論文 / 著書情報
Article / Book Information

論題(和文)	音声認識のためのスペクトル内挿を用いた話者適応化
Title(English)	
著者(和文)	篠田浩一, 磯健一, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	電子情報通信学会論文誌A, Vol. J77-A, No. 2, pp. 120-127
Citation(English)	, Vol. J77-A, No. 2, pp. 120-127
発行日 / Pub. date	1994, 2
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) 1994 Institute of Electronics, Information and Communication Engineers.

論　文

音声認識のためのスペクトル内挿を用いた話者適応化

正員　篠田　浩一[†]　　正員　磯　健一[†]　　正員　渡辺　隆夫[†]

Speaker Adaptation Using Spectral Interpolation for Speech Recognition

Koichi SHINODA[†], Ken-ichi ISO[†] *and* Takao WATANABE[†], *Members*

あらまし　連続分布型HMMに基づく音声認識のための教師あり話者適応化の手法を提案する．使用者の負担を軽くするために適応化用単語の発声数を少なくすると，その中に現れない半音節（認識単位）の割合が増える．本手法では，適応化用データに含まれる半音節のHMMパラメータを補正した後で，適応化用データには含まれていない半音節のHMMパラメータをパラメータ空間における内挿（スペクトル内挿）により補正する．更に，適応化用データセットに依存した偏ったパラメータが推定されるのを避けるため，多数話者の多数発声データに基づく補正を行う．5,000単語大語い音声認識をシミュレートした類似100単語認識実験を行い，提案手法を評価した．不特定話者HMMによる認識率81.2%のところ，50単語を適応化に用いたとき，85.2%まで認識率が上昇し，本手法の有効性が確かめられた．

キーワード　話者適応化，不特定話者認識，HMM，スペクトル内挿，語い依存性

1. まえがき

本論文では，連続分布型隠れマルコフモデル(Hidden Markov Model；HMM)に基づく音声認識のための話者適応化法を提案する．

現在，音声認識の分野では，HMMを用いた認識方式が主流である．HMMは，音声の発声を確率的事象とみなすことにより，さまざまな要因から起きる音響特徴量の揺らぎを比較的容易に扱うことができる．また，Baum-Welchアルゴリズムと呼ばれるパラメータ推定アルゴリズムが存在し，推定に用いる音声データが十分にある場合には，高い認識性能を示す．しかしながら，実用に用いる場合には，必要なデータ量が多く，使用者の発声の負担が大きい．近年盛んに研究されている不特定話者認識方式は，多数の話者の音声を用いてパラメータを推定し，話者間の発声の揺らぎに対しても頑強である．使用者の事前発声は不要であるという利点がある．しかし，認識性能の極端に悪い特異話者が存在する，認識性能が特定話者方式に及ばないなどの問題がある．そこで，本論文では，不特定話者認識方式をベースとした話者適応化を試みる．

過去にHMMの話者適応化法としてさまざまな手法が検討されてきた．それらは大きく2種類に分けられる．適応化に用いる単語あるいは文が既知である教師あり適応化[(1)~(3)]と，任意の発声を許す教師なし適応化[(4)~(6)]である．一般に，適応化用の発声データ量が同じであれば，教師あり適応化の方が認識性能が高い．本論文では教師あり適応化手法を検討する．話者の負担を軽減するために話者適応化に用いる音声データ量はなるべく少ないことが望ましい．しかしながら，音声データが少量の場合，適応化後の認識性能が適応前に比べ，変わらないか，あるいは，しばしば劣化することが問題点として指摘されている．原因としては以下のようなことが考えられる．

例えば，半音節単位の連続分布型HMMを用いた認識方式において適応化用単語数10単語で適応化する場合を考えると，10単語中には全半音節のうち3分の1程度の種類の半音節が含まれているに過ぎず，残り3分の2の半音節のパラメータは適応化されない．これらのパラメータは，適応化前HMMのパラメータが適応化されずに残り，認識時にかえって悪影響を与える可能性がある．このような，適応化用データが少量のとき未適応のパラメータが出てくるという問題は，他の認識方式についても起こる可能性がある．

また，対応する適応化用データがあり，適応化の可能なパラメータにおいても，適応化用データ量が少数であるために，適応化後のパラメータが適応化用デー

† 日本電気株式会社情報メディア研究所，川崎市
Information Technology Research Laboratories, NEC Corporation, Kawasaki-shi, 216 Japan

タのコンテキストに強く依存したものになるという問題がある．ここでのコンテキストとは，発声における韻律，アクセント，音素文脈などを指す．本論文では，この意味でのコンテキストを特に語いコンテキストと呼ぶ．一般に，音声の音響的特徴量は，同じ音韻を発声する場合でも，その語いコンテキストに依存して異なるものになる．適応化用データが十分にある場合には，その中にさまざまな語いコンテキストにおける発声が多数存在するために，それらを統計的に処理し，語いコンテキストの違いに対し頑強なパラメータを推定することが可能である．しかし，適応化用データが少ない場合，それら少ない適応化用データの語いコンテキストにのみに適するようにパラメータが適応化され，それ以外の語いコンテキストをもつ発声を認識する際に逆に性能が劣化する可能性がある．ここで述べた問題は，従来の話者適応化の研究では指摘されていない点であるが，少数発声を用いた教師あり話者適応化を行う際には，不可避な問題である．本論文では，これをパラメータの語い依存性と呼ぶ．

