

メル一般化ケプストラム分析に基づく CELP 音声符号化

小石田和人<sup>†</sup> 徳田 恵一<sup>††</sup> 小林 隆夫<sup>†</sup> 今井 聖<sup>†††</sup>

CELP Speech Coding Based on Mel-Generalized Cepstral Analysis

Kazuhito KOISHIDA<sup>†</sup>, Keiichi TOKUDA<sup>††</sup>, Takao KOBAYASHI<sup>†</sup>, and Satoshi IMAI<sup>†††</sup>

あらまし 本論文では、メル一般化ケプストラム分析に基づいた CELP 音声符号化方式を提案する。メル一般化ケプストラムによって音声スペクトルを表現することにより、スペクトルの極と零の表現精度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、更に人間の周波数分解能に関する聴覚特性を考慮することも可能となる。提案方式では、聴覚重み付けとポストフィルタリングの操作をメル一般化ケプストラムを介して行うことができることから、人間の聴覚特性にあった処理が期待される。ビットレート 4 kbit/s において、提案方式と従来の線形予測に基づく CELP 方式を主観的に比較した。その結果、提案方式は従来方式を等価  $Q$  値で 2 dB 程度上回ることが示された。

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, 聴覚特性, CELP 音声符号化

1. ま え が き

電話回線のデジタル化, 移動体通信やボイスメールなどの需要の高まりに伴い, 低ビットレートかつ高品質な音声符号化方式の研究が数多く行われている。その中で最も有力な手法の一つに CELP (Code Excited Linear Prediction) 符号化方式 [1] がある。CELP 方式では, 音声信号を線形予測分析により得られるスペクトル包絡成分と励振源に分離して符号化処理が行われる。これまでに, CELP 方式を基にした LD-CELP [2], CS-ACELP [3], VSELTP [4], PSI-CELP [5] などが勧告・標準化されている。これらの符号化系の多くは, 線形予測分析の枠組の中で, 主として励振源の効率的な構造化により品質の向上を達成してきた。しかし, ビットレートが低くなるにつれ, 励振源のベクトル長が長くなり, 励振源に割り当てられるビット数も減少することなどから, 励振源の構造化だけでは十分な品質が得られなくなる。従って, 更

に品質を向上させるためには, 励振源のみならず, スペクトル包絡をより効率的に表現することも重要な課題の一つとなる。

CELP 方式など多くの音声符号化方式で用いられている線形予測分析法は, 低演算量という利点をもつが, 全極モデルを仮定していることから鼻音などのスペクトルの零を表現することができないという問題がある。これに対して, メルケプストラム分析法 [6] を用いることにより, スペクトルの極と零とを同等に表現することができ, 更に低周波数域で細い分解能, 高周波数域で粗い分解能をもつ人間の音の高さに対する聴覚特性を考慮することが可能となる。実際に筆者らは, メルケプストラム分析を ADPCM 方式や CELP 方式に適用することによって, 主観的な品質を改善できることを報告した [7], [8]。

更に, 音声スペクトルを少ないパラメータ数でより効率的に推定する手法に, メル一般化ケプストラム分析法 [9] がある。メル一般化ケプストラム分析法では, 極と零との表現精度を全極形からケプストラム形の間で任意に選択することができ, その上, メルケプストラム分析法と同様に人間の周波数軸に対する聴覚特性を考慮することも可能となる。

以上の観点から本論文では, メル一般化ケプストラム分析に基づく CELP 符号化方式を提案する。提案する CELP 符号化系では, 合成フィルタ, 聴覚重み付け

<sup>†</sup> 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, 226-8503 Japan

<sup>††</sup> 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, 466-8555 Japan

<sup>†††</sup> 千葉工業大学電子工学科, 習志野市  
Department of Electronics, Chiba Institute of Technology, Narashino-shi, 275-0016 Japan

フィルタ, ポストフィルタをメル一般化ケプストラムを介して実現できるという特徴をもつ。ビットレート 4 kbit/s の同一符号化条件下において, 提案符号化系と従来の線形予測に基づく CELP 符号化系とを主観評価実験により比較し, 提案法の有効性を示す。

## 2. メル一般化ケプストラム分析

### 2.1 スペクトルモデルと評価関数

メル一般化ケプストラム分析 [9] では, 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係数  $\mathbf{c} = [c(0), \dots, c(M)]^T$  によって,

$$H(z) = s_\gamma^{-1}(\tilde{z}^T \mathbf{c}) \quad (1)$$

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

とモデル化し, 対数スペクトルの不偏推定法 [10] における評価関数を適用する。但し,  $s_\gamma^{-1}(w)$  は一般化対数関数  $s_\gamma(w)$  の逆関数であり,

