論　文

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

正　員　徳田　恵一[†]　　正　員　小林　隆夫[††]
正　員　深田　俊明[†††]　　正　員　今井　聖[††]

Speech Coding Based on Adaptive Mel-Cepstral Analysis and Its Evaluation

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

あらまし　短期予測のパラメータとしてメルケプストラムを用いるADPCM音声符号化系を提案する．提案符号化系では，適応メルケプストラム分析法に基づいてバックワード形の適応予測を行い，ノイズシェイピング，ポストフィルタリングなどの操作をメルケプストラムを介して行っているため，ノイズシェイピング，ポストフィルタリングの効果は人間の聴覚特性に合ったものとなっていることが期待できる．ここでは，特にピッチ予測器の付加などの改良を施した符号化系の構成を示し，符号化音声の品質評価を行っている．その結果，提案する符号化系はアルゴリズム遅延がないにもかかわらず，16 kbit/s で 32 kbit/s ADPCM 相当の音声品質をもつことが示された．

キーワード　メルケプストラム，音声符号化，適応メルケプストラム分析，聴覚特性

1. まえがき

従来の音声符号化方式では，線形予測法[(1)]に基づいた適応予測器が広く用いられている．しかし，線形予測法では，全極モデルで音声スペクトルを表現するため，鼻音などの零点を含んだ音声スペクトルを正確に予測することは困難であると考えられる．これに対して，ケプストラム[(2)]をパラメータとしてスペクトルを表現すれば，極と零を同等に表現することができる．更に，低い周波数で細かい分解能を，高い周波数で粗い分解能をもつという人間の聴覚特性を考慮したメルケプストラム[(3)]をパラメータとしたスペクトルモデルによる予測器が実現できれば，音声符号化の分野において有効なものになると予想される．実際に，これまでケプストラムを利用する音声符号化法としてCELPに基づく手法[(4)]が，また人間の聴覚特性を表す線形周波数軸を近似するように周波数変換されたARモデルを用いたADPCM系[(5)]が，それぞれ提案されている．

本論文では，以上のような観点から，音声の適応分析として有効な手法の一つである適応メルケプストラム分析[(6)]に基づいた適応予測器を提案し[(7)]，メルケプストラムを利用する音声符号化法について検討を行う．この符号化系では，ケプストラムを利用するCELP符号化法[(4)]と異なり，(1) メルケプストラムをパラメータとした適応分析に基づいてADPCMの枠組みを利用している，(2) 予測誤差エネルギーを最小化するようにメルケプストラム係数を決定している，(3) MLSAフィルタを利用することにより，IIRシステムとして合成フィルタを実現している等の特徴をもっている．更に，ノイズシェイピング，ポストフィルタリングをメルケプストラムを介して行うことができるため，聴覚的な性能の向上が期待できる．

本論文では，特にピッチ予測器，およびピッチ予測に基づくノイズシェイピング，ポストフィルタリングを付加した符号化系[(8)]を示し，品質評価実験によりこの符号化系の有効性を示す．

以下，2. で適応メルケプストラム分析アルゴリズムおよびその実現法について簡単に述べた上で，3. で適応メルケプストラム分析に基づいてバックワード形の適応予測を行うADPCM符号化系を示す．4. では符号化音声の品質評価を行った結果を示している．

† 東京工業大学工学部，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
††† キヤノン株式会社情報システム研究所，川崎市
Information and Systems Laboratories, Canon Incorporated, Kawasaki-shi, 211 Japan

2. 適応メルケプストラム分析

2.1 適応メルケプストラム分析アルゴリズム

メルケプストラム分析法[9]では，M 次までのメルケプストラム $\tilde{c}(m)$ によって表されたスペクトルモデル

$$D(z)=\exp\sum_{m=0}^{M}\tilde{c}(m)\tilde{z}^{-m} \tag{1}$$

但し，

$$\tilde{z}^{-1}=\frac{z^{-1}-\alpha}{1-\alpha z^{-1}} \tag{2}$$

を考え，対数スペクトルの不偏推定法[10]のスペクトル評価を適用している．$\tilde{z}^{-1}=e^{-j\tilde{\omega}}$ の位相特性 $\tilde{\omega}$ は，例えば標本化周波数が 8kHz のとき α を 0.31 に選べば，メル尺度[11]をよく近似する[3]ことから，このスペクトルモデルは人間の聴覚特性と同様の周波数分解能をもつと言える．評価関数の最小化は，$D(z)$ のゲイン，つまりインパルス応答の時刻 0 の値が 1 という条件のもとで，

$$\varepsilon=E\left[e^2(n)\right] \tag{3}$$

を最小化することに等価であることが示される[9]．但し，$e(n)$ は信号 $x(n)$ を逆フィルタ $1/D(z)$ に通したときの出力とする．

ここで，$D(z)$ を

$$D(z)=\exp\sum_{m=0}^{M}b(m)\,\Phi_m(z) \tag{4}$$

但し，

$$\tilde{c}(m)=\begin{cases} b(M), & m=M \\ b(m)+\alpha b(m+1), & 0\leqq m\leqq M-1 \end{cases} \tag{5}$$

$$\Phi_m(z)=\begin{cases} 1, & m=0 \\ \dfrac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}}\,\tilde{z}^{-(m-1)}, & m\geqq 1 \end{cases} \tag{6}$$

のように変形することを考える．$\{b(m)\}_{m=0}^{M}$ は $\{\tilde{c}(m)\}_{m=0}^{M}$ を線形変換したものであることから，式(3)の $\{\tilde{c}(m)\}_{m=0}^{M}$ に関する最小化は $\{b(m)\}_{m=0}^{M}$ に関する最小化に等価である．更に，$D(z)$ のゲインは $b(0)$ のみに依存していることから，$D(z)$ を

