

TR-I-0096

ファジィベクトル量子化に基づく話者適応化  
Speaker Adaptation through Fuzzy Vector Quantization

中村 哲, 鹿野清宏  
Satoshi NAKAMURA, Kiyohiro SHIKANO

1989. 8

概要

ベクトル量子化が、スペクトル空間の離散表現になることを利用して、未知話者のスペクトログラムを標準話者のスペクトログラムに写像する話者正規化方式について検討を行う。本論文では、ベクトル量子化を用いたスペクトログラムの話者間での正規化にファジィベクトル量子化の考えを導入する。ファジィベクトル量子化は、既存の各コードベクトルに対する帰属度を用いて入力ベクトルを表現するため、量子化歪を小さくできる。さらにファジィベクトル量子化にk近傍則を導入し高速化、高精度化を行い、スペクトログラムの話者間正規化に適用する。また、話者間の対応関係を示す変換コードブックの学習において、ファジィベクトル量子化の帰属度を確率として使い、変換コードブックの高精度化を行う。音韻バランス216単語、男女計3名分の音声データに対し、スペクトル歪、及び話者間正規化の正規化誤差を用いて評価を行い、ファジィベクトル量子化を用いたスペクトログラム正規化方式の有効性を示す。

ATR 自動翻訳電話研究所

ATR Interpreting Telephony Research Laboratories

© (株)ATR 自動翻訳電話研究所 1989

© 1989 by ATR Interpreting Telephony Research Laboratories

## 目次

まえがき .....	1
2. ファジィベクトル量子化 .....	2
2.1 ファジィ集合 .....	2
2.2 ファジィベクトル量子化(ファジィクラスタリング) .....	2
3 ファジィベクトル量子化による量子化歪の低減 .....	4
3.1 ファジィベクトル量子化による量子化歪の低減方法 .....	4
3.2 ベクトル量子化歪による比較 .....	4
3.3 k近傍則による高速化 .....	5
3.4 比較実験 .....	5
4. スペクトログラム正規化 .....	6
4.1 話者正規化のアルゴリズム .....	6
4.2 ファジィベクトル量子化による話者正規化の高精度化 .....	7
4.2.1 ファジィベクトル量子化を用いた変換コードブック生成法 .....	7
4.2.2 ファジィベクトル量子化による入力系列のスペクトル正規化方法 ..	8
4.3 実験 .....	8
5. まとめ .....	11
6. 謝辞 .....	11
7. 文献 .....	11

近年ベクトル量子化の手法が音声認識の分野にも導入され始めている<sup>(1,2)</sup>。ベクトル量子化は本来効率的な符号化の方法であり、情報圧縮に効果を発揮している<sup>(3,4)</sup>。さらに、ベクトル量子化をスペクトル空間の離散表現としてとらえることにより、新しい話者正規化の方法の可能性が開かれた<sup>(5~13)</sup>。この方法では、話者正規化の問題を有限個のスペクトルの話者間での写像問題として捉えている。

文献(5)で提案された学習に基づく方法では、図1に示すように、まず発声者毎に音声をベクトル量子化してコードブックを生成する。その後、話者間で同一学習単語の動的計画法による非線形マッチング(DTW)の最適パスからベクトルコードの対応づけヒストグラムを求める。このヒストグラムにより話者間でのコードブックの対応関係を求めて話者正規化を行うものである。この手法が英単語の認識実験に適用され良好な結果を収めている。文献(6)では、標準パターンのスペクトルをDTWによる対応づけで置き換えて、話者の適応化を行っている。文献(7)では、話者に共通のコードブックを作成した後、単音節の学習から部分空間法により共通コードブックを話者に適応させる方法を試み良好な結果を収めている。文献(9),(10)では、テキストに依存せずに、歪最小の基準で共通コードブックを話者に適応させる方法を試みている。文献(11)では話者毎のベクトルの線形写像問題として変換行列を求め話者正規化を試みている。また、文献(12)では、話者に共通のコードブックを用いてHMMモデルの出力確率にベクトル間の変換確率を込めて学習するという方法を提案し、良好な結果を収めている。

文献(5)の方法は、話者間のスペクトル空間の対応を有限個のスペクトルの対応として扱え、対応するスペクトルの類似性をさほど必要としない利点を有している。本論文では、文献(5)の方法に基づいて、スペクトログラムリーディングベースの音声認識やHMMベースの音声認識の前処理としての話者間のスペクトログラムの正規化を検討する。しかし、この観点からは上述の方法には、スペクトル歪尺度に起因するスペクトログラムの再現性の問題、コードブックサイズの問題、ラーニングサンプルの問題があった。文献(13)では、この問題を解決するために特徴毎に別々の量子化系を構成するセパレートベクトル量子化を用いた方法が試みられて、良好な結果を収めている。しかし、まだ次の問題が残っている。

(a)ベクトル量子化のコードブックサイズが不十分である。

(b)話者間のコードブックの対応を表わす変換コードブックの精度が不十分である。

これらの問題に対しての解決法としては、ベクトル量子化のコードブックサイズを単純に増加させる方法が考えられるが、コードブック生成に莫大な単語数が必要であり、かつ変換コードブックを生成するためにも学習単語数の増加が必要となる。

本論文では、ベクトル量子化のコードブックサイズ、及び変換コードブック学習用単語数を増加させずに上記問題を解決する方法としてファジィベクトル量子化を用いた方法について検討を行う。ファジィ集合は、文献(14)により提案され、クラスタリング問題に適用され効果を収めている<sup>(15~17)</sup>。音声認識の分野ではHMMの出力確率のパラメータの平滑化に用いられて、良好な結果を収めている<sup>(18)</sup>。

## 2. ファジィベクトル量子化

本章では、ファジィ集合について簡単に述べ、さらにファジィベクトル量子化について説明する。

### 2.1 ファジィ集合

ファジィ集合の概念は文献(14)により提案されたものである。この方法は従来の集合に曖昧さの概念を導入したものである。任意の集合を $X$ 、その要素を $x$ とする。一般の部分集合は、関数を用いて記述できる。 $X$ のべき集合を $P(X)$ とすると、部分集合 $A \in P(X)$ に対して、関数 $u_A(x) : X \rightarrow \{0,1\}$ が次のように定義される。

$$\begin{aligned} u_A(x) &= 1 : x \in A \\ &= 0 : \text{otherwise} \end{aligned}$$

$u_A(x)$ を級関数(Membership function)と呼ぶ。

文献(14)にしたがって、関数 $u : X \rightarrow [0,1]$ とすることにより曖昧性を考慮した、ファジィ部分集合を定義できる。また、パターン認識のようなクラス分離問題を考えるとき、 $u_A$ と $u_{\bar{A}}$ は次の関係を満たすものとする<sup>(16)</sup>。

[定義] 
$$\begin{aligned} u_A + u_{\bar{A}} &= 1 \\ 0 < u_A < 1 \end{aligned}$$

### 2.2 ファジィベクトル量子化(ファジィクラスタリング)

2.1節で定義したファジィ集合に基づいた用いたクラスタリング問題を考える。ファジィ集合は、一般の集合と異なり要素の属性 $u_A$ は2値でなく0から1の連続値で表現されるため、曖昧さ、確信度を含むクラスタリングを行うことができる。2つのクラスタを考える。ある有限集合 $X$ の要素 $x$ に対して、従来のクラスタリングを行うと図2-(a)のようにどちらかのクラスタに一意に帰属してしまうことになる。これに対して、ファジィ集合を用いることにより、図2-(b)のように連続的な帰属度を持つ曖昧なクラスタリングを表わすことができる。これを基本にして、曖昧さを考慮したファジィクラスタリングが構成される。ここでは、文献(15)、文献(17)により提案された、ファジィC-meansクラスタリングアルゴリズムを用いる。この方法は2乗誤差最小の規範のもとで、クラスタリングを行うものである。このファジィC-meansクラスタリングは次のように定義されている。

