論文 / 著書情報
Article / Book Information

論題(和文)	音声認識における自律的なモデル複雑度制御を用いた話者適応化
Title(English)	
著者(和文)	篠田浩一, 渡辺隆夫
Authors(English)	Koichi Shinoda
出典(和文)	電子情報通信学会和文論文誌D-II, Vol. J79-D-II, No. 12, pp. 2054-2061
Citation(English)	The Transactions of IEICE D-II, Vol. J79-D-II, No. 12, pp. 2054-2061
発行日 / Pub. date	1996, 12
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) 1996 Institute of Electronics, Information and Communication Engineers.

論　文

音声認識における自律的なモデル複雑度制御を用いた話者適応化

篠田　浩一[†]　　渡辺　隆夫[†]

Speaker Adaptation Using Autonomous Model Complexity Control for Speech Recognition

Koichi SHINODA[†] and Takao WATANABE[†]

あらまし　連続分布 HMM を用いる音声認識における話者適応化法を提案する．本手法は混合ガウス分布の各要素分布の平均ベクトルの適応前後の差分ベクトル（適応化ベクトル）を推定する．適応化ベクトルを複数のガウス成分間で共有し，その共有の度合（共有度）をモデルの複雑さとみなし，データ量に応じてその制御を行うことで，適応化に用いるデータ量の多少にかかわらず高い性能を示すことを特徴とする．本手法においては，まず，共有度の異なる多くの適応化ベクトルの集合（モデル）が準備され，次に，適応化用に与えられたデータに対し最適な共有度をもつモデルが選択される．モデル選択の基準としてデータ量しきい値を用いる方法と情報量基準の一つである MDL 基準を用いる方法を比較・検討した．5000 単語認識実験により評価を行い，適応化用音声が 50 単語のとき，エラーの削減率が 40%と，良好な結果を得た．

キーワード　話者適応化，不特定話者認識，連続分布 HMM，モデル複雑度，MDL 基準

1. まえがき

近年，連続分布隠れマルコフモデル（Continuous Density Hidden Markov Models; CDHMM）を用いた不特定話者（Speaker-Independent; SI）認識システムが盛んに研究されている．モデルパラメータが多量の学習データを用いて学習されるため話者の声質の多様性に対し頑健であり，また，特定話者（Speaker-Dependent; SD）システムに比べ，モデルパラメータ学習用に使用者が発声する必要がないという利点がある．しかしながら，十分に学習された特定話者認識システムに比べ，一般に認識性能が低い．更に，ある一部の話者に対し極端に認識性能が低くなる現象が見られる．これらの問題を解決するために，話者の少量の発声を用いてシステムを話者に対して適応させる技術である，話者適応化法が数多く提案され，不特定話者認識システムに対し適用されてきた．

一般に，話者適応化法は，以下の条件を満たすことが要求される．

- 適応化用データ量が極めて少量の場合でも認識性能が改善される．
- 適応化用データ量が増加するに従って認識性能が向上する．

しかしながら，従来法で両方の条件を満たすものは極めて少ない．

この問題を，適応化に用いるモデルの複雑さという観点から考えてみる．ここでモデルがより複雑であるとは，推定すべき自由パラメータ数がより多いということを意味する．今，推定すべき自由パラメータ数のより少ないモデルを “Coarse model” と呼び，自由パラメータ数のより多いモデルを “Fine model” と呼ぶ．従来の話者適応法を，この観点から便宜的にこの二つのカテゴリーに分けてみよう．一般に各音韻に共通なパラメータしかもたないモデルは “Coarse model” であり，音韻ごとにパラメータをもつモデルは “Fine model” であると言える．例えば，Schwartz らのスペクトル写像 [1] や，小坂らの話者モデル選択 [2] などは前者に属し，C.H. Lee らの出力確率分布の自己共役分布を事前確率分布として用いる事後確率推定法 [3] は後者に属する．Coarse model は適応化用データが少量のとき効果的である．しかしながら，データ量が多くなるに従い，認識性能の改善が少なくなる．これは，パラメータ数が少ないため，データに内包される音韻に依存した特徴を十分反映したモデルを作成でき

