

TR-I-0095

セパレートベクトル量子化に基づく話者適応化

Speaker Adaptation through Separate Vector Quantization

中村 哲, 鹿野清宏

Satoshi NAKAMURA, Kiyohiro SHIKANO

1989. 8

### 概要

ベクトル量子化の手法が、スペクトル空間の離散表現になるという観点から、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正規化する話者適応方法について検討を行う。ここで述べる正規化方法では、ベクトル量子化により有限個のベクトルで話者毎のスペクトル空間を表現した後、予め学習により求めた話者間でのベクトルの対応関係を用いて正規化を行う。本論文では、従来のスペクトルとパワーによる複合スペクトル歪み尺度に基づいた単一のコードブックによるベクトル量子化に対し、スペクトルとパワー毎にコードブックを用いるセパレートベクトル量子化を導入することで、話者内のベクトル量子化歪み及び話者間のスペクトログラム正規化誤差を小さくすることを試みる。評価をスペクトル歪みおよび認識実験で行い、セパレートベクトル量子化が有効であることを示す。

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

1. まえがき .....	1
2. ベクトル量子化 .....	2
2.1. セパレートベクトル量子化 .....	2
2.2. 量子化歪みの検討 .....	3
3. 話者適応 .....	5
4. 話者適応評価実験 .....	6
4.1. 学習の収束について .....	6
4.2. スペクトル歪みによるスペクトログラム正規化の評価 .....	7
4.3. 単語認識実験による評価 .....	9
5. まとめ .....	9
6. 謝辞 .....	10
文献 .....	10

## 1. まえがき

特定話者の音声認識では、前もって発声者がすべての単語を発声、登録することによって、数百語程度の孤立単語認識、および数十単語からなる連続単語音声認識が実現されつつある。ところが、不特定話者の音声認識では、百単語程度の孤立単語音声認識が実現されている程度で、技術的に未解決の問題を多く含んでいる。

これらの不特定話者音声認識の問題は、大別すると主に次のような発声者間の音声の違いに起因すると思われる。

(1) 発声機構のちがいによる差(スペクトルの差)。

(2) 発声方法、調音のちがいによる差。

上記の2つの問題は、非常に多くの問題を含み、互いに深く関連している。(1)の発声機構の違いを正規化するアプローチとしては、声道長の正規化(1)や、決定論的な特徴変換、母音定常部の学習(2)、平均化などが行われている。

また、(2)の発声方法、調音の正規化に関しては、マルチテンプレート法(3)、ネットワークモデルなどが研究されている。その他、理解のプロセスを人間の認知過程として捕らえるアプローチ、たとえばスペクトログラムリーディングによる特徴ベースの連続音声認識の研究(4)(5)も行われている。

一方、ベクトル量子化の手法や、Hidden Markov Model(HMM)が音声認識に適用されるようになり、音声の離散的取扱いの新しい可能性が開かれた(14~20)。ベクトル量子化は、スペクトル歪み最小のもとでスペクトル空間を有限個のスペクトルで表現する手法である。このベクトル量子化を用いた新しい話者適応の方法が提案され始めた(14)(15)。これは、話者適応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とらえる方法である。文献(15)で提案された学習に基づく方法では、まず発声者ごとに音声をベクトル量子化しコードブックを生成する。その後、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コードブック間の対応関係を求めて、話者適応を行うものである。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。また、文献(19)では話者に共通のコードブックを用いてHMMモデルの遷移確率および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、良好な結果を得ている。また文献(16)では話者毎のベクトルの線形変換問題として変換行列を求め話者適応を試みている。

本論文では、文献(15)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ベースの音声認識等の前処理としてのスペクトログラムの話者正規化について述べる。スペクトログラムの正規化という観点からい

ば、文献(15)のベクトル量子化を介した話者適応の方法には、次の問題点がある。

- (1) スペクトルとパワーからなる複合歪み尺度によるベクトル量子化を用いており、スペクトログラムの再現精度に問題がある。
- (2) 複合歪み尺度のために歪みを小さくしようとする、コードブックが大きくなる。

本論文では、これらの問題を解決すべく音声を特徴毎に別々の量子化(スカラーまたはベクトル量子化)を行うセパレートベクトル量子化を介した話者適応のアルゴリズムを提案する。さらに、この話者適応化によるスペクトログラムの正規化の有効性を離散発声の単語を対象に確かめる。

本アルゴリズムは、次の構成となっている。

- (a) 複数のコードブックをもつセパレートベクトル量子化を用いて、話者毎のスペクトル空間を有限個のベクトルで離散表現する。
- (b) 学習を用いて未知話者から標準話者への変換コードブックを求め、スペクトログラムの正規化を行う。

この概略のブロック図を図1に示す。

## 2. ベクトル量子化

本章では、セパレートベクトル量子化の有効性を示すため、パワーを考慮したWLRスペクトル歪み尺度(PWLR)(10)を用いた従来のベクトル量子化との比較実験を行い、量子化歪みについて検討する。

### 2.1. セパレートベクトル量子化

ベクトル量子化の手法は、計算時間及びメモリーを大幅に削減する方法として知られている。この方法ではスペクトル歪み最小の原理により信号を量子化するものであり、現在までに種々の改良が試みられてきた。これらの改良はFull Search Coding (FSC)(6)に対して計算量を減少させるBinary Search Coding(7)など、また情報源の構造を積極的に利用したものなど数多くのもが見られる。音声認識の分野においても近年多用されており、多くの成果を収めている。しかし、これらのほとんどは、音声認識に適したスペクトル尺度をつくり、これらにより音素、単語をベクトル列に変換し、マッチングを行うといった形態であった。

本稿では、これらの研究の流れに対し複数のコードブックを生成し、これらを用いることによりさらに小さなスペクトル歪みを実現する方法を提案する。この様な方法については、Product Codeのベクトル量子化の中のGain Shape

Vector Quantization(7)や Separating Mean Vector Quantization(8)の考えの中に類似の方法を見ることができる。本方法は、上記の方法と同様、Product Codeのベクトル量子化の一種と考えられるが、縦続構成となっておらず特徴毎に全く独立な量子化を行う並列構成となっている点が異なっている。ここでは、このような特徴量の取扱いを重視し、本方法をセパレートベクトル量子化(SPVQ)と呼ぶ。

コードブック生成のアルゴリズムとしてはLindeなどによって提案された準最適だが強力なFull Search CodingのLBGアルゴリズム(6)を用いる。また、スペクトル歪みの尺度として音声認識ではユークリッド距離や、ケプストラム距離なども用いられるが、ここではピーク重み付け距離尺度について考える。文献(9)で提案されているピーク重み付け距離尺度WLRは、音声のフォルマント特徴を強調する尺度となっており、単語音声認識において高い性能を示している。さらにパワー込み重み付け距離尺度PWLRを用いればさらに高い性能で音声認識が行えることが示されている(10)。

