TR-I-0097

ベクトル量子化話者適応のHMM音韻認識への適用

VQ-based Speaker Adaptation Applied to HMM Phoneme Recognition

中村　哲,花沢利行,鹿野清宏
Satoshi NAKAMURA, Toshiyuki HANAZAWA,Kiyohiro SHIKANO

1989. 8

概要

本稿では、既に提案した話者適応化アルゴリズムをHMM音韻認識に適用する。HMMに適用する際には、動的特徴を考慮したセパレートベクトル量子化、ファジィベクトル量子化、ファジィヒストグラム、ファジィマッピングを用いる。さらに、HMMとの効率的な整合のために、対応づけヒストグラムを標準話者のファジィ級関数としてファジィHMMを計算する話者適応アルゴリズムを用いる。音韻バランス216単語、重要語5240単語を用いて有声破裂音/b,d,g/および全音韻の音韻認識実験を男女計3名の話者について行なった結果、次の事柄が確かめられた。(1)動的特徴を考慮したセパレートベクトル量子化を用いることにより有声破裂音の認識率が6.4%改善できる、(2)ファジィベクトル量子化を用いることにより有声破裂音の認識率が3.4%改善できる、(3)ファジィヒストグラムを用いることにより話者適応化の学習に必要な単語数を100単語から25単語に削減しても認識率の低下を0.4%に抑えられる。また、有声破裂音の認識率は、男性間で83.1%、男女間で76.5%で、従来法[M.Feng et al. ICASSP 88]との比較では11.7%の認識率の改善となること、全音韻の認識では、男性間で75.6%、男女間で71.8%で、上位3位までの累積認識率では、男性間、男女間いずれの場合にも約91%を達成できることがわかった。

ATR 自動翻訳電話研究所
ATR Interpreting Telephony Research Laboratories

目次

1 はじめに

自動翻訳電話等の入力技術として、不特定の話者が発声した連続発声の音声認識技術が最終的に必要となる。特定話者の音声認識については、連続音声についても高いレベルが達成されているが、不特定話者の音声に対しては1000語程度のタスクに対して実験システムが研究されている[1]段階で、実際の大規模なタスクの不特定話者の連続音声を十分な精度で認識することは非常に難しい。このため、不特定話者への第一歩として、100単語以内程度の少数語彙を用いて話者適応化を行うことが実際的な方法であると考えられる。

一方、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た[2,3]。ベクトル量子化は、特徴空間の分布を考慮して分割し、代表ベクトル符号を伝達することにより効率的な伝送を行うものである[4,5]。また、ベクトル量子化により特徴空間の離散表現が行えるという観点から、話者毎の音声の特徴空間を有限個の離散点で表現した後、話者間でその有限個の離散点の対応関係を見いだす事で話者適応が行えることが文献(6,7)に示された。文献(6)の方法では、未知話者の学習用音声を用いてベクトル量子化のコードブックを生成し、ベクトル量子化する。次に学習単語と同一の標準パターンと動的計画法による非線形マッチング(DTW)を行い、最適パスからベクトルコードの対応付けヒストグラムを求める。その後、ヒストグラムにより与えられるコードブック間の対応関係により話者適応を行うものである。ベクトル量子化に着目して話者適応化を行う方法としては、他に、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[10]、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、話者適応を行う方法[11]、ベクトル量子化のコードブックの構造の類似性に着目して、学習単語を規定せずに話者適応を行う方法[12,13]等が提案されている。

HMMにおける話者適応化に関しても、いくつかの研究が行われている。HMMでは、モデルのパラメータの推定に多量の学習データが必要となり、不特定の話者に対応するためには話者適応化が不可欠となる。現在研究の行われている話者適応化として、標準話者のコードブックを用いて未知話者の音声をベクトル量子化し、ベクトル間の変換確率を込めてHMMモデルの遷移確率及び出力確率を学習するもの[14]、ベクトル量子化のコードベクトルの適応をおこなった後、変換確率を学習するもの[15]、学習単語と標準パターンのDTWにより文献(14)の方法の変換確率の高精度化を図ったもの[16]がある。

文献(6)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。筆者らは、文献(6)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、音声合成における声質変換など一般の話者正規化方式としてベクトル量子化を用いた話者間のスペクトログラム正規化

の研究を行ってきた[8,9]。本論文では、このベクトル量子化を用いた話者適応方式をHMMによる音声認識系へ適用し、HMMの確率モデルに効率的に整合するHMM話者適応化方式について検討を行い、従来法[16]に比較して音韻認識率が改善できることを示す。

2 ベクトル量子化を用いた話者適応方法

図1に今回用いたHMM話者適応方法のブロック図を示す。アルゴリズムは2つのステップ、学習と認識により構成される。

(1) 学習

1)未知話者の学習用音声を用いてベクトル量子化コードブックを生成する。

2)未知話者のコードブックを用いて未知話者の学習用音声のベクトル量子化を行う。

3)あらかじめ準備した標準話者の学習単語の標準パターンとDTWを行い、同一単語間における最適パスを求める。

4)DTWの最適パスにしたがってコードベクトルの対応回数を求め、対応付けヒストグラムを求める。

5)対応付けヒストグラムの値を重みとして、新しく標準話者の空間に写像されたコードベクトルを求める。

$$V_i^{(A)} \rightarrow V_i^{(A \rightarrow B)} = \sum_{j=1}^{c} h_{ij} \cdot V_j^{(B)} \Big/ \sum_{j=1}^{c} h_{ij} \qquad (1)$$

ここで、

$V^{(A)}{}_i$: 未知話者Aのコードベクトル

$V^{(B)}{}_j$: 標準話者Bのコードベクトル

$V^{(A \rightarrow B)}{}_i$: 写像後のコードベクトル

h_{ij} : 対応づけヒストグラム

c : コードブックのサイズ

6) 5)のコードベクトルを未知話者のコードベクトルとして、ステップ1)-5)を繰り返し、スペクトル歪が収束するまで繰り返す。最終的に得られた5)におけるコードベクトルよりなるコードブックを変換コードブックと呼ぶ。

(2) 認識

1)未知話者の入力音声を未知話者のコードブックでベクトル量子化する。

2)標準話者のHMMモデルと、学習の時に求めた対応付けヒストグラムを用いて話者適応化、認識を行う。

この方法では、未知話者を標準話者に適応するので、標準話者のHMMモデルを変更する事なく未知話者の音声を認識することができる。本論文では、さらにベクトル量子化による話者適応化アルゴリズムの高精度化を図るため、セパレー