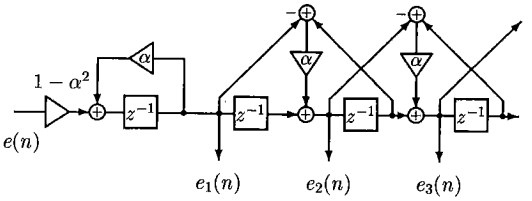


図1 フィルタ $\Phi_m(z)$
Fig. 1 Filter $\Phi_m(z)$.

$$D(z)=\exp\sum_{m=1}^{M}b(m)\,\Phi_m(z) \tag{7}$$

と置き換えれば，$D(z)$ のゲインは常に 1 となり，式(3)の条件付き最小化は，$\boldsymbol{b}=[b(1),b(2),\ldots,b(M)]^T$ に関する最小化に帰着される[9]．なお，$D(z)$ からゲインをくくり出すための操作は一意ではなく，ここで示した方法以外にもいくつかが考えられる[3]．

ε は $\boldsymbol{b}$ に関して凸であることが示される[9]ので，最急降下法を適用して，この最小化問題を解くことができる．つまり，i 番目の近似値を $\boldsymbol{b}^{(i)}$ として，$i+1$ 番目の近似値 $\boldsymbol{b}^{(i+1)}$ を

$$\boldsymbol{b}^{(i+1)}=\boldsymbol{b}^{(i)}-\mu^{(i)}\nabla\varepsilon \tag{8}$$

で得る．ε のこう配 $\nabla\varepsilon$ は

$$\nabla\varepsilon=-2\cdot E\left[e(n)\,\boldsymbol{e}_{\Phi}^{(n)}\right] \tag{9}$$

で与えられる．但し，

$$\boldsymbol{e}_{\Phi}^{(n)}=[e_1(n),e_2(n),\ldots,e_M(n)]^T \tag{10}$$

であり，$e_m(n)$ は，$e(n)$ 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[12]，適応ケプストラム分析法[13]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 $\hat{\nabla}\varepsilon^{(n)}$

$$\hat{\nabla}\varepsilon^{(n)}=-2\,e(n)\,\boldsymbol{e}_{\Phi}^{(n)} \tag{11}$$

を考え[6]，これに指数減衰窓を掛けた

$$\hat{\nabla}\varepsilon_{\tau}^{(n)}=-2\,(1-\tau)\sum_{i=-\infty}^{n}\tau^{n-i}\,e(i)\,\boldsymbol{e}_{\Phi}^{(i)} \tag{12}$$

によって式(9)のこう配 $\nabla\varepsilon$ を推定する[14]．ここで，τ は $0\leqq\tau<1$ の定数である．式(12)で $\hat{\nabla}\varepsilon_{\tau}^{(n)}$ が与えられるとき，これは時刻 $n-1$ における推定値 $\hat{\nabla}\varepsilon_{\tau}^{(n-1)}$ から

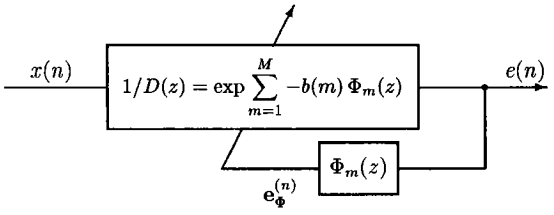


図2 適応メルケプストラム分析のブロック図
Fig. 2 Block diagram of the adaptive mel-cepstral analysis.

$$\hat{\nabla}\varepsilon_{\tau}^{(n)} = \tau\hat{\nabla}\varepsilon_{\tau}^{(n-1)} - 2(1-\tau)\,e(n)\boldsymbol{e}_{\Phi}^{(n)} \tag{13}$$

で得ることができる．このとき，時刻 n から $n+1$ への係数 $\boldsymbol{b}$ の更新アルゴリズムは，

$$\boldsymbol{b}^{(n+1)} = \boldsymbol{b}^{(n)} - \mu^{(n)}\,\hat{\nabla}\varepsilon_{\tau}^{(n)} \tag{14}$$

但し，

$$\mu^{(n)} = \frac{a}{M\,\varepsilon^{(n)}} \tag{15}$$

$$\varepsilon^{(n)} = \lambda\varepsilon^{(n-1)} + (1-\lambda)\,e^2(n) \tag{16}$$

で与えられる[6]．ここで，$\varepsilon^{(n)}$ は，時刻 n における ε の推定値であり，λ は $0<\lambda<1$, a は $0<a<1$ の定数である．以上の適応分析系の構成を図2に示す．

2.2 指数形伝達関数 $D(z)$ の実現と安定性

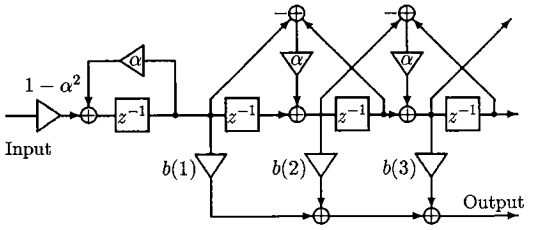
実際に適応メルケプストラム分析を行うには，図2に示されるように，指数形の伝達関数をもつ逆フィルタ $1/D(z)$ が必要であるが，次節に示す符号化系では $1/D(z)$ ではなく $D(z)$ あるいは $D(z)-1$ を用いることになるため，まず $D(z)$ を MLSA フィルタ[3]により実現することを考える．