次に、今回用いたセパレートベクトル量子化の原理を示す。図2のようにセパレートベクトル量子化では、入力音声に対し、パワー、及びLPCパラメータを抽出し、この後、別々にコードブック生成する。さらに、2つのコードブックで別々に符号化を行い、2つのベクトルコード列を出力する。LPCパラメータについては、WLRのスペクトル尺度によるベクトル量子化を行なう。パワーの量子化についてもLBGアルゴリズムによるコードブック生成手法を用いるので、不均一スカラー量子化となる。また、このセパレートベクトル量子化により、入力音声の量子化を行い、PWLRによるスペクトル歪みと、このセパレートベクトル量子化のパワー部とWLRのスペクトル部の歪みの和との比較を行なう。以後、PWLRによる従来のベクトル量子化を[PWLR]、今回提案するセパレートベクトル量子化の場合を[P][WLR]と表すことにする。

## 2.2. 量子化歪みの検討

今回の実験に用いた分析条件を表1に示す。

音声データとして、可能な2音素連鎖をすべて含み音韻バランス(13)を考慮した216単語を訓練された女性3名が発声したものをを用いる。

ベクトル量子化の比較実験では、上記216単語のうちの最初の100単語から5000フレームを取り出し、コードブック生成を行い、スペクトル歪みの観点から比較実験を行なう。

また、比較に用いたPWLR スペクトル尺度は次の式で表される。

$$d = \sum (C_n - C'_n)(R_n - R'_n) + \alpha \cdot (P/P' + P'/P - 2) \cdots \cdots (1)$$

但し、 $\alpha=0.01$  とする。

また、 $C_n, C'_n$  :  $n$ 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

$P, P'$  : パワー

伝送ビット数、記憶容量の条件を変えた時の平均の量子化歪みを比較する。セパレートベクトル量子化のスペクトル歪みとしてパワー、WLRとも同じコードサイズとして各々のベクトル量子化で求めたスペクトル歪みの和を用いる。この結果を表2に示す。ここで、closeとはベクトル量子化のコードブック生成に用いた学習データに対するもの、openとはそれ以外の100単語に対するものである。

この結果から次のことがわかる。

- (1) 伝送ビット数で見た場合は、8ビットの伝送ビット数の256コードサイズの[PWLR]と、各々4ビットの16コードサイズの[P][WLR]との比較となり、[P][WLR]の方が学習データ及び未知データに対して2倍程度歪みが大きい。
- (2) コードブックの記憶容量で見た場合は、13.3kB必要な256コードサイズの[PWLR]と各々256コードサイズの[P][WLR]の比較となり、[P][WLR]の方が、学習データに対して19.8%、未知データに対して20.1%歪みが小さい。また、128コードサイズの[P][WLR]を用いれば[PWLR]と同じ歪みで記憶容量を半分の6.6kBまで小さくできる。

ベクトル量子化は、符号化の立場からすると伝送ビット数が重要な要素となり、伝送ビット数の増加は重大な問題となる。しかし、音声認識の立場では、むしろ記憶容量や計算量が重要な要素となる。セパレートベクトル量子化[P][WLR]は、(2)の結果が示すようにコードブック記憶容量の削減が可能になるので、従来の[PWLR]ベクトル量子化によりも有効と言える。

さらに、PWLR スペクトル尺度では、その有効性はパワー項とWLR項の相関の存在に基づいて伝送ビット数を小さくできる点にあった。しかし、この方法ではコードブックの記憶容量はスペクトルに要する記憶容量とパワーに要する記憶容量の積になり、精密にスペクトルを量子化するには、大きいコードブックが必要となる。よって、コードブックの記憶容量、コードブック生成に必要な学習データおよび計算時間は莫大になる。これに対して、セパレートベクトル量子化では、記憶容量はパワーに要する記憶容量とスペクトルに要する記憶容量の和になる。このため、コードブックの記憶容量、コードブック生成の学習データ、計算時間を少なくできる。

次に図3にスペクトログラムの出力例を示す。出力例は単語/ikioi/の/kioi/の部分である。ベクトル量子化の際のコードサイズは、従来法[PWLR]が256、セパレート量子化[P][WLR]に対してはパワー、スペクトルとも同じ256とした。こ

れをみると、従来のベクトル量子化[PWLR]による方法では、不連続な部分が目立っている。スペクトル歪みによる比較では、セパレートベクトル量子化[P][WLR]の方が[PWLR]より1/3程度歪みが小さい。

### 3. 話者適応

本章ではベクトル量子化を用いた話者適応の方法として、文献(15)により提案されたText Dependentな学習アルゴリズムによる方法に基づいて検討を行う。

指定された単語を未知話者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(11)を用いてベクトルコード列のレベルでマッチングを行い、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、未知話者Aのコードブックと、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、図4に示すように、未知話者Aのベクトル $a_i$ を、対応づけのヒストグラム $w_{ij}$ 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 $b'_i$ へ対応づける。この線形結合で構成されたベクトル $b'_i$ 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。

本方式のブロック図を図5に示す。本方式は、大きく2つのステップ、つまり、学習のステップと適応のステップより構成されている。この各々の動作について説明を行う。

#### (1) 変換コードブックの学習(学習のステップ、図5(a))

学習では、まず未知話者Aの入力音声をベクトル量子化する。学習の過程は、次のアルゴリズムに従う。

- (step1) 未知話者Aの音声のコードブックを生成する。所定の学習サンプル量に達するまで音声を取り込み、LBGアルゴリズムに従ってコードブックを生成する。
- (step2) 学習に用いる単語についてstep1で生成したAのコードブックを用いてAの音声をベクトル量子化する。
- (step3) 学習用単語の全てについて、未知話者Aと標準話者Bのベクトルコード列を端点フリーのスタガードアレイ型DPマッチング(12)をDouble Split法により用い、ベクトル間の対応を表すヒストグラムからAからBへの変換コードブックを求める。
- (step4) Aのベクトルを更新し、所定回数、またはある収束条件を満たすまでstep3の処理を繰り返す。

#### (2) 話者適応(適応のステップ、図5(b))

適応の第一段階として時間軸整合なしのスペクトルの適応を考える。さらに、評価のために未知話者Aの発声単語を既知として固定の標準話者Bの同一単語

とマッチングし、時間軸を整合したスペクトログラムも出力する。

スペクトルの正規化のための話者適応の過程は次のアルゴリズムに従う。

- (step1) 上記の学習過程により未知話者Aと標準話者B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) ベクトル量子化により出力されたベクトルコード列をAからBへ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(スペクトルの正規化)。
- (step4) 評価のため、標準話者Bの同じ単語とDPマッチングし、時間軸整合を行った正規化スペクトログラムを出力する(時間軸の正規化)。