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

で定義される。また,  $\tilde{z} = [1, \tilde{z}^{-1}, \dots, \tilde{z}^{-M}]^T$  において,  $\tilde{z}^{-1}$  は 1 次オールパス関数

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

で与えられ, その位相特性は, 標準化周波数 8 kHz のとき  $\alpha$  を 0.31 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 [11] をよく近似することが知られている [12]。

ここで,  $H(z)$  からゲイン項  $K$  を括り出し, ゲインに依存しないパラメータにより  $H(z)$  を表現することを考える。式 (2) を  $\boldsymbol{\alpha} = [1, (-\alpha), (-\alpha)^2, \dots, (-\alpha)^M]^T$  を用いて

$$H(z) = K \cdot D(z) \quad (5)$$

但し,

$$K = s_\gamma^{-1}(\boldsymbol{\alpha}^T \mathbf{c}) \quad (6)$$

$$D(z) = s_\gamma^{-1}(\tilde{z}^T \mathbf{c}') \quad (7)$$

と変形する。フィルタ  $D(z)$  のゲインは常に 1 となることから, 係数  $\mathbf{c}' = [c'(0), \dots, c'(M)]^T$  は,  $\boldsymbol{\alpha}^T \mathbf{c}' = 0$  の関係を満たす。また, 係数  $\mathbf{c}'$  と  $\mathbf{c}$  の関係は,

$$\mathbf{c}'(m) = \begin{cases} \frac{c(0) - \boldsymbol{\alpha}^T \mathbf{c}}{1 + \gamma \boldsymbol{\alpha}^T \mathbf{c}}, & m = 0 \\ \frac{c(m)}{1 + \gamma \boldsymbol{\alpha}^T \mathbf{c}}, & 1 \leq m \leq M \end{cases} \quad (8)$$

となる。係数  $\mathbf{c}$  に関する対数スペクトルの不偏推定法における評価関数

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\exp R(\omega) - R(\omega) - 1\} d\omega \quad (9)$$

の最小化問題は,  $D(z)$  のゲインが 1 という条件のもとで, 係数  $\mathbf{c}'$  に関する

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|D(e^{j\omega})|^2} d\omega \quad (10)$$

の最小化, および  $K$  に関する  $E$  の最小化に帰着される。ここで,

$$R(\omega) = \log I_N(\omega) - \log |H(e^{j\omega})|^2 \quad (11)$$

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。この最小化問題は, 時間領域では入力信号  $x(n)$  を逆フィルタ  $1/D(z)$  でフィルタリングしたときの出力, つまり予測誤差の最小平均 2 乗問題と解釈することができる [13], [14]。

式 (10) の  $\varepsilon$  は,  $\mathbf{c}'$  に関して凸であり, 効率的な繰返し計算により大域的最適解を容易に求めることができる。また, 得られる伝達関数  $H(z)$  の安定性が保証される。

### 2.2 合成フィルタ

式 (7) の合成フィルタ  $D(z)$  は,  $\gamma = 0$  のときには MLSA フィルタ [12] により, それ以外の  $\gamma$  に対しては MGLSA フィルタ [15] により実現可能である。特に  $n$  を自然数として,  $\gamma = -1/n$  の場合のみを考えると, 合成フィルタは

$$D(z) = \left\{ \frac{1}{C(\tilde{z})} \right\}^n \quad (12)$$

但し,

$$C(\tilde{z}) = 1 + \gamma \tilde{z}^T \mathbf{c}' \quad (13)$$

と表すことができる。このとき  $D(z)$  は図 1 のように

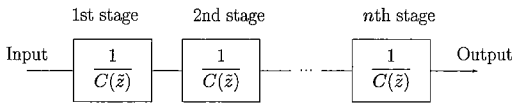


図1 合成フィルタ  $D(z)$  の構成 ( $\gamma = -1/n$ )  
Fig.1 Synthesis filter  $D(z)$  ( $\gamma = -1/n$ ).

$1/C(\bar{z})$  の  $n$  段連続構成となる [16]. しかし, 式 (13) を用いて合成フィルタを構成しようとするとき, デイレーブリーフをもつために直接実現することはできない. そこで, 文献 [12] や [17] と同様にして,  $C(\bar{z})$  を

$$C(\bar{z}) = 1 + \gamma \bar{z}^T \mathbf{A} \mathbf{A}^{-1} \mathbf{c}' \quad (14)$$

$$= 1 + \gamma \bar{\Phi}^T \mathbf{b} \quad (15)$$

のように変形する. 但し, 行列  $\mathbf{A}^{-1}$  は  $\mathbf{A}$  の逆行列であり, それぞれ

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha & 0 & \cdots & 0 \\ 0 & 1 & \alpha & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & & & \ddots & \alpha \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad (16)$$

