの位相特性 $\bar{\omega}$ は、標準化周波数 8 kHz のとき α を 0.31、10 kHz のときに

α を 0.35 に選べば、人間の音の高さに対する聴覚特性を表すメル尺度 [4] をよく近似することが知られている [13]。

メル一般化ケプストラム分析法では、式 (2) のスペクトルモデルに対数スペクトルの不偏推定法 [14] における評価関数

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{ \exp R(\omega) - R(\omega) - 1 \} d\omega \quad (5)$$

を適用する。但し、

$$R(\omega) = \log I_N(\omega) - \log |H(e^{j\omega})|^2 \quad (6)$$

であり、 $I_N(\omega)$ は、長さ N の時間窓 $w(n)$ 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である。ここで、 $H(z)$ からゲイン K をくり出すことを考える。式 (2) を

$$H(z) = K \cdot D(z) \quad (7)$$

但し、

$$K = s_\gamma^{-1} \left(\sum_{m=0}^M (-\alpha)^m c(m) \right) \quad (8)$$

$$D(z) = s_\gamma^{-1} \left(\sum_{m=0}^M c'(m) \bar{z}^{-m} \right) \quad (9)$$

のように変形する。係数 $c'(m)$ は

$$\delta = \sum_{m=0}^M (-\alpha)^m c(m) \quad (10)$$

とおくと、 $c(m)$ から

$$c'(m) = \begin{cases} \frac{c(0) - \delta}{1 + \gamma\delta}, & m = 0 \\ \frac{c(m)}{1 + \gamma\delta}, & 1 \leq m \leq M \end{cases} \quad (11)$$

の関係式で変換することができる。このとき、式 (5) の $\mathbf{c} = [c(0), \dots, c(M)]^T$ に関する最小化問題は、 $D(z)$ のゲインが 1 という条件のもとで $\mathbf{c}' = [c'(0), \dots, c'(M)]^T$ に関して

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \quad (12)$$

を最小化する問題、および K に関する E の最小化に帰着される。この最小化問題は、逆フィルタ $1/D(z)$

に $x(n)$ を入力したときの出力, つまり予測誤差 $e(n)$ の最小平均 2 乗問題と解釈することができる。

式 (12) の ϵ は, 係数 c' に関して凸であることから, 解は一意的に存在し, 効率的な繰返し計算により, これを容易に求めることができる。また, 得られた伝達関数 $H(z)$ の安定性が保証される。

メル一般化ケプストラム分析法は, 特別な場合として種々のスペクトル推定法を含む統一的な手法となっている。例えば, $\alpha = 0$ とした場合には一般化ケプストラム分析法に等価であり, 更に $\gamma = -1$ で線形予測法, $\gamma = 0$ で対数スペクトルの不偏推定法と等価となる。また, $\gamma = 0$ で α を $|\alpha| < 1$ の任意の値とした場合には, メルケプストラム分析法と等価になる。

3. メル一般化ケプストラムの量子化

メル一般化ケプストラム分析法を用いて音声の分析合成系や符号化系を構成する場合, スペクトル包絡情報としてメル一般化ケプストラムを量子化する必要がある。また, 量子化後のシステムの安定性が, 理論的に保証されることが望ましい。本章では, まずメル一般化ケプストラムによるスペクトル表現についてシステムの安定性の立場から考察する。更に, 提案するスペクトル表現の分布特性, スペクトル感度を示し, パラメータ変換のための計算精度について検討する。

3.1 量子化したシステムの安定性

$\gamma = 0$ のとき, 式 (9) の $D(z)$ は係数 c' の値によらず最小位相となるため, 因果性を仮定するならば, c' にどのような値を与えても安定となる。従って以下では, $-1 \leq \gamma < 0$ の場合について議論する。

$-1 \leq \gamma < 0$ では式 (9) の $D(z)$ を

$$D(z) = \left\{ \frac{1}{C(\tilde{z})} \right\}^{-1/\gamma} \quad (13)$$

但し,

$$C(\tilde{z}) = 1 + \gamma \sum_{m=0}^M c'(m) \tilde{z}^{-m} \quad (14)$$

と表すことができ, $1/C(\tilde{z})$ が z^{-1} に関して安定ならば, 式 (13) の $D(z)$ も安定となる [10]。

一般に, 伝達関数 $1/F(z)$ が安定, つまり

$$F(z) = \sum_{m=0}^M f(m) z^{-m} \quad (15)$$

の零点が単位円内にあるための必要十分条件は,

$$F(z) = F_1(z) + F_2(z) \quad (16)$$

のように $F(z)$ を偶多項式 $F_1(z)$ と奇多項式 $F_2(z)$ の和に分解すれば, 次のように与えられる [15], [16]。

- $F_1(z), F_2(z)$ のすべての零点は単位円上にある
- $F_1(z), F_2(z)$ の零点はすべて異なる
- $F_1(z), F_2(z)$ の零点は互いに他を隔離する