以上により、正規化されたスペクトログラムを得ることができる。

#### 4. 話者適応評価実験

本章では適応アルゴリズムの実験結果の評価、検討をスペクトル歪み、単語認識実験により行う。

##### 4.1. 学習の収束について

この節では、ベクトル量子化を用いた適応アルゴリズムの収束について検討する。話者適応アルゴリズムは、AのベクトルをBの空間に写像し、Bのベクトルの線形結合で表し、このベクトルを新たにAのベクトルとして再度対応を求めていく繰り返し学習アルゴリズムとなっている。

繰り返しアルゴリズムによる学習の効果を評価する基準としてDTW後のDTW距離を正規化誤差として用いる。正規化誤差は、学習単語すべてについて、標準話者Bのベクトル量子化後の単語パターンと、学習によって得られた変換コードブックで正規化された未知話者Aの単語パターンとをSADPでマッチングする。このSADP距離の平均を平均正規化誤差として評価に用いる。

実験では、セパレートベクトル量子化に於けるパワー[P]およびスペクトル情報[WLR]は各々256コードサイズ、従来法についてもコードサイズ256としてベクトル量子化を行なう。

未知話者と標準話者が同性の場合と異性の場合の効果を比較するために、学習用100単語を用いた実験を行なう。未知話者を女性と男性各1名、標準話者に女性を選んだ場合の学習の繰り返しに対する収束度を平均正規化誤差でみた場合の結果を図6に示す。

平均正規化誤差は、1回目の学習で女性で学習前の57.7%に、男性で23.6%に減



少した。4回目までに女性で56.2%に、男性で20.7%に減少した。平均正規化誤差は1回の学習で大きく減少し、2回目以降ではゆるやかに減少する。全話者の平均では2回目までに未適応の場合の40%程度に減少し、その後緩やかに収束するという結果を得た。1回目の学習では未知話者Aのコードブックのベクトルが標準話者Bの線形結合で表される変換コードブックに写像されるため大きな正規化誤差の減少が見られる。2回目以降の学習からは変換コードブック内での収束ということで、1回目とは意味が違うが、変換コードブックを用いてさらにDTWによる学習を行うことにより、DTWマッチングの対応づけが正確になり、変換コードブックの性能を向上させることができる。この結果は、2回以降の変換コードブック内の繰り返しでも4回目まで程度で十分収束することを示している。

#### 4.2. スペクトル歪みによるスペクトログラム正規化の評価

この節では、ベクトル量子化を用いたスペクトログラム正規化結果について検討する。未知話者の発声した単語音声に対し、まず未知話者の音声から生成したベクトル量子化のコードブックを用いてベクトル量子化を行い、ベクトルのコード列を求める。次に、学習アルゴリズムで求めた変換コードブックの中で対応するベクトルのコードを捜し、変換コードブックの特徴ベクトルとフレームごとに入れかえる。ここでは、ベクトル量子化を用いたスペクトログラムの正規化法により、効果的にスペクトログラムの話者間正規化ができること、従来のベクトル量子化を用いた場合に比べて、セパレートベクトル量子化を用いることでさらに効果的にスペクトログラムの話者間正規化ができることを示す。

実験では、コードブックの記憶容量を同じにするため、セパレートベクトル量子化におけるパワー[P]およびスペクトル情報[WLR]を各々256コードサイズ、従来のベクトル量子化を256コードサイズでベクトル量子化する。学習については、十分に学習が収束した変換コードブックを用いることが望ましいので、繰り返し回数を一様に10回とする。また、話者については、ベクトル量子化比較実験で用いた女性話者3名の他に男性話者3名(アナウンサー)を加え、同性間、異性間の正規化効果の検討も行う。

評価は、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換えを行ない、スペクトログラムの正規化を行う。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。この正規化処理の流れと正規化誤差を図7に示す。

男女各3名から未知話者、標準話者2名を選び全ての組合せについて、正規化誤差を求め平均した結果を表3に示す。正規化後では学習データでは[PWLR]では正規化前の37.6%に、[P][WLR]では37.7%に、未知データでは[PWLR]では正規化前の51.4%に、[P][WLR]では50.2%に正規化誤差が減少した。

このように、この方法により平均正規化誤差を半分程度まで縮めることができる。また、従来のベクトル量子化による方法と、セパレートベクトル量子化の比較では学習データで2.3%、未知データで7.7%だけ平均正規化誤差が、セパレートベクトル量子化を用いた方法の方が小さくなり、効果が認められた。これは、話者内でのベクトル量子化を精度よく行うことで、話者間の正規化に用いる変換コードブックも正確なものとなっていることを示しており、これにより話者間の正規化においてもセパレートベクトル量子化の効果を認めることができる。

次に表4に未知話者、標準話者を性別に分類した結果を示す。まず、正規化前では、未知話者と標準話者が異性の場合には同性の場合に比べて2倍程度正規化誤差が大きい。これは、男女間のフォルマント周波数の違いが同性間に比べて大きいことから考えて当然の結果である。

正規化処理の効果としては正規化前に比べて同性間の場合の30.5%の正規化誤差の減少に対し、異性間の場合には57.5%の正規化誤差減少がみられ、特に異性間において大きな効果があることがわかった。正規化後では、異性間の場合でも正規化誤差が同性間と同じ程度まで小さくなっている。これにより、異性間の差を同性間の差と同程度まで縮めるということから本方法の正規化効果は非常に高いといえる。

次に、実際の正規化出力をランニングスペクトルによる出力例によって正規化の効果を検討する。図8は未知話者を男性とし、標準話者を女性とした場合の単語/koonyuu/の例である。各図は次の様なものである。

- (a) 未知話者Aの音声を256コードサイズのセパレートベクトル量子化し標準話者Bへ時間軸変換したランニングスペクトル。
- (b) 未知話者Aの音声を変換コードブックを用いてスペクトル変換した後、標準話者Bへ時間軸変換した正規化ランニングスペクトル。
- (c) 標準話者Bの音声をLPC分析し256コードサイズのセパレートベクトル量子化したランニングスペクトル

この結果をまとめると、次の様になる。

- (1) 未知話者Aの音声をDTWによる時間軸変換だけ行った(a)は、基準となる標準話者Bのスペクトログラムとはかなり異なったものとなっている。
- (2) セパレートベクトル量子化し、適応を施した(b)では、基準の(c)に非常に近く変換されている。特に、第2モーラの/u/で、第2フォルマント周波数が標準話者側に上がり、第1モーラの/o/で、第3、第4フォルマントが標準話者側に上がり、つぶれて1つになっているところまで変換されている。

