

# スペクトル分析のための尺度を用いたメルケプストラム係数のベクトル量子化

高橋 徹<sup>†</sup>      徳田 恵一<sup>†</sup>      小林 隆夫<sup>††</sup>      北村 正<sup>†</sup>

Vector Quantization of Mel-Cepstral Coefficients Using Distortion Measure for Spectral Analysis

Toru TAKAHASHI<sup>†</sup>, Keiichi TOKUDA<sup>†</sup>, Takao KOBAYASHI<sup>††</sup>,  
and Tadashi KITAMURA<sup>†</sup>

あらまし 本論文では、スペクトル分析のための尺度に基づいてメルケプストラム係数のベクトル量子化を行う手法を提案する。本手法は、フレーム内の入力信号とコードブックに収められたメルケプストラム係数との適合の度合をスペクトル分析のための統計的尺度で評価し、量子化されたメルケプストラム係数を得るものである。また、提案法のためのコードブック学習アルゴリズムを示す。更に、提案法を多段ベクトル量子化に適用する方法についても示す。音声データベースを用いた実験により、コードブックのサイズが 9 bit の提案法により、12 bit の従来法（メルケプストラム係数間のユークリッド距離に基づくベクトル量子化手法）と同等の量子化ひずみが達成されることを示す。

キーワード 多段ベクトル量子化, ベクトル量子化, 統計的尺度, メルケプストラム分析

## 1. ま え が き

近年、携帯電話や、インターネット電話などの音声符号化技術を用いたアプリケーションが社会的に重要なインフラストラクチャとして普及しつつある。これらの応用分野において、音声信号を効率的に符号化することは通信コストや音声情報蓄積に必要なコストを削減するために極めて重要であると考えられる。このような背景から、種々の音声符号化方式が提案されているが、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、どのようにスペクトルを表現し、スペクトル分析パラメータをどのように量子化・伝送するかが、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。

従来、スペクトル分析パラメータをベクトル量子化 [1] する場合、分析対象の時系列信号は、分析過程と

量子化過程の 2 段の過程を経て量子化パラメータに変換されることが一般的であった。この場合、それぞれの過程で異なる基準が存在することになる。分析対象の時系列信号を、分析過程と量子化過程で一貫した基準に基づいて処理することにより、より効率的な量子化パラメータを得ることができると考えられる。そこで、本論文では、メルケプストラム分析 [2] で用いられる指数関数表現されたスペクトルモデル（指数型スペクトルモデル）に対し、分析過程、量子化過程とも一貫した尺度を適用するベクトル量子化手法を提案する。また、このような量子化に適したコードブック学習アルゴリズムを示す。同様の考え方は、LPC-VQ [3], [4] に見ることができる。しかし、提案法には、LPC-VQ にない以下のような利点がある。

- LPC-VQ は、スペクトル包絡を表現するために線形予測係数表現された全極型スペクトルモデルを用いるため、零点に相当するスペクトル包絡の谷部の表現が困難であるが、指数型スペクトルモデルを用いる提案法は、極と零とを同等に扱うため、スペクトル包絡の山部と谷部とを同じ精度で表現できる。

- LPC-VQ は、コードワードを線形予測係数によって表すが、線形予測係数が表すスペクトル包絡の

<sup>†</sup>名古屋工業大学知能情報システム学科, 名古屋市  
Department of Computer Science, Nagoya Institute of Technology, Nagoya-shi, 466-8555 Japan

<sup>††</sup>東京工業大学大学院総合理工学研究科, 横浜市  
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama-shi, 226-8502 Japan

周波数分解能は周波数軸上で一定である。一方、提案手法は、コードワードをメルケプストラム係数によって表すため、低域で高い解像度をもつという人間の聴覚特性に合ったスペクトルを効率的に表現可能である。

• LPC-VQ で困難であった多段ベクトル量子化を容易に処理することができる。

本論文では、音声データベースを用いた実験により、提案手法、LPC-VQ、従来法（分析過程でメルケプストラム分析を行い、量子化過程で係数間のユークリッド距離を基準として量子化を行う従来のベクトル量子化手法）の各手法を平均ひずみの観点から客観評価する。

以下、2. では、メルケプストラム分析法について述べる。3. では、提案手法によるベクトル量子化法と、コードブックの学習アルゴリズムを示し、4. では、3. で示した提案法を多段ベクトル量子化へ拡張する。5. では、3. と 4. とで提案した手法、LPC-VQ、従来のベクトル量子化手法の 3 種類の量子化手法を用いた量子化実験の結果を示す。最後に 6. では、まとめと今後の課題について述べる。

## 2. メルケプストラム分析

スペクトル包絡を表現する合成フィルタ  $H(z)$  はメルケプストラム係数によって

$$H(z) = \exp \sum_{m=0}^M c(m) \tilde{z}^{-m}, \quad (1)$$

$$\tilde{z} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (2)$$

と表すことができる。ただし、 $\alpha$  は周波数ワーピングに関する定数で、

$$|\alpha| < 1 \quad (3)$$

であり、 $\alpha = 0$  のとき、 $c(m)$  はケプストラム係数 [5] と等価となる。 $\alpha < 0$  のとき高周波数領域で高い解像度をもつスペクトル表現となり、 $\alpha > 0$  のときには、低周波数領域で高い解像度をもつスペクトル表現となる。 $\alpha$  に適切な値を設定することにより、低域で高い解像度をもつという人間の聴覚特性に合わせた表現が可能となる。

入力信号のあるフレーム  $x = [x(0), \dots, x(N-1)]^T$  が与えられたとき、メルケプストラム係数  $c = [c(0), \dots, c(M)]^T$  は、スペクトル評価関数 [5]

$$E(x, c) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{R(\omega) - \log R(\omega) - 1\} d\omega, \quad (4)$$