これらの性質により $F(z)$ を $F_1(z)$ と $F_2(z)$ の零点の位置で表現することにより, 伝達関数 $1/F(z)$ の安定性の条件が明確になる。つまり, 量子化・補間後の $F_1(z)$ と $F_2(z)$ の零点が上記の条件を満たしていれば, $1/F(z)$ の安定性が保証されることになる。線形予測係数の LSP 表現 [12] は, このような音声スペクトル表現法の一つとして知られている。

ところで, 式 (4) により \tilde{z} 平面の単位円内の点は z 平面の単位円内に 1 対 1 で写像される。従って, 伝達関数 $1/C(\tilde{z})$ の z^{-1} に関する安定性は, $1/C(\tilde{z})$ の \tilde{z}^{-1} に関する安定性の問題に帰着される。そこで, \tilde{z} 領域で $C(\tilde{z})$ を

$$C(\tilde{z}) = C_1(\tilde{z}) + C_2(\tilde{z}) \quad (17)$$

但し,

$$C_1(\tilde{z}) = \frac{1 + \gamma c'(0)}{2} \left\{ 1 + \tilde{z}^{-(M+1)} + \gamma \sum_{m=1}^M (c''(m) + c''(M-m-1)) \tilde{z}^{-m} \right\} \quad (18)$$

$$C_2(\tilde{z}) = \frac{1 + \gamma c'(0)}{2} \left\{ 1 - \tilde{z}^{-(M+1)} + \gamma \sum_{m=1}^M (c''(m) - c''(M-m-1)) \tilde{z}^{-m} \right\} \quad (19)$$

ここで,

$$c''(m) = \frac{c'(m)}{1 + \gamma c'(0)}, \quad 1 \leq m \leq M \quad (20)$$

と分解し, $C_1(\tilde{z})$ と $C_2(\tilde{z})$ の \tilde{z} に関する零点によってスペクトルを表現することを考える。このとき, \tilde{z} 平面における単位円上の零点は, メル周波数軸上の周波数値として表現されることになる。区間 $[0, \pi]$ のみに注目すれば, これらの零点を $e^{j\tilde{\omega}_i}$ ($i = 0, 1, \dots, M+1$) と表すことができる。但し, $\tilde{\omega}_i$ は

表1 音声の分析条件
Table 1 Speech analysis conditions.

| | |
|-----------|-------|
| サンプリング周波数 | 8 kHz |
| 分析次数 | 10 |
| 分析窓 | ハミング窓 |
| 分析窓長 | 32 ms |
| 分析周期 | 10 ms |

$$0 = \tilde{\omega}_0 < \tilde{\omega}_1 < \dots < \tilde{\omega}_M < \tilde{\omega}_{M+1} = \pi \quad (21)$$

の関係をもつ。零点 $\tilde{\omega}_0 (= 0)$ と $\tilde{\omega}_{M+1} (= \pi)$ は固定値であるため、 M 個のパラメータ $\tilde{\omega}_i (i = 1, 2, \dots, M)$ により音声スペクトルが表現されることになる。

なお、 $\gamma = 0$ の場合にはこのような周波数領域のパラメータ $\tilde{\omega}_i$ は定義されない。しかし、前述のとおり、この場合には常に安定性が保証されているため、周波数領域のパラメータに変換する必要はなく、ケプストラム若しくはメルケプストラムを直接量子化すればよい。

3.2 統計的性質

実際の音声データ（男女各8名の短文章、約50秒）を用いてスペクトルパラメータ $\tilde{\omega}_i$ の諸特性を明らかにする。音声データの分析条件を表1にまとめる。対数スペクトルひずみは128次のケプストラム距離から算出し、ケプストラムの値はメル一般化ケプストラムにメル一般化対数変換[17]を施して求めている。また、以下では対数スペクトルひずみをスペクトルひずみと呼ぶことにする。