入力ベクトルの要素数を $n$ 、クラスタ数を $c$ とし、入力ベクトル $x$ のコードベクトル $a_i$ に対する級関数を $u_i$ とすると、目的関数 $J_m$ を次のように定義する。

$$J_m = \sum (u_i)^m \cdot d_i$$

$d_i = \|x - a_i\|$  :  $\|\cdot\|$  は ユークリッド距離

$m$ : 重み係数 :  $m \in [1, \infty)$

$\{a_1, \dots, a_c\}$  :  $c$ 個のコードベクトル

この目的関数 $J_m$ を最小化するように $u_i$ を決定すればよい。ただし、ファジィ級関数の定義から、

$$\sum u_i = 1$$

のもとで、 $J_m$ を最小化する。これは、ラグランジュの未定乗数 $\lambda$ を用いて、

$$D = J_m - \lambda \cdot (\sum u_i - 1)$$

の最小化から求めることができ、この結果、 $u_i$ が $J_m$ を最小化する条件は、次のようになる。

$$u_i = 1 / [\sum (d_i / d_j)^{1/(m-1)}]$$

ただし、入力ベクトルがコードベクトルと一致したときは、一致したコードベクトルへの級関数値 $u_i$ を1.0とし、それ以外の級関数値を0とする。

ここで、 $m$ を曖昧度(fuzziness)と呼び、 $m \rightarrow \infty$ の時、 $u_i \rightarrow 1/c$ で、soft decision、 $m \rightarrow 1$ の時、従来のhard decisionによるクラスタリングとなる。従って、 $m \rightarrow 1$ の時に、2乗誤差最小に基づく本クラスタリングはベクトル量子化に近

づく。文献(18)では、この方法を $m \in [1, \infty)$ に拡張した場合も含めてファジィベクトル量子化と呼んでいる。

### 3 ファジィベクトル量子化による量子化歪の低減

本章では、ファジィベクトル量子化を用いた量子化歪の低減効果について述べ、従来法との比較を行う。

#### 3.1 ファジィベクトル量子化による量子化歪の低減方法

2.2節で述べたファジィベクトル量子化の方法は、入力ベクトル $\mathbf{x}$ から既存のコードベクトルへの帰属度を求める符号化の手続きといえる。しかし、スペクトログラム再現のためには復号化の手続きが必要である。そこで、目的関数 $J_m$ を、入力ベクトルについて解いて、次式を得る。

$$\mathbf{x}^{(A)} = \Sigma[(u_i)^m \cdot \mathbf{a}_i] / \Sigma(u_i)^m \quad \dots\dots(1)$$

(1)式は、コードベクトル $\{\mathbf{a}_1, \dots, \mathbf{a}_c\}$ の線形結合で、新しいベクトル $\mathbf{x}^{(A)}$ が合成され、近似的にコードブックサイズを増加させたことを意味している。図3にベクトル量子化による量子化の様子を概念的に示す。太い破線は、従来の方法による量子化を示し、細い破線がファジィベクトル量子化の場合を示す。とくに従来法では、過渡部分で歪が大きくなってしまいが、ファジィベクトル量子化によりこの歪を改善することができる。

#### 3.2 ベクトル量子化歪による比較

ファジィベクトル量子化と従来のベクトル量子化のスペクトル歪の比較を行う。分析条件は、12kHzサンプリング、分析窓長21.3msec、フレーム周期3msecで14次のLPC分析を行う。データベースは可能な2音素連鎖をすべて含む音韻バランス<sup>(21)</sup>216単語を女性1名が発声したものをを用いる。

ファジィベクトル量子化と従来のベクトル量子化の比較をスペクトル歪を用いて行う。ファジィベクトル量子化の場合には、目的関数 $J_m$ が最小化されるように級関数 $u_i$ が求められる。一般に、 $J_m$ は従来のベクトル量子化のスペクトル歪とは異なる。このため、一旦ファジィベクトル量子化した後、式(1)を用いて $\mathbf{x}^{(A)}$ を合成し、入力ベクトル $\mathbf{x}$ との距離を求めてスペクトル歪とする。ここで、いずれのベクトル量子化においても、コードベクトルは、データベースの前半百単語の中から5000フレームをサンプルとして取り出してLBGアルゴリズム<sup>(3)</sup>によって、生成したものである。

また、実験には、距離尺度として、近傍でユークリッド距離に近似的に等しく、音声認識において高い性能を示すWLR尺度<sup>(19)</sup>を用いる。WLR尺度は次式で表される。

$$WLR = \sum (C_n - C'_n) \cdot (R_n - R'_n)$$

$C_n, C'_n$  :  $n$ 次のLPCケプストラム係数

$R_n, R'_n$  :  $n$ 次のLPC自己相関係数

### 3.3 $k$ 近傍則による高速化

2.2節で述べたファジィベクトル量子化の方法では、入力 $x$ に対しすべてのコードベクトルに対し級関数を計算し、記憶しなければならないので、すべてのコードベクトルに関する級関数の計算は膨大なものとなる。

この問題を解決するため、 $k$ 近傍則を導入する。これは、入力 $x$ に対して最も距離の小さかったコードベクトルから $k$ 番目までに対してのみ級関数を計算するものである。これにより、級関数の計算が $k$ 個のコードベクトルに対してのみで済む。また、より小さな部分空間に分割できるため、スペクトル歪も小さくできる。

### 3.4 比較実験

実験は、女性話者1名の発声した音韻バランス216単語を用いて行う。まず、曖昧度 $m$ と近傍数 $k$ を決定するため、予備実験を行った。実験は、女性話者1名の未知単語20単語について行った。

一般に $k$ の値と曖昧度の値は無相関ではない。このため、スペクトル歪最小になるように繰り返しアルゴリズムを構成して最適な $k$ と曖昧度を求める。まず、 $k=256$ の全てのコードを用いた場合にスペクトル歪が最小になる曖昧度の値を求める。次に、曖昧度の値を固定にしてスペクトル歪が最小になる $k$ の値を求める。この手続きを繰り返すことにより最適な $k$ と曖昧度の組合せを求める。この結果を表1に示す。 $k$ が6、曖昧度が1.6の時スペクトル歪がほぼ収束した。曖昧度が、1.6の時の $k$ に対する変化を図4に示す。また、 $k$ が6のときの曖昧度に対する変化を図5に示す。最終的に、曖昧度が1.6で、 $k$ が6のとき従来法の28.8%だけスペクトル歪が減少した。

次に、学習データとして、コードブック生成に用いた100単語、未知データとしてそれ以外の100単語に対して、スペクトル歪の比較を行う。この結果を表2に

示す。ファジィベクトル量子化を用いることにより、全てのコードベクトルを用いた場合( $k=256$ )、学習データで、スペクトル歪が22.0%だけ、未知データでは、24.0%だけ減少し有効性が確認された。さらに、 $k$ 近傍則を用いて $k=6$ にした場合は、学習データで、スペクトル歪が、25.5%だけ、未知データでは、27.7%だけ減少し、計算量の上からばかりでなくスペクトル歪の上からも $k$ 近傍則が有効であることが確認された。

図6に従来のベクトル量子化を用いた場合と、ファジィベクトル量子化を用いた場合のスペクトログラムの出力例を示す。これは、単語「勢い」の「きおい」の部分である。ファジィベクトル量子化により量子化歪が改善されており、特に過渡部分の連続性がよく表現されていることがわかる。

#### 4. スペクトログラム正規化

本章ではベクトル量子化を用いた話者正規化のアルゴリズムと、ファジィベクトル量子化による高精度化について述べる。

##### 4.1 話者正規化のアルゴリズム

話者正規化のアルゴリズムは、文献(5)により提案された学習に基づくアルゴリズムに基づいている。この方法では、指定された学習用単語を未知話者Aに発声させ、これを標準話者Bの発声した標準パターンとDouble Split法<sup>(2)</sup>を用いてベクトルコード列のレベルでマッチング<sup>(20)</sup>を行い、ベクトルコードの対応づけを行う。これを所定の単語数について行うことにより、標準話者Bのコードブックの対応づけを知ることができる。ここでは、未知話者Aのベクトル $a_i$ を、対応づけのヒストグラム $h_{ij}$ を重みとして標準話者Bのベクトル $b_j$ から線形結合で構成したベクトル $b'_i$ へ1対1に対応づける。この様子を図7に示す。この線形結合で構成されたベクトルによりなるコードブックを変換コードブック(Mapped Codebook)と呼ぶ。

スペクトル正規化のための話者正規化の過程は、次のアルゴリズムに従う。

step1) 未知話者Aから標準話者Bへの変換コードブックを求める。

step2) 未知話者Aの音声を学習時に生成したAのコードブックを用いてベクトル量子化する。

step3) 未知話者Aの音声に対し、ベクトル量子化により求められたベクトルコード列をAからBの変換コードブックを用いて変換し正規化されたスペクトログラムを出力する(スペクトログラムの正規化)。



