T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	マルチテンプレートと教師なし話者適応化による音声認識
Title(English)	
著者(和文)	古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会論文誌 A, Vol. J72-A, No. 10, pp. 1476-1483
Citation(English)	, Vol. J72-A, No. 10, pp. 1476-1483
発行日 / Pub. date	1989, 10
URL	http://search.ieice.org/
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1989 Institute of Electronics, Information and Communication Engineers.

Powered by T2R2 (Science Tokyo Research Repository)

論　文

マルチテンプレートと教師なし話者適応化による音声認識

正　員　古井　貞熙†

Speech Recognition Based on Multiple-Templates and Unsupervised Speaker Adaptation Techniques

Sadaoki FURUI†, *Member*

あらまし　ベクトル量子化を基礎とする音声認識系において，不特定話者の音声を高い精度で認識することを目的として，マルチテンプレートと教師なし話者適応化の技術を組み合わせて用いる種々の方法を検討した．本方法によれば，発声内容を限定しない比較的少数の音声サンプル（認識すべき未知音声でもよい）を用いて，広い範囲の話者の変動に対して自動的に符号帳を適応させたり，未知の入力音声を符号帳に適応させることができる．話者適応化は，学習音声に含まれるスペクトルと符号帳の要素を，それぞれ階層的にクラスタ化し，対応するクラスタ間のずれを用いて，隣接クラスタ間の連続性を考慮しながら，符号帳あるいは入力音声を変化させることによって行う．男性20名が発声した100都市名単語音声の認識において，適応化なしのときに4.9％であった誤り率が，10単語を用いた適応化によって2.6％まで低下すること（特定話者認識の誤り率は2.2％）が確認された．

1. まえがき

近年HMM等を用いた不特定話者の音声認識の研究が活発に行われ，認識性能が向上しつつあるが，いかなる話者に対しても高い性能を発揮するためには，認識系に話者適応機能をもたせることが必要であると考えられる．一般に話者適応の方法は，学習音声の発声内容が既知であることが前提になっている教師ありの方法と，任意の音声を用いて適応化することができる教師なしの方法に分類できるが，これまでに試みられた方法は教師ありの方法がほとんどである．教師なしの方法でも，学習の過程で音声認識を行って陽に音素を決定し，その情報を用いて適応化を行う場合が多く，その場合には誤認識を生じたときに適切な適応化が行われない問題がある．

完全に発声内容に依存しない適応化方法が確立できれば，その用途は大きく広がると思われる．これにより，認識システムのユーザにとっては，学習用の音声の発声を必要とせず，あたかも不特定話者の認識システムであるかのように動作し，システムの側は入力された認識用の未知音声を用いて，自動的にその話者の音声に適応することが可能になる．この際更に，発声者がときどき変わっても，そのことをシステムに陽に教えることなく，自動的に適応していくことができれば望ましい(自己適応系)．このようなことができれば，話者の変動だけでなく，電話機やマイクロホンの種類，環境騒音，室内音響特性，周波数帯域制限などの変動に対しても，ある程度対応できるようになると期待される．

本論文では，このような観点から新しい教師なしの話者適応化アルゴリズムについて検討した結果について述べる．

2. マルチテンプレートと適応化による認識系

2.1 概　　要

ここでは，ベクトル量子化（VQ），すなわち特徴パラメータセットのクラスタ化によって作成した符号帳を用いて，その要素の時系列によって音声を表現する方法をベースとする認識系における話者適応化について検討する．この際，話者適応化は符号帳の要素（瞬間的なスペクトルの分布の話者による変動）に関してのみ行い，その時系列（時間構造・発声の仕方）の個人差に関しては，マルチテンプレートとDPマッチングによって対処することにする．マルチテンプレート

† NTTヒューマンインタフェース研究所，武蔵野市
NTT Human Interface Laboratories, Musashino-shi, 180 Japan

法は，不特定話者を対象とした音声認識における有力な方法の一つであり，これにより，新しい話者に関して未学習の段階でも，ある程度の認識性能を確保することができる．

ここで用いる話者適応化法は，セグメント量子化による音声符号化において，学習音声の発声内容を特定せずに，新しい話者に符号帳を適応化することを目的として，白木らが提案した方法[1]に基づいている．この話者適応化の基本的方法の音声認識における有効性についても，既に種々の実験で確認されている[2],[3]．

2.2 3種類の認識系の構成[4],[5]