提案する話者適応化は，これら二つの問題に対し，それぞれ解決案を示している．まず，適応化用データを用いてパラメータの適応化前後の差分（適応化ベクトル）を求め，適応化用データのないパラメータの適応化ベクトルをパラメータ空間における内挿により求める処理を行う[11]~[13]．この処理を本論文ではスペクトル内挿と呼ぶ．また，あらかじめ用意された多数話者の発声データから話者に独立な語い依存性を補正ベクトルとして抽出し，少量発声によるスペクトル内挿後に，語い依存性の補正に用いる[14]．

提案手法の効果を確認するため，半音節を認識単位とした連続分布型HMM音声認識方式[15]を用いて5,000単語大語い認識をシミュレートした類似100単語認識実験を行った．

本論文は以下のように構成されている．**2.**では，提案手法の評価を行った，半音節連続分布型HMMを用いた音声認識方式について述べる．続く**3.**ではスペクトル内挿話者適応化法について，**4.**では語い依存性を補正する手法について述べる．最後に，大語い音声認識をシミュレートした認識実験による評価実験の結果を示す．

2. 半音節連続分布型HMMによる音声認識

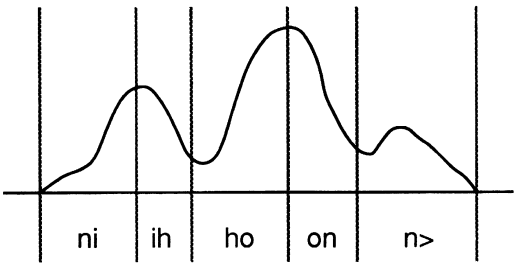
この章では，半音節を認識単位として用いた連続分布型HMMに基づく音声認識方式について述べる[15],[16]．

半音節は基本的には音節をその母音中心で半分に分割した単位である．日本語の音節は，基本的にC+Vの構造をもっているので，半音節は，基本的にはCV，VCセグメントとなる．これに加え，連続母音，長母音，促音，無音を表現するセグメントを用いる．半音節セグメントの例を図1に示す．

半音節には，次の二つの特徴がある．まず，音素認識において重要な，音素間の遷移の部分を中心に含んでいるために，調音結合による発声変形を効果的に扱うことができる．また，半音節は，VCV，CVCなどの3音素環境単位と比べると著しく種類が少なく，比較的小規模な発声データでパラメータを精度良く推定できる．

各々の半音節は図2に示すようなleft-to-right HMMで表される．大部分の半音節HMMは4状態をもつ．長母音と無音のモデルは1状態である．各々の状態の出現確率は混合ガウス分布で表される．パラメータ数を減らすため，混合ガウス分布の各々の成分ガウス分布の共分散行列の非対角成分は0としてある．また，無声化，長母音化などの大きな発声変形に対しては，別の変形用モデルを用意する．全モデル数は241個である．

図3に半音節HMMを用いた音声認識システムの



"n i h o n"

図1 半音節単位
Fig. 1 Demi-syllable unit.

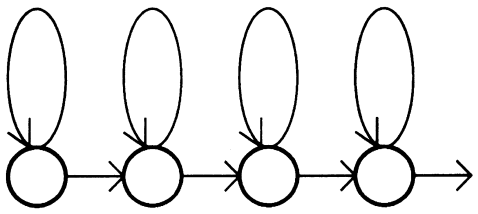


図2 半音節HMM
Fig. 2 Demi-syllable HMM.

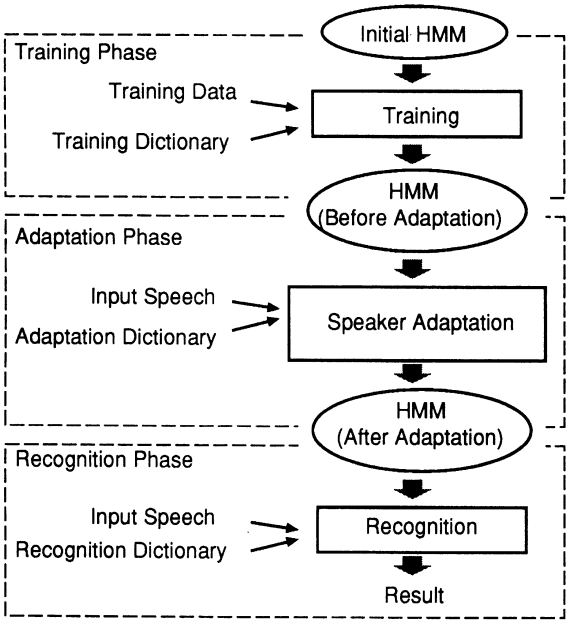


図 3 認識系の構成
Fig. 3 Recognition system diagram.