を  $c$  に関して最小化することによって決定される [2]。ただし、

$$R(\omega) = \frac{I(\omega)}{|H(e^{j\omega})|^2}, \quad (5)$$

$$I(\omega) = \frac{1}{W} \left| \sum_{n=0}^{N-1} w(n)x(n)e^{-j\omega n} \right|^2, \quad (6)$$

$$W = \sum_{n=0}^{N-1} w^2(n) \quad (7)$$

であり、 $w(n)$  は窓長  $N$  の窓関数である。メルケプストラム係数  $c$  は、 $E(x, c)$  が  $c$  に関して凸な関数であることから、反復アルゴリズム（例えば、Newton-Raphson 法）を使って高速に求めることができる。

$x(n)$  が平均 0 のガウス過程であると仮定すると、 $x$  に対する  $c$  の対数ゆわ度  $\log P(x|c)$  は、

$$\begin{aligned} \log P(x|c) \simeq & -\frac{N}{2} \left[ \log(2\pi) \right. \\ & \left. + \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \log |H(e^{j\omega})|^2 + \frac{I(\omega)}{|H(e^{j\omega})|^2} \right\} d\omega \right] \end{aligned} \quad (8)$$

と近似できる。このとき、 $E(x, c)$  の  $c$  に関する最小化は、 $\log P(x|c)$  の  $c$  に関する最大化と等価であることがわかる（付録）。また、メルケプストラム分析に用いるスペクトル評価関数は、線形予測分析 [6] のそれと同様な形式で表されていることもわかる。更に、 $H(e^{j\omega})$  からゲイン項  $g$  をくり出すことによって、 $c$  に関する  $E(x, c)$  の最小化は、予測残差エネルギーの最小化、あるいは予測ゲインの最大化と等価であることもわかる。

## 3. スペクトル分析のための尺度を用いたメルケプストラム係数のベクトル量子化

本章では、LPC-VQ 同様に一貫して統計的基準に基づき音声信号をベクトル量子化する手法について述べる。LPC-VQ では、音声信号を全極モデルでモデル化するが、提案法では、音声信号を指数関数表現された極零スペクトルモデルでモデル化する。また、このような量子化に用いるコードブックは、学習用音声信号から一貫して統計的基準に基づき学習することによって提案手法に適合したコードブックを得ることが

できる．本章では，そのようなコードブック学習アルゴリズムについても示す．

### 3.1 LPC-VQ

LPC-VQ は，線形予測モデルによってモデル化されている．量子化を含まない通常の線形予測分析において，スペクトル包絡を表現する合成フィルタ  $G(z)$  は線形予測係数によって [6]

$$G(z) = \frac{\sigma}{A(z)} \quad (9)$$

$$A(z) = 1 + \sum_{m=1}^M a(m)z^{-m} \quad (10)$$

と表すことができる．ただし， $\sigma$  はゲイン項を表すものとする．入力信号のあるフレーム  $x = [x(0), \dots, x(N-1)]^T$  が与えられたとき， $a = [\sigma, a(1), \dots, a(M)]^T$  は，スペクトル評価関数

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

を  $a$  に関して最小化することによって決定される．ただし，

$$R(\omega) = \frac{I(\omega)}{|G(e^{j\omega})|^2}, \quad (12)$$

$$I(\omega) = \frac{1}{W} \left| \sum_{n=0}^{N-1} w(n)x(n)e^{-j\omega n} \right|^2, \quad (13)$$

$$W = \sum_{n=0}^{N-1} w^2(n) \quad (14)$$

であり， $w(n)$  は窓長  $N$  の窓関数である．この最小化問題は解析的に解くことができる．このように  $a$  を定義すると，スペクトル評価関数 (11) が小さい値であるほど最適な線形予測係数であると考えられることができる．したがって，入力フレーム  $x$  に対してスペクトル評価関数 (11) を最小化するコードワードをコードブック内から選択することで，入力フレームを線形予測分析と同一の基準で量子化することができる．このとき，入力フレーム  $x$  の量子化インデックスは，

$$i = \arg \min_{j \in \{1, 2, \dots, I\}} E(x, a_j) \quad (15)$$

と表される．ただし， $a_j$  は， $j$  番目のコードワードであり， $I$  はコードブックサイズ (コードワード数) である．このとき，与えられたコードブック

$A = \{a_1, a_2, \dots, a_I\}$  に対する入力信号  $X = \{x_1, x_2, \dots, x_T\}$  の平均ひずみは，

$$D(X, A) = \frac{1}{T} \sum_{t=1}^T \min_{i \in \{1, 2, \dots, I\}} E(x_t, a_i) \quad (16)$$

$$= \frac{1}{T} \sum_{i=1}^I D_i(X, a_i), \quad (17)$$

ただし，

$$D_i(X, a_i) = \sum_{t \in \{t | i=i_t\}} E(x_t, a_i), \quad (18)$$

$$i_t = \arg \min_{j \in \{1, 2, \dots, I\}} E(x_t, a_j) \quad (19)$$

と表される．

### 3.2 メルケプストラム係数のベクトル量子化

従来ベクトル量子化では，メルケプストラム係数からなるベクトルは，分析過程と量子化過程の二つの過程を経て量子化されるとみなすことができる．従来法によるベクトル量子化は，図 1 に示される構成となる．ある入力音声のフレーム  $x \in R^N$  に対応するメルケプストラム係数  $c \in R^{M+1}$  は，スペクトル評価関数 (4) を基準として

$$c = \min_{c'} E(x, c') \quad (20)$$

と分析された後に， $R^{M+1}$  上にあらかじめ定義された何らかの距離尺度  $E'(c, \hat{c})$  によって，量子化インデックスが，

$$i = \arg \min_{j \in \{1, 2, \dots, I\}} E'(c, c_j) \quad (21)$$

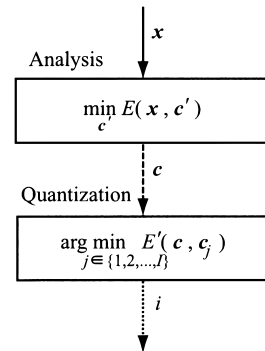


図 1 従来法によるベクトル量子化のブロック図  
Fig. 1 The block diagram of conventional vector quantization.