トベクトル量子化、ファジィベクトル量子化、ファジィベクトル量子化による対応付けヒストグラム生成法、及び入力ベクトルを連続的に写像するファジィマッピングを用いる。

2.1 セパレートベクトル量子化(SPVQ)[8]

ベクトル量子化の際に単一の複合尺度を用いる方法には、特徴空間の次元数が大きくなりベクトル量子化における量子化歪が大きくなる点、コードブック生成に必要な学習データが増加する点などの欠点がある。特に、話者適応では学習データの量を増加させることは大きな問題となる。そこで、特徴毎に別々のベクトル量子化系を構成するセパレートベクトル量子化を用いる。本論文では、従来のスペクトル、及びパワーの特徴量に加えて式(2)に示す48msec遅れのスペクトルの回帰係数$DCEP_i(l)$を用いる[17]。

$$DCEP_i(l)=\left[\sum_{n=-8}^{8}\left(C_{i+n}(l)\cdot n\right)\right] / \sum_{n=-8}^{8} n^2 \qquad (2)$$

ここで、$C_{i+n}(l)$は、$(i+n)$フレームにおけるl次のLPCケプストラム係数である。

また、距離尺度としては次のものを用いる[18]。

$$Spectrum:\ D_{i,j}^{(spect)}=\sum_{l=1}^{16}\left(C_i(l)-C'_j(l)\right)\cdot\left(R_i(l)-R'_j(l)\right) \qquad (3)$$

$$Power:\ D_{i,j}^{(pow)}=P_i/P'_j+P'_j/P_i-2 \qquad (4)$$

$$Difference\ Cepstrum:\ D_{i,j}^{(dcep)}=\sum_{l=1}^{16}\left(DCEP_i(l)-DCEP'_j(l)\right)^2 \qquad (5)$$

但し、$R_i(l)$, $C_i(l)$, $DCEP_i(l)$, P_iは、未知話者Aのiフレームにおけるl次のLPC自己相関係数、LPCケプストラム係数、LPCケプストラムの回帰係数とパワーである。同様に、$R'_j(l)$, $C'_j(l)$, $DCEP'_j(l)$, P'_jは、標準話者Bのjフレームにおけるl次のLPC自己相関係数、LPCケプストラム係数、LPCケプストラムの回帰係数とパワーを表す。

また、話者適応化の学習でのDTWの距離尺度は、次式に示す各特徴量の線形和により定義される距離を用いる[19]。

$$D_{i,j}^{(total)}=D_{i,j}^{(spect)}+0.01\cdot D_{i,j}^{(pow)}+0.3\cdot D_{i,j}^{(dcep)} \qquad (6)$$

2.2 ファジィベクトル量子化(FZVQ)[9]

セパレートベクトル量子化を用いた話者適応方式を用いて話者適応化の性能を改善する場合にも、さらに量子化歪を小さくする必要がある。この問題の解決方法として、ファジィベクトル量子化[20]を導入する。ファジィベクトル量子化で

は入力ベクトルに対して、式(7)に従ってファジィ級関数u_iが計算される。再生ベクトルは、このファジィ級関数を用いて、式(8)にしたがって、既存のコードベクトルからの線形和で求められる。ここで、ファジィ級関数u_iは、距離に反比例して決まる1以下の正の値で、距離が0に近いほど1に近く、遠いと0に近くなる。mはファジィネスといわれる値で、量子化の曖昧さを表しており、本論文では、1.6を用いた。また、ファジィ級関数は全てのコードベクトルV_iに対して求めるため計算量が膨大なものとなる。そこで、k-近傍則を用いて入力ベクトルに近いものからk個を選択し、それらが構成する部分空間の中でファジィ級関数を求める[9]。本論文では、$k=6$とした。

$$u_i = 1 / \left[\sum_{j=1}^{k} \left(d_i / d_j \right)^{1/(m-1)} \right] \qquad (7)$$

$$X' = \sum_{i=1}^{k} \left[\left(u_i \right)^m \cdot V_i \right] / \sum_{i=1}^{k} \left(u_i \right)^m \qquad (8)$$

2.3 ファジィベクトル量子化による対応付けヒストグラムの生成(Fzhist)[9]

話者間におけるベクトル量子化のコードベクトルの対応関係は、未知話者と標準話者の同一単語のDTWの最適パスを求めることによりなされる。このため、学習単語への依存度が高く、出現しないコードベクトルのヒストグラムの推定が困難になる。この依存性を改善するために、ファジィベクトル量子化を用いて対応付けヒストグラムを生成する。この方法では、図2に示すように未知話者と標準話者の学習用単語をファジィベクトル量子化しk-近傍のコードベクトル$a_i, b_j (1 \leqq i,j \leqq k)$とファジィ級関数$u^{(a)}{}_i, u^{(b)}{}_j$を求めた後、同一単語間のDTWを行って最適パスで対応するフレームt_a, t_bを求める。次に、図3に示すように、ファジィ級関数を確率とみなしてk^2の組合せについてコードベクトルの対応回数を積算しヒストグラムの推定を行うものである。但し、P_{ij}は文献(9)に示されている対応づけの信頼度である。

2.4 ファジィマッピングによる話者適応化[9]

ベクトル量子化を用いた話者適応化は、図4に示すように未知話者の音声Xをベクトル量子化して$V^{(A)}{}_i$を求め、$V^{(A)}{}_i$に対応する変換コードブックのコードベクトル$V^{(A \to B)}{}_i$を出力することにより行なう。ファジィベクトル量子化の場合も同じで、入力ベクトルXをファジィベクトル量子化してX'を求め、式(9)のように級関数を保存したまま対応する変換コードブックのコードベクトルを用いて写像することにより変換後の出力ベクトルX''が求められる。2つの特徴空間で1対1にコードベクトルが対応し、ファジィ級関数が保存されるような写像をファジィマッピング[22]と呼んでいるが、本論文ではファジィベクトル量子化に用い

るk近傍則の近傍が構成する部分空間の未知話者Aと標準話者Bの間のファジィマッピングを考える。

$$X \rightarrow X' = \frac{\sum_{i=1}^{k} \left(u_i^{(A)}\right)^m \cdot V_i^{(A)}}{\sum_{i=1}^{k} \left(u_i^{(A)}\right)^m} \rightarrow X'' = \frac{\sum_{i=1}^{k} \left(u_i^{(A)}\right)^m \cdot V_i^{(A \rightarrow B)}}{\sum_{i=1}^{k} \left(u_i^{(A)}\right)^m} \qquad (9)$$