ここで提案する方法は，いずれも原理的に，単語音声認識から連続音声認識まで，ベクトル量子化に基づくあらゆる種類の音声認識に適用可能であるが，ここでは簡単のため，SPLIT 法[6]に基づく単語音声認識系を想定する．ここで検討した3種類の話者適応化法を含む単語音声認識系のブロック図を図1（構成(a)～(c)）に示す．ここでは，単語辞書（マルチテンプレート）はすべての話者に共通とし，符号帳を各話者に適応化するか，入力音声のスペクトルを符号帳に適応化する方法をとる．図中 N は，話者間変動を代表する標準話者の数であり，実験では男性・女性それぞれについて，$N=4$ とした．それらの標準話者の音声を用いて，男性，女性別に共通符号帳と単語辞書（各単語4テンプレート）を作成する．

共通符号帳の作成法は次のとおりである．まず各単語について，4標準話者の音声の全音声フレームを LBG 法[7]によってクラスタ化し，32の要素からなる符号帳を作成する．次にこれらの符号帳を100単語分合わせて3,200要素の集合を作り，これに再び LBG 法を適用して，1,024の要素からなる共通符号帳を作成する．単語辞書は，各標準話者が発声した各単語の学習音声を，共通符号帳を用いてベクトル量子化することによって作成する．男性の場合，各標準話者の全学習音声の量子化に用いられる符号帳要素の数は，682, 719, 655, 717であり，1,024の全要素に対する割合は各話者平均67.7％である．1,024の要素のうち，4名の標準話者に共通して用いられる要素は357，3名に用いられる要素は270，2名に用いられる要素は190であった．共通符号帳の各要素は音素としてのラベル付けはされていないので，当然ながら，話者が変わると同じ要素が異なる音素区間に用いられることがある．

話者適応化法の具体的方法としては，符号帳を入力音声に適応させる方法と，入力音声を符号帳に適応させる方法が考えられる．符号帳を入力音声に適応される方法に比べて，入力音声を符号帳に適応させる方法には，次のような特徴がある．すなわち，前者の場合は入力に適応化する処理を繰り返していくうちに極端に偏った符号帳になってしまわないように，標準的な符号帳を常に記憶しておいて，ときどきリセットする必要があると思われるが，後者の場合はその必要がない．更に後者の場合は，話者が変わったときに，そのことを陽にシステムに教えなくても，後に説明する構成(b)を用いて，最も近い標準話者の選択と，その符号帳と入力音声との対応関係の抽出を周期的にやり直すことにより，容易に対応できる．入力音声を適応させる場合は，各フレームの特徴パラメータについて適応化処理を行う必要があるが，符号帳の適応化と同程度の処理量であり，後の認識時の処理量に比べれば極めてわずかである．

適応化法に用いる各話者の学習音声としては，ここで提案する方法では原理的に，任意の内容の複数の単語あるいは短い文章を発声した音声を用いることができ，未知入力音声をそのままプールしておいて学習に用いることもできる．後に述べる実験では，簡単のために学習音声として単語音声を用いる．

2.3 構成（a）

構成(a)（図1 Structure a）は，共通符号帳をそのまま用いて，入力音声と共通符号帳の要素間の関係に基づいて話者適応化を行うもので，符号帳を入力音声に適応化する場合を(a-1)，入力音声を符号帳に適応化する場合を(a-2)と呼ぶ．(a-1)は，文献(2)で報告した方法と同じである．

2.4 構成（b）

構成(b)（図1 Structure b）では，共通符号帳の部分集合である各標準話者の符号帳（重なりあり）をそれぞれ用いて入力音声のベクトル量子化を行い，量子化ひずみに基づいて，入力音声に最も近い標準話者の符号帳を選択する．次に，この選択された符号帳に入力音声を適応させる．上述のように，共通符号帳の要素は，異なる話者間で異なる音素として使われることがあるので，選択された符号帳の要素を入力に適応させると，他の標準話者によるマルチテンプレートに不都合な変形を生ずる可能性がある．このため構成(b)においては，入力音声の方を動かして，符号帳に適応化する方法をとる．