以上より、実際のセパレートベクトル量子化スペクトログラムでも正規化が良く働くことがわかった。過渡部については、ベクトル量子化のコードサイズの増加である程度の対処が可能と思われる。

#### 4.3. 単語認識実験による評価

ここでは、スペクトログラムの話者間正規化の評価として、離散単語の認識率を用いて性能評価を行う。

スペクトログラムの正規化処理の効果を実際の離散単語の認識により評価する。実験条件は4.2節と同じもので、[PWLR]に対し256のコードサイズ[P][WLR]についてはそれぞれ256のコードサイズとし、Double SplitによるDTWを用いる。認識対象は音韻バランス216単語の前半100単語を学習に用い、その他100単語を評価用の未知データとした。話者は、男女各3名である。

表5にこの結果を示す。これは、未知話者、標準話者を全ての組合せで求めた平均の認識率である。正規化後では正規化前に比べて学習データでは[PWLR]で15.7%、[P][WLR]で16.6%、未知データでは[PWLR]で12.9%、[P][WLR]で13.9%だけ高い単語認識率を得られている。正規化を行うことによりセパレートベクトル量子化では学習データに対し96.7%、未知データに対し94.1%の単語認識率を得られている。また、従来のベクトル量子化との比較ではセパレートベクトル量子化は従来の[PWLR]によるベクトル量子化に比べて学習データで1.1%だけ、未知データで1.0%高い単語認識率を得られている。これは、話者内のスペクトル空間をセパレートベクトル量子化で精密に表現することによって、コードブックのベクトルの対応づけがさらに精密になり話者間の正規化においても高い認識率を得られることを意味している。

次に表6に未知話者、標準話者を性別に分類した結果を示す。未知データに対する結果については、正規化前では同性間の場合平均で93.5%、異性間の場合67.0%の単語認識率であった。正規化後の認識率については同性間の場合平均で95.9%であり、正規化による効果は2.4%であった。これに対し、異性間の場合平均で92.3%の認識率を得、正規化により25.3%の認識率向上がみられ、同性間に匹敵する認識率を得た。

#### 5. まとめ

ベクトル量子化を用いた話者適応の手法をスペクトログラムの正規化に適用した。従来の複合尺度により単一のコードブックを生成する方法に対して、特徴ベクトル毎に複数のコードブックを用いるセパレートベクトル量子化を用いる方法を提案し、比較検討を行った。さらに、これを用いて、話者適応、スペクト

プログラムの正規化を行った結果についても検討を行い、この結果次のことがわかった。

- (1) ベクトル量子化では、従来のPWLR スペクトル尺度で単一のコードブックを生成しベクトル量子化する方法に対し、パワーとスペクトル(WLR尺度)を別々にベクトル量子化するセパレートベクトル量子化を用いることで、コードブックの記憶容量が同じ13.3KBの場合、学習データで19.8%、未知データで20.1%だけ平均スペクトル歪みが減少することがわかった。
- (2) セパレートベクトル量子化を用いてスペクトログラムの正規化を行ない、実際の正規化スペクトログラム出力の観察の結果、定常スペクトルについては、うまく変換されることがわかった。また、DTWを用いて時間整合した後の距離によって話者適応アルゴリズムを用いたスペクトログラムの正規化評価を行ったところ、正規化効果により学習データで正規化前の37.7%に、未知データで50.2%に平均誤差が減少し、従来のベクトル量子化を用いた場合に比べて、正規化後で学習データで2.3%、未知データで7.7%だけ正規化誤差減少効果があった。また、単語認識率では正規化処理により学習データで96.7%、未知データで94.1%の認識率を得、また従来のベクトル量子化を用いた場合に比べて正規化後で学習データで1.1%、未知データで1.0%高い認識率を得、本方式の有効性を確認した。

また、今後、動的特徴の利用、音韻情報の利用を行って、スペクトログラムの正規化、話者適応のアルゴリズムの高精度化をさらに進める予定である。

## 6. 謝辞

日頃御指導頂くATR自動翻訳電話研究所 樽松社長、御討論頂いた音声情報処理研究室の皆様、感謝いたします。

## 文献

- (1) 古井 貞熙:”音声波に含まれる個人性情報の研究” 東京大学学位論文(1976-04)
- (2) 杉山雅英、鹿野清宏:”母音標準パターンの教師なし学習法” 音響学会 音声研資S83-48(1983)
- (3) L.R.Rabiner, S.E.Levinson, A.E.Rosenberg, J.G.Wilpon :”Speaker-Independent Recognition of Isolated Words Using Clustering Techniques” IEEE Transactions on Acoustics Speech Signal Processing ASSP-27, pp336-349 (1979-08)
- (4) V.W.Zue, L.F.Lamel :”An Expert Spectrogram Reader:A Knowledge Based Approach to Speech Recognition” Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-86,23.2 (1986-04)

- (5) R.Mizoguchi, K.Tsujino, O.Kakusho : "A Continuous Speech Recognition System Based on Knowledge Engineering Techniques" Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-86,23.8(1986-04)
- (6) Y.Linde,A.Buzo,R.M.Gray:"An algorithm for Vector Quantizer Design" IEEE Transactions on Communications COM-28,pp84-95(1980-01)
- (7) A.Buzo, A.H.Gray, R.M.Gray, J.D.Markel : "Speech Coding Based Upon Vector Quantization" IEEE Transactions on Acoustics Speech Signal Processing ASSP-28,pp562-574 (1980-10)
- (8) R.L.Baker, R.M.Gray : "Differential Vector Quantization of Automatic Imagery" Proceedings of the International Picture Coding Symposium (1983-04)
- (9) 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)
- (10) 相川清明、鹿野清宏、古井貞熙 : "パワー情報で重みづけた距離による単語音声認識" 音響学会 音声研資S81-59(1981-12)
- (11) 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)
- (12) 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)
- (13) 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)
- (14) 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)
- (15) K.Shikano, K.Lee, 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)
- (16) K.Choukri, G.Chollet, 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)
- (17) 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)

- (18) 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)
- (19) R.Schwartz, Y.Chow, 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)
- (20) M.Nishimura, K.Toshioka :”HMM-Based Speech Recognition Using Multi-dimensional Multi-labeling” Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing ICASSP-87,27.11 (1987-04)

Table 1. Analysis Conditions

Sampling Frequency	12kHz
Analysis Window	21.3 msec
Analysis Interval	3 msec
Window Function	Hamming Window
Analysis	14-order LPC