本論文では、文献(13)において提案したセパレートベクトル量子化の場合の話者正規化のアルゴリズムにさらにファジィベクトル量子化を導入し、高精度化をはかる。

## 4.2 ファジィベクトル量子化による話者正規化の高精度化

ここでは、ファジィベクトル量子化を用いた場合の話者正規化方法について述べる。

### 4.2.1 ファジィベクトル量子化を用いた変換コードブック生成法

4.1節に述べたように話者正規化は、学習単語のマッチングによるベクトルコードの対応づけにより行われる。ベクトルコードの対応づけはDTWマッチング時の累積距離が最小になるパスの上でのコードの1対1対応を積算しヒストグラムを作ることによりなされる。この方法では、マッチングパス上での1対1対応の積算を基本としているので、学習単語への依存度が高い。これに対し、ファジィベクトル量子化では、級関数の値を確率として用いることで、ファジィベクトル量子化の符号系列から、一度に複数のベクトルコードの対応を取ることが可能となり、学習単語への依存度を軽減できる。また、従来法では、マッチングパスで対応したものに対して一意に1回と積算しているのに対し、ファジィベクトル量子化の場合には、級関数の値をそのまま確からしさの連続値として積算することができる。ここでは、図8に示す(a)(b)(c)の3種類のヒストグラム積算方法を考える。k近傍則による部分空間を設定しているので、近傍の数をkとすれば、学習時の1フレーム対応が(a)では1個、(b)ではk個、(c)では、k<sup>2</sup>個のヒストグラム積算が可能となり、学習サンプル数低減の効果が期待できる。実験では、次に示す3つの方法を検討する。

(a)従来法におけるヒストグラム積算方法

$$h_{ij} \leftarrow h_{ij} + 1$$

(b)入力側のみファジィベクトル量子化を行なうヒストグラム積算方法

$$h_{ij} \leftarrow h_{ij} + u^{(A)}_i$$

(c)入力、標準パターンともファジィベクトル量子化を行なうヒストグラム積算方法

$$h_{ij} \leftarrow h_{ij} + u^{(A)}_i \cdot u^{(B)}_j$$

ここで、 $u^{(A)}_i, u^{(B)}_j$ は、DTWの最適パス上で対応するフレームにおけるファジィベクトル量子化の未知話者Aおよび標準話者Bのコードベクトルへの級関数値である。

#### 4.2.2 ファジィベクトル量子化による入力系列のスペクトル正規化方法

話者正規化は、式(2)に示すよう変換コードブックを用いて入力ベクトルを入れ換えることにより行う。式(1)に示すように、ファジィベクトル量子化では、入力ベクトル $\mathbf{x}$ を、ファジィ級関数を用いて既存のコードベクトル $\mathbf{a}_i$ からの合成 $\mathbf{x}^{(A)}$ で表わす。次に図9に示すように、ベクトル $\mathbf{x}^{(A)}$ を表すコードベクトル $\mathbf{a}_i$ を、変換コードブックの対応するコードベクトル $\mathbf{b}'_i$ に一对一に入れ換え、級関数の値はそのままにして標準話者の空間でベクトル $\mathbf{x}^{(B)}$ を生成する。スペクトルの変換は、未知話者の入力ベクトル $\mathbf{x}$ をこのベクトル $\mathbf{x}^{(B)}$ に変換することにより行われる。ここで、話者間の空間構造の異なりが問題となるが、 $k$ 近傍則による十分小さな部分空間の選択と、ヒストグラムに基づく変換コードブックによる1対1対応の写像により、級関数の値が部分空間の間で保存されて写像されるものとする。

$$\begin{array}{ccc} \mathbf{x} & \rightarrow \mathbf{x}^{(A)} = \frac{\sum(u_i)^m \cdot \mathbf{a}_i}{\sum(u_i)^m} & \rightarrow \mathbf{x}^{(B)} = \frac{\sum(u_i)^m \cdot \mathbf{b}'_i}{\sum(u_i)^m} \\ \text{Fuzzy VQ} & & \text{Mapping} \end{array} \dots\dots\dots(2)$$

### 4.3 実験

実験には、男性2名、女性1名を用いる。また、学習時には、ベクトル量子化のコードブック生成に用いたのと同じ100単語を用いる。評価尺度には、スペクトログラム正規化に於ける正規化誤差を用いる。まず、未知話者のスペクトログラムに対しベクトル量子化を行い、変換コードブックを用いてフレーム毎のスペクトルの入れ換えを行う。正規化誤差は、正規化したスペクトログラムとターゲットとなる標準話者のスペクトログラムとのDTW距離である。本稿では、正規化時に入力をファジィベクトル量子化するか否か、変換コードブック生成時にファジィベクトル量子化するか否かについて表3に示す6通りについて、学習単語数をパラメータとした実験を行った。未知話者、標準話者を同性間、異性間の組合せについて、学習外の100単語で実験し、平均を行った結果を図10に示す。但し、ファジィベクトル量子化における曖昧度の値は、1.5とした。これにより次の結果を得た。

- (1) 入力をファジィベクトル量子化することによる効果が最も大きい。スペクトログラム正規化誤差は、従来のベクトル量子化に比べて学習単語が25単語の場合平均で9.5%、100単語の場合4.6%減少する。
- (2) 変換コードブック生成は、入力、標準パターン共にファジィベクトル量子化したIII,VIが効果的であり、入力がファジィでない場合、Iの場合に比べて学習単語が25単語の時IIで5.4%、IIIで10.8%だけの減少である。学習単語が100単語の場合は、IIIでも1.6%だけの減少で、変換コードブック生成については、学習単語数の少ないところで効果がある。
- (3) ファジィベクトル量子化により学習単語の削減が可能となる。入力に従来のベクトル量子化を用いた場合、学習単語が100単語の時のIの正規化誤差と、IIIの約70単語の正規化誤差が同等となり、30単語程度削減できる。入りにファジィベクトル量子化を用いればさらに約20単語削減できる。
- (4) 正規化及び変換コードブック生成にファジィベクトル量子化を用いた場合VIは、学習単語が50単語以上でIV,Vより正規化誤差が大きくなる。これは、VIのヒストグラム積算方法で、ヒストグラムに過度のスムージングが行われているのが原因と思われる。

入力をファジィベクトル量子化する効果が大きいですが、これはベクトル量子化が精密にできること、及び4.2.2節で述べたファジィベクトル量子化の場合の写像方法が妥当であったことを示している。

変換コードブックをファジィベクトル量子化により生成した場合、学習単語の少ないところでは効果が大きいですが、100単語程度になると、余り大きな効果が得られない。これは、学習単語の増加による効果が飽和してくるためである。この原因としては、未知話者の入力ベクトルのファジィベクトル量子化による級関数値を確率とみる方法に問題があると考えられる。実際には、話者内のスペクトル空間にコードベクトルは分布しているのに対し、この方法では距離情報のみにより対応づけを行うので、距離が近くても位置が異なったコードベクトルとの対応づけが行われるなど、不適當なコードベクトルの対についてヒストグラムの積算が行われる可能性がある。そこで、未知話者Aと標準話者Bのコードベクトルの対応づけの確からしさを求めてヒストグラムの積算の拘束条件とする次のアルゴリズムを構成する。

- step 1) 全学習単語について4.2.1で述べた(c)のヒストグラム積算方法に従って、ヒストグラムを生成する。
- step 2) ヒストグラムの値から次式に従って、対応づけの確からしさ $p_{ij}$ を求める。 $p_{ij}$ は、未知話者と標準話者の $k$ 近傍による部分空間の対応におけるコードベクトルの対応づけ確率である。

$$p_{ij} = (h_{ij} + \theta) / \sum (h_{ij} + \theta)$$

sub(i) :  $a_i$ とDTWで対応した話者Bのフレームのk近傍に含まれるbのコードベクトルの集合

$\theta$  :  $p_{ij}$ が0にならないための定数

step 3) step 2 で求めた $p_{ij}$ を用い、全学習単語について次式に従ってヒストグラムの積算を行う。

$$h_{ij} \leftarrow h_{ij} + u^{(A)}_i \cdot u^{(B)}_j \cdot p_{ij}$$

step 4) step 3 で求めたヒストグラムに従って、変換コードブックを生成する。

step 5) 全学習単語について平均正規化誤差を求める。

step 6) 平均正規化誤差が収束していれば終了。

収束していなければ、未知話者のコードブックを変換コードブックに入れ換え、step 2)へ。

このアルゴリズムに従って求めた未知単語に対する平均正規化誤差を図10のVIIに示す。平均正規化誤差は、Iに比べて100単語の場合で、9.7%の平均誤差の低減であった。また、話者間のコードブックの対応を求めるための学習単語数からいえば、従来法の平均正規化誤差を約30単語の学習で実現でき、70単語の削減になることがわかった。このアルゴリズムにより不適切な対応づけが除かれ、学習単語を増加しても正規化誤差の低減に効果があることが示された。