構成(b)は構成(a)に比べてやや複雑になるが，次

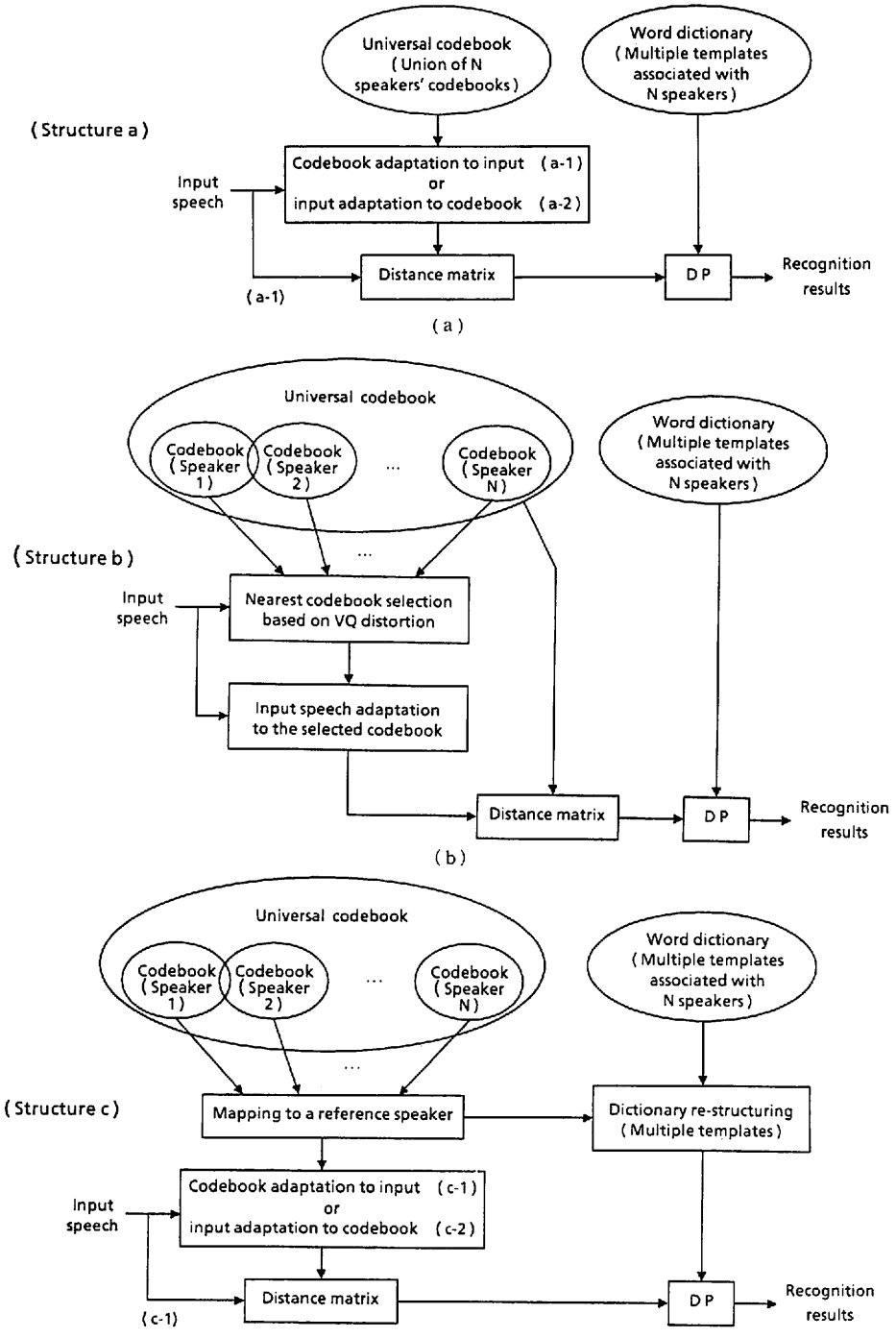

図 1 マルチテンプレートと話者適応化による単語音声認識系のブロック図

Fig. 1 Block diagram of word recognition systems with multiple-templates and speaker adaptation mechanism.

のような特徴がある．すなわち，男性，女性を含む広い範囲の話者を対象とする場合，すべての標準話者の共通符号帳を用いたのでは，各音素のスペクトルの広がりと重なりが大きく，個々の音素について適切に適応させるのが難しいと考えられる．構成(b)を用いれば，入力音声に最も近い標準話者に入力音声を合わせることになるので，広い話者の変動に対処できると期待される．

2.5 構成(c)

構成(c)(図1 Structure c)は，共通符号帳を構成する各標準話者の符号帳を，あらかじめ1人の標準話者に適応（マッピング）しておく方法である．このマッピングは，階層的クラスタ化に基づく適応化法を，標準話者の音声相互間に適用することによって行うことができる．具体的には，4人の標準話者のうち，比較的中心に近い位置にある話者を選択し，この話者の符号帳に，他の各標準話者の全学習音声を適応させる(マッピングする)．このようにして適応化された全標準話者の学習音声を用いて，共通符号帳と単語辞書を作り直す．これにより，音声に含まれる個人性情報のうち，スペクトル分布の個人差と時間構造の個人差が近似的に分離され，共通符号帳内の各音素の話者による広がりが小さくできるので，構成(b)のように入力に最も合った符号帳を選択しなくても，共通符号帳全体を用いて，入力音声を適応化することができるようになると期待される．