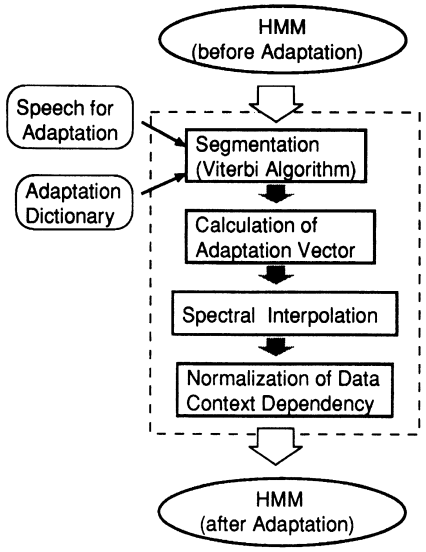


図 4 話者適応化の構成
Fig. 4 Speaker adaptation system diagram.

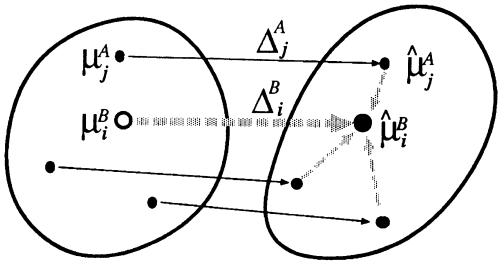


図 5 スペクトル内挿
Fig. 5 Spectral interpolation.

構成を示す．学習は Baum-Welch アルゴリズムを用いる．認識処理では，認識用辞書の各単語について，半音節モデルを連結して単語モデルを生成し，最大ゆう度を与える単語を認識結果として求める．

3. スペクトル内挿話者適応化法

混合ガウス分布を用いた連続分布型 HMM では，4種類のパラメータが適応化の対象となる．状態間の遷移確率，状態内の各ガウス分布の重みを表す重み係数，ガウス分布の平均ベクトルと分散である．適応化に用いるデータ量が少ない場合，分散や遷移確率の推定精度が悪いことが報告されている(例えば文献(3))．この点を考慮して，適応化するパラメータは平均ベクトルのみとした．図4に話者適応化の構成を示す．

まず，すべての状態の連続確率分布が単一ガウス分布の場合について説明する．適応化の過程は2段階に分けられる．第1の段階では，話者の発声した適応化用データを用いて適応化を行う．まず，適応化前の HMM を用いて，適応化用データをビタビアルゴリズムを用いてセグメンテーションする．そして，HMM の各状態に対応づけられた特徴ベクトルを平均して，その状態の平均ベクトル $\hat{\mu}^A$ とする．

第2の段階では，適応化用データ語い中に含まれない半音節に対応する HMM をスペクトル内挿と呼ぶ手法を用いて適応化する．適応化用データに含まれる HMM の各状態の平均ベクトルの集合を集合 A，含まれない HMM の各状態の平均ベクトルの集合を集合 B とする．まず，集合 A のすべての状態について適応化ベクトル Δ^A が計算される．適応化ベクトルは，適応化後の平均ベクトル $\hat{\mu}^A$ と，適応化前の平均ベクトル μ^A の差として定義される．次に，集合 B の状態の適応化ベクトルを適応化するために，集合 A の状態の適応化ベクトルを内挿することにより求める．従来，VQ コードブックの話者適応化[7],[8]，あるいは，VQ コードブックを用いた教師なし話者適応化[5] で用いられた内挿式を，平均ベクトルの内挿に適用する．図5はスペクトル内挿の原理を表す図である．アルゴリズムは以下のとおりである．

(1) 集合 A の状態 j においては，適応化後の平均ベクトル $\hat{\mu}_j^A$ は既に求められている．適応化ベクトル Δ_j^A は以下の式で与えられる．

$$\Delta_j^A = \hat{\mu}_j^A - \mu_j^A \qquad (1)$$

ここで，A は状態 j が集合 A に属することを示す添字である．適応化ベクトル Δ_j^A は集合 A におけるすべ

ての状態について計算される．

（2） 集合 B の状態 i に対して，適応化ベクトル Δ_i^B は，集合 A の状態 j の適応化ベクトルを内挿することにより求める．

$$\Delta_i^B=\sum_j w_{ij}\Delta_j^A \qquad (2)$$

適応化ベクトル Δ_j^A への重み w_{ij} は μ_i^B と μ_j^A との距離 d_{ij} の関数として定義される．例えば，w_{ij} は以下のように定義される．

$$w_{ij}=\frac{d_{ij}^{-m}}{\sum_{j'} d_{ij'}^{-m}} \qquad (3)$$

ここで m は重み w_{ij} の距離 d_{ij} への依存度を表す定数である．適応化ベクトル Δ_i^B は集合 B に属するすべての状態について計算される．