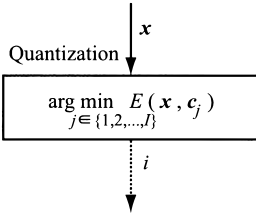


図2 提案手法によるベクトル量子化のブロック図  
Fig. 2 The block diagram of proposed vector quantization.

と決定されるとみなすことができる。ただし、 $c_j$  は、 $j$  番目のコードワードとする。このように二つの過程を経て量子化するとき、分析と量子化の各段階で異なる尺度 ( $E$  と  $E'$ ) が混在し、一貫した基準で量子化されないという問題がある。

2. では最適なメルケプストラム係数は、スペクトル評価関数 (4) を最小化する係数として定義した。つまり、線形予測分析と同様に、スペクトル評価関数 (4) が小さい値であるほど最適なメルケプストラム係数に近いと考えることができる。したがって、入力フレーム  $x$  に対してスペクトル評価関数 (4) を最小化するコードワードをコードブック内から選択することで、入力フレームをメルケプストラム分析と同一の基準で量子化できる。提案手法によるベクトル量子化は、図 2 に示される構成となる。入力フレーム  $x$  の量子化インデックスは、

$$i = \arg \min_{j \in \{1, 2, \dots, I\}} E(x, c_j) \quad (22)$$

と表される。ただし、 $c_j$  は、 $j$  番目のコードワードであり、 $I$  はコードブックサイズ (コードワード数) である。この量子化手法では、入力フレームに対して、メルケプストラム分析に用いたスペクトル評価関数を直接適用し、量子化を行うため、分析過程と量子化過程間における評価尺度の一貫性に関する従来手法の問題を解消できる。

### 3.3 コードブック学習アルゴリズム

本節では従来のベクトル量子化手法におけるコードブック学習アルゴリズムと、統計的尺度に基づくベクトル量子化に用いる最適なコードブック学習アルゴリズムについて述べる。

従来のベクトル量子化では、コードブック学習時に分析時と同じ基準が適用されないという問題がある。コードブック学習用ベクトルの集合を  $X = \{x_1, x_2, \dots, x_T\}$  と表すと、各学習用ベクトル

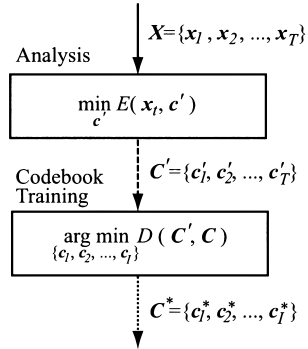


図3 従来のベクトル量子化におけるコードブック学習アルゴリズムのブロック図

Fig. 3 The block diagram of codebook training algorithm for conventional vector quantization.

ルは、スペクトル評価関数 (4) を基準とした式 (20) に従ってケプストラム係数からなるベクトル集合  $C' = \{c'_1, c'_2, \dots, c'_T\}$  に変換される。

$$c'_t = \min_{c'} E(x_t, c') \quad (23)$$

ただし、

$$t = 1, 2, \dots, T \quad (24)$$

このとき、学習用ベクトル集合に対して最適なコードブック  $C^* = \{c_1^*, c_2^*, \dots, c_I^*\}$  は、コードブック  $C = \{c_1, c_2, \dots, c_I\}$  の平均ひずみと、 $i$  番目のクラスタのクラスタひずみと、量子化インデックスとをそれぞれ

$$D(C', C) = \frac{1}{T} \sum_{t=1}^T \min_{i \in \{1, 2, \dots, I\}} E'(c'_t, c_i) \quad (25)$$

$$= \frac{1}{T} \sum_{i=1}^I D_i(C', c_i), \quad (26)$$

$$D_i(C', c_i) = \sum_{t \in \{t | i=i_t\}} E'(c'_t, c_i), \quad (27)$$

$$i_t = \arg \min_{j \in \{1, 2, \dots, I\}} E'(c'_t, c_j) \quad (28)$$

と表すと、

$$C^* = \arg \min_{\{c_1, c_2, \dots, c_I\}} D(C', C) \quad (29)$$

と表すことができる。従来のベクトル量子化におけるコードブック学習アルゴリズムは、図 3 に示される構成となる。この学習アルゴリズムは、学習用ベクトル

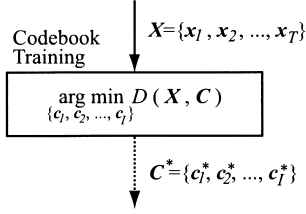


図4 提案ベクトル量子化におけるコードブック学習アルゴリズムのブロック図

Fig. 4 The block diagram of codebook training algorithm for proposed vector quantization.