すなわち，複素指数関数 $\exp w$ を L 次有理式 $R_L(w)$ で近似することにより，

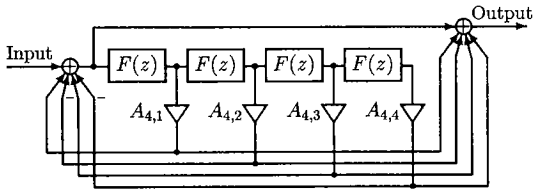
$$D(z) = \exp F(z) \simeq R_L(F(z)) = \frac{1+\sum_{l=1}^{L} A_{L,l}\,\{F(z)\}^l}{1+\sum_{l=1}^{L} A_{L,l}\,\{-F(z)\}^l} \tag{17}$$

但し，

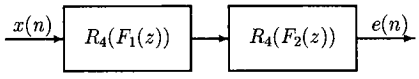
$$F(z) = \sum_{m=1}^{M} b(m)\,\Phi_m(z) \tag{18}$$



(a) Basic filter $F(z)$



(b) $R_L(F(z)) \simeq D(z) \quad L=4$



(c) Tow-stage cascade structure
$R_4(F_1(z)) \cdot R_4(F_2(z)) \simeq D(z)$

図3 指数形伝達関数 $D(z)$ の実現
Fig. 3 Realization of the exponential transfer function $D(z)$.

と $D(z)$ を実現する．$F(z)$ の構成を図3 (a) に，$L=4$ としたときの MLSA フィルタ $R_L(F(z))$ の構成を図3 (b) に示しておく．$F(z)$ がディレイフリーパスをもたないことから，MLSA フィルタ $R_L(F(z))$ はディレイフリーループをもたない．

ところで，$F(z)$ の極はすべて $z=\alpha\ (=0.31)$ にあるので，$F(z)$ は安定である．従って，$b(m)$ が有限の値をとるならば，$\left|F(e^{j\omega})\right|$ の値も有限であり，

$$\left|F(e^{j\omega})\right| \leqq r \tag{19}$$

となるような定数 r を選ぶことができる．このとき，$|w| \leqq r$ における $\exp w$ と $R_L(w)$ との対数近似誤差 $|\log\exp w - \log R_4(w)|$ の最大値を最小化するように，$A_{L,l},\ l=1,2,\ldots,L$ を最適化することができる．文献 (13) では，$r=4.5$, $L=4$ とした場合に最適化された $A_{L,l},\ l=1,2,\ldots,L$ の値が示されている．また，この係数値を用いた場合の，$|w| \leqq r$ における対数近似誤差の最大値は 0.24 dB であることが示されている．従って，

$$\left|F(e^{j\omega})\right| \leqq r = 4.5 \tag{20}$$

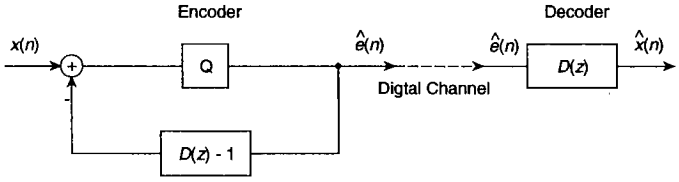


図4 符号化系の基本構成
Fig. 4 Basic structure of the coder.

の条件のもとで，対数近似誤差 $|\log D(e^{j\omega}) - \log R_4(F(e^{j\omega}))|$ は 0.24 dB 以下となることがわかる．また，文献 (13) において，$R_4(w)$ の w 平面上で極および零点の最小半径は 6.2 より大きいことが示されている．このことから，$|F(z)| \leqq 6.2$ の条件のもとでは，$R_4(F(z))$ は極および零点をもち得ないことがわかる．従って，$R_4(F(z))$ が z 平面上の単位円外（単位円を含む）に極も零点ももたない条件は，

$$|F(z)| \leqq r_{max} = 6.2, \qquad |z| \geqq 1 \tag{21}$$

となる．この条件は，$F(z)$ が単位円内にしか極をもたないことから，z 平面上の単位円外（単位円を含む）を閉領域として最大値の原理を適用すれば，

$$\left|F(e^{j\omega})\right| \leqq r_{max} = 6.2 \tag{22}$$

に等価である．つまり，式 (22) の条件のもとで，$R_4(F(z))$ は安定，かつ最小位相となる．

ここで基礎フィルタの伝達関数 $F(z)$ を

$$F(z) = F_1(z) + F_2(z) \tag{23}$$

と和の形に分解することを考える．このとき指数形伝達関数は

$$\begin{aligned} D(z) &= \exp F(z) = \exp F_1(z) \cdot \exp F_2(z) \\ &\simeq R_L(F_1(z)) \cdot R_L(F_2(z)) \end{aligned} \tag{24}$$

と縦続接続構成とすることができ，$\left|F_1(e^{j\omega})\right|$, $\left|F_2(e^{j\omega})\right|$ の最大値がそれぞれ $\left|F(e^{j\omega})\right|$ 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．$F_1(z)$, $F_2(z)$ の分解の仕方はいくつかあるが，以下では，$F(z)$ を

$$F_1(z) = b(1)\,\Phi_1(z) \tag{25}$$

$$F_2(z) = \sum_{m=2}^{M} b(m)\,\Phi_m(z) \tag{26}$$

と分解し，それぞれ $L = 4$ として近似した図 3 (c) に示す構成で $D(z)$ を実現することにする．

実際に適応メルケプストラム分析を行う際に必要なフィルタ $1/D(z)$ は，式 (17) における $F(z)$ の符号を変えることで実現できる．このようにして実現された $1/D(z)$ も安定かつ最小位相となり，更に指数形伝達特性との近似誤差が 0.24 dB 以下となる．実際の音声信号では，$\left|F_1(e^{j\omega})\right|$, $\left|F_2(e^{j\omega})\right|$ の最大値は 4 程度であり，4.5 を超えないことが実験的に確かめられている[14]．