Table 2. Spectral Distortion Comparisons between Conventional and Separate VQ

Vector Quantization Method	Bits	Memory	distortion close	distortion open
PWLR Conventional VQ[256]	8 bits	13.3 kB	0.08015	0.09102
Separate VQ [16][16]	4 + 4 8 bits	0.8 kB	0.16778	0.17138
Separate VQ [32][32]	5 + 5 10 bits	1.6 kB	0.12618	0.12890
Separate VQ [64][64]	6 + 6 12 bits	3.3 kB	0.09661	0.10153
Separate VQ [128][128]	7 + 7 14 bits	6.6 kB	0.07985	0.08724
Separate VQ [256][256]	8 + 8 16 bits	13.3 kB	0.06428	0.07276

Table 3. Comparisons of Average Normalization Errors between VQ and SPVQ

NL error	[PWLR]close	[PWLR]open	[P][WLR]close	[P][WLR]open
Before Adaptation	0.8251	0.9020	0.8030	0.8525
After Adaptation	0.3103	0.4636	0.3031	0.4277

Table 4. Average Normalization Errors by SPVQ

NL error		Before Adaptation		After Adaptation	
Sample	sex	female	male	female	male
close	female	0.4657	1.1204	0.2494	0.4095
close	male	1.0442	0.5816	0.2689	0.2846
open	female	0.4343	1.3426	0.3556	0.6130
open	male	1.0960	0.5371	0.4231	0.3191

Table 5. Word Recognition Rates by Speaker Adaptation(%)

R.R	[PWLR]close	[PWLR]open	[P][WLR]close	[P][WLR]open
Before Adaptation	79.9	80.2	80.1	80.2
After Adaptation	95.6	93.1	96.7	94.1

Table 6. Word Recognition Rates by SPVQ(%)

R.R		Before Adaptation		After Adaptation	
Sample	sex	female	male	female	male
close	female	89.0	56.4	96.0	96.0
close	male	79.7	95.2	96.4	98.5
open	female	90.8	54.1	93.5	93.6
open	male	79.8	96.2	90.9	98.3



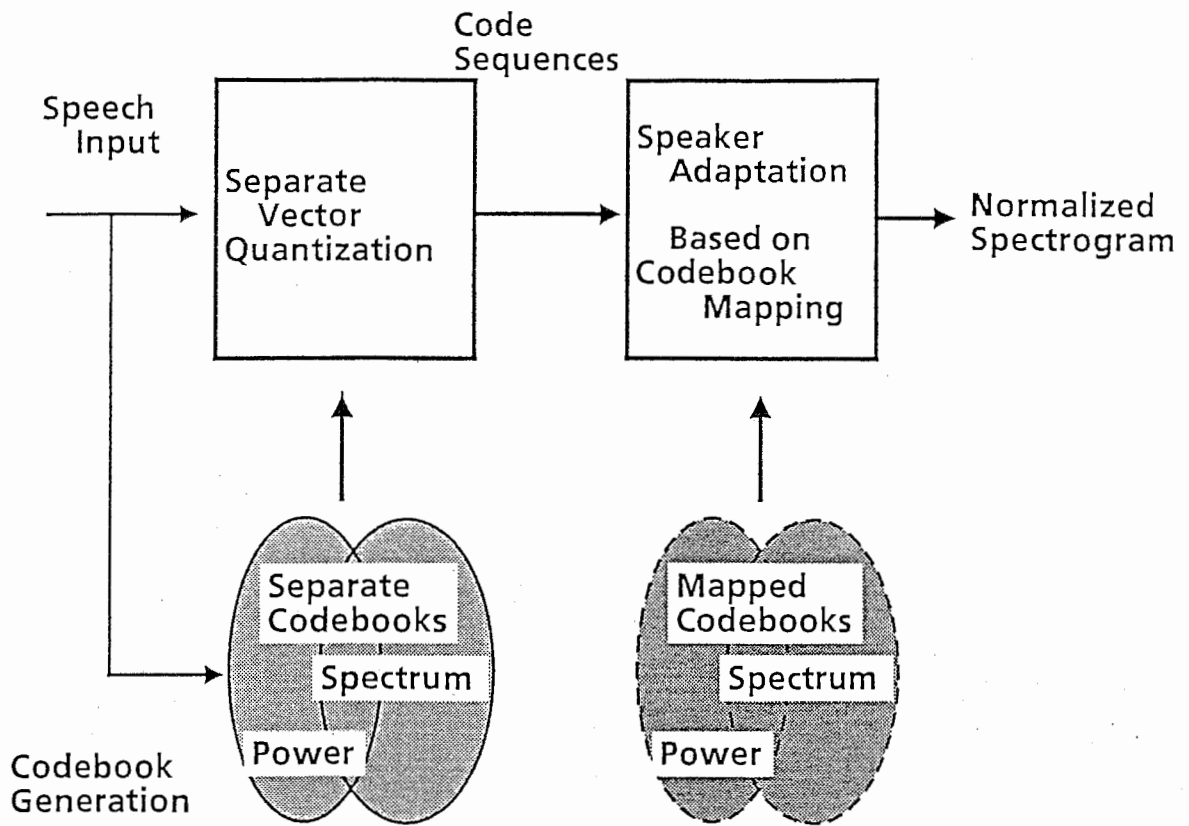


Fig.1. Block Diagram of Spectrogram Normalization based on Separate Vector Quantization