図11にスペクトログラム正規化例を示す。学習外単語「購入」を発声したもので、未知話者は男性、標準話者は女性である。図で、(A)は未知話者の音声をファジィベクトル量子化したスペクトログラム、(B)は従来のベクトル量子化を用いて正規化を行った正規化スペクトログラム、(C)は、ファジィベクトル量子化を行い、上述のVIIの方法により生成した変換コードブックを用いて求めた正規化スペクトログラム、(D)はターゲットの女性の標準話者の音声をファジィベクトル量子化したスペクトログラムである。母音部のホルマント構造は、ターゲットの構造に近いところまで変換されている。過渡部分についてもファジィベクトル量子化を用いることで、かなりなめらかに表現されることが観察できる。

このように入力のファジィベクトル量子化がスペクトログラム正規化の高精度化に効果的である、変換コードブック生成に用いる学習単語数の削減に効果があることがわかった。

## 5. まとめ

ベクトル量子化を用いたスペクトログラムの正規化方法に、ファジィベクトル量子化の考えを導入し高精度化を試みた。ファジィベクトル量子化では、各コードベクトルに対する帰属度を用いて入力ベクトルを表現するため、同じコードブックを用いてコードブックサイズを等価的に増加させる効果がある。本論文では、さらにファジィベクトル量子化に高精度化、高速化を目的としたk近傍則の考えを導入し、高精度なベクトル量子化を実現した。これに関して、異話者間に於けるスペクトログラム正規化実験を行い、次の事を示した。

- ファジィベクトル量子化により、話者内のベクトル量子化のスペクトル歪みが約28%低減でき、過渡部分の歪みが減少する。
- 話者間のスペクトログラム正規化もファジィベクトル量子化により改善でき、変換コードブックの生成にファジィベクトル量子化を用いることにより正規化誤差が10%程度減少する。
- 学習単語数について評価した結果ファジィベクトル量子化により変換コードブックの学習単語数が100単語から30単語程度に削減できる。

ファジィベクトル量子化により、コードベクトル間にあるベクトルを表現し、写像する方法を提案し、その有効性をスペクトル歪の観点から確認した。今後の課題としては、声質変換、認識実験等への適用を行っていく予定である。また、本論文では、フレーム毎のベクトル量子化により話者正規化を試みたが、さらに時間長を考慮した、いわゆるセグメント量子化にも適用可能である。

## 6. 謝辞

日頃、ご指導頂く樽松社長、御討論いただいた音声情報処理研究室の皆様に感謝致します。

## 7. 文献

- (1) A.Buzo and H.Martinez and C.Rivea, "Discrete Utterance Recognition Based Upon Source Coding Techniques", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-82, pp539-542 (1982-05)
- (2) 鹿野清宏, "入力音声のベクトル量子化による単語音声認識", 音響学会 音声研資S82-60(1982-12)

- (3) Y.Linde and A.Buzo and R.M.Gray, "An algorithm for Vector Quantizer Design", IEEE Transactions on Communications COM-28,pp84-95(1980-01)
- (4) A.Buzo and A.H.Gray and R.M.Gray and J.D.Markel, "Speech Coding Based Upon Vector Quantization", IEEE Transactions on Acoustics Speech Signal Processing ASSP-28,pp562-574 (1980-10)
- (5) K.Shikano and K.Lee and R.Reddy, "Speaker Adaptation through Vector Quantization", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-86,49.5(1986-04)
- (6) 中島邦夫、高橋真哉、"大語彙音声認識における話者適応化法", 音講論集1-1-6(1983-10)
- (7) 新美康永、小林豊、"ベクトル量子化のコードブックの話者適応化", 音講論集2-5-13(1987-10)
- (8) 石毛俊一、佐野浩司、間野一則、白井克彦、"スペクトルの量子化分布による話者性の記述と話者識別", 音講論集1-3-26(1986-10)
- (9) 山下泰樹、松本弘、"単語音声認識におけるベクトル量子化誤差を利用した話者適応", 信学技報SP87-118(1988-01)
- (10) 古井貞熙、"スペクトル空間のクラスタ化に基づく教師なし話者適応化法", 音講論集2-2-16(1988-03)
- (11) K.Choukri and G.Chollet and Y.Grenier, "Spectral Transformations through Canonical Correlation Analysis for Speaker Adaptation in ASR", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-86,49.9(1986-04)
- (12) R.Schwartz and Y.Chow and F.Kubala, "Rapid Speaker Adaptation using a Probabilistic Spectral Mapping", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-87,15.3 (1987-04)
- (13) 中村哲、鹿野清宏、"ベクトル量子化を用いたスペクトログラムの正規化", 信学技報SP87-17(1987-06)
- (14) L.Zadeh, "Fuzzy Sets", Information Control, Vol8(1965)
- (15) J.Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum 1981
- (16) E.Ruspini, "Numerical Methods for Fuzzy Clustering", Information Sciences, Vol.2(1970)

- (17) J.Dunn, "A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well Separate Clusters", Journal of Cybernetics Vol 3(1974)
- (18) H.Tseng and M.Sabin and E.Lee, "Fuzzy Vector Quantization Applied to Hidden Markov Modeling", Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-87,15.5 (1987-04)
- (19) 杉山雅英、鹿野清宏、"ピークに重みをおいたLPCスペクトルマッチング尺度", 信学論(A)J64-A5 (1981-05)
- (20) 鹿野清宏、相川清明、"Staggered Array DP マッチング", 音響学会 音声研資S82-15(1982-06)
- (21) 鹿野清宏、"エントロピーによる音韻バランスリストの作成", 音講論集3-3-10(1984-10)

Table 1. Optimization of Number of Nearest Neighbor,  $k$  and Fuzziness,  $m$

Iteration	1	2	3	4	5
$k$	256	10	10	6	6
Fuzziness	1.3	1.3	1.5	1.5	1.6
Fuzzy VQ Distortion	0.0598	0.0586	0.0574	0.0571	0.0569



Table 2. Distortion Comparison between VQ and Fuzzy VQ

Method	Close	Open
VQ	0.07192	0.08372
Fuzzy VQ(k = 256,m = 1.3)	0.05610	0.06361
Fuzzy VQ(k = 6, m = 1.6)	0.05356	0.06053

Table 3. Experiment Conditions for Spectrogram Normalization

NO.	VQ Method for Normalization	VQ Method for Mapped Codebook Generation	
		Input	Reference
I	VQ	VQ	VQ
II	VQ	FZVQ	VQ
III	VQ	FZVQ	FZVQ
IV	FZVQ	VQ	VQ
V	FZVQ	FZVQ	VQ
VI	FZVQ	FZVQ	FZVQ