（3） 新しい話者の状態 i の平均ベクトル $\hat{\mu}_i^B$ は，次式で与えられる．

$$\hat{\mu}_i^B=\mu_i^B+\Delta_i^B \qquad (4)$$

ここで，μ_i^B は標準話者の HMM の平均ベクトルである．

（4） (2)～(3)の過程を集合 B のすべての状態について繰り返す．

上の手続きは，出力確率分布が混合ガウス分布である HMM にも，状態内の複数の成分分布を別々に扱うことにより，適用することができる．第1段階のビタビアルゴリズムによるセグメンテーションにおいては，状態内の成分分布のうち，対応する特徴ベクトルの出力確率に重み係数を乗じた値が最大になるものを選び，集合 A に分類する．対応する適応化用データのない成分分布は集合 B に分類される．第2段階のスペクトル内挿は，集合 B の成分分布に対して行われる．すなわち，集合 B の成分分布の適応化ベクトルは，すべての状態にわたる集合 A の成分分布の適応化ベクトルを用いたスペクトル内挿で求められる．

なお，本方式の提案後に，服部ら[9]，大倉ら[10]により提案された移動ベクトル場平滑化方式は，本章で述べたスペクトル内挿の処理後に，特徴ベクトル空間上で適応化ベクトルのスムージングを行う手法である．

4. 語い依存性の補正

半音節を認識単位とした連続分布型 HMM における語い依存性の補正手法について述べる．

十分な量の発声データを使った Baum-Welch アルゴリズムを用いた推定では，さまざまな語いコンテキストにおける発声データが存在し，異なる語いコンテキストにおける発声の音響的特徴量の違いを，統計的に処理し，その他の原因から起きる発声の揺らぎと共に，連続確率出力分布の形で表すことができる．ところが，少数の適応化用発声を用いて各状態の平均ベクトルを適応化する場合には，各半音節の出現頻度は小さく，各状態の平均ベクトルは適応化用単語の語いコンテキストに強く依存したものになる．この語い依存性は，話者によらない話者共通の傾向をもつと考えられ，そして，あらかじめ用意した多数話者の発声データからこの傾向を抽出できれば，それを用いて語い依存性を補正できる．補正には，スペクトル内挿の場合と同様，平均ベクトル空間における補正ベクトルを用いる．

まず，多数話者の多数語いの発声で不特定話者 HMM M_{CI} を作成する．この HMM は通常の不特定話者モデルに相当するが，多くの種類の語いコンテキストを含む発声データを用いてパラメータを推定してあり，そのパラメータは特定の語いコンテキストに依存していないとみなすことができる．次に，多数話者の適応化用単語の発声を用いて不特定話者 HMM M_{CD} を作成する．このとき，適応化用単語以外の条件はモデル M_{CD} 作成時の条件と同一にする．このモデル M_{CD} は，適応化用単語の語いコンテキストに依存したパラメータをもつ．これら二つの HMM の相違は語いコンテキストの違いのみを反映していると考えられる．そこで，M_{CD} から M_{CI} への写像を作成し，その写像を用いて教師あり適応化後のモデルを補正する．

適応化データの存在するすべての分布 i について，適応化ベクトルを求めた後，以下の処理を行う(図6)．

（1） 多数話者の適応化用単語の発声データを用いて，適応化用単語の語いコンテキストに依存した平均ベクトル μ_i^{CD} を適応化する．分散，遷移確率，重み係数は固定し，平均ベクトルのみを適応化する．適応化前 HMM として不特定話者モデル M_{CI} を用いる．

（2） 平均ベクトル μ_i^{CD} とモデル M_{CI} の平均ベク

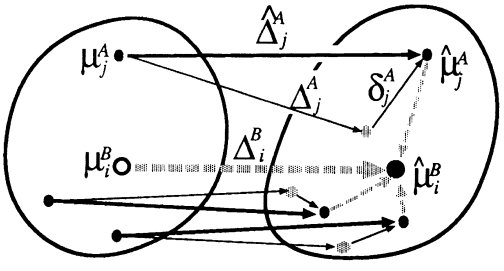


図 6 語い依存性の補正
Fig. 6 Normalization of data context dependency.