3. Hidden Markov モデルへの適用

3.1 HMM話者適応化法

2章において述べたように、話者適応化は未知話者の入力音声を未知話者と標準話者のコードベクトルの対応関係に基づいて標準話者の空間に写像することによりなされる。ところがこの方法では、離散モデルのHMMなどに入力する場合、変換後の出力ベクトルを再度ベクトル量子化しなければならず効率が悪いばかりでなくベクトル量子化歪みを被ってしまう。そこで、次のようなアルゴリズムを用いる。

1) ベクトル量子化による話者適応法を用い、ヒストグラムh_{ij}を用いて未知話者の入力コードベクトル$V^{(A)}{}_i$を標準話者の空間の中間的な点$V^{(A \rightarrow B)}{}_i$に写像したと仮定する。

2) 1)で得られたヒストグラムh_{ij}を写像された点$V^{(A \rightarrow B)}{}_i$から標準話者のコードベクトル$V^{(B)}{}_j$へのファジィ級関数とみなす。

3) 標準話者内で、ファジィVQによるHMMを構成し、音韻認識を行う。

この方法によりベクトル量子化を重複して行う事なく効率的に話者適応化が実現できる。以後、このアルゴリズムをファジィヒストグラムマッピング(Fzhmap)と呼ぶ。ファジィベクトル量子化によるHMMは、離散モデルのHMMにおける学習データの不足を補うための方法として提案されている[21]。この様子を図4(a)に示す。また、式(10)に計算方法を示す。この方法は、変換ヒストグラムh_{ij}を確率とみれば、文献(14)の方法と形式的に同一となる。しかし、本方法では未知話者の入力ベクトルがファジィベクトル量子化された場合にも図4(b)のように容易に拡張でき、式(11)のように出力確率を計算することができる。

$$\omega_{i,t}(s) = \sum_{j=1}^{c} h_{ij} \cdot O_{sj} \qquad (10)$$

$$\omega_t(s) = \sum_{j=1}^{c} \left(\sum_{i=1}^{k} u_i \cdot h_{ij} \right) \cdot O_{sj} \qquad (11)$$

ここで、$\omega_{i,t}(s)$は、未知話者Aの音声がコードベクトルiに量子化され標準話者Bの空間に写像されたベクトルの時刻t、ステートsにおける出力確率である。$\omega_t(s)$は、未知話者Aの音声がファジィVQ、ファジィマッピングを経て標準話者

Bの空間に写像されたベクトルの時刻t、ステートsにおける出力確率である。O_{sj}は、標準話者Bのコードベクトルjのステートsにおける出力確率である。また、セパレートベクトル量子化を用いる場合は、式(12)に示すように各々の出力確率の積を全体の出力確率とする。

$$\omega_t(s)=\omega_t^{(spect)}(s)\cdot\omega_t^{(pow)}(s)\cdot\omega_t^{(dcep)}(s) \qquad (12)$$

3.2 分析条件および音声資料

今回の実験に用いた分析条件を表1に示す。

音声データとしては、コードブック生成、および、話者適応の学習用データとして可能な2音韻連鎖をすべて含む音韻バランス216単語の前半100単語を用いる。HMMのパラメータの学習用データとしては、5,240単語の偶数番目の単語を、評価用のデータとして奇数番目の単語を用いる。また、音韻レベルのHMMモデルを用意し、コンテキストの影響を考慮して語頭と語中を別のモデルとする。実験に用いた話者は、男性2名、女性1名の計3名で、未知話者を男性とし、男性間で同性間の評価、男女間で異性間の評価とする。また、これらのデータは、すべて予め目視によって音韻境界が求められており以下の実験では、音韻境界は既知のものとして行う。

3.3 実験

本節では、2章および3.1節で述べた話者適応化の各方法の評価を有声破裂音/b,d,g/を用いて行い、最終的な全体の性能評価を日本語全音韻を用いて行なう。

3.3.1 話者適応アルゴリズムの比較

ファジィマッピングによるHMM話者適応化の効果を評価する。実験は、PWLR尺度を用い、3.2に述べた音声データのうち有声破裂音/b,d,g/に対して行う。ここでは、次の5つの方法の音韻認識率の比較評価を行う。但し、学習には通常のベクトル量子化およびヒストグラムを用いる。認識は、未知話者の入力音声を通常のベクトル量子化を用いて量子化した後、話者適応等を行い、1-1)、1-2)については、通常の離散HMMモデルを用い、1-3)、1-4)については、ファジィHMMモデルを用いて行う。

1-1) 話者適応なし (Without Adaptation)
未知話者の音声を標準話者のコードブックでベクトル量子化し標準話者のHMMモデルで認識を行う。

1-2) 変換コードブックによる方法 (Mapped Codebook)
予め学習により標準話者から未知話者への変換コードブックを生成する。未知話者の入力音声に対し、この変換コードブックを用いてベクトル量子化を行い、標準話者のHMMモデルで認識を行う。

1-3) ベクトル量子化による話者適応化を行った後、HMM音韻認識を行う方法(Fzmap)
通常のベクトル量子化を用いた話者適応化を行って変換後の出力ベクトル$V^{(A\to B)}{}_i$を求めた後、ファジィベクトル量子化を行いファジィHMMモデルで認識を行う。

1-4) Fzhmap-HMM話者適応化方法(Fzhmap)
3.1節で述べたFzhmap法によるHMM話者適応化アルゴリズムに従ってHMM話者適応化、認識を行う。

1-5) 特定話者認識
標準話者に対しHMMのパラメータ学習に用いなかったデータを用いて特定話者の認識を行う。

この結果を表2に示す。1-1)では話者適応なしの場合、男性間で69.0%の音韻認識率が得られている。これは、この2名の話者の有声破裂音の特徴量が近いことを示す。男女間の場合は、音韻認識率は低く約34%なので、ほぼ無作為に/b,d,g/を抽出したのに等しい。1-2)の変換コードブックを用いることにより特に男女間で大きな改善がみられる。しかしながらこの方法では、変換コードブックで未知話者の入力をベクトル量子化を行うためベクトル量子化歪が大きくなるという欠点がある。今回提案する1-4)のFzhmapによる方法を用いることにより、男性間で76.2%、男女間で67.0%の音韻認識率を得、男性間と男女間の平均で、話者適応無しの場合に比べて約20%の改善が得られる。また、1-3)の方法が1-4)より5%程度悪いのは、ヒストグラムを重みとして変換後の出力ベクトルを求め、再度標準話者内でファジィベクトル量子化を行うことによる情報欠落が原因であると考えられる。

3.3.2 セパレートベクトル量子化の効果

セパレートベクトル量子化を用いることによる効果の評価を行う。実験条件は、3.3.1と同じである。実験として次の4つの比較評価を行った。但し、変換方法はFzhmapによるHMM話者適応方法とした。