3. 符号化系の構成

3.1 基本構成

$D(z)$ あるいは $1/D(z)$ のゲイン，つまりインパルス応答の時刻 0 の値が規格化されているため，図 2 の $e(n)$ は予測残差とみなされる．従って，従来の線形予測法に替え，適応メルケプストラム分析法に基づいて音声符号化系のバックワード形適応予測器を構成することが可能である．適応メルケプストラム分析に基づく適応予測器は，メルケプストラムをパラメータとしていることから，人間の聴覚特性を考慮した適応予測器と考えることができる．

このような考えに基づいた符号化系の基本構成を図 4 に示す．$D(z)$ のインパルス応答の時刻 0 の値は 1 となるため，$D(z) - 1$ はディレイフリーパスをもたないことに注意する．$D(z)$ を MLSA フィルタで近似した場合の $D(z) - 1$ の具体的な構造は付録に示す．この基本構成では信号 $x(n)$，および量子化器 Q によって発生する量子化ノイズ $q(n)$ の z 変換をそれぞれ $X(z)$, $Q(z)$ とすると，復号器出力 $\hat{x}(n)$ の z 変換 $\hat{X}(z)$ は，

$$\hat{X}(z) = X(z) + Q(z) \tag{27}$$

となる．以下では，適応量子化器 Q は文献 (15) のものを用いた．$D(z)$ の係数 $b(m)$ は，バックワード適応によって，つまり $e(n)$ ではなく $\hat{e}(n)$ によって更新される．従って，この符号化系は，予測係数などのサイ

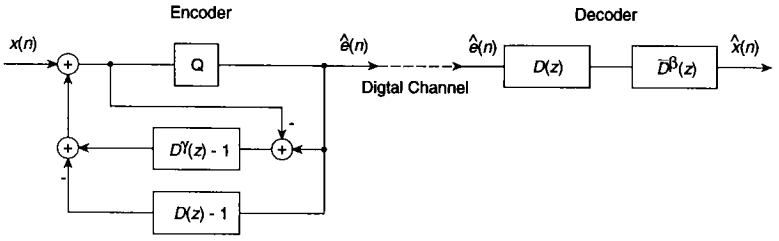


図5 ノイズシェイピング，ポストフィルタリング付き構成
Fig. 5 Structure with noise shaping and postfiltering.

ド情報を伝送する必要がない，フレーム単位の処理に起因する遅延を生じないなど，バックワード形特有の利点をもつ．

3.2 ノイズシェイピングとポストフィルタリング

図5にノイズシェイピング，ポストフィルタリングを付加した構成を示す．ここでは，ノイズノイズシェイピング，ポストフィルタリングとして，文献(16)なども参考に，それぞれ $D^{\gamma}(e^{j\omega})$, $\overline{D}^{\beta}(e^{j\omega})$ の形を考えている．つまり，信号 $x(n)$，および量子化器 Q によって発生する量子化ノイズ $q(n)$ の z 変換をそれぞれ $X(z), Q(z)$ とすると，復号器出力 $\hat{x}(n)$ の z 変換 $\hat{X}(z)$ は，

$$\hat{X}(z) = \{X(z) + D^{\gamma}(z)Q(z)\}\,\overline{D}^{\beta}(z) \tag{28}$$

となる．定数 γ, β はそれぞれノイズシェイピングおよびポストフィルタリングのためのパラメータであり，ノイズシェイピング，ポストフィルタリングとも行わないのが，$\gamma = 0, \beta = 0$ の場合である． $D^{\gamma}(z)$ は，式(1)において $\tilde{c}(m)$ を γ 倍した伝達関数であり，

$$D^{\gamma}(z) = \exp \sum_{m=1}^{M} \gamma b(m)\,\Phi_m(z) \tag{29}$$

のように $b(m)$ を γ 倍することにより，$D(z)$ と全く同様に実現することができる．また，$\overline{D}(z)$ は，式(7)において $\tilde{c}(1) = 0$ とおき，ゲインを1に正規化した伝達関数であり，

$$\overline{D}(z) = \exp \sum_{m=1}^{M} \overline{b}(m)\,\Phi_m(z) \tag{30}$$

$$\overline{b}(m) = \begin{cases} b(m), & 2 \leqq m \leqq M \\ -\alpha b(2), & m = 1 \end{cases} \tag{31}$$

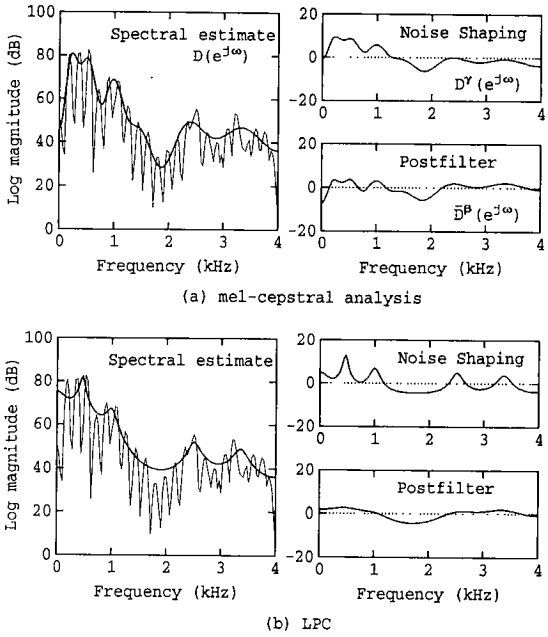


図6 ノイズシェイピング，ポストフィルタリングの効果
Fig. 6 Effect of noise shaping and postfiltering.