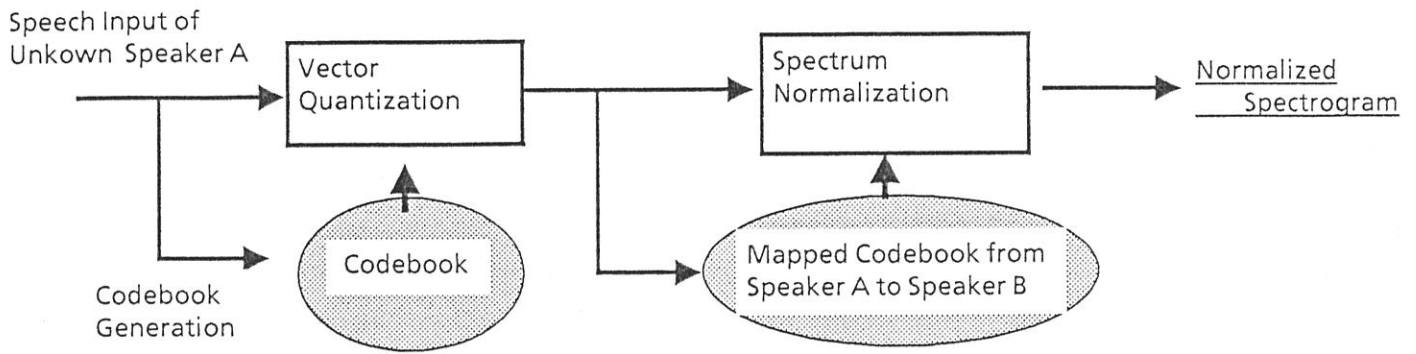


Fig. 1. Block Diagram of Speaker Adaptation through Vector Quantization



Fig. 2. Fuzzy Partitions

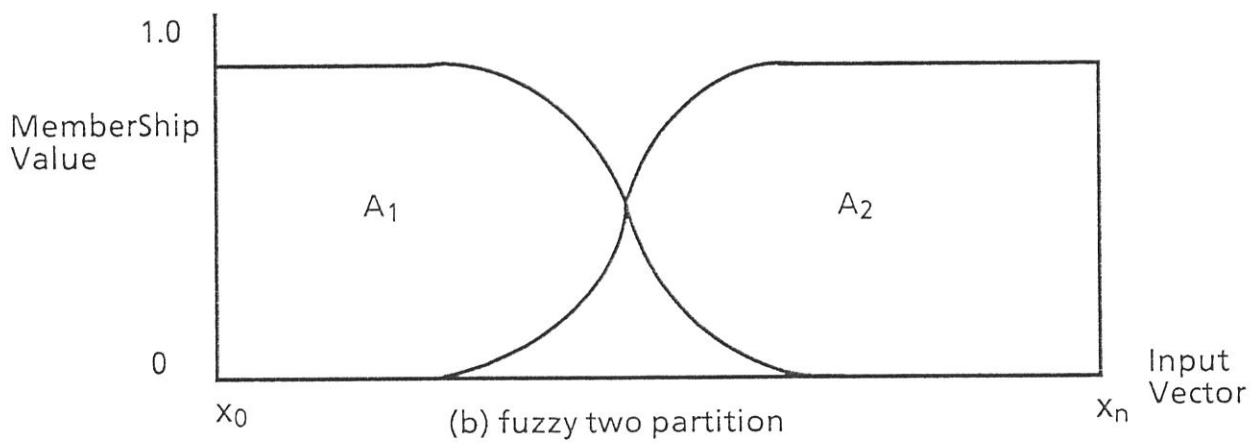
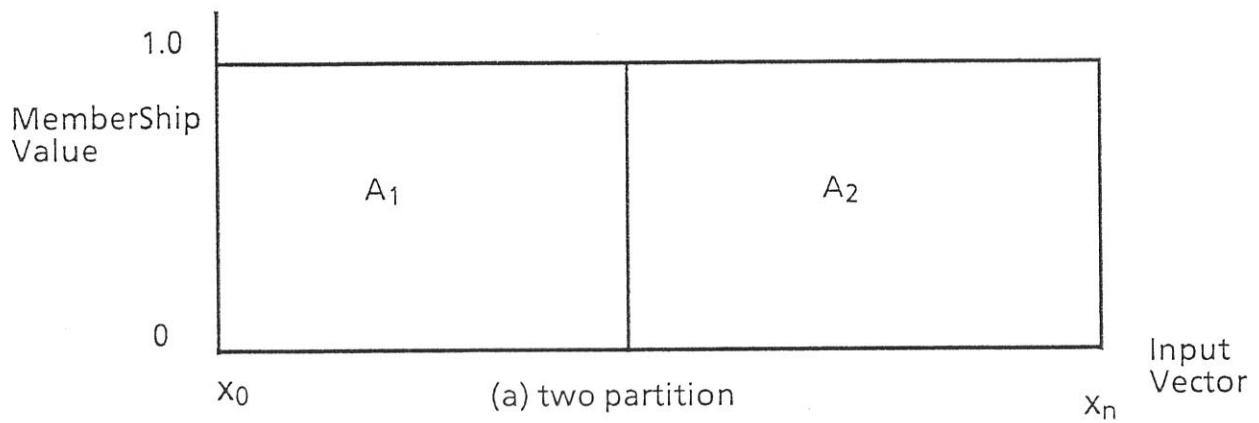