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & (-\alpha) & (-\alpha)^2 & \cdots & (-\alpha)^M \\ 0 & 1 & (-\alpha) & \ddots & \vdots \\ 0 & 0 & 1 & \ddots & (-\alpha)^2 \\ \vdots & & & \ddots & (-\alpha) \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} \quad (17)$$

で定義される. また,  $\bar{\Phi}$  は

$$\bar{\Phi} = \mathbf{A}^T \bar{z} = [1, \Phi_1(z), \dots, \Phi_M(z)]^T \quad (18)$$

但し,

$$\Phi_m(z) = \frac{(1 - \alpha^2)z^{-1}}{1 - \alpha z^{-1}} \bar{z}^{-(m-1)}, \quad m \geq 1 \quad (19)$$

により, 係数  $\mathbf{b}$  は

$$\mathbf{b} = \mathbf{A}^{-1} \mathbf{c}' = [0, b(1), \dots, b(M)]^T \quad (20)$$

により与えられる. ここで係数  $\mathbf{b}$  の第1項が0となるのは, 係数  $\mathbf{c}'$  が  $\alpha^T \mathbf{c}' = 0$  を満たすことによる. 式 (20) の行列演算は

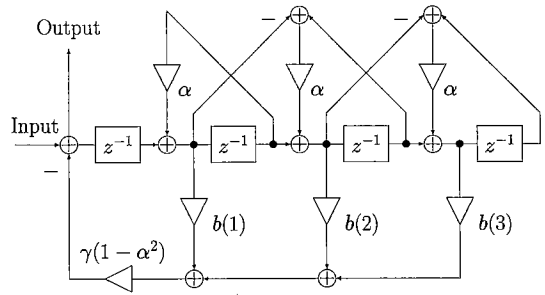


図2 フィルタ  $1/C(\bar{z})$  の構成 ( $M = 3$ )  
Fig.2 Structure of the filter  $1/C(\bar{z})$  ( $M = 3$ ).

$$b(m) = \begin{cases} c'(M), & m = M \\ c'(m) - \alpha b(m+1), & \\ 0 \leq m < M \end{cases} \quad (21)$$

により, 再帰的に計算することもできる. 図2には,  $M = 3$  のときの式 (15) に基づくフィルタ  $1/C(\bar{z})$  の構成を示す.

### 3. CELP 符号化方式

CELP 符号化方式は, 入力音声から線形予測分析法によって, スペクトル包絡に対応する線形予測係数を抽出し, これを符号帳に蓄えた励振ベクトルで駆動することで音声を合成する符号化方式である. 図3に適応符号帳を用いる CELP 符号化方式の構成を示す. 合成フィルタは  $M$  次の線形予測係数  $a(m)$  を用いて,

$$S(z) = \frac{1}{A(z)} \quad (22)$$

但し,

$$A(z) = 1 + \sum_{m=1}^M a(m)z^{-m} \quad (23)$$

により定義される. 聴覚重み付けは, マスキング効果によって聴覚上の雑音低減をはかる操作であり, そのフィルタには

$$S_w(z) = \frac{A(z/\delta_1)}{A(z/\delta_2)} \quad (24)$$

のように, ARMA 型の伝達関数を用いられる [18].

適応符号帳は, 音声の周期性を表現するために用いられ, 過去の励振源が蓄えられている. 一方, 雑音符号帳は, 適応符号帳では表現しきれない音声の変動部分の表現に用いられる. また, 各符号帳のベクトルは,

A-b-S (Analysis-by-Synthesis) 法による閉ループ探索により決定され、適応符号帳、雑音符号帳の順に探索が行れた後、ゲインをベクトル量子化で決定する。この場合、雑音符号帳ベクトルを適応符号帳ベクトルに直交化して雑音符号帳の探索を行うと、最終的にひずみが小さい雑音符号帳ベクトルを選択することができる [4]。

復号器では、伝送されてくる情報から励振源を生成し、これを合成フィルタで駆動し、得られた符号化音声信号をポストフィルタに通す。ポストフィルタリングはスペクトルのホルマントの部分強調することにより、符号化音声の聴覚的な品質を向上させる操作である。ポストフィルタは、ホルマントポストフィルタ  $A(z/\delta_3)/A(z/\delta_4)$  と傾き補正フィルタ  $(1 - \mu z^{-1})$  の縦続接続により

$$S_p(z) = \frac{A(z/\delta_3)}{A(z/\delta_4)}(1 - \mu z^{-1}) \quad (25)$$

で実現される [19]。また、ポストフィルタリング後の信号のパワーは、元の信号のそれと異なるため、パワーの正規化が行われる。

#### 4. メル一般化ケプストラム分析による CELP 符号化方式

メル一般化ケプストラム分析では、式 (2) のスペクトルモデルを  $(\alpha, \gamma)$  によって任意に選択することができるが、符号化系を実現するにはこれらの値を固定する必要がある。以下では、文献 [9] や [14] などの結果により、 $(\alpha, \gamma) = (0.31, -1/3)$  として符号化系を実現する。