と書ける．従って，$\overline{D}^{\beta}(z)$ は，$\overline{b}(m)$ を β 倍することにより，

$$\overline{D}^{\beta}(z) = \exp \sum_{m=1}^{M} \beta\,\overline{b}(m)\,\Phi_m(z) \tag{32}$$

で実現することができる．このように，$\tilde{c}(1)$ を 0 に置き換えているのは，ポストフィルタリングによってスペクトルの大域的な傾きが強調され，符号化音声の音質が変化するのを防ぐためである．

$\gamma = 0.3, \beta = 0.3$ の場合のノイズシェイピング，ポストフィルタリングの様子を示したのが図6 (a)である．参考のため線形予測法に基づくノイズシェイピン

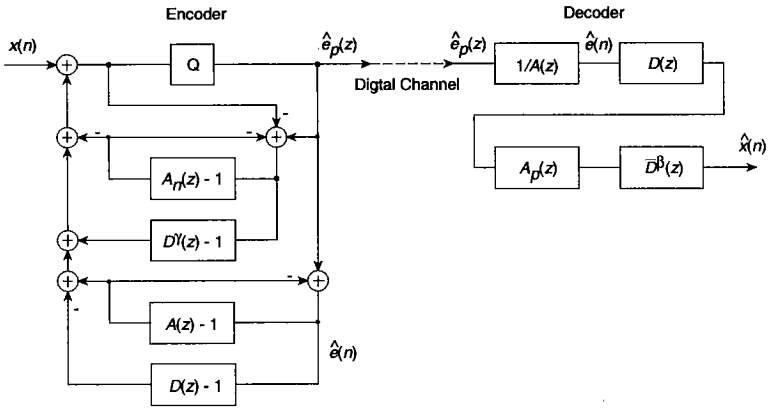


図7 ピッチ予測付き構成
Fig. 7 Structure with pitch predictor.

グ，ポストフィルタリングの様子を図6 (b) に示す†．ここでは，文献(17), (18) などをもとに，

$$\hat{X}(z) = \left\{ X(z) + \frac{B(z/0.85)}{B(z)} Q(z) \right\} \frac{B(z/0.5)}{B(z/0.8)} (1 - 0.5\, z^{-1}) \tag{33}$$

の形を用いた．但し，$B(z)$ を線形予測法によって得られる予測多項式とする．図からもわかるとおり，提案法では，低い周波数域ほど細かいスペクトル形状が保存されており，また対数スペクトルを定数倍する形の処理がされる．つまり，ノイズシェイピング，ポストフィルタリングいずれの操作も対数振幅-メル周波数軸上で行われており，人間の聴覚特性にあった処理がなされていることが期待できる．実際，ポストフィルタについては，$\alpha = 0.31$ とすることにより，$\alpha = 0$ とする場合に比べて聴感上大きな雑音低減効果をもつことが確かめられている[19]．

ポストフィルタリングを行った信号のパワーは，もとの信号のそれと異なるため，提案法でも文献(18)と同様のパワーの正規化を行っている．

3.3 ピッチ予測

図7にピッチ予測器を付加した符号化系の構成を示す．復号化器出力 $\hat{x}(n)$ の z 変換は，

$$\hat{X}(z) = \left\{ X(z) + \frac{D^{\gamma}(z)}{A_n(z)} Q(z) \right\} A_p(z)\, \overline{D}^{\beta}(z) \tag{34}$$

で与えられる．ここでは，

$$A(z) = 1 + \sum_{k=p-1}^{p+1} a(k)\, z^{-k} \tag{35}$$

の形をもつ3タップのピッチ予測器を用いた．ピッチ p およびピッチ予測係数 $a(k)$ は，再帰的指数形窓によって得られる $\hat{e}(n)$ の共分散から，各サンプルごとに求めている．但し，文献(20)の安定化法を適用し，また簡単な有声/無声判別により無声と判断されたサンプルでは $a(k) = 0,\ k = p-1, p, p+1$ としている．ピッチ予測に基づくノイズシェイピングおよびポストフィルタリングはそれぞれ文献(21), (22)を参考にしており，$A_n(z)$, $A_p(z)$ は

$$A_n(z) = 1 + \epsilon_n \sum_{k=p-1}^{p+1} a(k)\, z^{-k} \tag{36}$$

$$A_p(z) = \frac{\left(1 - \epsilon_p \sum_{k=p-1}^{p+1} a(k)\, z^{-k} \right)}{\left(1 - \epsilon_p \sum_{k=p-1}^{p+1} a(k) \right)} \tag{37}$$

で定義される．ϵ_n, ϵ_p はそれぞれピッチ予測に基づくノイズシェイピング，ポストフィルタリングに関するパラメータであり，$\epsilon_n = 0, \epsilon_p = 0$ がどちらも行わない場合にあたる．復号器側でも符号器と同様に $\hat{e}(n)$

† 図6 (a)，図6 (b) とも，概念図であり，符号化系に組み込まれた形で行われたバックワード形のスペクトル分析結果ではなく，入力音声を直接分析した結果に基づいて描かれている．図6 (a) では $M = 12$ のメルケプストラム分析[9]を，図6 (b) では10次の自己相関法を用いている．

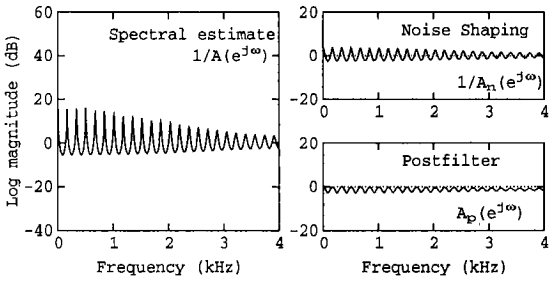


図8 ピッチ予測に基づくノイズシェイピング，ポストフィルタリングの効果

Fig. 8 Effect of noise shaping and postfiltering based on pitch predictor.