[†] 日本電気株式会社，川崎市
NEC Information Technology Laboratory, 4-1-1 Miyazaki, Miyamae-ku, Kawasaki-shi, 216 Japan

ないためである．一方，Fine model は，適応化用データが十分多いときには高い認識性能を示す．しかしながら，次の二つの問題のため，データ量が少ないときには，認識性能が劣化する．

- 少量のデータで推定されたパラメータはしばしば精度が低い．
- データに現れないカテゴリーに属するパラメータは更新されない．

以上の議論から，Coarce model も Fine model も上で掲げた二つの条件を満たしていないことは明らかである．

本論文では，この問題を解決する手法として，適応化用のモデルの複雑さをデータ量に応じ自動的に制御する手法を提案する．この手法をここでは，自律的モデル複雑度制御法（Autonomous Model Complexity Control; AMCC）と呼ぶ．AMCC は，

（1） 階層クラスタリングを用いたモデル集合の作成

（2） モデル集合からの最適なモデルの自動選択

の 2 段階から構成される．データ量が少ないときには Coarse model を，多いときには Fine model を用意する．

本論文は以下のように構成されている．次の章で，提案する話者適応化の概要を示し，**3.** で，階層的な共有構造を作成してさまざまな複雑度のモデルを用意する方法を述べる．**4.** で，最適な複雑度をもつモデルを選択する方法としてデータ量しきい値を用いる方法と，MDL 基準を用いる方法について述べる．**5.** で，認識実験による提案手法の評価結果を述べ，考察・議論を行う．

2. 適応化の枠組み

CDHMM の各々の状態 s の出力確率密度関数 $b_s(y)$ は，成分ガウス分布 $g_{s,v}(y|\mu_{s,v},\Sigma_{s,v}), v = 1,\ldots,V(s)$ の重み付き和で表される．

$$b_s(y) = \sum_{v=1}^{V(s)} w_{s,v} g_{s,v}(y|\mu_{s,v},\Sigma_{s,v}). \qquad (1)$$

ここで，$V(s)$ は状態 s の成分分布の数であり，$g_{s,v}$ は状態 s の v 番目の成分分布である．また，$\mu_{s,v}$，$\Sigma_{s,v}$ はそれぞれ $g_{s,v}$ の平均ベクトルおよび共分散行列，$w_{s,v}$ は成分分布 $g_{s,v}$ に対する重み係数を表す，本論文では，各々の成分分布の平均ベクトル $\mu_{s,v}$ の更新に焦点をあてる．以下，CDHMM におけるすべての認識単位のすべての状態のすべての成分分布の集合を記号 G で表す．そして，各々のガウス成分は，この集合 G において，$g_m(y|\mu_m,\Sigma_m), m = 1,\ldots,M$ と番号づけられる．ここで，M は，集合 G におけるすべての成分分布の個数である $(M = \sum_s V(s))$．提案手法では，成分分布 m の平均ベクトル μ_m に対する適応前後の差分ベクトル δ_m を推定する．すなわち，適応化後の平均ベクトルを $\hat{\mu}$ とすると，以下の式が成り立つ．

$$\hat{\mu}_m = \mu_m + \delta_m, \quad m = 1,\ldots,M \qquad (2)$$

以後，この δ_m を適応化ベクトルと呼ぶ．

ガウス成分の集合 G における成分分布数 M は一般に数千以上の数になる．例えば我々の認識システム [4] では 2000 程度である．適応化用のデータとしてほんの 2，3 の発声しか利用できない場合，すべての適応化ベクトルを推定すると認識性能に重大な劣化をもたらす可能性がある．この問題に対しては，異なる成分分布の間で同じ適応化ベクトルを共有することで，推定パラメータ数を減少させる方法が有効であることが知られている（e.g. [5]）．

提案する AMCC では，推定パラメータの共有の度合（共有度）を自律的に制御する．すなわち，データ量の少ないときには多くの成分分布間で同一の適応化ベクトルを共有することで適応化ベクトルの数は比較的少数に抑え，データ量が多くなるにつれ共有度を徐々に小さくし適応化ベクトルの数を増加させる操作を，外部からの制御なしに行う．AMCC については次章以下に詳しく説明する．