トル μ_i^{CI} のとの差ベクトル δ_i を求める．

$$\delta_i = \mu_i^{CI} - \mu_i^{CD} \quad (5)$$

δ_i を分布 i の語い補正ベクトルと呼ぶ．

（3） 語い補正ベクトルを，適応化ベクトルに加える．

$$\hat{\Delta}_i^A = \Delta_i^A + \delta_i^A \quad (6)$$

今までに本手法と同様にあらかじめ用意された多数話者の発声データを利用した話者適応化法がいくつか提案されている．例えば，Lee らは，多数話者のデータから事前確率分布を推定している[3]．また松岡らは，混合ガウス分布の重み係数のみを適応化し，残りのパラメータは不特定話者 HMM のものを用いている[17]．これらの手法では，周囲雑音，使用マイクなどの認識時の周囲環境が多数話者データ収録時の環境と同一である必要がある．それに対し，本手法は多数話者のデータから語い依存性のみを抽出して利用するため，多数話者データ収録時と認識時との環境が異なっている場合に対しても適用可能である．

5. 実　　験

5.1 スペクトル内挿話者適応化

5.1.1 実験条件

半音節 HMM 音声認識方式を用いて評価実験を行った．最初に，提案するスペクトル内挿話者適応化法(SA)を特定話者(SD) HMM を適応化前 HMM として用いて評価した．出現確率分布は単一ガウス分布を用いた．次に，2 ガウス混合分布を出力確率分布とする不特定話者(SI) HMM を適応化前 HMM として用い評価した．

この実験には以下の三つのデータセットを用いた．第 1 のデータセット(DB1)は 46 名の男性話者の発声からなる多数話者データベースである．第 2 のデータセット(DB2)は適応化用データであり，9 名の男性話者(S1, S2, S3, S4, S5, S6, S7, S8, S9)で構成される．これら 9 名の話者は多数話者データベース DB1 には含まれない．第 3 のデータセット(DB3)は，評価用で，DB2 と同じ 9 名の男性話者の発声からなる．各々のデータセットに対し，日本語辞書から音素のバランスを考慮した 250 単語を選んだ．データセット DB3 の単語は，DB1 および DB2 とは異なる．すべてのデータセットにおいて各単語は各々の話者により 1 回ずつ発声されている．

認識対象は，重要語 5,000 単語とした．認識実験においては，実験の効率化のため，各入力単語について，5,000 単語の中から，類似した 100 単語をあらかじめ選択しておき，それらを認識対象とする認識実験を行った．類似 100 単語はあらかじめ定義された音素間の類似度を用いて，音素列として表現された単語同士の類似度を DP マッチングで算出することにより求めた．類似 100 単語認識は 5,000 単語認識を良好にシミュレートする[16]．

表 1 単語数 W に対する集合 A に属する状態数 S_A と S_A の全状態数 S_T に対する比率

W	1	10	50	100	250	
S_A	42	268	613	701	771	919
S_A/S_T (%)	4.6	29.2	66.7	76.3	83.9	100.0

適応化用データセット DB2 に現れる半音節 HMM の状態数(S_A)の全状態数(S_T)に対する比の値を，ある話者 1 名の 1 回発声を用い，いくつかの単語数 W について調べたものを表 1 に掲げる．ここで単語数 W の状態数とは，単語番号 1 から W までの W 単語の発声に出現する状態数である．250 単語での状態数(771)が全状態数 S_T(919)と等しくないのは，発声変形用モデルの状態など，対応するデータが発声データ中に存在しない状態があるためである．10 単語で適応化用データに全状態の 30%が現れ，50 単語では 67%が現れることがわかる．

発声は，サンプリング周波数 16 kHz でディジタル化され，10 ms 周期で分析された．使用した特徴量は，21 次元で，その内訳は，パワー差分，メルケプストラム 10 次元，メルケプストラム差分 10 次元である．

5.1.2 特定話者モデルからの適応

最初に，スペクトル内挿適応化の効果を特定話者モデルを適応化前 HMM として評価した．適応化前 HMM は DB2 の 250 単語発声で学習した話者 S1 の特定話者 HMM を用い，評価話者として S2 を用いた．出現確率分布は単一ガウス分布である．スペクトル内挿には，式(3)の重み w_{ij} を用いた．ここで，d_{ij} は，μ_i^B と μ_j^A の間のユークリッド 2 乗距離，パラメータ m は 1.0 とした．図 7 は，適応化に用いたデータ量が 10 単語と 50 単語のときの，5,000 単語認識をシミュレートした類似 100 単語認識実験の結果を示す．図 7 において，方法 A では，Baum-Welch アルゴリズムを用いて，適応化用単語に現れる半音節モデルのガウス分布の平均ベクトルと分散および遷移確率を推定する．方法 B は，**3.** で述べたスペクトル内挿話者適応化におい

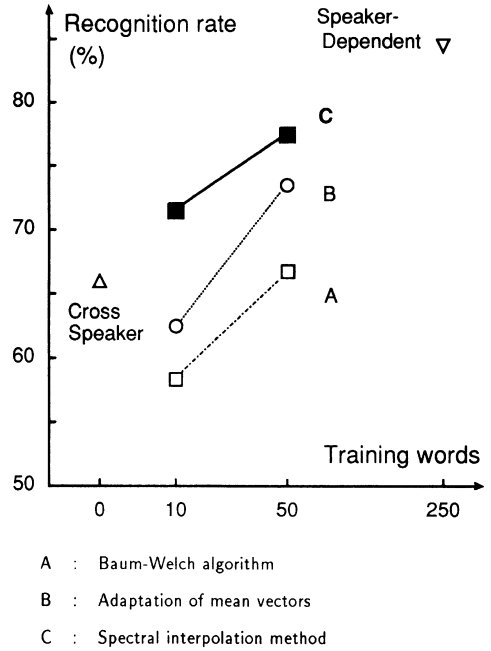


図 7 適応化手法の比較
Fig. 7 Adaptation method comparison.