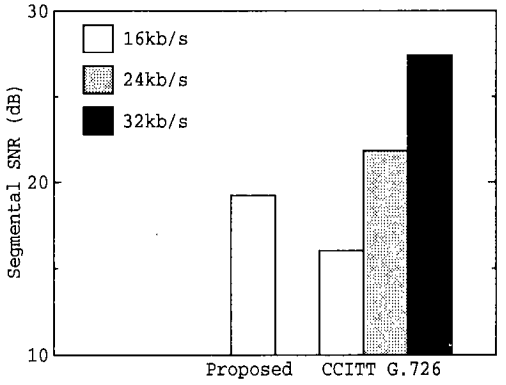


図9 セグメンタルSNRによる客観評価

Fig. 9 Objective performance assesment based on segmental SNR.

から $p, a(k)$ を求めており，つまりバックワード適応に基づくピッチ予測となっている．図8に $\epsilon_n = 0.4$, $\epsilon_p = 0.2$ としたときのピッチ予測に基づくノイズシェイピング，ポストフィルタリングの様子を示す．図6と図8は同じ音声データの同じ部分に対応しており，符号化系全体としてのノイズシェイピング，ポストフィルタリングは図6 (a)，図8を合わせたものとなる．

4. 音声品質評価

4.1 客観評価試験

図9にセグメンタルSNRによる評価結果を示す．男性話者5人，女性話者5人が発声した合計約40秒の音声資料（標本化周波数8kHz）を用いた．ノイズシェイピング，ポストフィルタリングを行った場合，セグメンタルSNRによる評価は劣化するため，ここでは $\beta, \gamma, \epsilon_n, \epsilon_p$ はすべて0としている．また適応メルケプストラム分析のパラメータは，$M = 12, a = 0.1, \lambda = 0.98, \tau = 0.96$ とした．式(17)の L は $L = 4$ とし，$A_{L,l}$ は文献(13)に示されたものを用いた．比較の基準とするため，CCITT G.726（高速適応モードに固定）による結果を併せて示した．図9より，16 kbit/sの提案法は，同ビットレートのCCITT G.726に対して，約3 dBの改善を達成していることがわかる．

4.2 主観評価試験

受聴試験は，被験者にレファレンス音（原音声）とテスト音を対にして聞かせ，得られたテスト音の5段階評価からMOS（mean opinion score）を算出する方法[23]をとった．各被験者はすべてのテスト音を順序を変えて3回受聴する．テスト音は，提案法（16 kbit/s）と，比較の基準とするためのCCITT G.726（16, 24, 32 kbit/s）による符号化音声の他に，等価Q値を計算するためのMNR信号（Q=12～48 dB）を加えた．但し，CCITT G.726は高速適応モードに固定とした．また，適応メルケプストラム分析のパラメータは**4.1**と同じである．表1に主観評価実験の諸元をまとめる．

表1 主観評価実験の諸元

音声資料	文音声（男声1, 女声1）
レファレンス信号	原音声（8 kHz sampling 14 bit linear）
テスト信号	提案法 16 kbit/s $(\gamma, \beta, \epsilon_n, \epsilon_p) = (0.0, 0.0, 0.0, 0.0)$ $(\gamma, \beta, \epsilon_n, \epsilon_p) = (0.3, 0.2, 0.3, 0.1)$ CCITT G.726 16, 24, 32 kbit/s MNR信号 $Q = 12, 18, 24, 30, 36, 42, 48$ dB
評価方法	オピニオン評価法（5段階）
被験者	5名（各人が順序をかえて3回）

受聴試験の結果を等価Q値によって表したものを図10に示す．図10の結果より，提案する符号化方式（16 kbit/s）は，ノイズシェイピング（$\gamma = 0.3$, $\epsilon_n = 0.3$），ポストフィルタ（$\beta = 0.2$, $\epsilon_p = 0.1$）を付加することにより，等価Q値で12 dB近い改善を達成しており，これらの効果が大きいことがわかる．非公式な受聴試験からは，特にメルケプストラムに基づくポストフィルタの効果が大きいことがわかっている．

CCITT G.726と比較した場合には，ノイズシェイピング，ポストフィルタリングなしの提案法（16 kbit/s）は，CCITT G.726（16 kbit/s）に対して等価Q値で約5dBの改善となっている．更にノイズシェイピング（$\gamma = 0.3$, $\epsilon_n = 0.3$），ポストフィルタリング（$\beta = 0.2$, $\epsilon_p = 0.1$）を付加したときには，CCITT G.726（16 kbit/s）に対して17 dB以上の改善を達成し

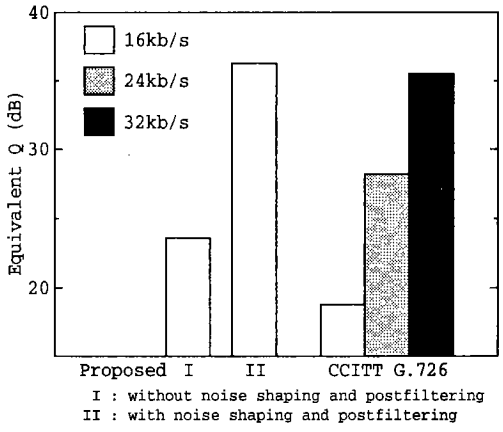


図10 等価 Q 値による主観評価

Fig. 10 Subjective performance assesment based on equivalent Q value.