2-1) WLR VQ(1コードブック)

2-2) PWLR VQ(1コードブック)

2-3) SPVQ (2コードブック)

2-4) SPVQ(DCEP)(3コードブック)

この結果を表3中の2-1)~2-4)に示す。但し、表3の最右列は、標準話者Male2、Female1の特定話者の認識率の平均、最下段はそれぞれ標準話者Male2、Female1の特定話者の認識率を示す。この表からスペクトルだけの2-1)に対しパワー情報を使用することにより、平均で3.8%認識率が改善されること、2-2)と2-

3)の比較でセパレートベクトル量子化を用いることにより認識率が1.6%改善されることがわかる。スペクトルの動的特徴である線形回帰係数の利用は不特定の話者の音声認識に効果のあることが報告されているが[17]、本実験でも2-3)に対し平均で1.0%の改善がみられる。最終的にパワーとスペクトルとスペクトルの回帰係数を用いる事により平均で74.2%の認識率を得、スペクトルだけの場合に対し6.4%の認識率の改善が得られる。

3.3.3 ファジィベクトル量子化の効果

ファジィベクトル量子化を用いることによる効果の評価を行う。実験条件は3.3.1節と同じである。評価のため次の比較実験を行った。但し、変換方法はFzhmapによるHMM話者適応方法とする。

2-4) SPVQ(DCEP) 対

2-5) SPVQ(DCEP)+FZVQ

結果を表3中の2-4),2-5)に示す。男性間男女間とも同程度の効果で、平均で約3.4%の改善がみられ、ファジィベクトル量子化が効果的であることが示されている。

3.3.4 ファジィヒストグラムの効果

ファジィヒストグラムを用いることによる効果の評価を行う。実験条件は3.3.1節と同じで次の比較をおこなった。変換方法はFzhmapによるHMM話者適応方法とする。

2-4) SPVQ(DCEP) 対

2-6)SPVQ(DCEP)+Fzhist

2-5) SPVQ(DCEP)+FZVQ 対

2-7)SPVQ(DCEP) +FZVQ +Fzhist

結果を表3中の2-4)~2-7)に示す。2-4)と2-6)の比較で、ファジィヒストグラムの使用によりコードベクトルの対応づけが精密になり音韻認識率が約4%改善されることがわかる。但し、ファジィヒストグラムとファジィベクトル量子化を併用した2-5)と2-7)の場合は逆に認識率が0.3%劣化し、平滑化の効果が相乗的に作用し改善効果がみられない。

3.3.5 学習単語数の効果

話者適応化のための学習単語数とHMM音韻認識率の関係を調べる。次の4つの場合について学習単語数を変えて実験を行なう。但し、変換方法はFzhmapによるHMM話者適応方法とする。

5-1) SPVQ

5-2) SPVQ(DCEP)

5-3) SPVQ(DCEP) + Fzhist

5-4) SPVQ(DCEP) + Fzhist + FZVQ

この結果を図5に示す。図の値は、男性間と男女間の音韻認識率の平均値である。これにより、ファジィヒストグラムを用いれば、学習単語を100単語から25単語に低減しても認識率は、5%程度しか劣化しないことがわかる。また、入力にファジィベクトル量子化を用いる事によりさらに学習単語への依存性が減り、0.4%しか低下しない。これに対し、5-1)のセパレートベクトル量子化とファジィヒストグラムマッピングだけを用いた方法では、学習単語を100単語から25単語に削減することにより約10%の劣化が見られた。

3.3.6 ヒストグラムの打ち切りによる高速化(Prune)

ファジィベクトル量子化を用いたHMMを行う場合すべてのコードベクトルへの級関数の値を用いて出力確率を求めるので計算量が非常に多くなる。この問題を回避するためにヒストグラムの打ち切りによる高速化を行なう。この方法はヒストグラムを値の大きいものからn'個で打ち切り、式(10)(11)(12)の計算を高速に行うものである。この結果を図6に示す。図は入力にファジィベクトル量子化を用いた場合の結果である。男性間、男女間ともほぼ同じ傾向を示しており、打ち切り数は20程度とれば十分であることがわかる。また、表3中の2-8),2-9)に示すように、ヒストグラム打ち切りを行うことによりそれぞれ男性間、男女間の平均で0.9%、2.2%認識率が改善される。これは、不適切なコードベクトルへの関係付けが省かれることによる改善と考えられる。

3.3.7 考察

表4に有声破裂音/b,d,g/に関する音韻認識率の結果をまとめる。最終的に男性間で83.1%、男女間で76.5%の音韻認識率を得た。男性間と男女間の平均では、100単語学習では入力にファジィベクトル量子化を使用した方がよく79.5%であった。話者適応無しの場合に比べると27.8%の改善であった。また、2-1)の方法は文献(16)に紹介されているものとほぼ同じであり、この方法と比較すると筆者らの方法ではさらに11.7%だけ高い音韻認識率が得られている。

また、未知話者の発声した話者適応用100単語を用いてHMMパラメータを学習し未知話者に対する特定話者の認識を行った。ただし、HMMのパラメータ学習には、予め与えられているラベル情報を用い、ベクトル量子化としては動的特徴を加えたセパレートベクトル量子化とファジィベクトル量子化を行った。この結果、有声破裂音/b,d,g/に対する認識率は75.2%で話者適応化に比べて4.3%低い認識率であった。実際には未知話者の学習用音声に音韻ラベルを付与することは

非常に困難であり、自動ラベリング、HMMモデルの連結学習等による対処を行ってもさらにかなりの認識率の低下が予想される。

3.3.8 全音素の適応評価

次に、本方法を用いて全音韻の認識実験を行なう。音韻は23種類の音韻であり、計算はヒストグラム打ち切りによる高速化手法を用いる。但し、ヒストグラムの打切り数k'は20とする。この結果を表5に示す。表中の認識率は、子音の平均認識率と母音の平均認識率の平均とする。最下段は、それぞれ標準話者Male2、Female1の特定話者の認識率を示す。この結果、日本語全音韻の話者適応化音韻認識率は、男性間で75.6%、男女間で71.8%であった。また、累積では男性間、男女間の平均で2位までで約85%、3位までで約91%の認識率であった。平均で本方法を用いることで話者適応無しの場合に比べて一位で26.5%の改善が得られた。男性間における各音韻毎の認識率を図7に示す。誤認識は、主に無声破裂音間や、有声破裂音、鼻子音間などの同じ音韻群で発生しており、特に後続音韻などのコンテキストの影響を受けやすい/z,k,g,m,n,u/の認識率が低い傾向がみられる。しかし、これらの音韻に関しても3位までの累積認識率ではかなり高い認識率を示している。