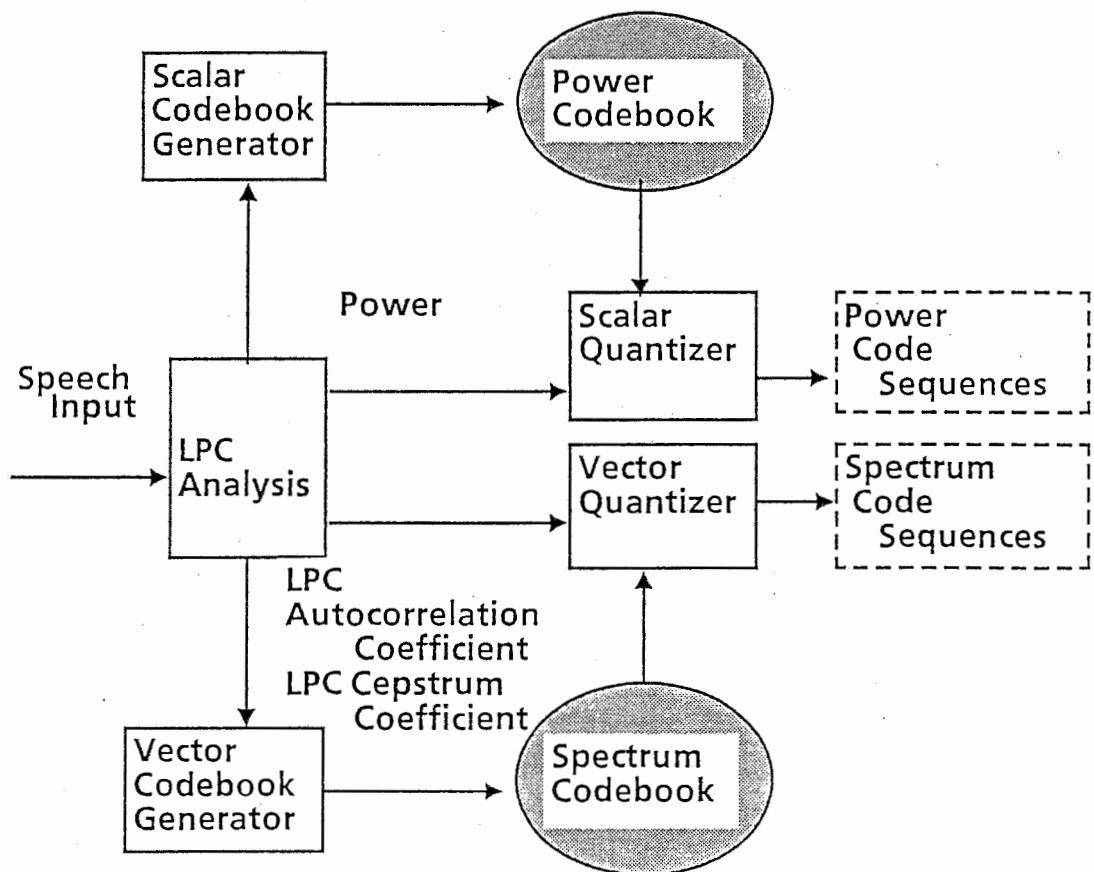
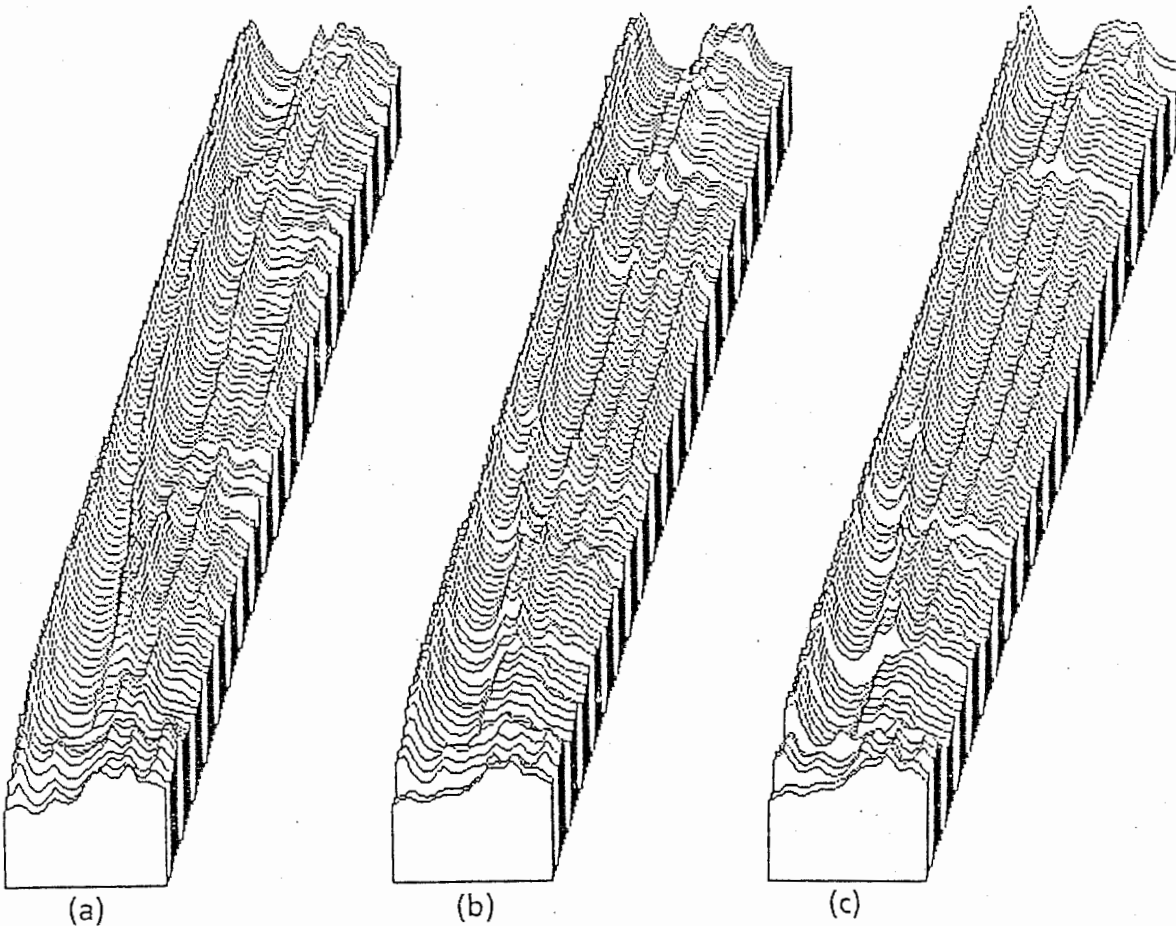