適応化の全体の流れは以下のとおりである（図 1）．まず最初に，入力音声が分析され，$x^N = x_1,\ldots,x_N$ で表される特徴ベクトルの時系列が得られる．次に特徴ベクトル列を構成する各ベクトル x_n がビタビアルゴリズムを用いて不特定話者（SI）HMM の各分布 $g_1,\ldots,g_M$ のうちの一つの分布と対応づけられる．通常のビタビアルゴリズムでは各特徴ベクトルは HMM の状態のうちの一つと対応づけられるが，ここでは，対応づけられた状態内での複数の成分分布のうち，出力確率が最大になる成分分布を選択し，特徴ベクトルをその分布と対応づけることとする．今，分布 g_m と対応づけられた特徴ベクトルの集合を $X_m = x_{m,1},\ldots,x_{m,P_m}$ とする．ここで，P_m は分布 g_m と対応づけられた特徴ベクトルの個数である．更に，各々の分布 g_m において，特徴ベクトルと分布の

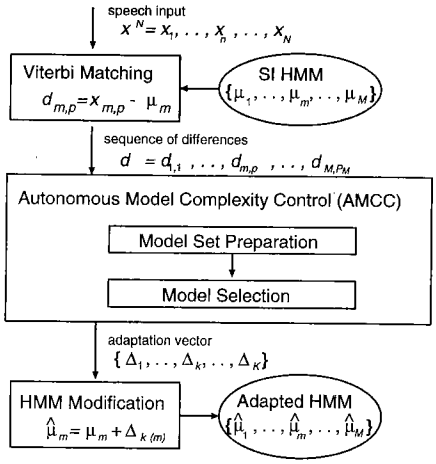

図1 適応化の枠組み
Fig. 1 Adaptation scheme.

平均ベクトルとの差分ベクトルが以下のように計算される．

$$d_{m,p} = x_{m,p} - \mu_m, p = 1, \ldots, P_m, m = 1, \ldots, M \tag{3}$$

次に，これら $d_{m,p}$ から適応化ベクトルの集合 $\Delta_1, \ldots, \Delta_K (1 \leq K \leq M)$ が，自律的モデル複雑度制御法（AMCC）により得られる．AMCC については次章以下に詳しく説明する．最後に，新しい平均ベクトル $\hat{\mu}_m$ が対応する $\Delta_{k(m)}$ を平均ベクトル μ_m に付加することにより得られる．

$$\hat{\mu}_m = \mu_m + \Delta_{k(m)} \tag{4}$$

ここで，$k(m)$ は，分布 m に対応する適応化ベクトルの番号を表す．

3. モデル集合の作成

3.1 階層的な共有構造の構築

適応化ベクトルの共有度を調節する仕組みとして，各ノードに適応化ベクトルが付随する木構造を導入する（図 2）．各々のリーフノードは，HMM の一つの成分分布 g_m と関連づけられており，式 (2) における，g_m が他のどの分布とも適応化ベクトルを共有しないときの適応化ベクトル δ_m が付随する．今，中間ノード Q_k を頂点とする部分木において，リーフノードに関連づけられた成分分布の集合を $G^k = g_1^k, \ldots, g_{R^k}^k$ とする．ここで，R^k は Q_k を頂点とする部分木におけるリーフノードの個数である．そのとき，ノード Q_k には，G_k に含まれる成分分布の適応化ベクトル $\delta_1^k, \ldots, \delta_{R^k}^k$ を共有したときの適応化ベクトルが付随する．このノード Q_k に付随する適応化ベクトルを Δ_k と書くことにする．ルートノードには，すべての成分分布間で一つの適応化ベクトルを共有した場合の適応化ベクトルが付随する．

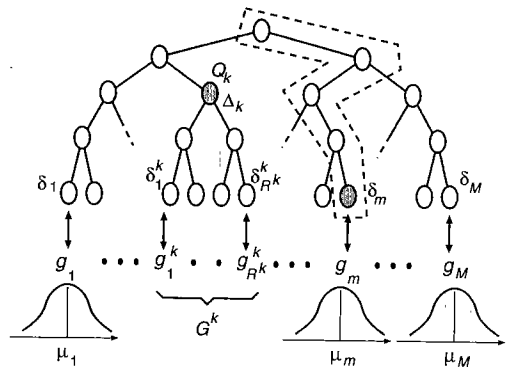

図2 木構造
Fig. 2 Tree structure.