4. まとめ

ベクトル量子化を用いた話者適応方法をHMM音韻に適用し評価を行った。また、ベクトル量子化を用いた話者適応方法をHMMへ適用する際にFzhmap法を用いることにより再度ベクトル量子化を行うことなく効率的にHMM話者適応化ができる事を示した。また、ベクトル量子化を改善するセパレートベクトル量子化及びファジィベクトル量子化、コードベクトルの対応関係を表すヒストグラムの生成にファジィベクトル量子化の級関数を用いる方法をHMM話者適応化に適用した。これに関して、男女計3名の話者のデータについて有声破裂音/b,d,g/及び全音韻に対する評価実験を行った。有声破裂音に対する実験では男性間の場合83.1%、男女間の場合76.5%で平均では79.5%であった。これは、従来法に比べて11.7%の改善であった。全音韻に対する認識率では、男性間の場合75.6%で男女間の場合71.8%であった。また、累積では男性間、男女間の平均で2位までで約85%、3位までで約91%の認識率であった。話者適応無しの場合に比べると平均で26.5%の改善であった。各々アルゴリズムの効果を以下にまとめる。

1)Fzhmap法によるHMM話者適応化により効率的なHMM話者適応が実現できる。

2)セパレートベクトル量子化により認識率が改善できる。さらに、スペクトルの線形回帰係数を用いることで認識率が向上する。

3)入力にファジィベクトル量子化を用いることで認識率が改善できる。

4)対応付けヒストグラムの生成にファジィベクトル量子化の級関数を用いる方法により学習単語数の削減ができる。

問題としては、有声子音などのコンテキストの影響を受けやすい音韻や話者間で発声様式の変動の大きい音韻の認識誤りが考えられるが、発声方法の学習による適応化や、複数の話者の学習データを標準話者の空間へ写像しHMMのパラメータ学習を行って発声のバリエーションを学習することでさらに認識率の改善を行ってゆく予定である。

5. 謝辞

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

文献

(1) K.F.Lee, H.W.Hon, "Large-Vocabulary Speaker-Independent Continuous Speech Recognition Using HMM," Proc. ICASSP 88 S3.7 (1988)

(2) A.Buzo,H.Martinez,C.Rivea,"Discrete Utterance Recognition Based Upon Source Coding Techniques," Proc. ICASSP 82 539-542 (1982).

(3) 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).

(4) Y.Linde,A.Buzo,R.M.Gray,"An Algorithm for Vector Quantizer Design," IEEE Trans. Commun. COM-28, 84-95 (1980)

(5) A.Buzo,A.H.Gray,R.M.Gray,J.D.Markel, "Speech Coding Based Upon Vector Quantization," IEEE Trans.Acoust. Speech Signal Process. ASSP-28 562-574 (1980)

(6) K.Shikano,K.F.Lee,R.Reddy,"Speaker Adaptation through Vector Quantization," Proc. ICASSP 86, 49.5 (1986)

(7) 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)

(8) 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," 音響学会誌 44, 595-602 (1988)

(9) 中村哲、鹿野清宏,"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," 音響学会誌 45, 107-114 (1989)

(10) 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)

(11) K.Choukri,G.Chollet,Y.Grenier,"Spectral Transformations through Canonical Correlation Analysis for Speaker Adaptation in ASR," Proc. ICASSP 86, 49.9 (1986)

(12) 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)

(13) 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(1988-3)

(14) R.Shwartz,Y.Chow,F.Kubala,"Rapid Speaker Adaptation Using a Probalistic Spectral Mapping," Proc. ICASSP 87 15.3 (1987)

(15) M.Nishimura,K.Sugawara,"Speaker Adaptation Method for HMM-Based Speech Recognition," Proc. ICASSP 88 S5.7 (1988)

(16) M.Feng,F.Kubala,R.Schwartz,J.Makhoul, "Improved Speaker Adaptation Using Text Dependent Spectral Mappings," Proc. ICASSP 88 S3.9 (1988)

(17) 古井貞熙,"音声スペクトルの動的特徴を用いた単語音声認識," 音響学会音声研資S84-65(1984)

(18) 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)

(19) K.Shikano,"Evaluation of LPC Spectral Matching Measures for Phonetic Unit Recognition," CMU Techinical Report (1986 - 02)

(20) E.Ruspini, "Numerical Methods for Fuzzy Clustering," Inf.Sci.,Vol.2 32-57 (1970)

(21) H.P.Tseng, M.J.Sabin, E.A.Lee ,"Fuzzy Vector Quantization Applied to Hidden Markov Modeling," Proc. ICASSP 87 15.5 (1987)

(22) 水本雅晴、ファジィ理論とその応用 (サイエンス社、東京、1988), pp.78-105.

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. HMM phoneme recognition rates vs. speaker adaptation algorithms for /b,d,g/ task

	Method	Recognition Rate (%)		
		Male1→ Male2	Male1→ Female1	Average
1-1	Without Adaptation	69.0	34.3	51.7
1-2	Mapped Codebook	71.6	55.7	63.7
1-3	Fzmap	75.3	57.4	66.4
1-4	Fzhmap	76.2	67.0	71.6
1-5	Dependent	95.2	94.6	94.9

Table 3. HMM phoneme recognition rates for /b,d,g/ task

	Method	Recognition Rate (%)			
		Male1→ Male2	Male1→ Female1	Average	Depend-ent
2-1	WLR VQ	72.5	63.1	67.8	94.7
2-2	PWLR VQ	76.2	67.0	71.6	94.9
2-3	SPVQ	76.8	69.5	73.2	95.8
2-4	SPVQ(DCEP)	78.6	69.7	74.2	96.6
2-5	SPVQ(DCEP)+FZVQ	80.8	74.4	77.6	95.8
2-6	SPVQ(DCEP)+Fzhist	80.0	76.2	78.1	96.6
2-7	SPVQ(DCEP)+Fzhist +FZVQ	79.8	74.8	77.3	95.8
2-8	SPVQ(DCEP)+Fzhist +Prune (k'=50)	81.5	76.5	79.0	96.6
2-9	SPVQ(DCEP)+Fzhist +FZVQ+Prune (k'=20)	83.1	75.9	79.5	95.8
2-10	Speaker Dependent	97.6	95.5	96.6	96.6