3.2.1 分布特性

各次数のスペクトルパラメータの度数分布を図1に示す。図1より、 γ が0に近づくにつれて、パラメータはほぼ等間隔に並び、その分布の範囲が狭くなっていくことがわかる。このような性質は、フィルタ $1/C(z)$ の表す対数スペクトルがなだらかで鋭いピークをもたなくなることによる。

3.2.2 スペクトル感度

パラメータの微小変化に対するスペクトルひずみは、スペクトル感度と呼ばれている。ここでは、各フレームで得られる M 次のパラメータのうちの一つのパラメータのみに微小変化 Δ を与え、そのときのスペクトル包絡の変化量をそのパラメータのスペクトル感度と定義する。微小変化として、 $\Delta = \pi/L$ を各次数のパラメータに加算した。図2には、 $L = 256$ におけるスペクトル感度を示す。この図から、 γ が0に近づくにつれて感度が高くなることが読み取れる。これは分布が狭くなっていくことに対応していると考えられ

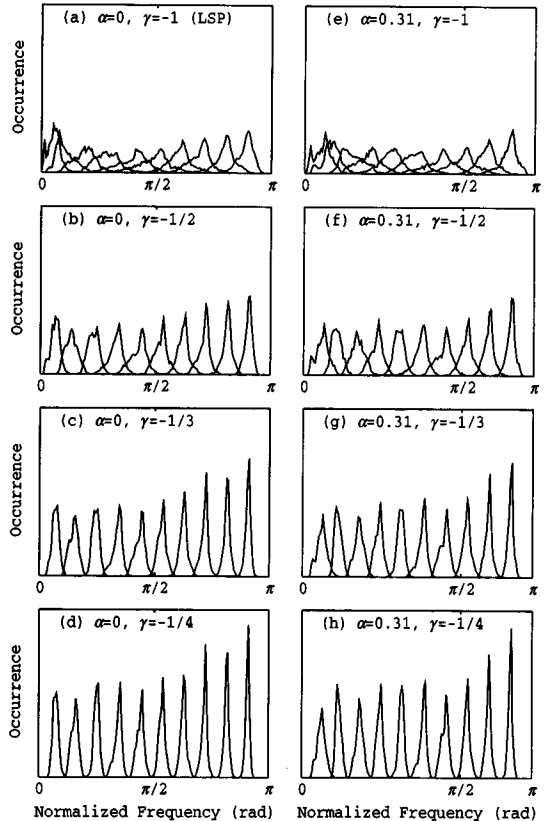


図1 提案するスペクトルパラメータの度数分布
Fig.1 Distribution plots of proposed parameters.

る。また、メル周波数軸では低域のパラメータの感度が低くなるという傾向をもつ。

3.2.3 計算精度

線形予測係数からLSP係数を計算する際には、ニュートンの反復法などにより直接解を求める方法、係数のDCTを利用する方法[18]、チェビシエフ多項式を利用する方法[19]などが用いられる。メル一般化ケプストラム係数をスペクトルパラメータ $\tilde{\omega}_i$ に変換する場合にも、これらの手法を利用することができる。ここではDCTによる方法を用い、その計算精度および変換誤差について検討する。この方法では、 $C_1(z)$ の係数から $\tilde{z}_k^{-1} = e^{-jk\pi/L} (k = 0, \dots, L)$ における値をDCTから求め、隣接するDCT値が異符号となる、つまり零点を含む区間を特定する。更に、特定された2点間に対して補間を繰り返し、零点を推定する。一方、 $C_2(z)$ の零点は、単位円上において $C_1(z)$ の零点には含まれていることから、繰り返し補間を行うことに

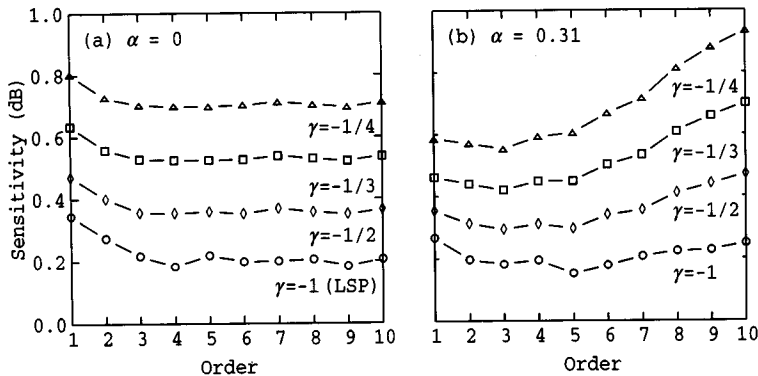


図2 提案するスペクトルパラメータのスペクトル感度
Fig. 2 Spectral sensitivity of proposed parameters.