このようにして集約化された共通符号帳を用いて，構成(a)の場合と同様に，符号帳を入力音声に適応化する方法を構成(c-1)，入力音声を符号帳に適応化する方法を構成(c-2)と呼ぶ．

2.6 符号帳を入力音声に適応化するアルゴリズム

ここで用いる適応化法の基本的アルゴリズムのポイントは，入力（学習）音声スペクトルと標準パターン(符号帳の要素)を階層的にクラスタ化し，対応する各クラスタ重心間のずれに基づいて，スペクトルの大局的な話者特性から局所的な話者特性へと徐々に適応化を進めていくところにある．

符号帳を入力音声に適応させる場合の手順は次の通りである（図2，詳細は文献(1)，(2)，(5)参照)．(i)学習音声全フレームの重心と全標準パターンの重心のずれを求める．(ii)全標準パターンをこのずれに相当するだけ移動させて重心を一致させる．(iii)クラスタ数を2倍に増やし，学習音声をLBG法によりクラスタ化する．(iv)各学習音声クラスタの重心を求め，

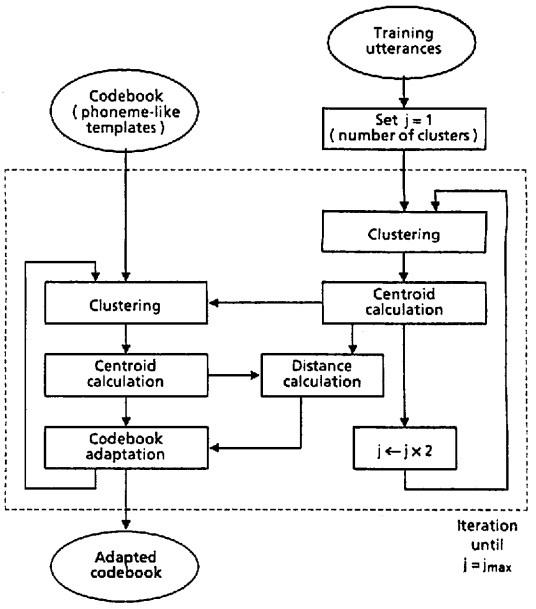

図2 符号帳を入力音声に適応化させる手順
Fig. 2 Block diagram of codebook adaptation to the input (training) speech spectra.

これらの重心を用いて全標準パターンをクラスタ化する．すなわち各標準パターンを，最も近い重心の学習音声クラスタに対応づける．(v)標準パターンの各クラスタの重心を求め，学習音声と標準パターンの対応するクラスタの重心間のずれを求める．(vi)これらのずれをもとに，隣接クラスタ間の連続性を保ちながら標準パターンを移動する．クラスタ数があらかじめ設定した数になるまで(iii)～(vi)の処理を繰り返し，標準パターンと学習音声の階層的マッピングを進める．最終状態の標準パターンを用いて，SPLIT法による入力単語音声の認識を行う．

隣接クラスタ間の連続性の保持は，標準パターンを話者に適応させて移動する際，その標準パターンが属すクラスタと，対応する学習音声のクラスタの重心間のずれに相当するだけ単純に移動するのではなく，以下のように，周辺の他のクラスタの重心のずれの重み付き平均値を用いることによって行う．

ある符号帳クラスタの重心を u_m，対応する学習音声クラスタの重心を v_m，このときのクラスタの総数を j とする．u_m と v_m のずれベクトルを p_m で表す．

$$p_m = v_m - u_m \quad (m=1, \cdots, j) \qquad (1)$$

各符号帳要素 c_i について，すべての符号帳クラスタの重心 u_m との距離を求め，その逆数の α 乗を重み $\{w_{im}\}$

としたすべてのずれベクトル $\{p_m\}$ の和を次のようにして求める．

$$\Delta_i=\frac{\sum_m w_{im}p_m}{\sum_m w_{im}} \quad (2)$$

$$w_{im}=1/\|c_i-u_m\|^{\alpha} \quad (3)$$

この Δ_i を用いて，c_i を

$$c_i'=c_i+\Delta_i \quad (4)$$

へ移動する．

以上の階層的クラスタ化に基づく方法を方法Ⅰとする．比較のために方法Ⅱとして，クラスタ化を階層的に行わず，学習音声を直接複数個のクラスタに分割し，標準パターンと学習音声のマッピングを行う方法も検討した．すなわち方法Ⅱでは，(iii)～(vi)の処理を繰り返さず，事前に定めたクラスタ数の条件下で1回のみ行う．