Table 4. HMM Phoneme recognition rates for /b,d,g/ task

	Method	Recognition Rate (%)			
		Male1→ Male2	Male1→ Female1	Average	Depen dent
1-1	Without Adaptation	69.0	34.3	51.7	94.9
2-1	WLR+Fzhmap	72.5	63.1	67.8	94.7
2-8	SPVQ(DCEP)+Fzhist +Prune (k'=50)	81.5	76.5	79.0	96.6
2-9	SPVQ(DCEP)+Fzhist+FZVQ +Prune (k'=20)	83.1	75.9	79.5	95.8

Table 5. HMM phoneme recognition rates for Japanese all phonemes

Method	Recognition Rates (%)								
	Male1 → Male2			Male1→ Female1			Average		
	1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
Without Adaptation	62.1	79.2	86.2	32.3	47.6	57.2	47.2	63.4	71.7
Speaker Adaptation	75.6	87.0	92.1	71.8	84.5	91.3	73.7	85.8	91.7
Speaker Dependent	92.7	96.9	98.6	92.7	96.8	98.6	92.7	96.9	98.6

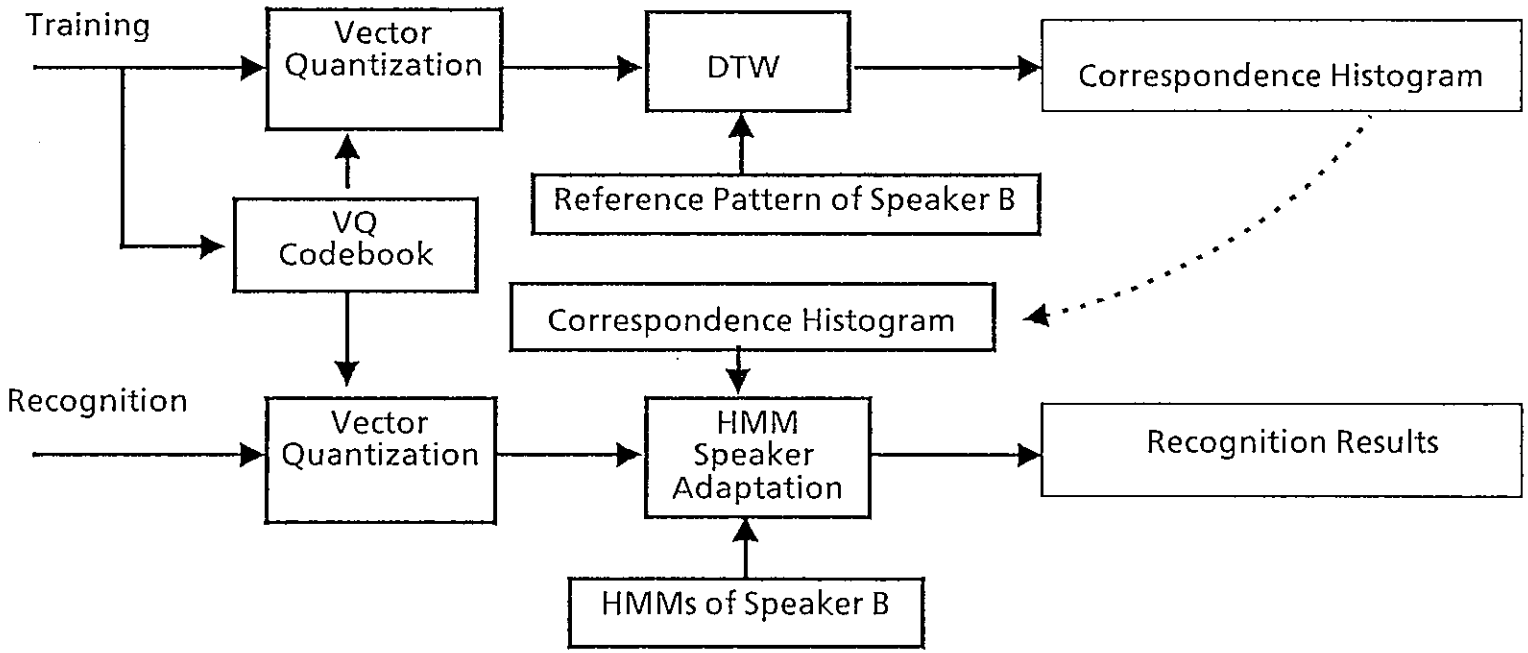

Fig.1. Block diagram of HMM speaker adaptation

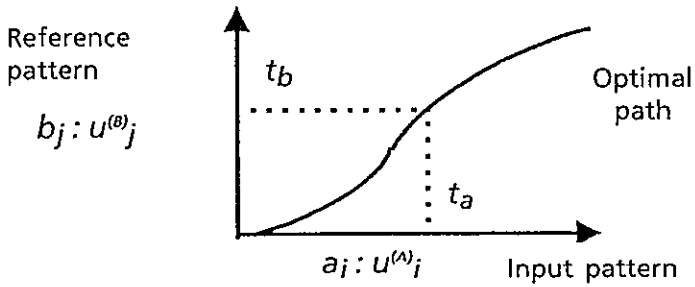

Fig.2. DTW optimal path

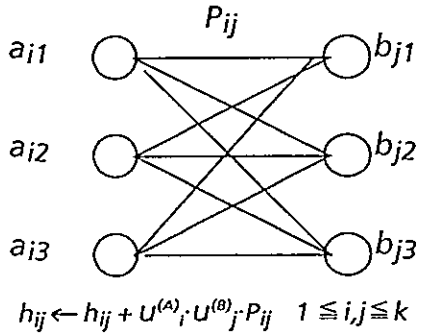

Fig.3. Histogram generation
using fuzzy membership function

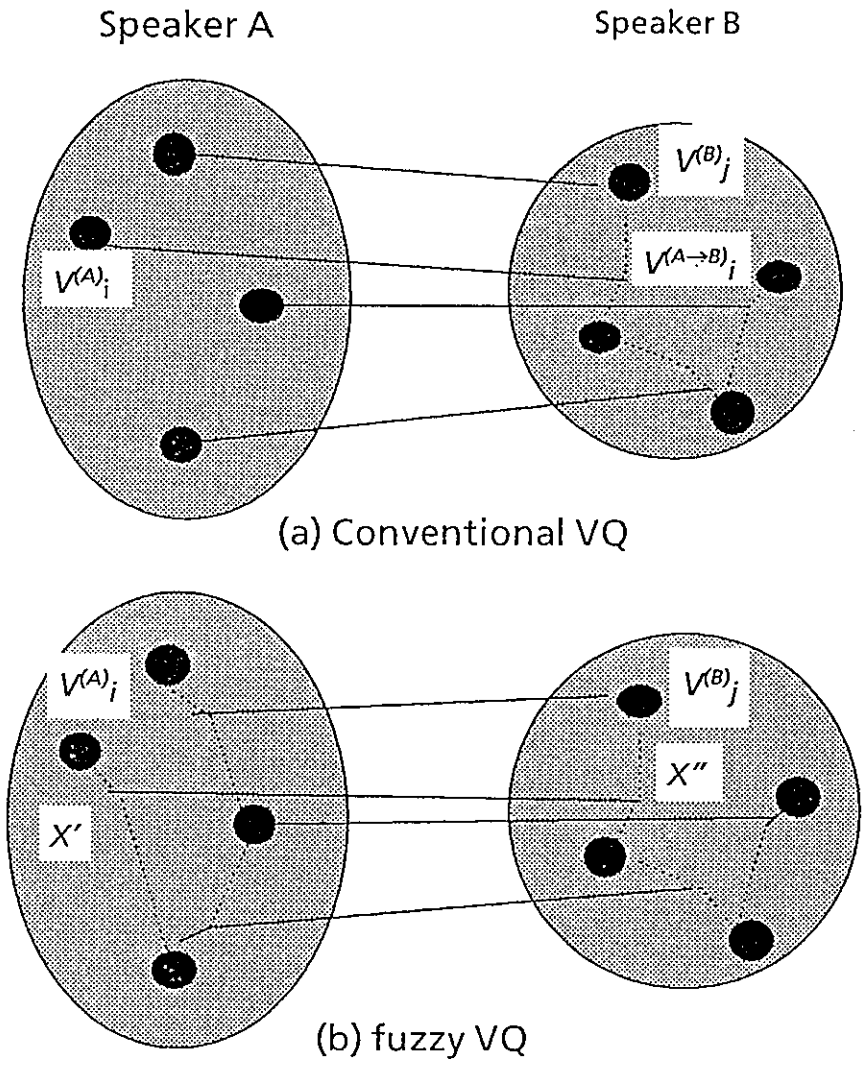

Fig.4. HMM speaker adaptation using fuzzy mapping

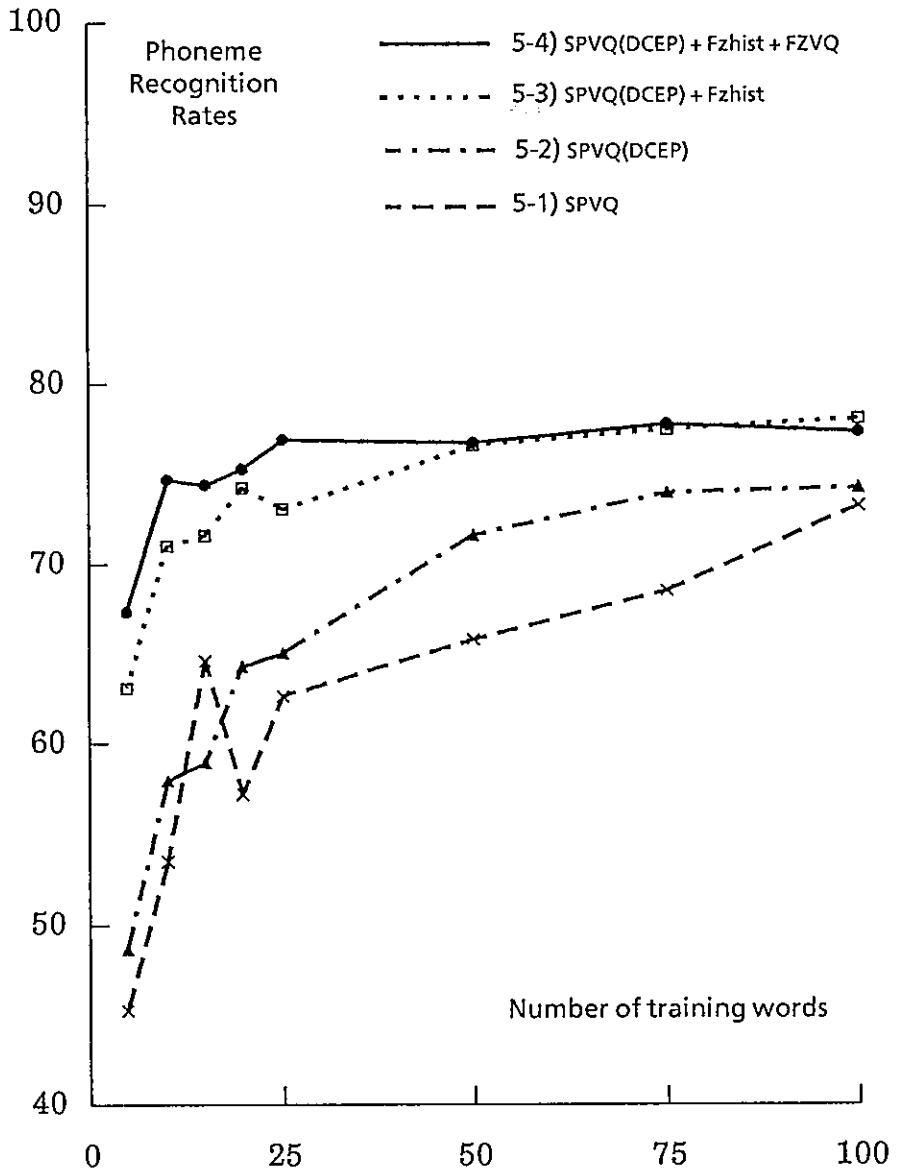

Fig.5. Phoneme recognition rates vs. number of training words

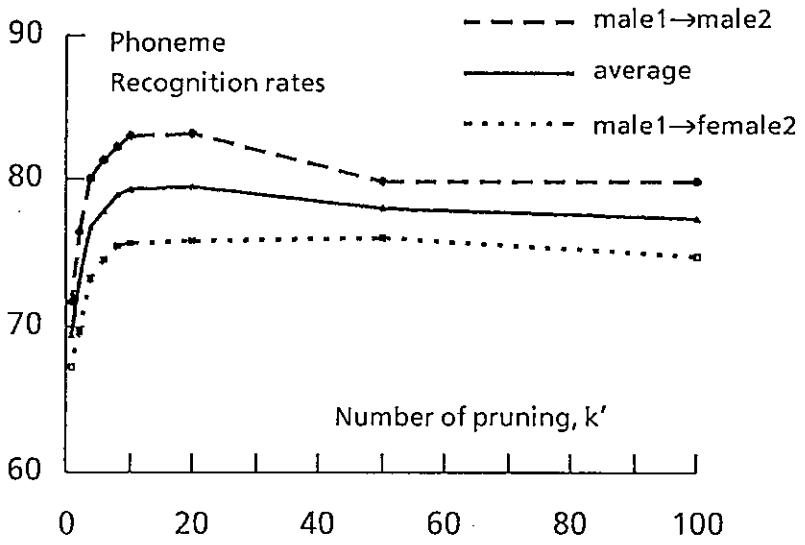

Fig.6. Phoneme recognition rates
vs. number of pruning,k'