##### 4.1 構成

2. で述べたメル一般化ケプストラム分析法とその合成フィルタを用い、図 3 と同じ枠組で、CELP 符号化を実現することができる。合成フィルタは、

$$S(z) = \left\{ \frac{1}{C(\tilde{z})} \right\}^3 \quad (26)$$

のように、 $1/C(\tilde{z})$  の 3 段縦続構成で与えられる。聴覚重み付けフィルタとポストフィルタの構成はいくつか考えられるが、ここではそれぞれ

$$S_w(z) = \frac{C(\tilde{z})C(\tilde{z})}{C(\tilde{z}/\beta_1)} \quad (27)$$

$$S_p(z) = \frac{C(\tilde{z}/\beta_2)}{C(\tilde{z})}(1 - \mu z^{-1})^p \quad (28)$$

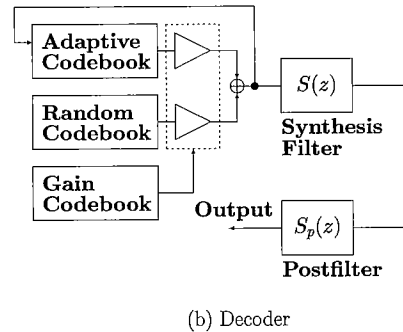
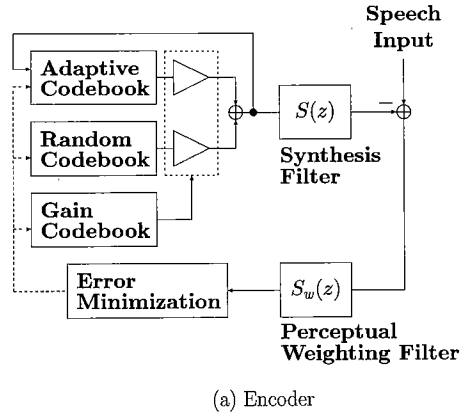


図 3 CELP 符号化方式の構成  
Fig. 3 Structure of a CELP coding system.

と定義する。式 (27) の聴覚重み付けフィルタでは、分子項を  $C(\tilde{z})$  の 2 段構成としている。これは、 $\gamma = -1/3$  の場合には  $1/C(\tilde{z})$  の表す対数スペクトルが合成フィルタ  $S(z)$  の対数スペクトルの  $1/3$  倍に相当しているため、分子項が  $C(\tilde{z})$  の 1 段では、4 kbit/s 程度の低ビットレートで聴覚重み付けの効果が十分に得られないことによる。

式 (27) と (28) のフィルタで共通に用いられている  $\tilde{z}$  領域での帯域幅拡張の操作を 4.2 で述べる。また、4.3 ではポストフィルタの傾き補正法について説明する。

##### 4.2 $\tilde{z}$ 領域での帯域幅拡張

従来の帯域幅拡張の操作は、式 (24) や (25) の  $A(z/\beta)$  のように  $z$  を  $z/\beta$  と置き換えるものである。周波数領域ではこの操作は、 $A(e^{j\omega})$  と  $1/(1 - \beta e^{-j\omega})$  の畳み込みと解釈することができる。これに対して、式 (27), (28) では、 $\tilde{z}$  を  $\tilde{z}/\beta$  と置き換えることにより、帯域幅拡張を行っている。これを  $\tilde{z}$  領域での帯域幅拡張とよぶことにする。この場合、非直線周波数軸上で畳み込

みの操作が行われることから、従来の帯域幅拡張の操作と比較して低周波数域のスペクトルの情報がよく保存されることになる。一方、 $C(\tilde{z})$  を  $z$  に対して帯域幅の拡張を行うことも可能である。しかし、この場合のフィルタの構成は図 2 において各遅延に  $\beta$  という乗算器を付加したものとなり、乗算回数が増加する。

フィルタ  $C(\tilde{z})$  を  $\tilde{z}$  領域で帯域幅拡張した  $C'(\tilde{z}/\beta)$  は

$$C'(\tilde{z}/\beta) = 1 + \gamma \tilde{z}^T F c' \quad (29)$$

但し、

$$F = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \beta & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \beta^M \end{bmatrix} \quad (30)$$