2.7 入力音声を符号帳に適応化するアルゴリズム

上で述べた方法は，符号帳を入力（学習）音声に適応させる場合であるが，入力音声を符号帳に適応させる場合もほぼ同様の手順で行うことができる．但しこの場合は，適応化前と後の学習音声の要素間の関係を求めるために，全学習音声をメモリに蓄えておく．まず符号帳の要素（標準パターン）をクラスタ化し，この重心を用いて学習音声をクラスタ化する．標準パターンと学習音声の対応するクラスタの重心間のずれを求め，このずれをもとに学習音声を標準パターンに合うように，隣接クラスタ間の連続性を保ちながら移動する．クラスタの数を2倍に増やしながらこの処理を繰り返して，学習音声を符号帳に適応させる．最終状態の学習音声の各フレームの特徴ベクトル y_n' ($n=1, \cdots, n_{\max}$, $n_{\max}$：学習音声のフレーム数)と，メモリに蓄えておいた適応前の学習音声の対応するフレームの特徴ベクトル y_n のずれベクトル

$$q_n=y_n'-y_n \quad (5)$$

を計算する．

入力音声の各フレームの特徴ベクトル x に関して，最も近い学習音声フレームの特徴ベクトル y_{n0} を探し，その学習音声フレームのずれベクトルを用いて，移動（適応化）処理を行う．

$$x'=x+q_{n0} \quad (6)$$

$$n_0=\underset{n=1,\,n_{\max}}{\operatorname{argmin}}\|x-y_n\| \quad (7)$$

x' を適応化後のフレームの特徴ベクトルとする．

なおこの方法の場合は符号帳は固定されているので，その要素のクラスタ化は適応化のたびに行う必要はなく，あらかじめ階層的にクラスタ化した結果を蓄えておけば，記憶容量は2倍になるが計算量を減らすことができる．

3. 認識実験

3.1 実験方法

日本語100都市名を認識対象とし，標本化周波数8 kHz，フレーム周期16 ms，分析次数10の条件で，LPCケプストラムにより分析を行った．特徴パラメータとして，ケプストラムのみを用いる場合(Cep)と，これにケプストラムおよび対数パワーの短区間(88 ms)の回帰係数[8]を組み合わせる場合(Cep & ΔCep)の2種類の実験を行った．(Cep & ΔCep)を用いる場合は，距離計算において両者のパラメータの寄与が同程度になるように，適当な重みを与えた．男性・女性それぞれについて，4名の標準話者の学習音声から共通符号帳と単語辞書（各単語4テンプレート）を作成した．未知音声として，この4名とは異なる男性20名，女性16名の音声（男性2,000サンプル，女性1,600サンプル）を用い，適応化のための学習音声には，各話者が別に発声した単語音声を用いた．

適応化法の有効性の評価は，入力音声に関するスペクトルひずみと認識率の両方を用いて行った．スペクトルひずみとしては，各入力音声に関し，(Cep)または(Cep & ΔCep)を用いて，その単語のテンプレートとの時間軸正規化後の全フレームにおける平均距離を算出し，4種類のマルチテンプレートとの距離の中の最小値を求めた．スペクトルひずみと認識率の，全話者の全未知入力音声に関する平均値を評価に用いた．

3.2 構成(a-1)に関する実験結果

基本形として，構成(a-1)において符号帳を入力に適応させる場合を取り上げ，種々の実験を行った．標準パターンの移動における連続性の度合を決めるパラメータ α の値は，予備実験の結果をもとに，最適値 $\alpha=1$ に設定した．図3に，男性・女性間で適応化を行った場合の，クラスタ数と，適応化を行う前の値で正規化したスペクトルひずみの関係を示す．特徴パラメータとして(Cep)を用いる場合について，方法Ⅰによってクラスタ数を増やしながら適応化を行ったときの推移と，方法Ⅱによってクラスタ数を種々に選択して適応化を行ったときの結果を示した．学習単語数は10単語とし，単語セットを2種類に変えた場合の，全未知音声に関する平均値を示した．