集合をスペクトル評価関数 (4) を用いてケプストラムベクトル集合に変換した集合に対して何らかの距離尺度  $E'$  を基準としてコードブックを学習しているととらえられる。つまり一貫した基準によるコードブック学習アルゴリズムではない。

一方、統計的尺度に基づくベクトル量子化では、コードブックを統計的基準によって学習する。コードブック学習用ベクトルの集合  $X = \{x_1, x_2, \dots, x_T\}$  に対する最適なコードブック  $C^* = \{c_1^*, c_2^*, \dots, c_I^*\}$  は、コードブック  $C = \{c_1, c_2, \dots, c_I\}$  の平均ひずみと、 $i$  番目のクラスタのクラスタひずみと、量子化インデックスとをそれぞれ

$$D(X, C) = \frac{1}{T} \sum_{t=1}^T \min_{i \in \{1, 2, \dots, I\}} E(x_t, c_i) \quad (30)$$

$$= \frac{1}{T} \sum_{i=1}^I D_i(X, c_i), \quad (31)$$

$$D_i(X, c_i) = \sum_{t \in \{t | i=i_t\}} E(x_t, c_i), \quad (32)$$

$$i_t = \arg \min_{j \in \{1, 2, \dots, I\}} E(x_t, c_j) \quad (33)$$

と表すと、

$$C^* = \arg \min_{\{c_1, c_2, \dots, c_I\}} D(X, C) \quad (34)$$

と表すことができる。提案ベクトル量子化におけるコードブック学習アルゴリズムは、図4に示される構成となる。また、図5に平均ひずみ  $D(X, C)$  に基づく  $k$ -means 版のコードブック学習アルゴリズムを示す。はじめにステップ0において、このアルゴリズムに、適当な初期コードブック  $C_0$  を与え  $m = 0$  とする。以下、ある停止条件（例えば、平均ひずみの変化量が十分小さいという条件）が満たされるまでコードブックの更新を続ける。コードブックを再推定するた

**Step 0.** Begin with an initial codebook  $C_0$ . Set  $m = 0$ .

**Step 1.** Given a codebook,  $C_m$ , partition the training data set into cluster set using

$$i_t = \arg \min_i E(x_t, c_i), \quad t = 1, \dots, T$$

**Step 2.** Compute the centroids for the cluster set to obtain the new optimal codebook  $C_{m+1}$ , using

$$c_i = \arg \min_{c'_i} \sum_{t \in \{t | i=i_t\}} E(x_t, c'_i), \quad i = 1, \dots, I$$

**Step 3.** Compute the average distortion,

$$D(X, C) = \frac{1}{T} \sum_{i=1}^I \sum_{t \in \{t | i=i_t\}} E(x_t, c_i),$$

for  $C_{m+1}$ . If it has changed by a small enough amount since the last iteration, stop. Otherwise set  $m+1 \rightarrow m$  and go to Step 1.

図5 コードブック学習アルゴリズム ( $k$ -means 版)

Fig. 5 Codebook training algorithm ( $k$ -means).

めに、ステップ1では、コードブック  $C_m$  と式 (33) とを用いて、学習用ベクトルの集合をクラスタ分割し、ステップ2では、ステップ1で求めたクラスタ分割に対する最適なコードブック  $C_{m+1}$  を再計算する。 $i$  番目のコードワードの更新値  $c'_i$  は、式 (32) によって表されるクラスタひずみを  $c_i$  に関して最小化することで求めることができる。つまり、

$$c_i = \arg \min_{c'_i} D_i(X, c'_i) \quad (35)$$

$$= \arg \min_{c'_i} \sum_{t \in \{t | i=i_t\}} E(x_t, c'_i) \quad (36)$$

$$= \arg \min_{c'_i} \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\bar{R}_i(\omega) - \log \bar{R}_i(\omega) - 1\} d\omega \quad (37)$$

ただし、

$$i_t = \arg \min_{j \in \{1, 2, \dots, I\}} E(x_t, c_j), \quad (38)$$

$$\bar{R}_i(\omega) = \frac{\bar{I}(\omega)}{|H_i(e^{j\omega})|^2}, \quad (39)$$

$$\bar{I}(\omega) = \frac{1}{T_i W} \sum_{t \in \{t | i=i_t\}} \left| \sum_{n=0}^{N-1} w(n) x_t(n) e^{-j\omega n} \right|^2, \quad (40)$$

$$W = \sum_{n=0}^{N-1} w^2(n), \quad (41)$$

$$H_i(z) = \exp \sum_{m=0}^M c_i(m) \tilde{z}^{-m}, \quad (42)$$

$w(n)$  は窓長  $N$  の窓関数,  $T_i$  は,  $i$  番目のクラスタに量子化された学習用ベクトルの総数である. 式 (37) は, 式 (4) と同様な形式であることから式 (37) の最小化問題は, 式 (4) と同様に反復アルゴリズム (例えば, Newton-Raphson 法) を用いて高速に解くことが可能である. 最後にステップ 3 では, アルゴリズムの停止条件を判定する. 停止条件が満たされなかった場合,  $m+1 \rightarrow m$  とし, ステップ 1 に戻りコードブックを再推定する. 停止条件が満たされた場合, アルゴリズムを停止する. このときの  $C_{m+1}$  が求めるべきコードブックを表している. 従来の  $k$ -means アルゴリズムと同様, ステップ 1, 2, 3 の繰返しにより, 平均ひずみは, 単調に減少することが示される. ここでは,  $k$ -means 版のコードブック学習アルゴリズムを示したが, このアルゴリズムは, 簡単に LBG 版 [1] のアルゴリズムに拡張可能である. 5. の実験では, コードワードの分割を伴う LBG 版のアルゴリズムを用いてコードブックを学習した.

#### 4. 多段ベクトル量子化

ベクトル量子化は, 量子化ビット数  $B$  を増加させることで, 量子化誤差を減少させることが可能であるが,  $B$  の増加に伴い, 量子化, 復号化, コードブック学習の計算量は, いずれも  $2^B$  に比例して増加する. このため, 現実的な計算時間で処理を行いながら, 更に量子化ひずみを減少させる必要がある場合には, 多段ベクトル量子化 [1] などの構造化ベクトル量子化手法が用いられる. ところが, LPC-VQ には, 多段ベクトル量子化を適用することは困難であるという問題があった. 提案法は, 指数型スペクトルモデルを用いるため, 容易に提案手法を多段ベクトル量子化に適用可能となる. 本章の以下の節では, 提案手法における多段ベクトル量子化について述べ, そのための学習アルゴリズムを示す.