と表すことができる。式 (29) を式 (14)、(15) と同様に変形していくと、

$$C'(\tilde{z}/\beta) = 1 + \gamma \tilde{z}^T A A^{-1} F c' \quad (31)$$

$$= 1 + \gamma \Phi^T b'_\beta \quad (32)$$

但し、

$$b'_\beta = A^{-1} F c' = [b'_\beta(0), \dots, b'_\beta(M)]^T \quad (33)$$

となる。通常、 $b'_\beta(0) \neq 0$  であり、 $C'(\tilde{z}/\beta)$  のゲインは 1 にならないことから、式 (5) と同様にして式 (32) からゲインを括り出すと、

$$C'(\tilde{z}/\beta) = K_\beta \cdot C(\tilde{z}/\beta) \quad (34)$$

但し、

$$K_\beta = 1 + \gamma b'_\beta(0) \quad (35)$$

$$C(\tilde{z}/\beta) = 1 + \gamma \Phi^T b_\beta \quad (36)$$

また、

$$b_\beta = [0, b_\beta(1), \dots, b_\beta(M)]^T \quad (37)$$

となる。ここで、係数  $b'_\beta$  と  $b_\beta$  との関係は次式で与えられる。

$$b_\beta(m) = \frac{b'_\beta(m)}{1 + \gamma b'_\beta(0)}, \quad 1 \leq m \leq M \quad (38)$$

以上を整理すると、 $\tilde{z}$  領域での帯域幅拡張の操作は、係数  $c'$  から式 (33) の行列演算  $A^{-1} F c'$ 、若しくは次の再帰式

$$b'_\beta(m) = \begin{cases} \beta^M c'(M), & m = M \\ \beta^m c'(m) - \alpha b'_\beta(m+1), & 0 \leq m < M \end{cases} \quad (39)$$

により係数  $b'_\beta$  を計算し、式 (38) を用いて  $b_\beta$  を求める。係数  $b_\beta$  を図 2 のフィルタ  $C(\tilde{z})$  の係数  $b$  と置き換えることにより、 $C(\tilde{z}/\beta)$  を  $C(\tilde{z})$  と全く同じ構成で実現することができる。更に、式 (33) は

$$b'_\beta = A^{-1} F A A^{-1} c' \quad (40)$$

$$= A^{-1} F A b \quad (41)$$

と変形することも可能であり、係数  $b$  から行列  $A^{-1} F A$  を介して  $b_\beta$  を求めることも可能である。

### 4.3 メルケプストラム領域でのポストフィルタの傾き補正

傾き補正フィルタは、ホルマントポストフィルタのローパス効果を打ち消すもので、式 (25) のように 1 次全零型フィルタが用いられる。その係数  $\mu$  には、固定値を用いるもの [19]、反射係数の 1 次項  $k_1$  を利用するもの [20] などが提案されている。後者の傾き補正は、ホルマントポストフィルタがローパス効果をもつ場合にはハイパス特性、ハイパス効果をもつ場合にはローパス特性をもち、ホルマントポストフィルタの傾きの大きさに応じて適応的に傾きを制御することができる。一方、筆者らは文献 [7] などでメルケプストラム係数によって定義されるポストフィルタにおいて、対数スペクトルの大域的な傾きを表すメルケプストラムの 1 次項を 0 とする適応的な傾き補正を提案している。

ここでは、メル一般化ケプストラムにより定義される式 (28) のポストフィルタ  $S_p(z)$  に対して、ポストフィルタ全体のメルケプストラムの 1 次項が 0 となるように、 $\mu$  を適応的に制御する手法を提案する。この手法では、ホルマントポストフィルタのメルケプストラムの 1 次項を  $\tilde{c}_f$ 、傾き補正フィルタのメルケプストラムの 1 次項を  $\tilde{c}_\mu$  として、条件  $\tilde{c}_f + \tilde{c}_\mu = 0$  を満たす  $\mu$  を求める。メルケプストラムの 1 次項  $\tilde{c}_f$  と  $\tilde{c}_\mu$  は、周波数・一般化対数変換の再帰式 [21] を用いてそれぞれ