これらの結果から，方法Ⅰの場合は適応化の進行と

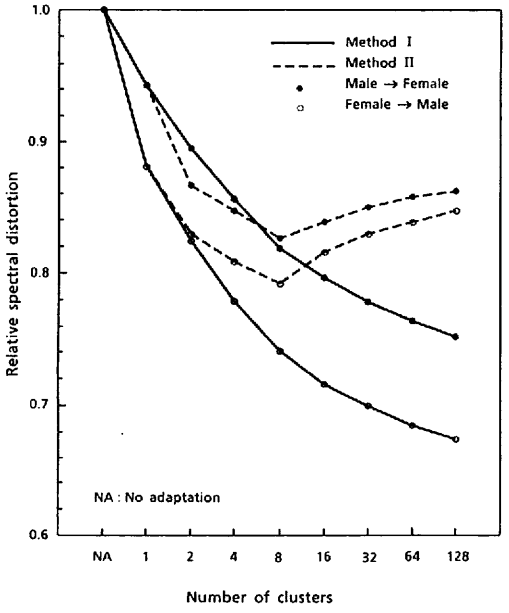

図 3　適応化による未知音声と単語標準パターンのスペクトルひずみの変化（男性・女性間の適応化）
Fig. 3　Change in spectral distortion between input speech and word reference templates as speaker adaptation progresses (adaptation between male and female speakers).

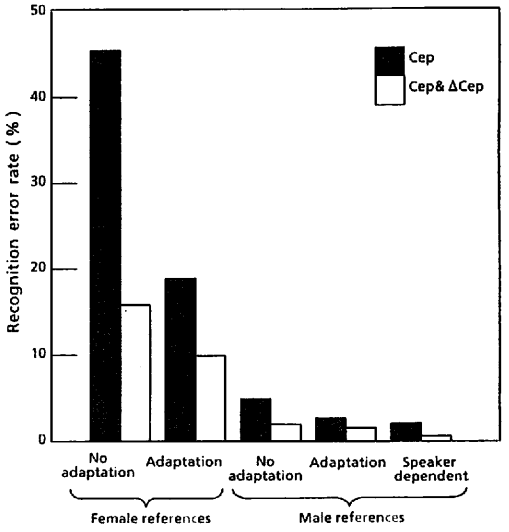

図 4　種々の条件における男性音声の認識誤り率
Fig. 4　Recognition error rates under various experimental conditions.

共にひずみが低下するが，方法IIの場合はクラスタ数とひずみの間には単純な関係はなく，条件をうまく選ばないとひずみが大きくなってしまうことがわかる．クラスタ数を最適値に選んでも，男女間のようにもとのスペクトルのずれが大きい場合には，方法 I と II で達成できるひずみの下限には大きな差がある．このことは，階層的方法がひずみを確実に低下させる上で有効性が高いことを示している．この結果から，以下では方法 I のみを用いて実験を行った．特徴パラメータに関しては，(Cep & Δ Cep)の場合よりも(Cep)の場合の方が，適応化の効果が大きい傾向がある．

図 4 は，男性の入力音声を男性または女性の符号帳および単語辞書を用いて認識した場合の，認識誤り率を示したものである．適応化なし(不特定話者)，方法 I を用いて 128 クラスタまで適応化，話者ごとに全 100 単語を別に発声した音声から各単語辞書を作成(特定話者，全単語学習)の条件の場合についての結果を示した．各条件について，学習単語セットを上述と同様に 2 種類に変えた場合の平均結果を示してある．ここで提案した適応化法が，認識誤り率を低下させる上で極めて有効性が高いことがわかる．標準パターンが男性で(Cep)を用いる場合，適応化前に 4.90 %であった誤り率が適応化により 2.88 %まで低下している(特定話者の場合の誤り率は 2.15 %)．(Cep & Δ Cep)の場合の適応化の効果の度合は，(Cep)の場合よりもやや小さいが，この場合も適応化により誤り率が 1.95 %から 1.55 %に低下している．

3.3　構成(a-1)による補足実験

これまでの実験では，クラスタ間の連続性を保持するためのパラメータ α の値を 1 に設定していたが，この値の大きさによって適応化の効果は変化する．学習単語数を 10，適応化を 128 クラスタまでとしたときの，α の値によるスペクトルひずみの変化を男性・女性別々に調べた結果を図 5 に示す．式(3)から明らかなように，α の値が小さくなるほど，等価的に連続性を考慮する範囲，あるいは標準パターンの移動量を平滑化する範囲が広くなる．$\alpha=\infty$ は連続性を考慮せず，その標準パターンが属すクラスタの重心のずれベクトルをそのまま用いることに相当する．

この結果から，連続性を考慮することは極めて重要であり，連続性を考慮しないと適応化の効果は全くなく，むしろひずみが増加してしまうことがわかる．α の最適値は男性，女性に共通して 0.5〜1 の範囲にあり，$\alpha=0.7$ 前後でひずみが最小値を示すが，この近傍での α の変化に対するひずみの変化は比較的小さい．この結果から，計算の容易さを考えると，$\alpha=1$ が最適と言