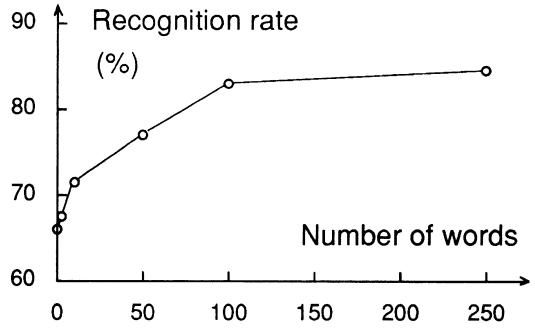


図 8 認識率と適応化に用いた単語数の関係
Fig. 8 Recognition rates vs. Number of words used for speaker adaptation.

表 2 特定話者(SD)モデルからの適応化による認識率と不特定話者(SI)モデルからの適応化による認識率の比較(%)

Initial Model		S4	S5	S6	S7	S8	S9	Av.	SI
No Adaptation	(1-Gauss)	58.4	48.8	59.2	50.0	52.0	32.0	50.1	
	(2-Gauss)	50.6	46.4	58.4	31.2	36.4	22.8	41.0	81.2
SA(W=50)	(1-Gauss)	81.6	78.4	79.6	72.8	79.6	74.4	77.7	
	(2-Gauss)	82.7	76.8	83.2	79.2	70.0	63.2	75.9	88.4

て第1段階のみに相当する，平均ベクトルのみを適応化する手法である．方法Cは，第1，第2段階とも行う，スペクトル内挿話者適応化である．また，図中の点 Cross Speaker は異話者認識実験の結果で，話者 S1 の HMM で話者 S2 の発声を認識した．学習データ量が 250 単語のときの話者 S2 の特定話者認識実験結果も併せて示す(Speaker Dependent)．

適応化用単語数が 10 単語のとき，方法 B が方法 A に比べ 4.0%認識率が高い．適応化用データが少ないときには，Baum-Welch アルゴリズムによるパラメータ推定よりも平均ベクトルのみの適応化の方が効果があることが確認された．これは，前述したように，Baum-Welch アルゴリズムによるパラメータ推定では，適応化に用いるデータ量に比してパラメータ数が多過ぎるために，偏った推定が行われるためであると考えられる．

更に，スペクトル内挿を行うことにより(方法C)，認識率が 9.2%改善された．この結果は，特に適応化用データ量が少ないとき，スペクトル内挿が効果的であることを示す．つまり，適応化用データのない HMM 状態数が適応化データのある HMM 状態数に比べより大きいとき，効果的である．

話者適応化のための適応化用データ量を変えたときの単語認識率を図 8 に示す．適応化用データがごく少量の場合でも認識率は向上する．十分な量のデータが用いられた場合には，特定話者認識と同程度の認識率まで達する．

5.1.3 不特定話者モデルからの適応

不特定話者認識方式におけるスペクトル内挿話者適応化方式を評価する実験を行った．不特定話者モデルは，分布数 2 の混合ガウス分布を出力分布としてもつ HMM を用いた．

まず，ある話者(S3)を未知話者として選び，不特定話者モデルからの適応化と 6 人の話者の特定話者モデルからの適応化を比較した．不特定話者モデルは，多数話者データベース DB1 を用いて学習した．6 人の話者の特定話者モデルは DB2 を用いて学習した．ともに学習単語数は 250 単語である．特定話者モデルからの適応化後の認識率を，出力確率分布が単一ガウス分布のとき(1-Gauss)とガウス分布数 2 の混合ガウス分布のとき(2-Gauss)とで調べた．認識実験にはデータベース DB3 を用いた．認識実験の結果を表 2 に示す．表 2 を見ると，すべての場合において，特定話者モデルからの適応化よりも不特定話者モデルからの適応化が優れていることがわかる．これは未知話者の発声に対する不特定話者モデルの認識性能が，他の話者の特定話者モデルよりも高く，適応化が安定に行われるためと考えられる．

次に，スペクトル内挿話者適応化を，7 人の未知話

表 3　7人の話者に対する認識率(%)

Speaker	S3	S4	S5	S6	S7	S8	S9	Av.
SI	81.2	88.3	76.8	90.4	87.6	76.8	85.2	83.8
SA(W=50)	88.4	89.6	82.0	89.6	87.6	83.2	86.4	86.7
SD(W=250)	92.0	94.4	91.2	94.4	93.2	89.6	88.4	91.9