$$\tilde{c}_f = -\gamma(1 - \beta_2)c'_q(1) \quad (42)$$

$$\tilde{c}_\mu = -\frac{(1 - \alpha^2)\mu}{1 - \alpha\mu} p \quad (43)$$

となり、条件  $\tilde{c}_f + \tilde{c}_\mu = 0$  のもとで  $\mu$  は

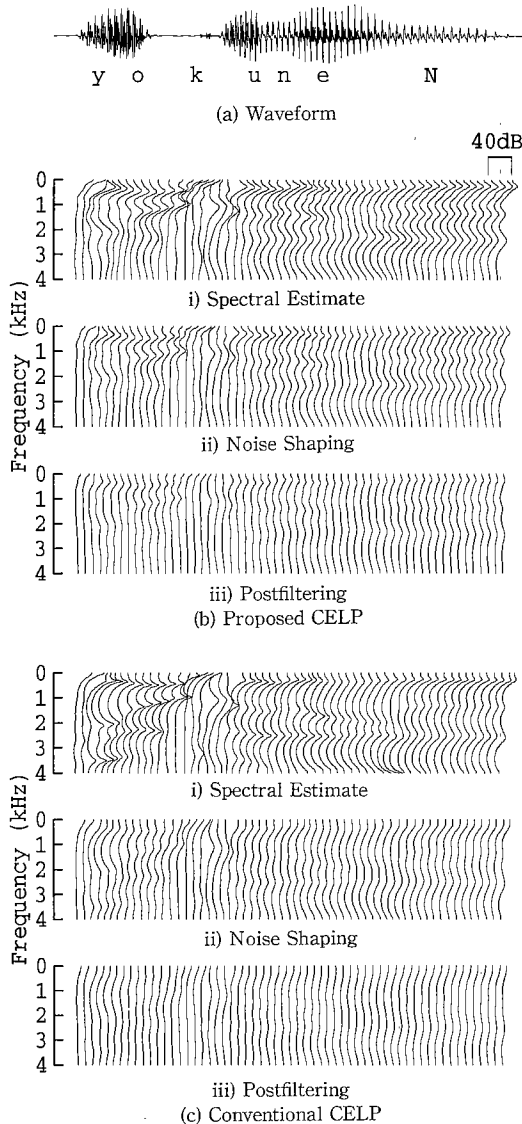


図4 聴覚重み付け, ポストフィルタリングのスペクトル  
Fig.4 Spectra of perceptual weighting and postfiltering.

$$\mu = \frac{-\gamma(1 - \beta_2)c'_q(1)}{-\alpha\gamma(1 - \beta_2)c'_q(1) + (1 - \alpha^2)p} \quad (44)$$

により与えられる。但し、 $c'_q(1)$  は  $c'(1)$  を量子化したものである。

また、式(28)の傾き補正フィルタは、従来のものを  $p$  段縦続としている。これにより、従来の1次の全零フィルタによる傾き補正と文献[20]で提案されている1次の全極フィルタによる傾き補正との中間的な特性

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

サンプリング周波数	8 kHz
分析次数	$M = 10$
分析窓	ハミング窓
分析窓長	32 ms
分析周期	20 ms

表2 ビット割り当て  
Table 2 Bit allocations.

	サブフレーム (80 サンプル)	フレーム (160 サンプル)
スペクトル情報	-	24
適応符号帳	8	16
雑音符号帳	11	22
ゲイン符号帳	9	18
合計	-	80

を得ることができる。

#### 4.4 聴覚重み付け・ポストフィルタリングの様子

提案符号化系の聴覚重み付けとポストフィルタリングのスペクトルを図4(b)に示す。各パラメータは予備的な試験により、 $(\beta_1, \beta_2, p) = (0.7, 0.0, 2)$  としている。比較のため、従来法のスペクトルを図4(c)に示した。従来法の聴覚重み付けフィルタとポストフィルタには、それぞれ式(24)と(25)を用い、各パラメータは文献[18]や[19]などを参考に、 $(\delta_1, \delta_2, \delta_3, \delta_4) = (0.9, 0.4, 0.5, 0.8)$  とした。傾き補正は反射係数によるものである。図4より、メル一般化ケプストラム分析によって推定される音声スペクトルは、線形予測法によるものと比較して、スペクトルの極だけでなく零も表現することができ、特に低周波数域での分解能が高くなっていることがわかる。また、提案符号化系の聴覚重み付け(図中では、ノイズシェイピングとして描かれている)とポストフィルタリングのスペクトルは、従来のものよりも低周波数域のスペクトル情報がよく保存されていることが読み取れる。

### 5. 主観評価実験

この章では、同一の符号化条件のもとで従来の線形予測に基づく CELP 符号化系と提案するメル一般化ケプストラム分析に基づく CELP 符号化系とを主観的に比較し、提案法の有効性を検討する。

#### 5.1 実験条件

フレーム長は 20 ms で、各フレームは二つの 10 ms サブフレームから構成されている。表1には音声の分析条件を、表2には符号化系のビット割り当てを示す。

表3 主観評価の実験条件  
Table 3 Conditions of subjective evaluation.

音声資料	文音声 (男性 6, 女性 6)
レファレンス信号	原音声 (14 bit linear)
テスト信号	従来法による符号化音声 提案法による符号化音声
MNR 信号	$Q=12\sim 40$ dB (4 dB ごと)
評価方法	オピニオン法 (5 段階)
繰り返し	2 回
被験者	6 名

ビットレートは 4 kbit/s である。

### 5.1.1 スペクトル情報の符号化

スペクトルパラメータとして従来法は LSP 係数 [22] を、提案法はメル一般化ケプストラムを文献 [23] で提案されているメル周波数領域のパラメータに変換し、これを量子化する。スペクトルパラメータの量子化において、LSP には重み付きユークリッド距離 [24] を、提案法ではユークリッド距離を用いた。符号帳には各段 12 bits の 2 段ベクトル量子化 [25] を採用し、LBG アルゴリズム [26] により符号帳を学習した。