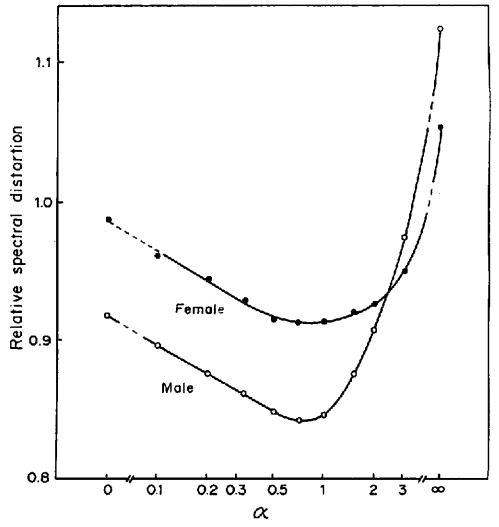

図 5 クラスタ間の連続性を保持するための パラメータ α の値と適応化の効果の関係

Fig. 5 Relative spectral distortion after speaker adaptation as a function of the smoothing coefficient α.

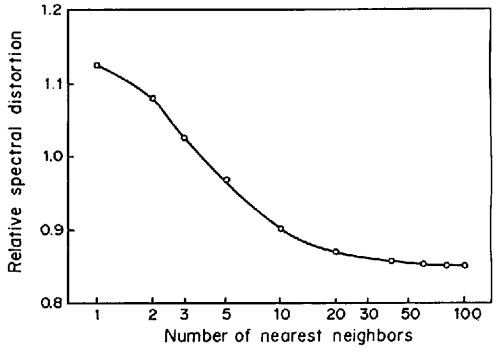

図 6 平均化を行う近傍ずれベクトルの数と適応化の効果の関係

Fig. 6 Relative spectral distortion after speaker adaptation as a function of the number of nearest neighbor deviation vectors used in the averaging process.

えよう．

図6は，男性の音声を用いて，連続性を考慮する近傍重心点の数，すなわち式(2)において平均化を行うずれベクトルの数を変えたときの，スペクトルひずみの変化を示したものである．この結果は，かなり広い範囲のずれベクトルを用いて平均化を行うことが必要であることを示している．

3.4 入力を符号帳に適応させる場合の結果（構成(a-2)(b)）

構成(a-2)と(b)を用いて，入力音声を符号帳に適応化させる方法で行った認識実験の結果を，適応化を行う前の結果と併せて表1に示す．実験には男性の音声を用い，特徴パラメータとしては(Cep)を用いた．構成(b)については，入力音声を適応化させる符号帳として，VQのひずみ量によって入力音声に最も近い標準話者の符号帳を選ぶ場合と，無作為に選ぶ場合の比較を行った．

表1 入力音声を符号帳に適応させる場合の認識誤り率

適応なし	構成 (a-2)	構成(b) (無作為 話者選択)	構成(b) (最適 話者選択)
4.90%	3.40%	3.15%	2.85%

表2 集約された共通符号帳を用いる場合の認識誤り率

適応なし	構成 (a-2)	構成 (c-1)	特定話者
4.90%	2.88%	2.60%	2.15%

この結果から，入力音声を符号帳に適応化する場合，多数話者の共通符号帳に適応化するよりも，1名の標準話者を選んで適応化した方がよく，VQのひずみ量によって入力音声に近い符号帳を選んで適応化するのが更によいことがわかる．但し入力音声を符号帳に適応化する場合の認識誤り率は，符号帳を入力音声に適応化する場合よりもやや大きくなる傾向があるため，構成(b)によって改善された後の誤り率は，構成(a-1)によって共通符号帳を入力音声に適応させる場合の結果とほぼ同程度である．

ここで得られた結果は男性音声のみの場合のものであるので，男性と女性が混在する場合や，より広い範囲の声質の変動に対処する場合には，構成(b)の特徴が更に明確に出るものと考えられる．

3.5 構成(c-1)による実験結果

構成(c-1)，すなわち集約化された共通符号帳を入力音声に適応化する方法による認識実験の結果を，適応化前，集約化を行わなかった場合（構成(a-1)）と比較して表2に示す．この結果から，複数標準話者の音声から作成した共通符号帳を，あらかじめ話者間で適応化処理を行って集約化することにより，認識誤り率を低下させることができることがわかる．ここで検討しなかったような，更に広い範囲の話者変動に対しては，構成(b)と構成(c)の組合せが良い結果をもたらすと期待される．