ており，CCITT G.726 (32 kbit/s) 以上の等価 Q 値を示していることがわかる．

4.3 検　　討

提案する符号化系における演算の多くを占めるのは，MLSA フィルタおよび適応アルゴリズムの計算と，ピッチ予測のために使用する共分散係数の計算であり，これらに要する標本点当りの乗算回数は約500回となる．従って，提案する符号化系を汎用 DSP などで実現する際に要する演算量は，適応量子化器，ピッチ予測係数およびピッチ予測フィルタの計算など，その他の部分の演算があることなどを考慮すると，G.726 ADPCM よりは，かなり多くの演算量であるが，CCITT により勧告されている 16 kbit/s の音声符号化方式 LD-CELP [22] と同程度かそれ以下と見込まれる．

但し，提案する符号化系では，演算量の削減については検討を行っていない．例えば，$D(z)$ を近似する際には $L=4$ としているのに対し，$D^{\gamma}(z)$，$D^{\beta}(z)$ を近似する際には，式 (17) の有理式における $|F(e^{j\omega})|$ が，$D(z)$ の場合に比べ小さくなることから，近似次数を $L=2$ とすることも可能である．また，ピッチの適応も各サンプルごとに行っているが，数十点おきにピッチの適応を行っても性能はほとんど変わらないと予想される．

16 kbit/s LD-CELP は，CCITT G.726 32 kbit/s ADPCM と同等以上の音声品質をもつ．従って，受聴試験の結果より，提案法の音声品質は LD-CELP と同程度と見込まれる．一方，LD-CELP ではベクトル量子化を行っているため，コードブック探索のための繰り返し計算を必要とするのに対し，提案手法ではスカラ量子化器を用いているため，このような探索を要しない．また，それに伴い，LD-CELP では符号・復号による遅延を生じるのに対し，提案手法ではアルゴリズム遅延が 0 である．

しかし，提案する符号化系では符号の伝送路誤りに対する耐性，およびタンデム接続に関する考慮がなされていないことに注意する必要がある．但し，短期予測器の伝送路誤りに対する対策については，CCITT G.726 など従来のバックワード形 ADPCM で用いられているものと同様の手法を用いることにより，よい見通しが得られている [24]．この際，ピッチ予測に関しては，バックワード適応を用いるピッチ予測は伝送路誤りに弱いと言われているため，高次の短期予測をバックワード形で行う [22]，あるいはフォワード形のピッチ予測を行う，特別な工夫をしたピッチ適応法を用いる [25]，などの方法をとる必要があると考えられる．

5. むすび

メルケプストラムを介して，適応予測，ノイズシェイピング，ポストフィルタリングを行うバックワード形 ADPCM 系を提案した．主観評価実験より，提案法は従来の 32 kbit/s ADPCM 相当の品質を 16 kbit/s で達成していることを示し，提案符号化系の有効性を明らかにした．

提案法における各パラメータの最適値に関する調査は今後の課題である．また多くの音声符号化方式は伝送誤りに対するロバスト性を強く意識して設計されていることから提案法における伝送誤りへの対策については検討を要す．ここでは，ADPCM 系にメルケプストラムを導入することを考えたが，CELP など他の符号化系でのメルケプストラム利用については改めて報告したい．

謝辞　プログラム作成，受聴試験の実施など本研究に協力頂いた綾誠司君，松村英俊君に感謝します．

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 Schafer R.W.: "Descrete-Time Signal Processing", Prentice-Hall, Englewood Cliffs, N.J. (1989).
(3) 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).

(4) Chung J.H. and Schafer R.W., "A 4.8 Kbps homomorphic vocoder using analysis-by-synthesis excitation analysis", Proc. **ICASSP-89**, pp.144–147 (1989).
(5) Krüger E. and Strube H.W.: "Linear prediction on a warped frequency scale", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-36**, pp.1529–1531 (Sept. 1988).
(6) 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），**J74-A**, 8, pp.1249–1256 (1991–08).
(7) 深田俊明，徳田恵一，小林隆夫，今井　聖："適応メルケプストラム分析による適応予測器"，1991 信学春季全大，A-148, 1-148.
(8) 徳田恵一，松村英俊，小林隆夫，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系とその評価"，信学技報，**SP93**-62 (1993–08).
(9) 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).
(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) Fant G.: "Speech sound and features", MIT Press, Cambridge (1973).
(12) Widrow B. and Stearns S.D.: "Adaptive Signal Processing", Prentice-Hall, Englewood Cliffs, N.J. (1985).
(13) 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).
(14) 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 回ディジタル信号処理シンポジウム講論集，B-5-1, pp.279–284 (1989–12).
(15) Jayant N.S.: "Adaptive quantization with a one-word memory", Bell Syst. Tech. J., **52**, pp.1119–1144 (Sep. 1973).
(16) 北村　正，早原悦郎，太田悦生，謝　詠珪："ケプストラム合成音声の品質改善の一方法"，信学論（A），**J76-A**, 9, pp.1373–1375 (1993–09).
(17) Atal B.A. and Schroeder M.R.: "Predictive Coding of Speech Signals and Subjective Error Criteria", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-27**, pp.247–254 (Jun. 1979).
(18) Chen J.H. and Gersho A.: "Real-time vector APC speech coding at 4800 bps with adaptive postfilter", Proc. **ICASSP-87**, pp.2185–2188 (1987).
(19) 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 (1993–03).
(20) Ramachandran R.P. and Kabal P.: "Stability and Performance Analysis of Pitch Filters in Speech Coders", IEEE Trans. Acoust., Speech & Signal Process., **ASSP-35**, pp.937–946 (Jul. 1987).
(21) Gerson I.A. and Jasiuk M.A.: "Techniques for Improving the Performance of CELP-Type Speech Coders", IEEE Journal of Selected Areas in Communications, **10**, pp.858–865 (Jun. 1992).
(22) Chen J.H., Cox R.V., Lin Y.C., Jayant N. and Melchner M.J.: "A Low-Delay CELP Coder for the CCITT 16kb/s Speech Coding Standard", IEEE Journal of Selected Areas in Communications, **10**, pp.830–849 (Jun. 1992).
(23) 高橋　玲，長渕裕実："ATM 網におけるセル廃棄により劣化した広帯域音声の品質評価"，信学論（A），**J74-A**, 8, pp.1232–1239 (1991–08).
(24) 小石田和人，徳田恵一，小林隆夫，今井　聖："伝送路誤りを考慮した適応メルケプストラム音声符号化系"，日本音響学会講論集 (1993–03).
(25) 片岡章俊，守谷健弘："CELP 方式に基づく 8 kbit/s 低遅延音声符号化"，信学論（A），**J75-A**, 11, pp.1657–1665 (1992–11).