表2 $C_1(z)$ の零点探索のための平均補間回数の比較
Table 2 Comparison of the number of interpolation for searching the roots of $C_1(z)$.

| α | γ | L | | | | |
|----------|----------|------|------|------|------|------|
| | | 16 | 32 | 64 | 128 | 256 |
| 0 | -1 | - | - | 2.86 | 2.32 | 1.93 |
| | -1/2 | 3.66 | 2.91 | 2.49 | 2.11 | 1.87 |
| | -1/3 | 3.23 | 2.67 | 2.33 | 2.01 | 1.83 |
| | -1/4 | 2.98 | 2.49 | 2.23 | 1.96 | 1.79 |
| 0.31 | -1 | - | - | 2.92 | 2.36 | 1.96 |
| | -1/2 | 3.74 | 3.03 | 2.56 | 2.15 | 1.89 |
| | -1/3 | 3.36 | 2.76 | 2.38 | 2.04 | 1.85 |
| | -1/4 | 3.10 | 2.59 | 2.27 | 1.98 | 1.81 |

よって、効率的に探索を行うことができる。

分割数 L に対して $C_1(z)$ の零点推定の際の平均補間回数を表 2 に示す。補間の終了条件は、 $|C_1(e^{j\omega})| < 10^{-5}$ とした。表中の“-”は、単位円の分割数不足のために零点を分離できなかったことを示している。提案するスペクトルパラメータは、 $\gamma = -1$ の場合を除いて、分割数 $L = 16 \sim 32$ 程度でも零点が分離可能であるという特徴をもつ。この特徴は、先に述べたようにパラメータの分布範囲が狭くなり、ほぼ均一に並ぶためである。参考のため、 $C_2(z)$ の零点探索のための平均補間回数はいずれの場合も約 16 回程度でほぼ等しい。また、パラメータ変換（メル一般化ケプストラム \rightarrow スペクトルパラメータ $\tilde{\omega}_i \rightarrow$ メル一般化ケプストラム）によって生じるスペクトルひずみは 10^{-4} dB 以下で、LSP 係数の場合のそれと同等以下になることが確認されている。

4. 評価実験

音声の分析合成や符号化のシステムを実現する場合、

スペクトルパラメータを量子化・補間して表現する必要がある。このとき、量子化・補間誤差によって、スペクトル包絡が変化してしまい、品質劣化の一要因となる。この章では、メル一般化ケプストラム係数によるスペクトルパラメータの量子化・補間特性をスペクトルひずみと合成音による主観評価を通して検討し、LSP 係数のそれと比較する。

4.1 実験条件

学習データには、男女各 20 名の短文章、合計約 650 秒を、評価用には学習外の男女各 8 名の短文章、約 50 秒を用いた。また、分析は表 1 と同じ条件で行った。

符号帳には、多段ベクトル量子化 [20] を採用した。本実験では各段 12 bits の 2 段構成を用いることとする。符号帳の学習には LBG アルゴリズム [21] を用い、符号帳は各段ごとに学習した。また、符号帳の学習・探索のための距離尺度には、ユークリッド距離を用いた。但し予備的な実験により、 $\gamma = -1$ の場合のみユークリッド距離よりも重み付きユークリッド距離 [22] を用いた方が客観・主観性能ともに改善が得られたことから、 $\gamma = -1$ の場合には重み付きユークリッド距離を用いることにする。

4.2 スペクトルひずみによる客観評価

まずはじめに、補間特性を調べた。補間特性とは、パラメータの更新周期を長くしていった場合に、伝送されないパラメータ値を伝送されてきたパラメータの値からどの程度近似できるかというものである。ここでは、補間特性のみに注目するため量子化は行わず、補間は直線補間とした。更新周期を 20 ms から 40 ms まで変化させて、10 ms ごとの平均スペクトルひずみを表 3 に示す。次に、パラメータの量子化周期を 10 ms

表3 補間特性 (dB)
Table 3 Interpolation properties (dB).