##### 4.1 メルケプストラム係数の多段ベクトル量子化

入力フレーム  $x$  の  $K$  段ベクトル量子化は,  $k$  段目のコードブックを  $C^{(k)} = \{c_1^{(k)}, c_2^{(k)}, \dots, c_I^{(k)}\}$  とすると, 各段の量子化インデックス  $i^{(k)} \in \{1, 2, \dots, I\}$  は, 次のように定式化できる.

$$i^{(k)} = \begin{cases} \arg \min_j E(x, c_j^{(k)}) & (k=1) \\ \arg \min_j E(x, \tilde{c}^{(k-1)} + c_j^{(k)}) & (k \geq 2) \end{cases} \quad (43)$$

ただし,

$$\tilde{c}^{(k-1)} = c_{i^{(1)}}^{(1)} + c_{i^{(2)}}^{(2)} + \dots + c_{i^{(k-1)}}^{(k-1)}. \quad (44)$$

LPC-VQ における多段ベクトル量子化は, 提案手法と形式的に同様に定式化可能である. しかし, 各段のコードブックを量子化基準において最適に学習することは容易ではなく, また, LPC-VQ においては, 各段の量子化結果の和が安定なシステム関数を表す必要があり, このような条件下で最適化を行うことは更に困難である.

一方, 提案手法において,  $K$  段ベクトル量子化から再構成される合成フィルタのシステム関数  $H(z)$  は, 多段ベクトル量子化の定義により,

$$c(m) = \sum_{k=1}^K c^{(k)}(m) \quad (45)$$

であることから,

$$\log H(z) = \sum_{m=0}^M \sum_{k=1}^K c^{(k)}(m) \tilde{z}^{-m} \quad (46)$$

$$= \sum_{k=1}^K \sum_{m=0}^M c^{(k)}(m) \tilde{z}^{-m} \quad (47)$$

$$= \sum_{k=1}^K \log H^{(k)}(z) \quad (48)$$

と表すことができ, つまり

$$H(z) = H^{(1)}(z) \cdot H^{(2)}(z) \cdots H^{(K)}(z) \quad (49)$$

と書くことができる. したがって,  $k$  段目の量子化器の入力は,  $k-1$  段目までのシステム関数  $H^{(1)}(e^{j\omega}) \cdot H^{(2)}(e^{j\omega}) \cdots H^{(k-1)}(e^{j\omega})$  により  $x$  を逆フィルタリングした信号  $x^{(k)}$  であることとらえることができる. ただし,  $x^{(1)} = x$  とする. 逆フィルタリングに基づく提案手法の多段ベクトル量子化は, 図 6 に示す構成となる.

##### 4.2 コードブック学習アルゴリズム

$K$  段ベクトル量子化において, コードブック学習用ベクトル集合を

$$\mathbf{X} = \{x_1, x_2, \dots, x_T\} \quad (50)$$

$k$  段目の入力を

$$\mathbf{X}^{(k)} = \{x_1^{(k)}, x_2^{(k)}, \dots, x_T^{(k)}\} \quad (51)$$

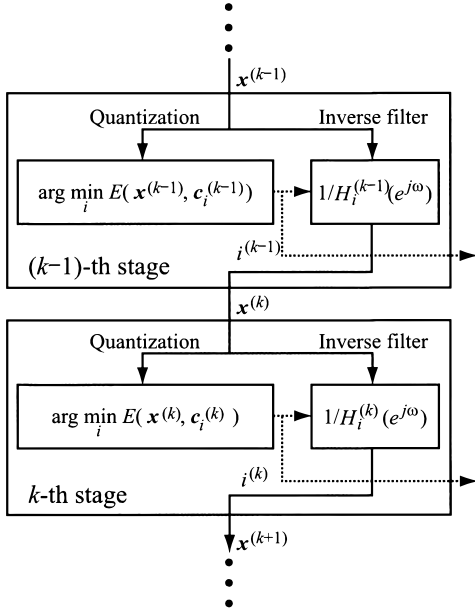


図 6 多段ベクトル量子化のブロック図  
Fig. 6 The block diagram of multi-stage vector quantization.

とすると、学習ベクトル集合に対する平均ひずみ、 $i$  番目のクラスターのクラスターひずみ、量子化インデックスは、それぞれ、式 (31), (32), (33) と同様に

$$D^{(k)}(\mathbf{X}^{(k)}, \mathbf{C}^{(k)}) = \frac{1}{T} \sum_{t=1}^T \min_{i \in \{1, 2, \dots, I\}} E(\mathbf{x}_t^{(k)}, \mathbf{c}_i^{(k)}) \quad (52)$$

$$= \frac{1}{T} \sum_{i=1}^I D_i^{(k)}(\mathbf{X}^{(k)}, \mathbf{c}_i^{(k)}) \quad (53)$$

$$D_i^{(k)}(\mathbf{X}^{(k)}, \mathbf{c}_i^{(k)}) = \sum_{t \in \{t | i_t = i^{(k)}\}} E(\mathbf{x}_t^{(k)}, \mathbf{c}_i^{(k)}) \quad (54)$$

$$i_t^{(k)} = \arg \min_i E(\mathbf{x}_t^{(k)}, \mathbf{c}_i^{(k)}), \quad (55)$$

と表すことができ、 $k$  段目の最適なコードブックは、

$$\mathbf{C}^{*(k)} = \arg \min_{\{\mathbf{c}_1^{(k)}, \mathbf{c}_2^{(k)}, \dots, \mathbf{c}_I^{(k)}\}} D^{(k)}(\mathbf{X}^{(k)}, \mathbf{C}^{(k)}) \quad (56)$$

で与えられる。式 (56) は、式 (34) と同様の形式で示されるため、各段のコードブック学習アルゴリズムとして、3.3 で示したアルゴリズムをそのまま利用可能であることがわかる。

表 1 実験条件  
Table 1 Experimental conditions.