Fig. 2. Fuzzy Partition

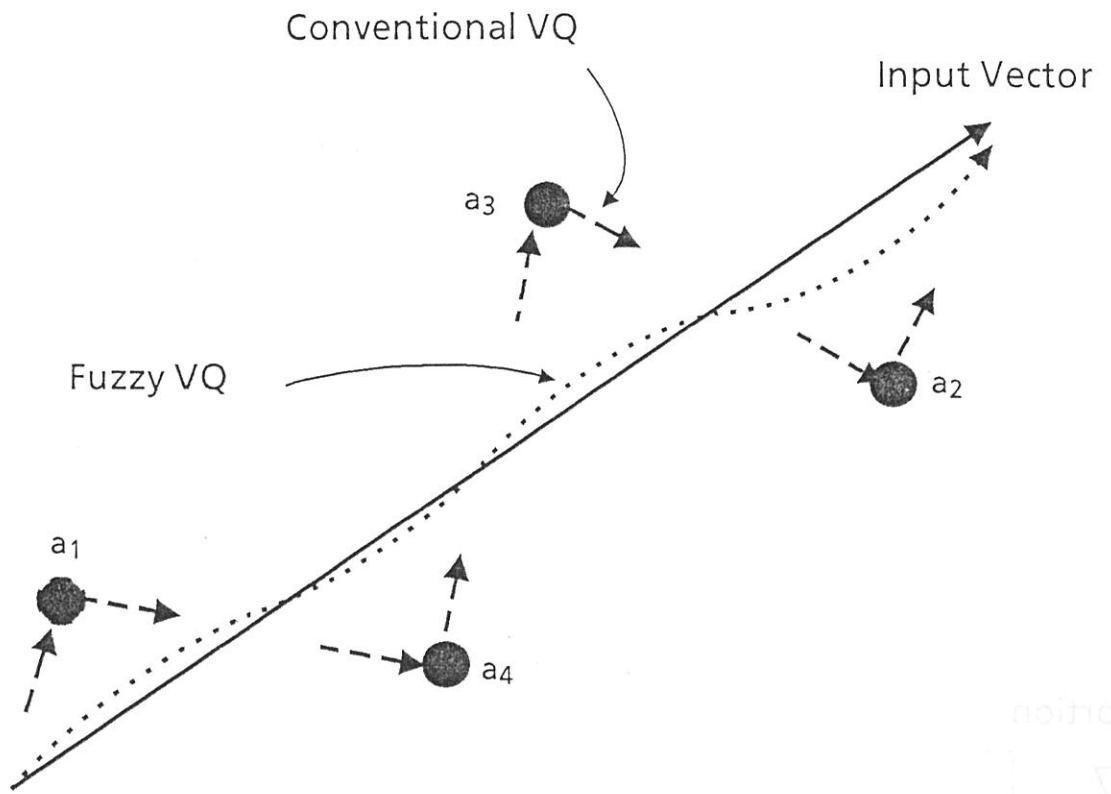


Fig. 3. Comparison between VQ and Fuzzy VQ



Fig. 4. Fuzzy VQ Distortion vs. Number of Nearest Neighbors

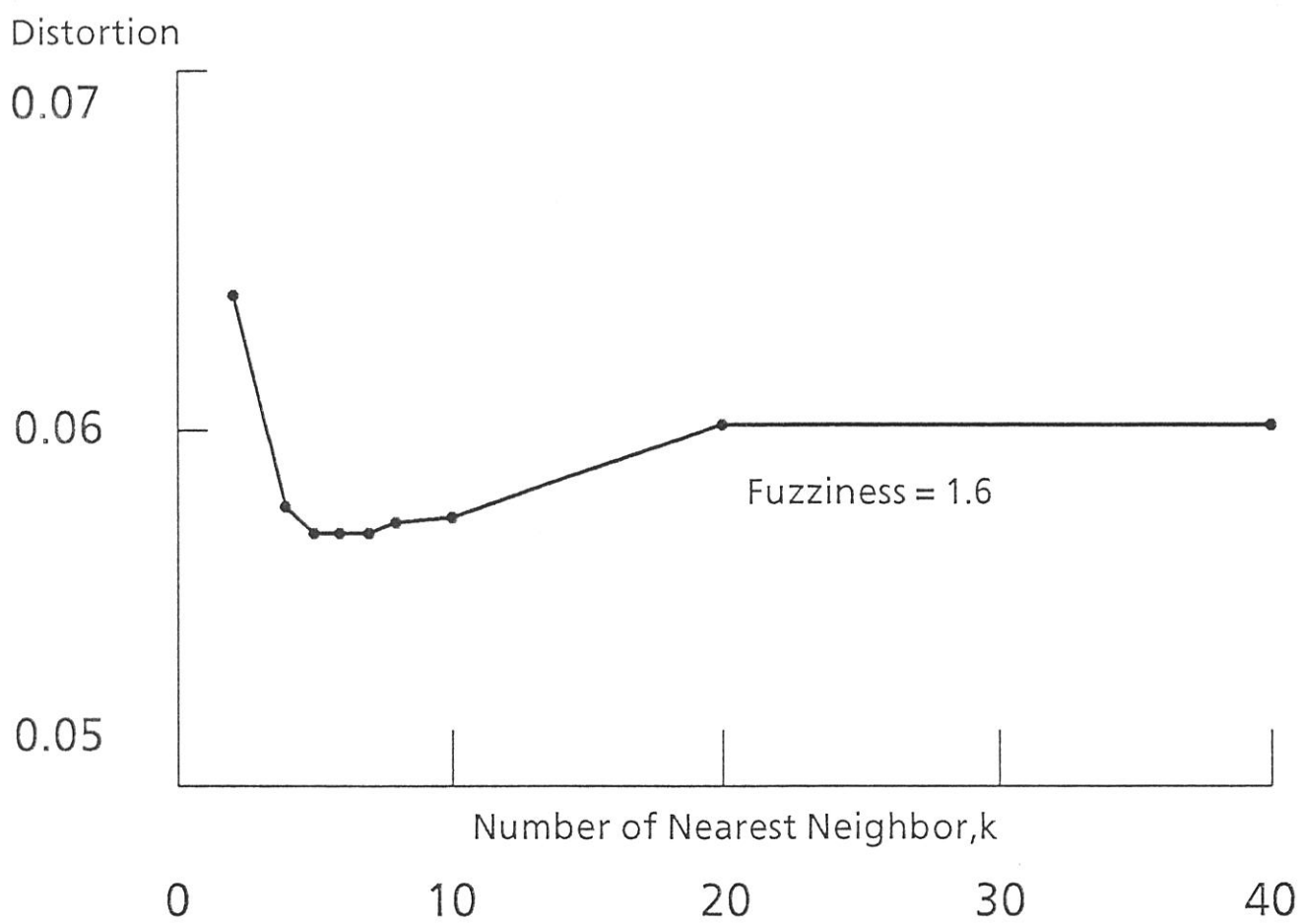


Fig. 4. Fuzzy VQ Distortion vs. Number of Nearest Neighbor,  $k$

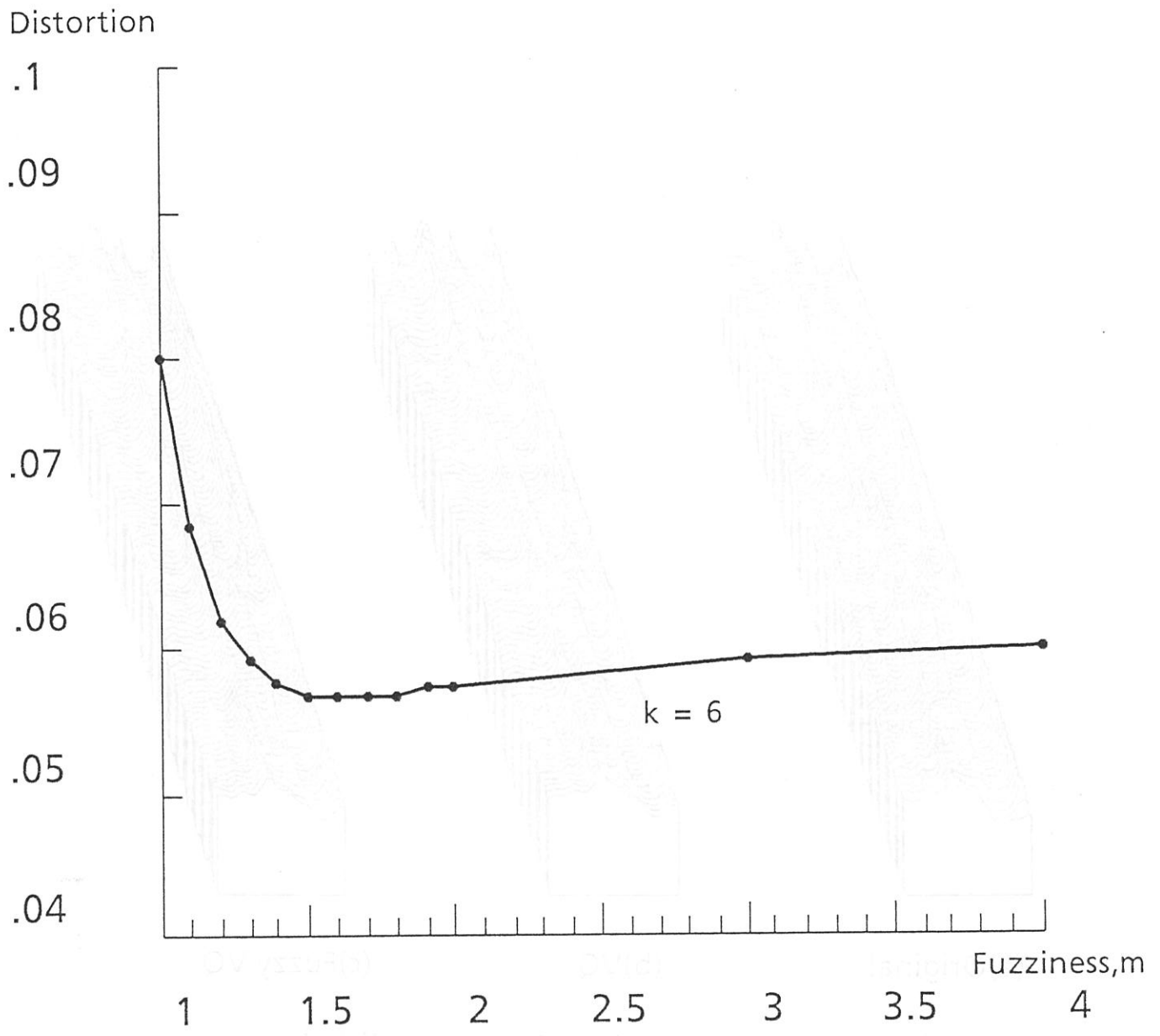
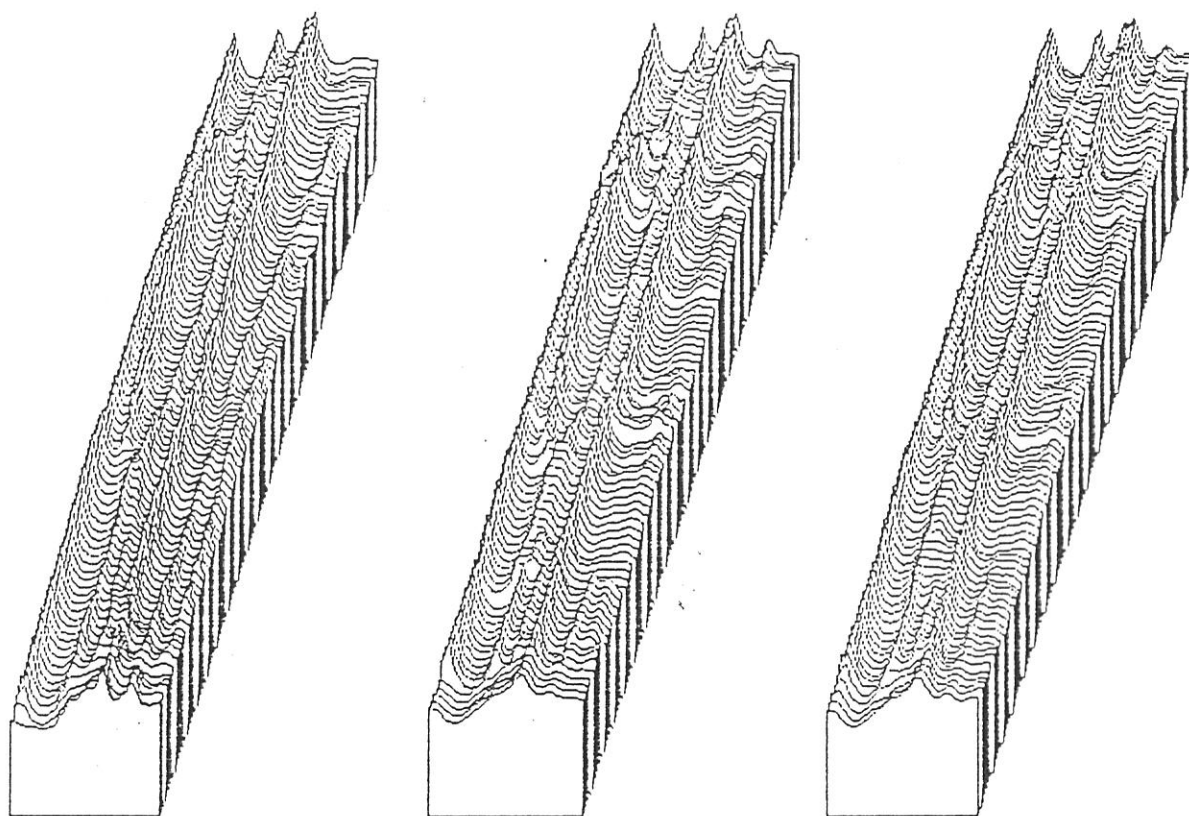