付　　録

$D(z)-1$ の構成

$D(z)$ を式 (17) のように近似すれば，$D(z)-1$ は，

$$D(z)-1 \simeq R_L(F(z))-1 = 2\frac{\sum_{l=1}^{[L/2]} A_{L,2l-1}\{F(z)\}^{2l-1}}{1+\sum_{l=1}^{L} A_{L,l}\{F(z)\}^{l}} \quad \text{(A·1)}$$

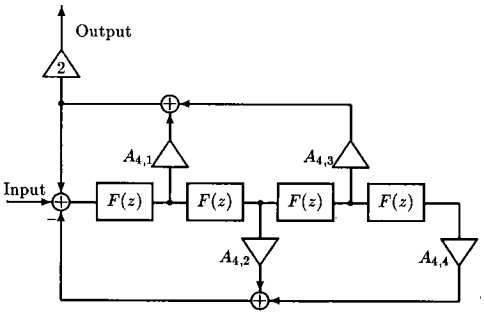
で実現することができる．但し，$[L/2]$ は $L/2$ を超えない最大の整数を表す．図 A·1 (a) に $L=4$ としたときの $R_L(F(z))-1 \simeq D(z)-1$ の構成図を示す．$F(z)$ は入力から出力へのディレイフリーパスを含まないので，図 A·1 (a) のように構成された $R_L(F(z))-1$ はディレイフリーループをつくらず，そのまま実現することができる．また，$R_4(F(z))$ が安定であることから，$D(z)-1$ を近似する式 (A·1) の伝達関数も安定となる．

2.2 で論じたように，$D(z)$ を $D_1(z)\cdot D_2(z)$ と 2 段に分割し，それぞれを $R_L(F_1(z))$, $R_L(F_2(z))$ によって近似した場合の $D(z)-1$ の構成法を考える．このとき，$D_1(z)-1$, $D_2(z)-1$ は，$D(z)-1$ と全く同様に近似的実現ができる．従って，$D(z)-1$ は

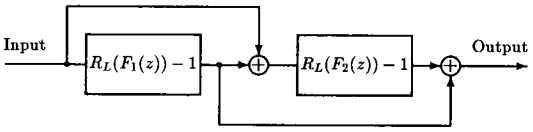
$$D(z)-1 = D_1(z)\cdot D_2(z)-1 = D_1(z)-1+(D_2(z)-1)\{(D_1(z)-1)+1\} \quad \text{(A·2)}$$

と書かれることから，

$$D(z)-1 \simeq R_L(F_1(z))-1 + (R_L(F_2(z))-1)\{(R_L(F_1(z))-1)+1\} \quad \text{(A·3)}$$



(a) $R_L(F(z)) - 1 \simeq D(z) - 1 \quad (L = 4)$



(b) $R_L(F_1(z)) \cdot R_L(F_2(z)) - 1 \simeq D(z) - 1 \quad (L = 4)$

図 A·1 $D(z) - 1$ の構成図 $(L = 4)$
Fig. A·1 Structure of $D(z) - 1$ $(L = 4)$.

と近似することができる．このときの構成図を図 A·1 (b) に示す．

なお，$D(z)-1$ と同様，$D(z)-1$ を近似する式 (A·3) もディレイフリーパスをもたないので，ここから z^{-1} を容易にくくり出すことができることに注意する．従って，図 4, 5, 7の符号化系もディレイフリーループをつくらず，実現可能である．

(平成 6 年 2 月 15 日受付，5 月 27 日再受付)

徳田　恵一

昭 59 名工大・工・電子卒．平 01 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．ディジタル信号処理，音声情報処理の研究に従事．日本音響学会，IEEE 各会員．

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．同年東工大精密工学研究所助手．現在同助教授．工博．ディジタルフィルタ，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 02 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）情報メディア研究所勤務．日本音響学会会員．

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．ディジタル信号処理，音声の合成および音声認識の研究に従事．昭 45 年度精機学会論文賞受賞．著書「ディジタル信号処理」など．計測制御学会，日本音響学会，IEEE，ASA 各会員．