Sampling rate	8 (kHz)
Frame length	32 (ms)
Frame period	10 (ms)
Window function	Blackman
Feature vector	LP coefficients (1st,...,10th) + gain, mel-cepstral coefficients (0th,...,10th)
Warping parameter	$\alpha = 0.31$
Database	ATR Japanese speech database
Training data	216 words (1,134,720 frames) b001-b216 40 speakers (male:20, female:20) m101-m120, f101-f120
Test data	520 words (2,511,210 frames) wi001-wi520 40 speakers (male:20, female:20) m301-m320, f301-f320

## 5. 実験

提案手法を客観的に評価する目的で、三つのコードブック (CB1, CB2, CB3) を作成し、平均ひずみを比較する。線形予測分析とベクトル量子化との基準を統一した手法を LPC-VQ と呼ぶのと同様に、メルケプストラム分析とベクトル量子化との基準を統一した提案手法を、Mel-Cepstrum-VQ と呼ぶことにし、実験では、Mel-Cepstrum-VQ を MC-VQ と略記する。

学習する 3 種類のコードブックは、それぞれ、式 (23) と  $E'$  をユークリッド距離関数とした式 (29) を用いてコードブックを学習する手法を従来手法とし、従来法に基づき学習されるコードブック (CB1)、MC-VQ 用に提案した学習アルゴリズムを用いて学習されるコードブック (CB2)、及び LPC-VQ 用の学習用アルゴリズムを用いて学習されるコードブック (CB3) である。

CB1 と CB2 では、メルケプストラム分析次数を 10 次とし、コードワードは、ゲインを表す 0 次の係数を含めた 11 次元ベクトルとした。LPC-VQ では、線形予測分析次数を 10 次とし、コードワードは、ゲイン項を含めた 11 次元ベクトルとした。コードブック学習用データは、ATR 日本語音声データベース C-set の 40 名 (男女各 20 名) がそれぞれ発声した 216 単語を用い、テスト用データは、学習外話者 40 名 (男女各 20 名) がそれぞれ発声した学習外単語 520 単語を用いた。表 1 に実験条件の詳細を示す。

平均ひずみを比較するにあたり、学習した 3 種類のコードブックの平均ひずみを 4 種類の方法で求め評価

表 2 学習基準とデコード基準

Table 2 Training criterion and decoding criterion.

	Spectral model	Codebook training criterion	Decoding criterion	Codebook
Baseline	Exponential	Euclidean	Euclidean	CB1
MC-VQ'	Exponential	Euclidean	$E(\cdot, \cdot)$	CB1
MC-VQ	Exponential	$E(\cdot, \cdot)$	$E(\cdot, \cdot)$	CB2
LPC-VQ	All-pole	$E(\cdot, \cdot)$	$E(\cdot, \cdot)$	CB3

する．表 2 に，各手法 Baseline, MC-VQ', LPC-VQ, MC-VQ のコードブック学習基準, デコード基準, 用いられるスペクトルモデル, 利用したコードブックの一覧を示す．従来法として位置づけられる手法 Baseline の平均ひずみは，コードブック CB1 を用い，式 (26), (27), (21) によって求められる．手法 MC-VQ' の平均ひずみは，コードブック CB1 を用い，提案手法と同様に式 (31), (32), (33) によって求められる．Baseline と MC-VQ' は，同一のコードブックを用いて平均ひずみが求められている．しかし，量子化インデックスの求め方が異なり，双方を比較することによって量子化インデックスの求め方の違いが平均ひずみに与える影響を比較できる．

提案手法として位置づけられる手法 MC-VQ の平均ひずみは，コードブック CB2 を用い，式 (31), (32), (33) を用いて求められる．手法 LPC-VQ の平均ひずみは，コードブック CB3 を用い，式 (17), (18), (19) を用いて求められる．手法 MC-VQ と LPC-VQ は，スペクトル評価関数を基準とした統一尺度を用いた量子化手法が用いられた場合である．MC-VQ と LPC-VQ とでは，スペクトルモデルのモデル化手法が異なり，双方を比較することによってモデル化の違いが平均ひずみに与える影響を比較することができる．

また，MC-VQ' と MC-VQ では，どちらも提案手法によって量子化インデックスが求められる．しかし，用いるコードブックの学習アルゴリズムが異なるため，提案量子化手法のためのコードブック学習アルゴリズムを用いることによる改善を確認できる．

5.1 量子化実験

図 7 に，学習データに対する平均ひずみ  $D(\cdot, \cdot)$  を示す．量子化ビット数が等しいとき，学習データに対する平均ひずみは Baseline, MC-VQ', LPC-VQ, MC-VQ の順に減少している様子がわかる．Baseline と MC-VQ' とを比較することによって提案手法による量子化インデックスの求め方の有効性が示されている．更に，MC-VQ' と MC-VQ とを比較することに

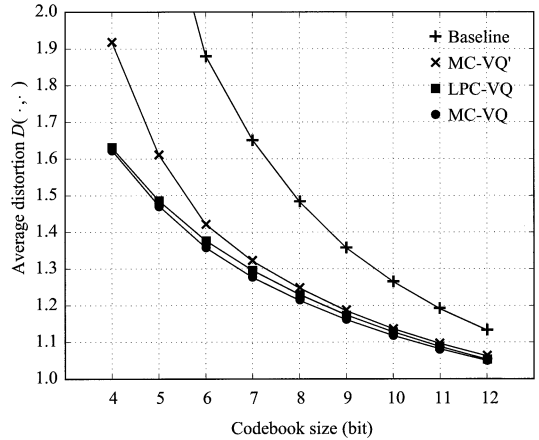


図 7 学習データに対する平均ひずみ  
Fig. 7 Average distortion for training data set.