Fig. 5. Fuzzy VQ Distortion vs. Fuzziness,  $m$



(a)Original

(b)VQ

(c)Fuzzy VQ

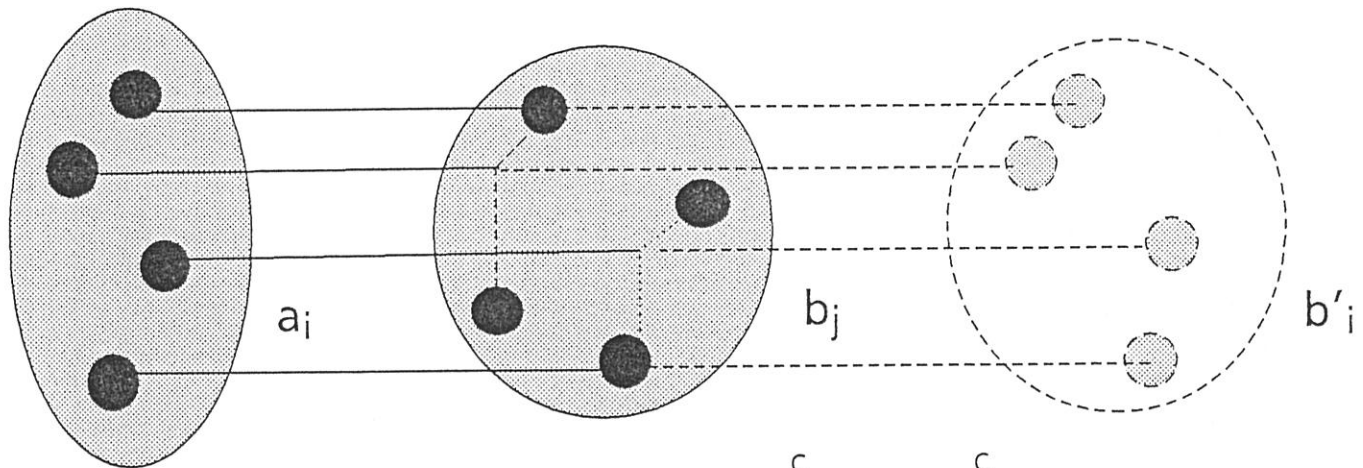
Fig. 6. VQ Running Spectra /ikioi/



Codebook for  
Unknown  
Speaker A

Codebook for  
Standard  
Speaker B

Mapped Codebook for  
Speaker A→B: B'



$$a_i \rightarrow b'_i = \frac{\sum_j h_{ij} \cdot b_j}{\sum_j h_{ij}}$$

- $a_i$  : Vectors of codebook of speaker A
- $b_j$  : Vectors of codebook of speaker B
- $b'_i$  : Vectors of mapped codebook of speaker A→B
- $h_{ij}$  : Histogram obtained by training
- $C$  : VQ codebook size

Fig.7. Basic concept of spectrogram normalization using vector quantization

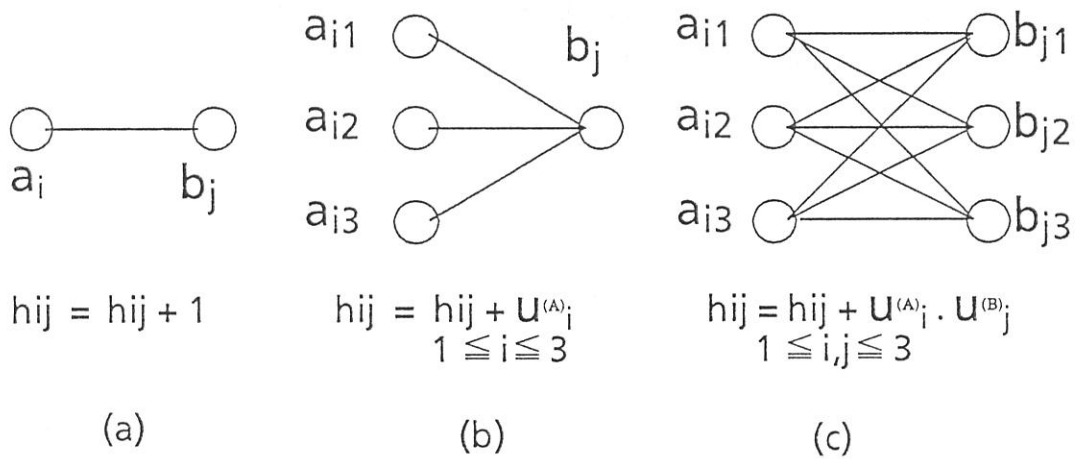


Fig. 8. Histogram Accumulation Using Fuzzy VQ  
 (k = 3)