Fig.2. Block Diagram of Separate Vector Quantization

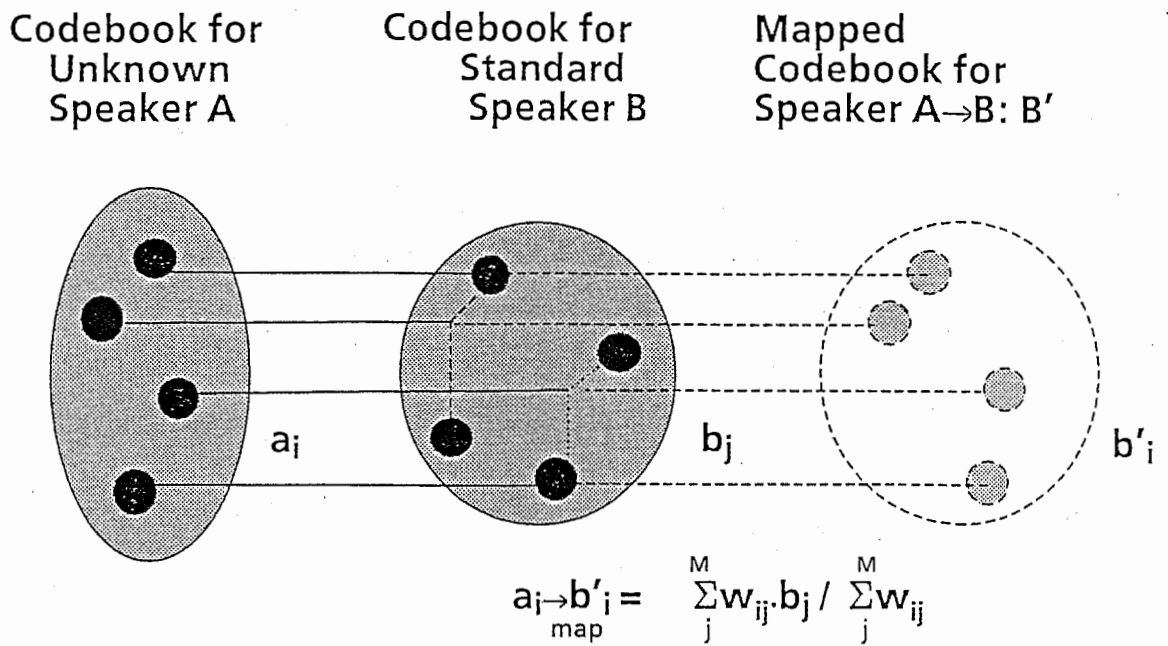


**Fig.3. Running Spectra (/kioi/)**

(a) Original Running spectra

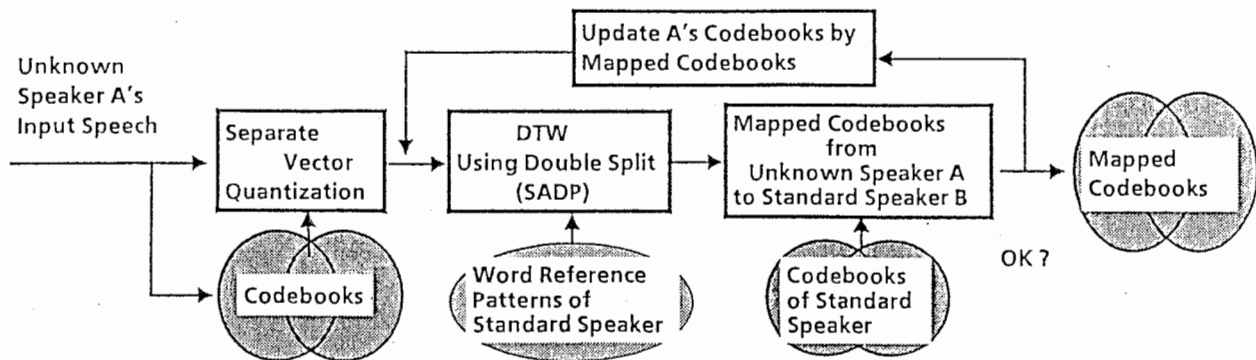
(b) VQ Running spectra by SPVQ(distortion:0.4764)

(c) VQ Running spectra by conventional VQ  
(distortion:0.6152)

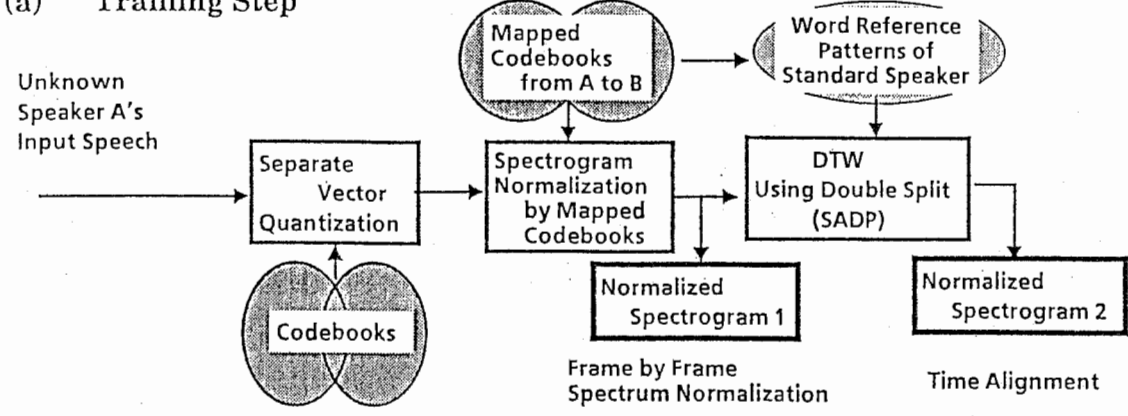


- $a_i$  : Vectors of codebook of speaker A
- $b_j$  : Vectors of codebook of speaker B
- $b'_i$  : Vectors of mapped codebook of speaker A→B
- $w_{ij}$  : determined from a histogram obtained by Training
- $M$  : Codesize of VQ codebook