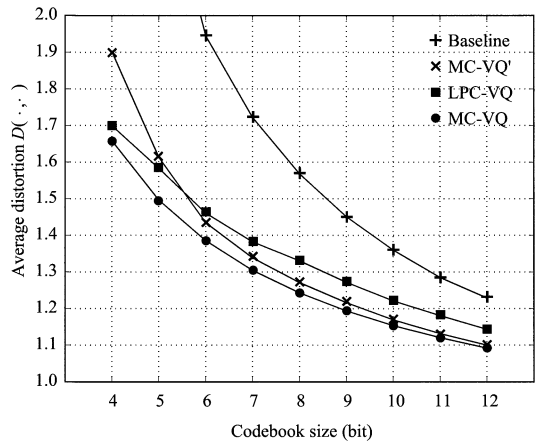


図 8 テストデータに対する平均ひずみ  
Fig. 8 Average distortion for testing data set.

よって，提案手法のためのコードブック学習アルゴリズムを適用することで平均ひずみを更に減少させることが可能であることが示されている．しかし，図 7 のように LPC-VQ の量子化ひずみに対して MC-VQ の量子化ひずみは小さいが，両者の差は小さく，学習データに対する量子化ひずみに限れば同程度の量子化ひずみを達成していると考えられる．

図 8 にテストデータに対する平均ひずみ  $D(\cdot, \cdot)$  を示す．量子化ビット数が等しいとき，5 bit 以下の量子化では平均ひずみは Baseline, MC-VQ', LPC-VQ, MC-VQ の順に減少しており，学習データに対する平均ひずみと同順で平均ひずみが減少している様子がわかる．一方，6 bit 以上の量子化では Baseline,



LPC-VQ, MC-VQ', MC-VQ の順に減少しており, LPC-VQ と MC-VQ' の平均ひずみが逆転している. しかし, Baseline, MC-VQ', MC-VQ の 3 種に対しては順序が保存されており, 提案手法による量子化インデックスの求め方の有効性と, 提案手法のためのコードブック学習アルゴリズムの有効性が, テストデータに対しても示されている. また, 学習データに対する MC-VQ の量子化ひずみは, LPC-VQ の量子化ひずみに比べて同程度の量子化ひずみであったのに対し, テストデータに対して, Baseline, LPC-VQ における 12 bit 量子化時の量子化ひずみを, MC-VQ では, それぞれ 9 bit, 11 bit 量子化で達成していることから MC-VQ の有効性が十分に示されていると考えられる.

### 5.2 2 段量子化実験

図 9, 図 10 に 2 段ベクトル量子化についての実験結果を示す. LPC-VQ には, 多段ベクトル量子化を適用できないため, これらの実験結果に LPC-VQ の結果は示されていない.

図 9 に, 学習データに対する平均ひずみ  $D(X, C)$  を示す. 量子化ビット数が等しいとき, 学習データに対する平均ひずみは Baseline, MC-VQ', MC-VQ の順に減少している様子がわかる. この結果は, 前節の量子化実験の結果 (図 7) と同様の傾向であり, 多段量子化を適用しても提案手法は, 従来法に対して優位性を損なわないことを確認できる.

図 10 に, テストデータに対する平均ひずみ  $D(X, C)$  を示す. 量子化ビット数が等しいとき, テストデータに対する平均ひずみは Baseline, MC-VQ', MC-VQ の順に減少している様子がわかる. 更に, 24 bit(12+12)Baseline の平均ひずみを 22 bit(11+11)MC-VQ で達成しており, 提案手法の有効性が確認できる. この結果は, 学習データに対する平均ひずみ (図 9) と同様な傾向であり, 提案法の優位性が, 話者や語彙に依存しないことを示している.

図 8 における, 12 bit LPC-VQ と 12 bit Baseline の平均ひずみは, それぞれ 1.14 と 1.23 を示しているが, 図 10 において, 10 bit(5+5)MC-VQ と 12 bit(6+6)MC-VQ の平均ひずみは, それぞれ 1.23 と 1.15 を示している. 提案手法は, 量子化ビット数が等しい 2 段構成であるにもかかわらず, LPC-VQ 手法における 1 段構成の 12 bit 量子化の平均ひずみと同等のひずみを達成できることが示されている. 更に, 提案手法は, 量子化ビット数が 2 ビット少ない 2 段構成

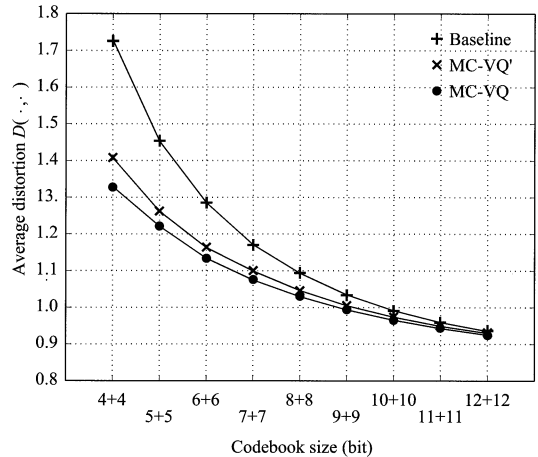


図 9 学習データに対する平均ひずみ (2 段ベクトル量子化)

Fig. 9 Average distortion for training data set (2-stage VQ).