4. む す び

マルチテンプレートと話者適応化の組合せによる不特定話者音声認識法を対象として，発声内容を限定しない比較的少数の音声サンプルによる，話者適応化法（構成(a)，(b)，(c)）を提案し，SPLIT 法による単語音声認識に適用して有効性を検討した．その結果，任意の 10 単語による学習で，本方法により誤認識が大幅に減らせることがわかった．本方法では，符号帳と単語辞書を複数の標準話者の音声から作成し，マルチテンプレート構成の単語辞書は，すべての話者に共通に用いる．符号帳に関しては，符号帳を入力音声に適応化するか，あるいは入力音声を各フレームごとに符号帳に適応化する．この処理は，学習単語に含まれるスペクトルと符号帳要素を，それぞれ階層的にクラスタ化し，対応するクラスタ間の重心のずれを用いて，大局的特徴から微細な特徴へ徐々に合わせていく方法によって行う．この方法では，クラスタ化を階層的に行うことと，クラスタ間の連続性を保持する，一種のぼけ効果が重要な働きをしている．

ここで得られた実験結果から，男性・女性・子供を同時に含む広い範囲の話者変動に対処するためには，例えばそれら 3 種類の符号帳を，それぞれ多数の話者から作成し，入力音声に最も近い符号帳を選択して，その符号帳に入力音声を適応化するのがよい（構成(b)）と考えられる．更に各符号帳に関しては，それを構成する多数話者の符号帳間でマッピングを行って集約しておく（構成(c)）のがよい．符号帳を適応化する方法に比べて，入力音声を適応化する方法はやや認識誤り率が高いが，メリットもあるので，今後更に検討を加えていく必要がある．

ここで検討した方法は原理的に，SPLIT 法による単語音声認識だけでなく，HMM をベースとする連続音声認識系など，種々の音声認識系に広く適用することが可能であり，雑音や伝送系によって変形したスペクトルへの適応などにも用いることが可能と考えられる．実際の認識系では，認識結果の確実性の度合により，教師ありの方法[9]を本方法に組み合わせて話者適応化を行うことも考えられる．

謝辞　貴重な討論をして頂いた，当所嵯峨山主幹研究員，基礎研究所・菅田主幹研究員および白木主任研究員に感謝する．

文　　献

(1) 白木善尚，菅田雅彰："セグメント符号化における話者適応化"，音響学会音声研資，SP87-67 (昭 62-10)．
(2) 古井貞熙："スペクトル空間の階層的クラスタ化による話者適応化"，音響学会音声研資，SP88-21 (昭 63-06)．
(3) 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"，音響学会音声研資，SP87-118 (昭 63-01)．
(4) 古井貞熙："マルチテンプレートと話者適応化による音声認識"，音響学会春季講論，2-6-10 (平 1-03)．
(5) S. Furui: "Unsupervised speaker adaptation method based on hierarchical spectral clustering", Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Glasgow, S 6.9 (May 1989).
(6) 菅村　昇，古井貞熙："擬音韻標準パターンによる大語い単語音声認識"，信学論(D)，**J65-D**, 8, pp. 1041-1048 (昭 57-08)．
(7) Y. Linde, A. Buzo and R. M. Gray: "An algorithm for vector quantizer design", IEEE Trans. Commun., **COM-28**, 1, pp. 84-95 (Jan. 1980).
(8) S. Furui: "Speaker-independent isolated word recognition using dynamic features of speech spectrum", IEEE Trans. Acoust. Speech & Signal Process., **ASSP-34**, 1, pp. 52-59 (Feb. 1986).
(9) K. Shikano, K-F. Lee and R. Reddy: "Speaker adaptation through vector quantization", Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Tokyo, 49.5 (April 1986).

(平成元年 5 月 8 日受付)

古井　貞熙

昭 43 東大・工・計数卒．昭 45 同大大学院修士課程了．同年 NTT 電気通信研究所入社．以後，同研究所において，音声認識，話者認識，音声知覚などの基礎研究に従事．昭 53～54 ベル研究所客員研究員．現在，NTT ヒューマンインタフェース研究所音声情報研究部長．工博．昭 50 米沢賞，昭 63 論文賞，昭 60，62 音響学会論文賞，平成元年科学技術庁長官賞，IEEE ASSP Society Senior Award 各受賞．著書「ディジタル音声処理」「Digital Speech Processing, Synthesis, and Recognition」「音声情報工学」(分担)，「音の科学」(分担)など．IEEE，日本音響学会各会員．