| α | γ | Frame Interval | | |
|----------|----------|----------------|-------|-------|
| | | 20 ms | 30 ms | 40 ms |
| 0 | -1 | 0.74 | 1.29 | 1.73 |
| | -1/2 | 0.73 | 1.28 | 1.72 |
| | -1/3 | 0.73 | 1.30 | 1.74 |
| | -1/4 | 0.74 | 1.31 | 1.76 |
| 0.31 | -1 | 0.71 | 1.25 | 1.67 |
| | -1/2 | 0.70 | 1.24 | 1.67 |
| | -1/3 | 0.71 | 1.26 | 1.69 |
| | -1/4 | 0.72 | 1.27 | 1.71 |

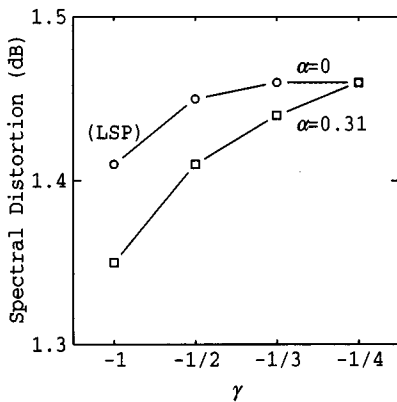


図3 2段ベクトル量子化によるスペクトルひずみ
Fig.3 Spectral distortion of 2-stage vector quantization.

として、量子化によって生じる平均スペクトルひずみを図3に示す。量子化・補間によるスペクトルひずみと (α, γ) の値には次のような傾向が見られる。

- γ を -1 に近づけるとひずみは減少する。
- α は 0 のときよりも 0.31 の方がひずみは小さくなる。

ここで、量子化によって生じるスペクトル包絡の変化の一例を図4に示す。この図から、スペクトルの形状、発生する誤差が (α, γ) の値によって異なることが読み取れる。量子化対象となるスペクトルの形状が同じ場合には、スペクトルひずみと主観性能との相関は大きく、スペクトルひずみから主観性能を推測することができる。しかし、本実験では量子化対象となるスペクトルの形状が異なるため、スペクトルひずみの大きさのみで量子化・補間性能を評価するのは十分ではない。そこで、次節では合成音による主観評価試験によって量子化・補間性能の評価を行うことにする。

4.3 合成音声の品質評価

分析合成による合成音声の品質を、5段階の Mean

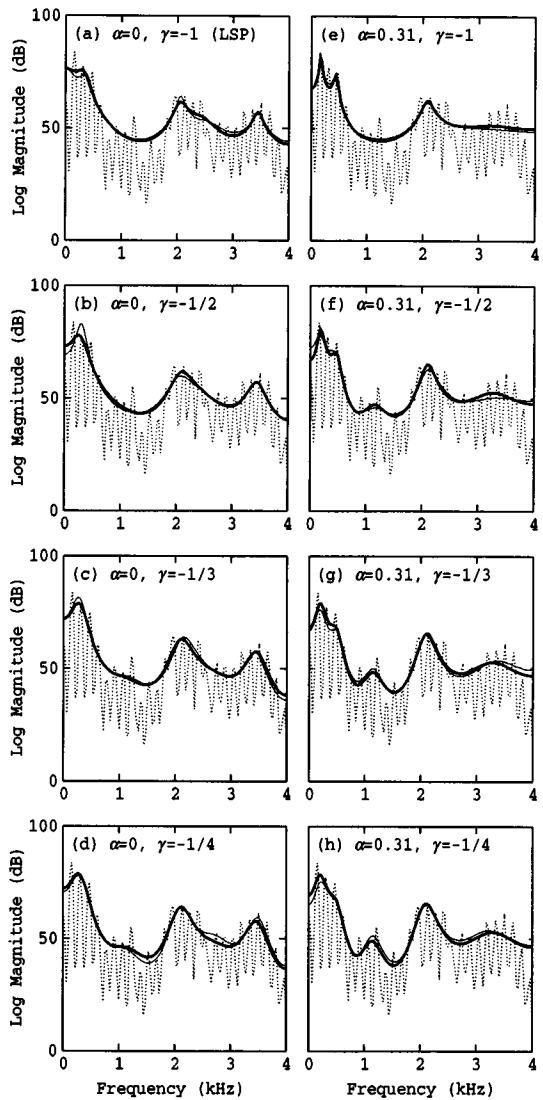


図4 量子化前後のスペクトル包絡
Fig.4 Original and quantized spectral envelopes.