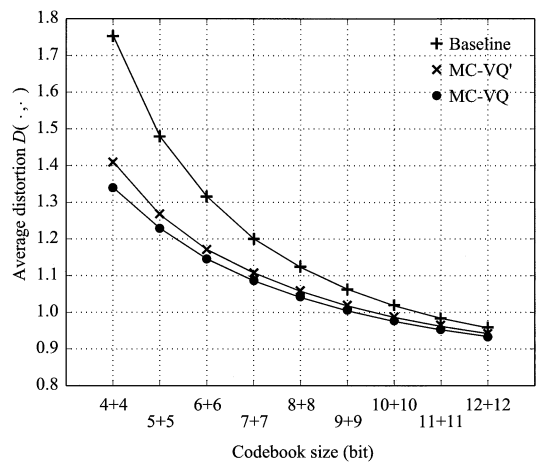


図 10 テストデータに対する平均ひずみ (2 段ベクトル量子化)

Fig. 10 Average distortion for testing data set (2-stage VQ).

であるにもかかわらず, 従来手法における 1 段構成の 12 bit 量子化の平均ひずみと同等のひずみを達成できることが示されている.

このように, 提案手法は多段ベクトル量子化手法を適用することで, 量子化ビット数の増加にも容易に対応することが可能である. また, LPC-VQ では多段構成とすることが困難なため, 現実的な計算時間で量子化やコードブックの学習を行うことは不可能であるのに対し, 提案手法は, 多段ベクトル量子化を用いることで現実的な演算量で量子化ひずみを最小限に抑える

ことが可能であるといえる。

これらの実験により、提案手法は、従来のメルケブストラム係数のベクトル量子化手法よりも少ないひずみを達成可能であることと、提案した多段ベクトル量子化手法の有効性が確認できた。更に、話者 80 名を用いた実験で、LPC-VQ を用いるよりも提案手法を用いた方が、量子化ひずみを抑制できることを確認できた。

## 6. む す び

統計的尺度に基づくメルケブストラム係数のベクトル量子化手法を提案し、提案手法に適合したコードブック学習アルゴリズムを示した。更に、提案手法を多段ベクトル量子化に適用できることを示した。提案手法により、LPC-VQ または従来のメルケブストラム係数のベクトル量子化手法よりも少ない量子化ビット数で、より少ない量子化ひずみを達成できることを客観評価実験により示した。

本論文で、評価に用いた式 (17), (31) は、予測残差エネルギー最小化など、音声符号化における性能改善に直結した評価となっているが、実際の応用における性能向上については別途、評価検討が必要である。今後、提案手法を高効率符号化系へ適用し、音声品質の評価を行うことで応用分野での有効性を確認したい。

## 文 献

- [1] A. Gersho and R.M. Gray, Vector quantization and signal compression, Kluwer Academic, 1992.
- [2] 徳田恵一, 小林隆夫, 深田俊明, 齋藤博徳, 今井 聖, “メルケブストラムをパラメータとする音声スペクトル推定,” 信学論(A), vol.J74-A, no.8, pp.1240-1248, Aug. 1991.
- [3] A. Buzo, A.H. Gray, Jr., R.M. Gray, and J.D. Markel, “Speech coding based upon vector quantization,” IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-28, no.5, pp.562-574, Oct. 1980.
- [4] B.H. Juang, D.Y. Wong, and A.H. Gray, Jr., “Distortion performance of vector quantization for LPC voice coding,” IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-30, no.2, pp.294-303, April 1982.
- [5] 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.
- [6] 板倉文忠, 齋藤収三, “統計的手法による音声スペクトル密度とホルマンと周波数の推定,” 信学論(A), vol.53-A, no.1, pp.35-42, Jan. 1970.

## 付 録

スペクトル評価関数の最小化と対数ゆ度関数の最大化との等化性

対数ゆ度関数  $\log P(x|c)$  は、スペクトル評価関数  $E(x, c) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{R(\omega) - \log R(\omega) - 1\} d\omega$  と、 $c$  を用いて

$$\log P(x|c) = -A(E(x, c) + B) \quad (\text{A-1})$$

と表すことができる。ただし、

$$A = \frac{N}{2} > 0 \quad (\text{A-2})$$

$$B = \log(2\pi) + \frac{1}{2\pi} \int_{-\pi}^{\pi} I(\omega) d\omega + 1 > 0 \quad (\text{A-3})$$

である。 $A, B$  は、 $c$  に無関係な定数であることから、 $\log P(x|c)$  の  $c$  に関する最大化は、 $E(x, c)$  の  $c$  に関する最小化に等価である。

(平成 13 年 7 月 18 日受付, 12 月 10 日再受付)



高橋 徹 ( 学生員 )

平 8 名工大・知能情報システム卒。平 10 同大学院博士前期課程了 ( 電気情報工學専攻 )。現在、同大学院博士後期課程在学中。音声分析の研究に従事。日本音響学会、情報処理学会各会員。



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学科助教授。工博。音声分析・合成・符号化・認識、デジタル信号処理、マルチモーダルインタフェースの研究に従事。平 13 電気通信普及財団賞、平 13 本会論文賞、猪瀬賞各受賞。日本音響学会、人工知能学会、情報処理学会、IEEE、ISCA 各会員。



小林 隆夫 (正員)

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。同助教授を経て平10 東工大大学院総合理工学研究科物理情報システム創造専攻教授。工博。デジタルフィルタ，音声分析・合成・合成・認識，マルチモーダルインタフェースの研究に従事。平13 電気通信普及財団賞，平13 本会論文賞，猪瀬賞各受賞。日本音響学会，情報処理学会，IEEE，ISCA 各会員。



北村 正 (正員)

昭48 名工大・工・電子卒。昭53 東工大大学院博士課程了。同年東工大精密工学研究所助手。昭58 名工大・工・電子工学科講師。昭59 同助教授。平7 名工大知能情報システム学科教授。工博。音声情報処理，マルチメディア情報処理の研究に従事。日本音響学会，情報処理学会，IEEE，ISCA 各会員。