### 5.1.2 励振符号帳とゲイン符号帳

適応符号帳は、非整数ラグ [27] を含む 8 ビット 256 状態を表現する。雑音符号帳には、各符号帳ベクトルを巡回シフトして使用するオーバーラップ型を採用し、ベクトル情報に 7 bits、シフト情報に 3 bits、極性に 1 bit を割り当てる。ゲイン符号帳は、適応ゲインと雑音ゲインの 2 次元で構成されている。また、雑音符号帳とゲイン符号帳を一般化ロイドアルゴリズムを用いて交互に 10 回学習している [28]。

### 5.1.3 聴覚重み付けとポストフィルタ

従来法、提案法ともに図 4 の場合と同じ聴覚重み付けフィルタとポストフィルタが用いられている。また、量子化した係数により聴覚重み付けフィルタを構成し、ハーモニク重み付けフィルタ [29] を聴覚重み付けフィルタと併用している。

## 5.2 結果

主観評価実験の条件を表 3 にまとめる。被験者は成人男性 6 名である。図 5 に等価  $Q$  値による主観評価実験の結果を示す。この図から、提案法は男女ともに従来法に対して性能が改善されており、平均で約 2 dB 程度の性能向上を達成した。この性能改善の要因を文献 [30] などの結果を踏まえて検討すると、聴覚重み付け、ポストフィルタリングが聴覚特性を考慮したパラメータであるメル一般化ケプストラムにより定義されているため、それらが主観品質に大きく貢献している

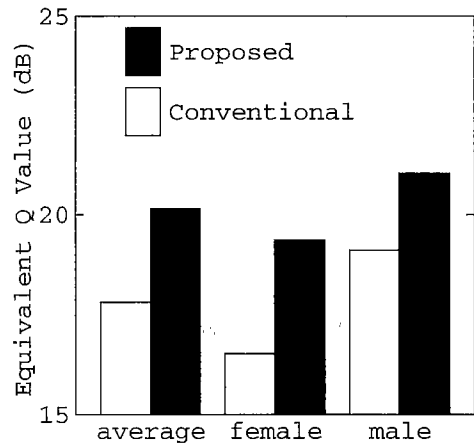


図5 主観評価実験の結果  
Fig.5 Results of subjective evaluation.

ものと考えられる。

## 5.3 検討

ここでは、フィルタ係数の計算と励振符号帳の探索に関わる演算量について考察する。インパルス応答に基づいて符号帳を探索する場合、その演算量は従来法、提案法ともに等しい。各サブフレームにおいて重み付き合成フィルタを直接駆動してインパルス応答を求めると、従来法は乗算  $3ML$  回と加算  $3ML$  回、提案法は乗算  $(4M+2)L$  回と加算  $(6M-2)L$  回の演算量となる ( $L$  はサブフレーム長)。また、スペクトルパラメータから予測係数 (従来法の場合  $a(m)$ , 提案法の場合  $c'(m)$ ) を求める演算量は等しく、予測係数からフィルタ係数の計算 (聴覚重み付けフィルタのための帯域幅拡張の操作など) には、従来法は  $2M$  回の乗算に対して、提案法は  $5M+1$  回の乗算、 $2M+2$  回の加算と 2 回の除算が必要となる。以上、フィルタ係数やインパルス応答の計算にかかる演算は、提案法は従来法の 2 倍程度になるが、符号帳探索の演算量はそれよりもはるかに大きいため、これらの演算量は提案法の方がやや多くなると考えることができる。

一方、提案法の符号化処理の中で、音声分析に関わる部分が従来法に対して最も演算量の増加が見込まれる [9]。しかし、予測係数からスペクトルパラメータへの変換では従来法の半分程度の演算量、スペクトルパラメータの符号帳探索では、従来法に比べ乗算回数が少なくなる (ユークリッド距離に重みを用いていないため) など、提案法の方が演算量の少ない処理もある。

また、本論文の主観評価実験では、従来法として線

形予測に基づく CELP を取り上げたが、CELP の枠組では、メルケプストラム分析や対数スペクトルによる不偏推定法(ケプストラム)よりもメル一般化ケプストラム分析が有効であることを確認している。

## 6. む す び

メル一般化ケプストラム分析を用いた CELP 音声符号化系を提案した。提案符号化系では、合成フィルタ・聴覚重み付けフィルタ・ポストフィルタをメル一般化ケプストラムを介して実現できるという特徴をもつ。ビットレート 4 kbit/s の符号化条件のもとで、従来の線形予測に基づく CELP 符号化系と提案符号化系を主観的に比較した。その結果、提案法は従来法に対して主観的性能が 2 dB 以上向上したことから、メル一般化ケプストラム分析に基づく CELP 符号化の有効性を確認した。