Opinion Score (MOS) により評価する。リファレンス信号は原音声、テスト信号はスペクトル包絡のみを10msごとに量子化した合成音である。但し、スペクトルパラメータを各点ごとに補間した。励振源には、有声部に対してはピッチ周期に対応したパルス列を、無声部に対してはM系列を用いた。また、ピッチやゲインなどの量子化は考慮していない。音声サンプルは男女各2名によって発話された4種類の短文章、被験者は6名、繰返しは2回である。

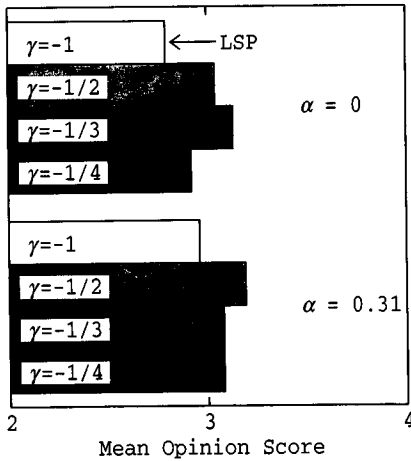


図5 受聴試験の結果
Fig. 5 Result of listening test.

図5に受聴試験の結果を示す。この図から、提案するスペクトルパラメータのMOS値はLSPよりも最大で0.3程度高く、主観的な量子化・補間性能が良いことが示された。実際に合成音を試聴した結果、次のような傾向が見られた。

- γ が0に近づくにつれ不連続感が軽減され、滑らかな感じの音質になる。
- γ が-1に近づくにつれ明るい感じの音質になる。

従ってこの実験結果を、被験者が音質の明るさを保持しつつ不連続感の比較的小さい $\gamma = -1/2$ 若しくは $\gamma = -1/3$ の合成音声に高いスコアを与えたと解釈できる。また、 α を0.31とすることにより若干の性能改善が見られる。

5. 検 討

メル一般化ケプストラム分析法は、 $(\alpha, \gamma) = (0, -1)$ のとき線形予測法と等価になることから、提案するスペクトル表現は特別な場合として線形予測係数のLSP表現を含む手法とみることできる。また、メル周波数軸で線形予測分析を行い、得られた係数をLSP表現するメル時間域LSP分析が提案されている[23]。この分析法は自己相関領域で周波数変換の処理が行われているため、メル一般化ケプストラム分析で $(\alpha, \gamma) = (0.31, -1)$ として推定された予測係数をLSP表現したものは異なったものになる。

提案法の応用の一つに広帯域音声符号化が挙げられ

る。分析次数を大きくとる必要がある広帯域音声符号化では、隣り合うLSPの周波数値が近接することがあり、LSPの変換に高い計算精度が要求される。文献[24]では、変換のための計算精度を緩和する手法が提案されている。この手法では、パワースペクトルのダイナミックレンジが1/2になるようなLPC係数をLPCケプストラムから求め、合成フィルタを全極フィルタの2段縦続構成として実現している。しかし、この全極フィルタの2段構成で得られるスペクトルは、もとの線形予測分析法から得られたスペクトルと異なったものとなり、いくぶんかのスペクトルひずみが発生する。一方、提案手法において $(\alpha, \gamma) = (0, -1/2)$ とした場合、つまり一般化ケプストラム分析法は、2段縦続の全極モデルにおいて予測誤差を最小にするという意味で最適であり、このようなスペクトル誤差は生じない。

6. む す び

音声の符号化や分析合成への利用を目的として、メル一般化ケプストラムによるスペクトル表現について述べた。提案するスペクトル表現パラメータにより合成フィルタの安定性を容易に保証できることを示し、パラメータの分布特性・スペクトル感度などを明らかにした。提案法をスペクトルパラメータの符号化へ応用するために、量子化・補間性能を評価した。合成音に対する受聴試験の結果、提案法はLSP以上の量子化・補間性能をもつことを確認した。

今後は、スペクトルパラメータの量子化効率を上げるために、フレーム間予測などの手法を取り入れていく予定である。また、電話音声など帯域制限された音声に対する検討も必要である。

謝辞 本研究の一部は、文部省科学研究費特別研究員奨励費、国際コミュニケーション研究奨励金の援助を受けて行われた。

文 献

- [1] 板倉文忠, 斉藤取三, “統計的手法による音声スペクトル密度とホルマント周波数の推定,” 信学論 (A), vol.J53-A, no.1, pp.35-42, Jan. 1970.
- [2] A.V. Oppenheim and R.W. Schaffer, “Digital signal processing,” Prentice-Hall, Englewood Cliffs, N.J., 1975.
- [3] 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), vol.J72-A, no.3, pp.457-465, March 1989.
- [4] G. Fant, “Speech sound and features,” MIT Press, Cambridge, 1973.
- [5] H.W. Strube, “Linear prediction on a warped frequency