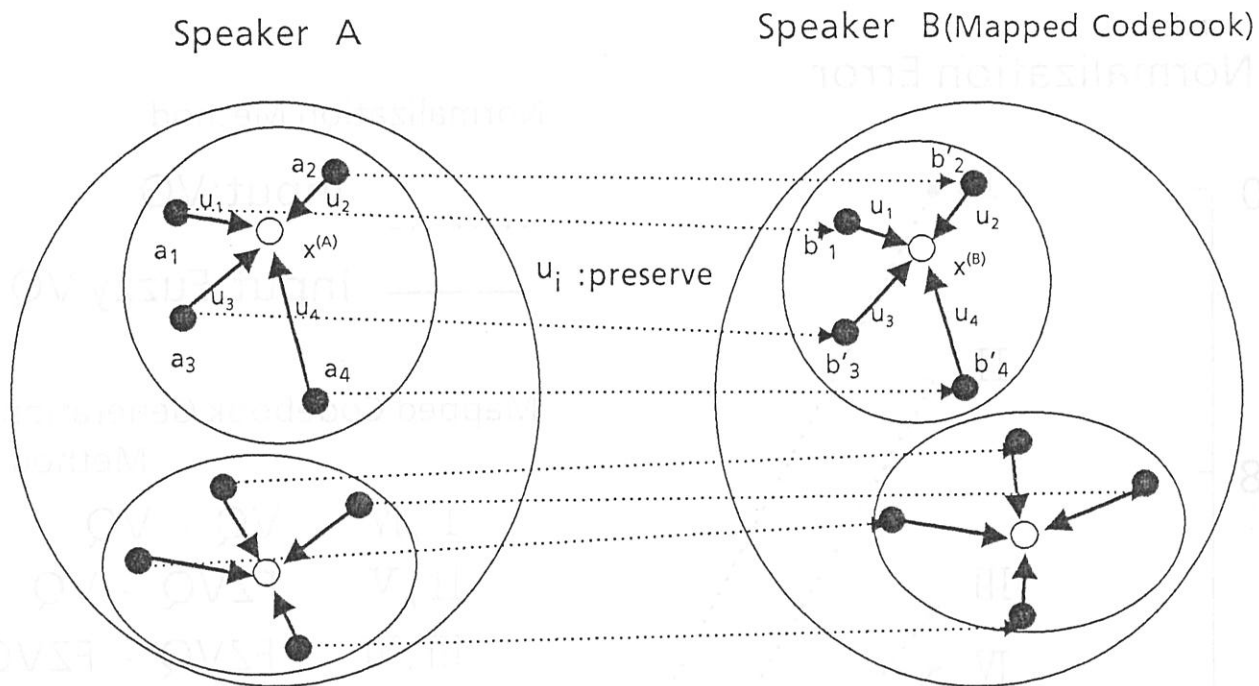


Fig. 9. Concept of Speaker Normalization Using Fuzzy Vector Quantization

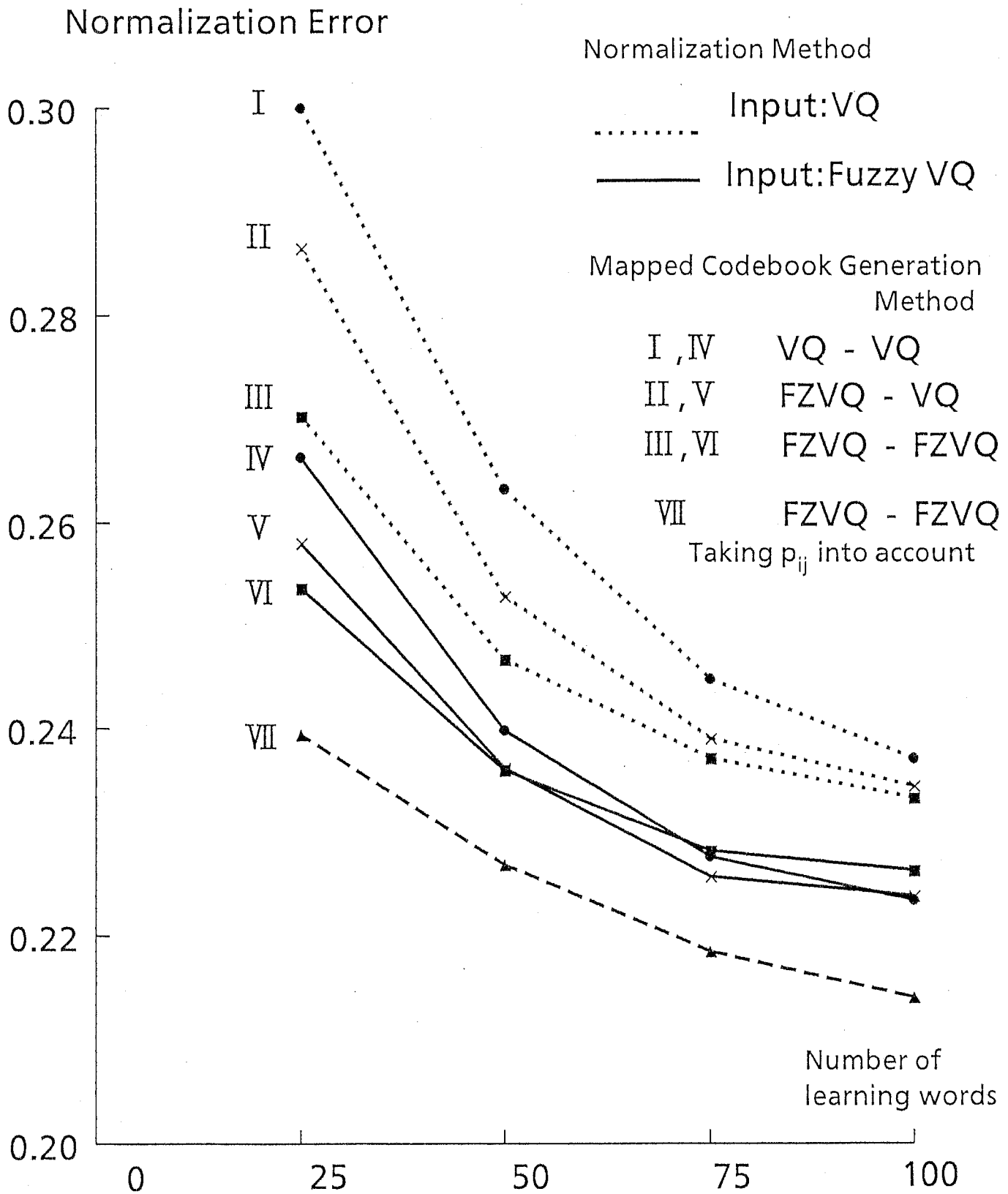
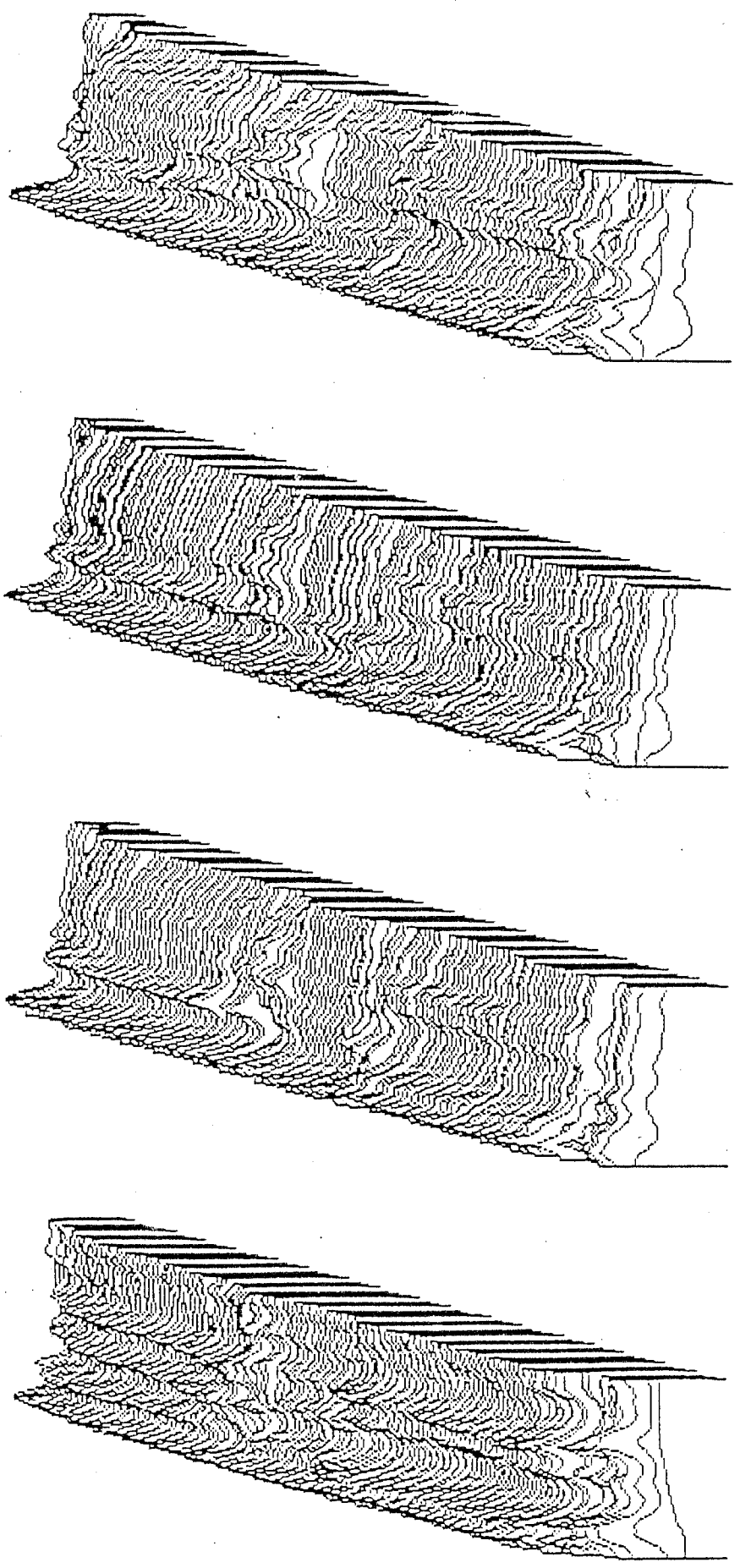


Fig. 10. Comparison of Normalization Methods



(A) Input Speaker (male)      (B) VQ (Method I)      (C) Fuzzy VQ (Method VIII)      (D) Standard Speaker (female)

Fig. 11. Normalized Running Spectra /koonnyuu/