さて，AMCC では，各成分分布に対して一つの適応化ベクトルが選択される．選択の範囲は，その成分分布に対応するリーフノード，およびそのリーフノードを支配するノードである（図 2 における点線で囲まれた部分）．選択された適応化ベクトルの集合をここでは「モデル」と呼ぶ．最も Coarse なモデルはルートノードのみのモデル，最も Fine なモデルは，すべてのリーフノードからなるモデルである．今，仮に各々の成分分布に対し選択の範囲となるノードの数がどのノードに対しても R 個であるとすると，全成分分布にわたるノードの選び方の組合せは全部で N^R とばく大な数になる．これら一つひとつが一つのモデルである．

この適応化ベクトルの共有構造は，話者によらないと仮定し，話者適応化を行う前に作成する．木構造は，より類似した適応化ベクトルをもつノード同士がより下層でマージされるように作成する必要がある．しかしながら，適応化前には異なる分布間の適応化ベクトルの類似度は未知であり，何らかの事前知識を用いて類似度を推定する必要がある．筆者らの以前の研究 [6] により，成分分布間の距離が近いほど適応化ベクトルの類似度が大きくなる傾向があることが報告されている．そこで木構造は，成分分布間の距離を用いて階層的にクラスタリングを行うことにより作成される．成

分分布間の距離としては分布間のダイバージェンスを用い，クラスタリングには k-means アルゴリズムを用いる [7].

3.2 共有パラメータの推定

木構造の各ノードの適応化ベクトルは話者の適応化用データを用いて以下のように推定される.

$$\Delta_k = \sum_{r=1}^{R^k}\sum_{p=1}^{P_r^k} d_{r^k,p} \Big/ \sum_{r=1}^{R^k} P_r^k \qquad (5)$$

ここで，P_r^k は G_k に含まれる r 番目の成分分布に対応づけられた特徴ベクトルの個数であり，$d_{r^k,p}$ は，r 番目の成分分布に対する p 番目の差分ベクトルである.

また，対角分散行列を仮定したときの適応化ベクトルの分散は以下の式で求められる.

$$\sigma_k^2 = \sum_{r=1}^{R^k}\sum_{p=1}^{P_r^k} (d_{r^k,p} - \Delta_k)^2 \Big/ \sum_{r=1}^{R^k} P_r^k \qquad (6)$$

この分散は **4.2** で用いる.

4. 最適な共有度をもつモデルの選択

ばく大な数のモデルの中から効率良く最適なモデルを選択する方法として，以下の二つの方法を検討した.

4.1 データ量しきい値を用いる選択

木構造におけるノードの選択の基準として，各ノードの適応化ベクトルの推定に用いた差分ベクトルの個数を用いる．木構造の全ノードに共通なしきい値 D を設定し，リーフノードからより上層のノードへ探索を行い，差分ベクトルの個数が初めて D を超えるノードを選択する．データ量が増加するに従い，各ノード当りの差分ベクトル数も増加し，選択されるノードはより木構造の下方へと移動する．すなわち，データ量が増加すると，より複雑なモデルが選択される（図 3）．簡便な手法ではあるが，しきい値 D の最適値を実験を通して決定する必要がある.

本方法を他の方法と比較する．Digalakis らは，共有された混合ガウス成分 "genones" を用いたアプローチを提案している [8]．筆者らの方法は，このアプローチにおいて与えられたデータ量に対して自動的に共有の度合を変える方法と位置づけられる．また移動ベクトル平滑化法（Vector Field Smoothing; VFS）[5] では，音響空間において適応化ベクトルの平滑化を行っている．このとき，平滑化の強度はガウス成分間の距離の関数で制御されている．それに対し本手法では，空間における距離の大小をあらかじめ木構造で表し，データ量の大小によって階層を上下に移動することでデータ量の変化に応じて平滑化の強度を変化させる手法となっている.