Fig.4. Basic Concept of Speaker Adaptation by Vector Quantization



(a) Training Step



(b) Normalization Step

Fig. 5. Block Diagram for Spectrogram Normalization Algorithm

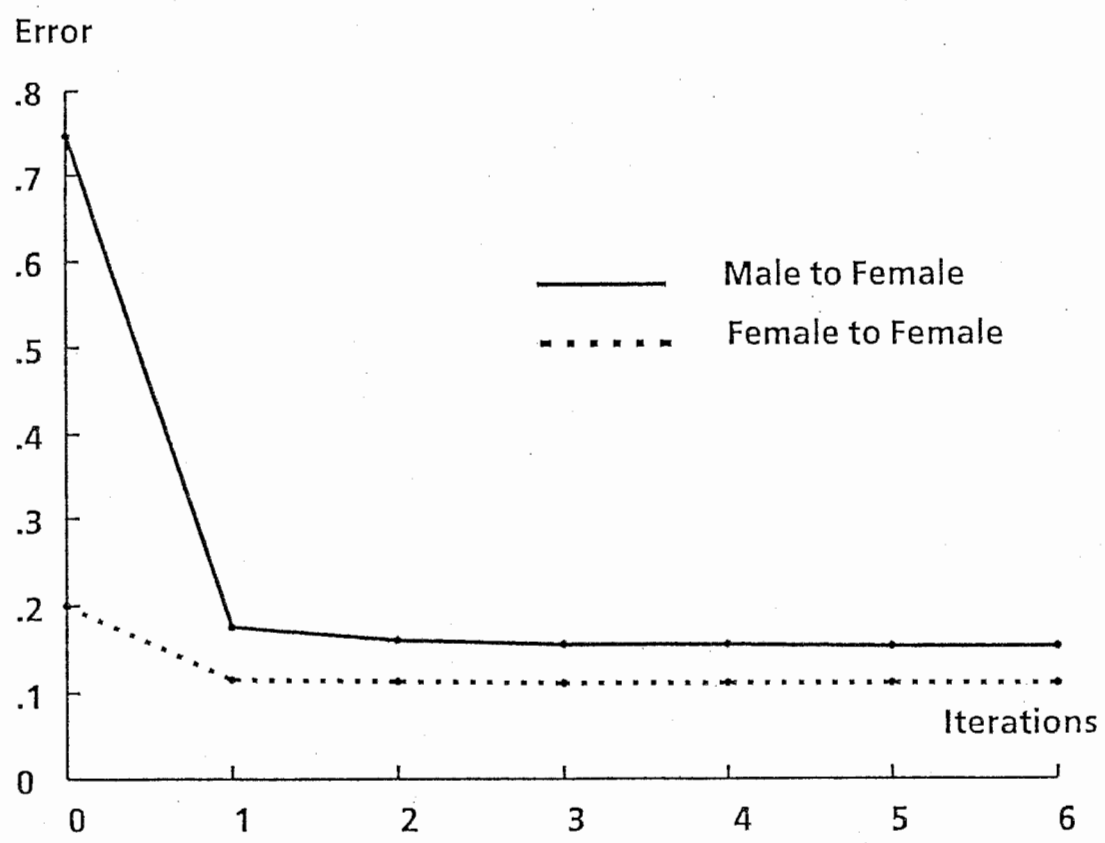


Fig.6. Average Normalization Errors vs. Iterations of Training

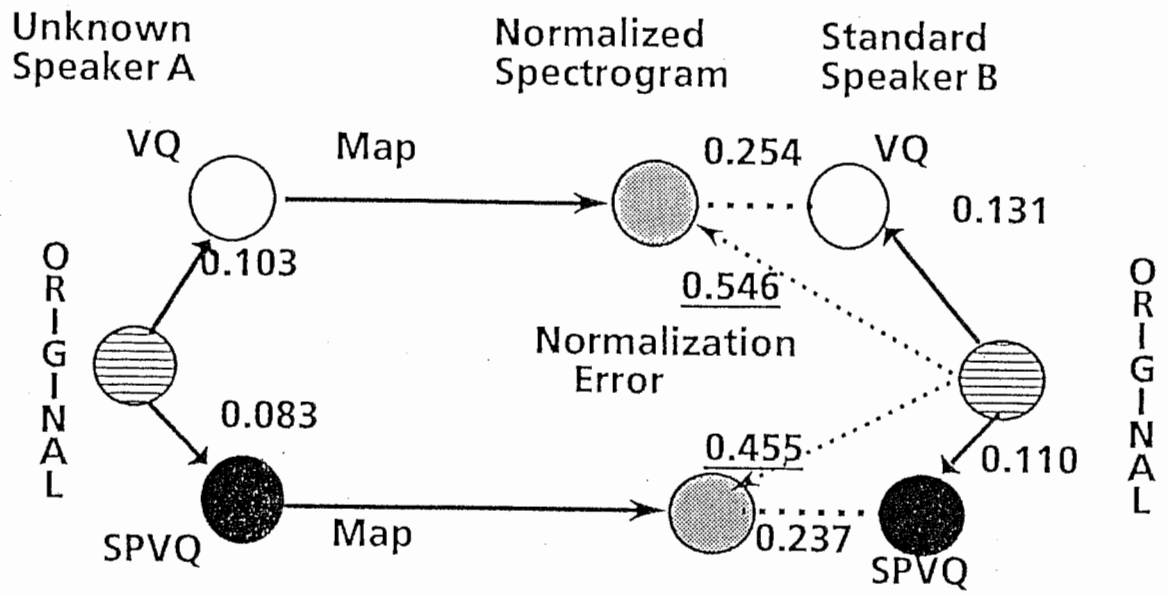


Fig.7. Distortion Comparison between VQ and SPVQ

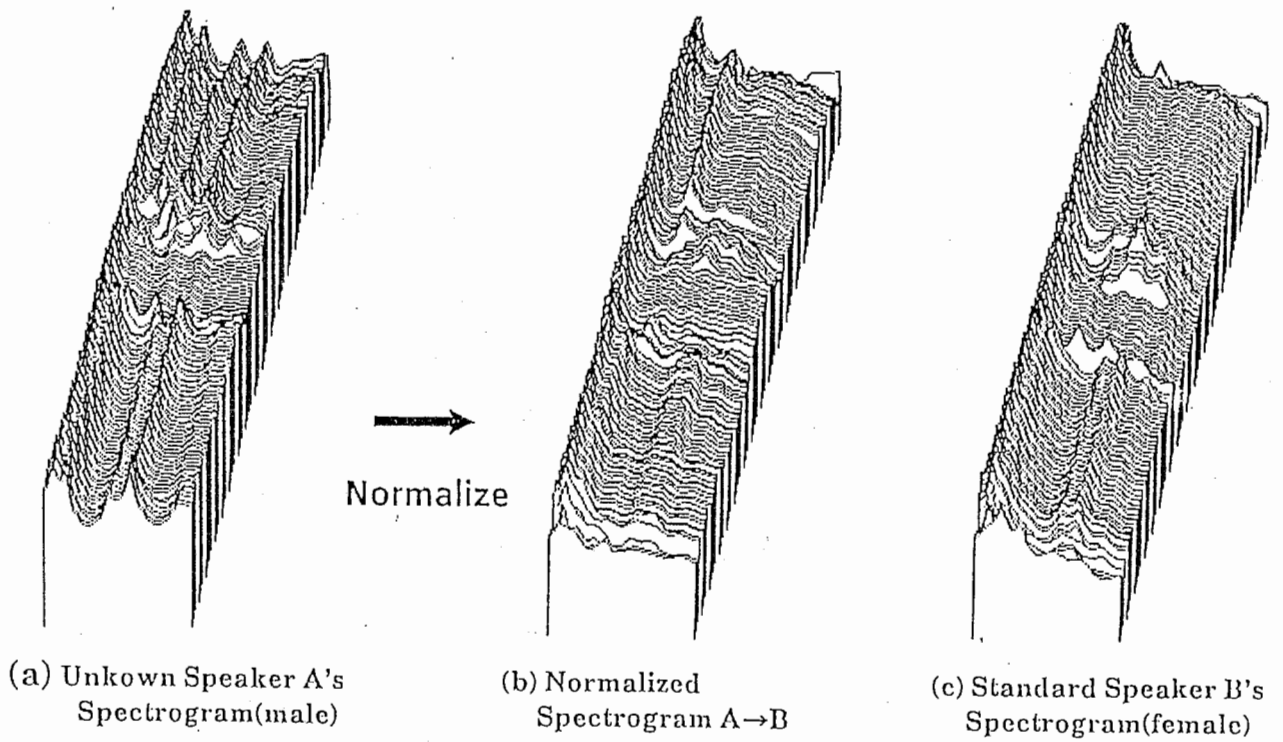


Fig.8. Normalized Spectrograms /koonyuu/