- scale," J. Acoust. Soc. Am., vol.68, no.4, pp.1071-1076, Oct. 1980.
- [6] 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), vol.J74-A, no.8, pp.1240-1248, Aug. 1991.
- [7] E. Krüger and H.W. Strube, "Linear prediction on a warped frequency scale," IEEE Trans. Acoust., Speech & Signal Processing, vol.ASSP-36, pp.1529-1531, Sept. 1988.
- [8] 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を利用した音声符号化とその評価," 信学論 (A), vol.J77-A, no.11, pp.1443-1452, Nov. 1994.
- [9] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "CELP coding based on mel-cepstral analysis," Proc. ICASSP-95, pp.33-36, May 1995.
- [10] 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分析による音声のスペクトル推定," 信学論 (A), vol.J75-A, no.7, pp.1124-1134, July 1992.
- [11] F. Itakura and S. Saito, "On the optimum quantization of feature parameters in the PARCOR speech synthesizer," Proc. Conf. Speech Commun. Process, pp.434-437, 1972.
- [12] 管村 昇, 板倉文忠, "線スペクトル対 (LSP) 音声分析合成方式による音声情報圧縮," 信学論 (A), vol.J64-A, no.8, pp.599-606, Aug. 1981.
- [13] 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), vol.J66-A, no.2, pp.122-129, Feb. 1983.
- [14] 今井 聖, 古市千枝子, "対数スペクトルの不偏推定," 信学論 (A), vol.J70-A, no.3, pp.471-480, March 1987.
- [15] H.W. Schussler, "A stability theorem for discrete systems," IEEE Trans. Acoust., Speech & Signal Processing, vol.ASSP-24, pp.87-89, Feb. 1976.
- [16] Y. Bistritz, "A discrete stability equation theorem and method of stable model reduction," Systems & Contr. Lett., vol.1, pp.373-381, May 1982.
- [17] 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), vol.J71-A, no.1, pp.128-131, Jan. 1988.
- [18] F.K. Soong and B-W. Juang, "Line spectrum pair (LSP) and speech data compression," Proc. ICASSP-84, pp.1.10.1-1.10.4, 1984.
- [19] P. Kabal and R.P. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-34, pp.1419-1426, Dec. 1986.
- [20] B.H. Juang and A.H. Gray, Jr., "Multiple stage vector quantization for speech coding," Proc. ICASSP-82, pp.597-600, 1982.
- [21] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., vol.28, pp.84-95, Sept. 1980.
- [22] N. Phamdo, N. Farvardin, 守谷健弘, "多段ベクトル量子化を用いた LSP パラメータの情報源-通信路結合符号

化," 信学技報, SP90-52, 1990.

- [23] 関 章光, 宮内 新, 荒井秀一, "メル LSP ベクトル量子化音声符号化方式," 信学技報, SP86-14, 1986.
- [24] 佐々木茂明, 片岡章俊, 守谷健弘, "演算精度及び処理量を考慮した広帯域 CELP," 音響学会講論集, 1-4-14, March 1995.

(平成 9 年 2 月 18 日受付, 5 月 30 日再受付)



小石田和人 (学生員)

平 6 東工大・工・電気電子卒。平 7 同大学院博士前期課程了。現在、同大博士後期課程在学中。音声の高効率符号化の研究に従事。日本学術振興会特別研究員、日本音響学会会員。



徳田 恵一 (正員)

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。平 8 名工大知能情報システム学科助教。工博。音声分析、音声合成・符号化、音声認識、デジタル信号処理の研究に従事。日本音響学会、IEEE 各会員。



小林 隆夫 (正員)

昭 52 東工大・工・電気卒。昭 57 同大学院博士課程了。同年東工大精密工学研究所助手。工博。現在同助教。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、日本神経回路学会、IEEE 各会員。



今井 聖 (正員)

昭 34 東工大・工・電気卒。昭 39 同大学院博士課程了。同年東工大精密工学研究所助手。昭 43 同大助教。昭 54 同大教授。平 9 東工大名誉教授、千葉工大電工工学科教授。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。