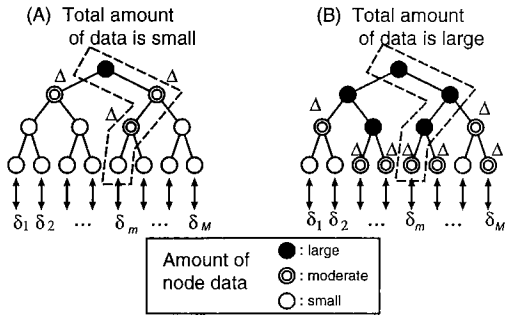

図 3 データ量しきい値を用いたモデル選択
Fig. 3 Model selection using data threshold.

4.2 MDL 基準を用いたモデル選択

MDL（Minimum Description Length）基準 [9] は情報量基準の一つであり，与えられたデータに対し最適なモデルを選択する問題において有効であることが知られている.

MDL 基準は，確率モデル $i = 1, \ldots, I$ の中で，データ $z^N = z_1, \ldots, z_N$ に対し，最も小さい記述長 L_R を与えるモデルを最適なモデルとする基準である．確率モデル i に対する記述長 $L_R^{(i)}$ は以下の式で与えられる.

$$L_R^{(i)} = -\log P_{\hat{\theta}^{(i)}}(z^N) + \frac{\alpha_i}{2}\log N + \log I \qquad (7)$$

ここで，α_i はモデル i の次元数（自由パラメータの個数）$\hat{\theta}^{(i)}$ はモデル i の自由パラメータ $\theta^{(i)} = (\theta_1^{(i)}, \ldots, \theta_{\alpha_i}^{(i)})$ の最ゆう推定量である．上式における第 1 項は，データに対する対数ゆう度であり，第 2 項は，モデルの複雑さを表す量である．第 3 項は，モデル i を選択するために要する記述長である．モデルが複雑になるにつれ，第 1 項の値は減少し第 2 項の値は増加する．記述長 L_R は適当な複雑さをもつモデルで最小になる.

式 (7) で容易にわかるように，MDL 基準では，モデルの集合とデータとが与えられれば最適なモデルを選択することができる．そこでは，データ量しきい値のような，実験を通して最適値を求める必要のある調節パラメータを必要としない．従って，MDL 基準を用いたモデル選択は，実験条件の変化に左右されることがない，より環境の変動に対し頑健な手法であるこ

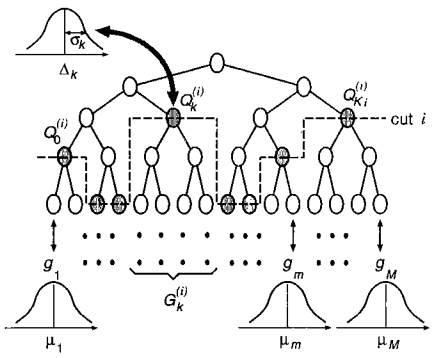

図4 MDL基準を用いたモデル選択
Fig. 4 Model selection using MDL criterion.

とが期待される．

さて，このMDL基準を，最適な共有度をもつ適応化モデルの選択に用いる方法を考える．まず，特徴量の差分ベクトルの集合 $\{d_{m,p}; p=1,\ldots,P_m, m=1,\ldots,M\}$ に対する確率モデルを定義する．ここで，簡単のため，差分ベクトル $d_{m,p}$ はお互いに独立であり，多次元ガウス分布を出力確率分布とする複数の情報源のうちのどれかから出力されると仮定する．更に各情報源のガウス分布は，平均ベクトル Δ_k，分散 σ_k^2 の単一ガウス分布であると仮定する．今，「カット」を木構造を上下に分断するノードの集合として定義する．カットの例が図4において点線で示してある．各々のカットが一つの「モデル」に対応している．

このとき，各々のカットは差分ベクトル d に対する一つの確率モデルであるとみなすことができる．差分ベクトル集合 $\{d_{r^k}^p; p=1,\ldots,P_r^k, r^k=1,\ldots,R^k\}$ に対する，ノードの集合 $Q_1^{(i)},\ldots,Q_{K_i}^{(i)}$ から構成されるカット i の記述長 $L_R^{(i)}$ は，以下のように計算される．

$$L_R^{(i)} = \sum_{k=1}^{K_i}\sum_{r=1}^{R^k}\sum_{p=1}^{P_r^k}\sum_{l=1}^{L}\left(\frac{1}{2}\log 2\pi\sigma_{k,l}^{2(i)} + \frac{(d_{r^k,p,l}-\Delta_{k,l}^{(i)})^2}{2\sigma_{k,l}^{2(i)}}\right) + K_i L\log N + \log I, \tag{8}$$