今後は、スペクトル情報、励振源などの符号化をより効率化し、ビット配分などの検討を加えて、符号化音声品質の更なる向上が課題となる。また、IRS 音声に対する評価も行う必要がある。

**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] M.R. Schroeder and B.S. Atal, "Code-excited linear prediction (CELP): high quality speech at very low bit rates," Proc. ICASSP-85, Tampa, USA, pp.937-940, March 1985.
- [2] J.H. Chen, R.V. Cox, Y.C. Lin, N. Jayant, and M.J. Melchner, "A low-delay CELP coder for the CCITT 16 kb/s speech coding standard," IEEE J. Sel. Areas Commun., vol.10, no.5, pp.830-849, June 1992.
- [3] R. Salami, C. Laflamme, J-P. Adoul, A. Kataoka, S. Hayashi, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Description of the proposed ITU-T 8 kb/s speech coding standard," IEEE Speech Coding Workshop, Annapolis, USA, Sept. 1995.
- [4] I.A. Gerson and M.A. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8 kbps," Proc. ICASSP-90, Albuquerque, USA, pp.461-464, April 1990.
- [5] 三樹 聡, 守谷健弘, 間野一則, 大室 伸, "ピッチ同期雑音励振源をもつ CELP 符号化 (PSI-CELP)," 信学論 (A), vol.J77-A, no.3, pp.314-324, March 1994.
- [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] 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.
- [8] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "CELP coding based on mel-cepstral analysis," Proc. ICASSP-95, Detroit, USA, pp.33-36, May 1995.
- [9] 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.
- [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), vol.J70-A, no.3, pp.471-480, March 1987.
- [11] G. Fant, "Speech sound and features," MIT Press, Cambridge, 1973.
- [12] 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.
- [13] 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.
- [14] K. Tokuda, T. Kobayashi, and S. Imai, "Generalized cepstral analysis of speech—unified approach to LPC and cepstral method," Proc. ICSLP-90, Kobe, Japan, pp.37-40, Nov. 1990.
- [15] 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), vol.J68-A, no.6, pp.610-611, June 1985.
- [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] 徳田恵一, 小林隆夫, 深田俊明, 今井 聖, "音声の適応メルケプストラム分析," 信学論 (A), vol.J74-A, no.8, pp.1249-1256, Aug. 1991.
- [18] B.S. Atal and M.R. Schroeder, "Predictive coding of speech signals and subjective error criteria," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-27, pp.247-254, June 1979.
- [19] J.H. Chen and A. Gersho, "Real-time vector APC speech coding at 4800bps with adaptive postfilter," Proc. ICASSP-87, Dallas, USA, pp.2185-2188, April 1987.
- [20] W.B. Kleijn, D.J. Krasinski, and R.H. Ketchum, "Improved speech quality and efficient vector quantization in SELP," Proc. ICASSP-88, New York, USA, pp.155-158, April 1988.
- [21] 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.
- [22] 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.
- [23] 小石田和人, 徳田恵一, 小林隆夫, 今井 聖, "メル一般化ケプストラムによる音声のスペクトル表現とその諸特性," 信学技報, SP95-49, Sept. 1995.
- [24] 大室 伸, 間野一則, 守谷健弘, "LSP パラメータのベクトル・マトリクス量子化," 信学技報, SP91-70, Oct. 1991.
- [25] B.H. Juang and A.H. Gray, Jr., "Multiple stage vector quantization for speech coding," Proc. ICASSP-82, Paris, France, pp.597-600, April 1982.



- [26] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., vol.28, no.1, pp.84-95, Sept. 1980.
- [27] P. Kroon and B.S. Atal, "Pitch predictors with high temporal resolution," Proc. ICASSP-90, Albuquerque, USA, pp.661-664, April 1990.
- [28] 守谷健弘, 三樹 聡, 大室 伸, 間野一則, "CELP 符号化における励振符号帳の学習法," 信学論 (A), vol.J77-A, no.3, pp.485-493, March 1994.
- [29] I.A. Gerson and M.A. Jasiuk, "Techniques for improving the performance of CELP-type speech coders," IEEE J. Sel. Areas Commun., vol.10, no.5, pp.858-865, June 1992.
- [30] 徳田恵一, 小林隆夫, 今井 聖, "メルケプストラム分析に基づく適応ポストフィルタによる符号化音声の品質改善," 音響学会講論集, 3-8-12, March 1993.

(平成9年4月22日受付, 8月5日再受付)



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同大助教授。昭54同大教授。平9東工大名誉教授, 千葉工大電子工学科教授。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。



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。音声の高速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授。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, 日本神経回路学会, IEEE 各会員。