者に対して適用した．適応化前 HMM は前述の実験で用いたものと同一の不特定話者モデルである．適応化および学習に用いられたデータベースは DB2 で，認識には DB3 を用いた．50 単語で適応化したときの認識結果を表 3 に示す．また，表 3 には 250 単語の学習データを用いて学習した特定話者モデルを用いたときの認識結果と，不特定話者モデルを用いたときの認識結果も併せて示す．

認識率は，1 名の話者 S6 を除いて，改善されていることがわかる．7 名平均で，不特定話者モデルによる認識率 83.8%のところ，50 単語の適応化で 86.7%と 2.9%の改善があり，スペクトル内挿話者適応化は，不特定話者モデルからの適応化においても効果があることが確認できた．

5.2　語い依存性の改善

5.2.1　実験条件

語い依存性の補正手法の効果を，不特定話者 HMM を適応化前 HMM として用いたスペクトル内挿話者適応化と組み合わせて評価した．実験に用いた音声分析条件，HMM の構造および認識対象はスペクトル内挿の評価実験と同一である．多数話者のデータとして，男性 46 名女性 39 名計 85 名の音素バランスを考慮した 250 単語 1 回発声を用いた．また，適応化用データとして，上の 85 名に含まれない話者男性 11 名女性 8 名計 19 名の，多数話者データベースと同じ 250 単語 1 発声を用い，評価用データとして，同じ話者グループの上の二つのデータセットには含まれない 250 単語 1 回発声を用いた．

不特定話者モデル M_{CI} は 85 名 250 単語発声のデータを用いて作成し，適応化用単語の語いコンテキストに依存した不特定話者モデル M_{CD} は，M_{CI} を適応化前 HMM とし，同じ 85 名のデータを用い，適応化用単語数 10，20，30，40，50，100 の場合について，作成した．単語は M_{CI} 作成に用いた 250 単語中から選択した．

5.2.2　実験結果

適応化前 HMM は不特定話者モデル M_{CI} を用いた．全評価話者平均の認識率を図 9 に示す．ここに CI はモデル M_{CI} による認識率，INTP はスペクトル内挿のみの，INTP＋NORM スペクトル内挿と語い依存性の補正を行ったときの認識率を表す．どの単語数でも，スペクトル内挿のみに比べ，認識性能が上昇している．スペクトル内挿と語い依存性の補正を組み合わせると，不特定話者認識率の 81.2%のところ，単語数 30 単語で，83.3%，50 単語で，85.2%とそれぞれ認識率が改善している．単語数 30 のときの各評価話者ごとの認識率を図 10 に示す．横軸はスペクトル内挿のみの認識率，縦軸はスペクトル内挿 ＋ 語い依存性補正の認識率である．話者によるばらつきもほとんどなく，適応化用単語の語いコンテキストによる音響的特徴量の変化は話者によらないことが裏づけられた．

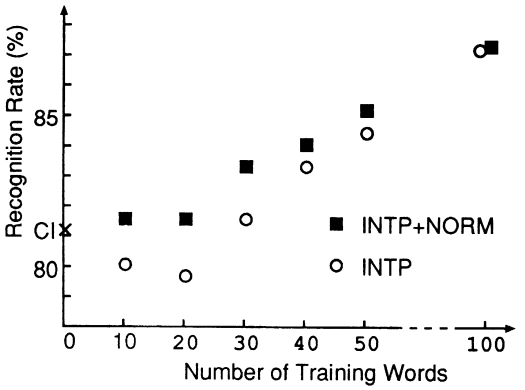


図 9　語い依存性改善手法の効果
Fig. 9　Normalization of data context dependency.

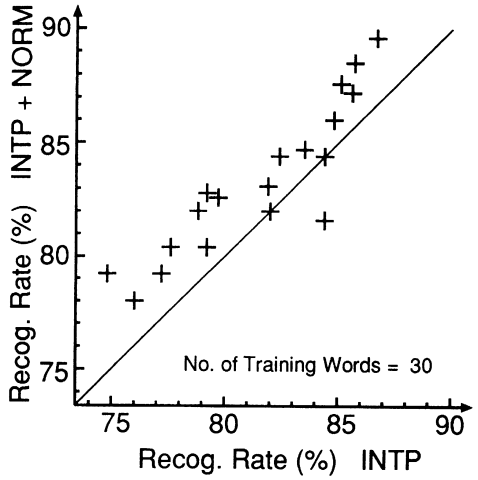


図 10　話者ごとの語い依存性改善手法の効果
Fig. 10　The effects of normalization of data context dependency for various speakers.