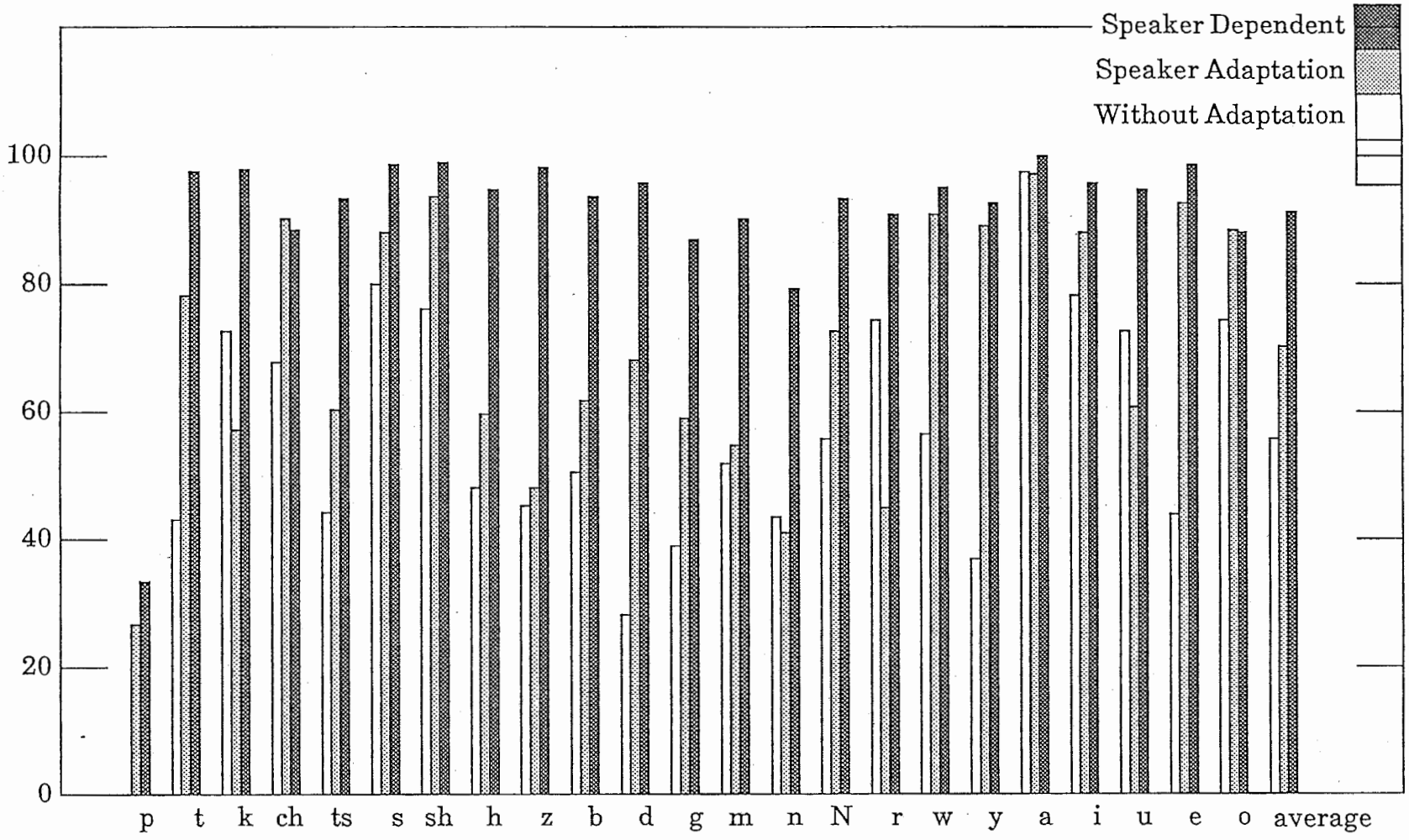

Fig.7 Phoneme recognition rates for all phonemes (male →male)

Table 5. Confusion matrix using HMM speaker adaptation (male →male)

	s	sh	h	z	ch	ts	p	t	k	b	d	g	m	n	N	r	w	y	a	i	u	e	o	data	err	1st	2nd	3rd
s	237	24			1	7																		269	32	88.1	96.3	100
sh		166			7	4																		177	11	93.8	97.7	100
h			158					50	25							4	2		21	1		2	2	265	107	59.6	80.4	89.1
z	35	33		109	33	14		1										1			1			227	118	48.0	68.3	76.2
ch		3		3	64	1																		71	7	90.1	94.4	100
ts	37	8		12	13	107																		177	70	60.5	86.4	91.5
p							4	8	2													1		15	11	26.7	40.0	66.7
t	3		4	1		1	2	172	24							8		2	3					220	48	78.2	86.8	91.8
k		1	42		3	2	6	41	133									2		1			2	233	100	57.1	81.5	92.7
b				3				2		140	17	7	20	7	1	25	1	2		2				227	87	61.7	80.2	87.7
d				3				2		20	122	2	2	5		15		5				3		179	57	68.2	86.0	90.5
g			1					2	1	10	21	149	20	3		17	1	10		9		8		252	103	59.1	75.0	82.1
m										12		47	132	22	9	14			1	4				241	109	54.8	83.4	91.7
n				4						4	7	45	86	109	3	7								265	156	41.1	76.2	89.1
N				1						2		31	12	6	177	1		2		9		3		244	67	72.5	87.7	91.4
r								2		3	2	16	8	5		108	10	10	21	8		48		241	133	44.8	66.8	78.0
w										3			1			1	71						2	78	7	91.0	94.9	97.4
y																1		155		6	2	10		174	19	89.1	99.4	100
a																	5		214			1		220	6	97.3	99.5	100
i		3	1	1		1			1									16		189		3		215	26	87.9	96.3	99.5
u	2			2		1		4		1		9	6	1	10	8		3		7	127	26	2	209	82	60.8	70.3	77.0
e																		12		4	1	209		226	17	92.5	97.8	100
o									1			1	1			1	18	1	2				189	214	25	88.3	94.9	95.8
cons																										65.8	82.3	89.8
vowel																										85.3	91.8	94.5
average																										75.6	87.0	92.1