6. むすび

連続分布型HMM向けの新しい話者適応化法を提案した．スペクトル内挿話者適応化により，適応化用データが少なく対応するデータが存在しないパラメータをも適応化することができる．また，適応化後のパラメータが適応化用単語の語いコンテキストに依存する傾向があること（語い依存性）を問題点として指摘し，あらかじめ用意された多数話者の発声データを利用して語い依存性を改善する手法を提案した．半音節連続分布型HMMを用いた不特定話者音声認識方式を用い，5,000単語認識をシミュレートした類似100単語認識実験で評価を行った．50単語を適応化に用いたとき，認識性能が4.0%向上し，本手法の効果を確認した．多数話者の発声データは，語い依存性以外にも適応化に用いることのできるさまざまな情報を保有していると考えられ，今後はこれらの情報を抽出し利用する方法について更に検討を続けたい．また，マイクの違い，周囲騒音の違いなど周囲環境の違いの適応化に対しても本手法の適用を試みる予定である．

謝辞 研究を進める上で貴重な御助言を頂いた音声言語研究部の諸氏に感謝致します．

文　献

(1) Shikano K., Lee K.-F. and Reddy R.: "Speaker Adaptation Through Vector Quantization", Proc. ICASSP-86, pp. 2643-2646(1986).

(2) Schwartz R. M., Chow Y. L. and Kubala F.: "Rapid Speaker Adaptation Using a Probabilistic Spectral Mapping", Proc. ICASSP-87, pp. 633-636(1987).

(3) Lee C.-H., Lin C.-H. and Juang B.-H.: "A Study on Speaker Adaptation of Continuous Density HMM Parameters", Proc. ICASSP-90, S3.4, pp. 145-148(1990).

(4) Furui S.: "Unsupervised Speaker Adaptation Method Based on Hierarchical Spectral Clustering", Proc. ICASSP-89, pp. 286-289(1989).

(5) 山下泰樹，松本　弘："単語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(1988).

(6) 平田好充，中川聖一："連続出力分布型HMMによる話者適応化の日本語音韻認識による評価"，信学技報，**SP90**-16(1990).

(7) Niimi Y. and Kobayashi Y.: "Speaker-Adaptation of a Code Book of Vector Quantization", Proc. of European Conf. on Speech Technology, **2**, p. 430(1987).

(8) 白木善尚，誉田雅彰："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(1987).

(9) 服部浩明，嵯峨山茂樹："移動ベクトル場平滑化話者適応の原理とアルゴリズム"，信学技報，**SP92**-16(1992).

(10) 大倉計美，杉山雅英，嵯峨山茂樹："混合連続分布HMMを用いた移動ベクトル場平滑化話者適応方式"，信学技報，**SP92**-17(1992).

(11) 篠田浩一，磯　健一，渡辺隆夫："半音節HMMによる音声認識のための話者適応"，音響学会論文集，1-8-12(1990-09).

(12) Shinoda K., Iso K. and Watanabe T.: "Speaker Adaptation for Demi-Syllable Based Speech Recognition Using Continuous HMM", ICSLP90, 7.16.1, pp. 261-264(1990).

(13) Shinoda K., Iso K. and Watanabe T.: "Speaker Adaptation for Demi-Syllable Based Continuous Density HMM", ICASSP91, S13.7, pp. 857-860(1991).

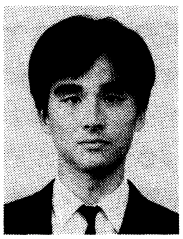
(14) 篠田浩一，渡辺隆夫："話者適応化における学習語い依存性の改善"，音響学会論文集，2-5-7(1992-10).

(15) 渡辺隆夫，吉田和永，古賀真二："半音節を単位としたHMMを用いた大語い認識"，信学論(D-II)，**J72-D-II**，8，pp. 1264-1269(1989-08).

(16) 渡辺隆夫，磯谷亮輔，塚田　聡："半音節を単位とするHMMを用いた不特定話者音声認識"，信学論(D-II)，**J75-D-II**，8，pp. 1281-1289(1992-08).

(17) 松岡達雄，鹿野清宏："混合ガウス分布不特定話者HMMをベースとした重み係数による話者適応化"，音響学会論文集，1-1-6(1992-03).

(平成5年7月5日受付，9月24日再受付)

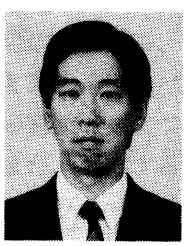


篠田　浩一

昭62東大・理・物理卒．平1同大大学院修士課程了．同年日本電気(株)入社．以来，音声認識の研究に従事．日本音響学会会員．

磯　健一

昭58東大・理・物理卒．昭60同大大学院修士課程了．昭61同大大学院博士課程中退．同年日本電気(株)入社．以来，音声認識の研究開発に従事．平3～4カーネギーメロン大学客員研究員．現在，情報メディア研究所勤務．日本音響学会会員．



渡辺　隆夫

昭47東大・工・計数卒．昭49同大大学院修士課程了．同年日本電気(株)入社．以来，音声認識の研究開発に従事．昭58～59マサチューセッツ工科大客員研究員．現在，C&C情報研究所勤務．日本音響学会会員．