ここで，L はガウス分布の次数，$d_{r^k,p,l}$ は差分ベクトル $d_{r^k,p}$ の第 l 次元の成分である．また，$\Delta_{k,l}^{(i)}$ および $\sigma_{k,l}^{2(i)}$ はそれぞれ，ノード $Q_k^{(i)}$ の第 l 次元の平均および分散である．式(7)の第3項は定数であるため省略する．簡単な式変形を行い定数項を除くと，以下の $L_R^{(i)'}$ を最小にするカットが最適であると言える．

$$L_R^{(i)'} = \sum_{k=1}^{K_i} h_k^{(i)} \sum_{l=1}^{L}\log\sigma_{k,l}^{2(i)} + 2K_i L\log N, \tag{9}$$

ここで，$h_k = \sum_{r=1}^{R^k} P_r^k$ である．

可能なカットの個数はばく大な数にのぼり，そのすべてを探索するのは困難である．そこでここでは，ダイナミックプログラミング手法に基づいた，最適なカット c_{opt} を効率的に求めるアルゴリズムを用いる[10]．カット c_{opt} は $c_{opt}=$ Find-MDL(k_0) の式で求められる．ここで，k_0 はルートノードを表す．Find-MDL(k) は以下に示すような再帰関数であり，ノード k をルートノードとした部分木における最適なカットを求めることができる．ここで，$[k_1,\ldots,k_J]$ はノード $k_1,\ldots,k_J$ から構成されるカットを表す．

```
Find-MDL(k)
1. if      k is a leaf node
2. then    return ([k])
3. else
4.         For each child k_j of k
5.             c_j := Find-MDL(k_j)
6.             c := append(c_j)
7.         if      L_R([k]) < L_R(c)
8.         then    return ([k])
9.         else    return (c)
```

5. 評価実験

5.1 実験条件

半音節を認識単位とした日本語音声認識システム[4]で提案手法の評価を行った．ここでは，データ量しきい値を用いる方法（AMCC-DT）およびMDL基準を用いる方法（AMCC-MDL）を評価した．

このシステムでは，入力音声は16 kHzの標本周波数でサンプリングされ，10ミリ秒ごとに分析される．分析の結果，各フレームの音声は21次元の特徴ベクトルとなる．各特徴ベクトルは，パワー差分，10次元のメルケプストラム係数，10次元のメルケプストラム差分から構成される．半音節単位は全体で260個ほどあり，それぞれ1個から4個の状態をもつ．各々の状態の成分ガウス分布の個数は2個である．成分ガウス分布のHMM全体での総数は2046である．適応化の

初期HMMとして用いられるSI HMMは，85名の話者の音素バランスを考慮した250単語セット1回発声を用いて学習された．

木構造は，4分木6階層と2分木14階層の2種類を作成した．ここで，N階層の木構造とは，ルートノードを1階層目とし，最下層のノードをN階層目とする木構造である．4分木6階層の木構造の作成方法を例として説明する．まず，1階層目から4階層目までは，親ノードからの分岐数が4になるよう，トップダウンにk-meansクラスタリングを行う．クラスタリングの過程で，その下の成分分布の個数が4以下になるノードについてはそれ以上の分割は行わない．5階層目からはクラスタリングを行わない．5階層目から6階層目（最下層）の分岐数は5階層目のノードに属する成分分布の数と等しく，6階層目のノードは成分分布に1対1に対応する．同様にして2分木14階層の木構造も作成する．予備実験の結果，AMCC-DTでは4分木6階層が，AMCC-MDLでは2分木14階層の性能が良かったため，認識実験はその組合せで行った．

評価実験として，大語彙離散単語認識実験を行った．認識辞書として，日本語辞書から選んだ重要語5000単語セットを用いた．評価用に学習用話者に含まれない男性5名女性5名計10名の音声を用いた．これらの話者は各々250単語を適応化用に発声し，これとは異なる250単語を認識実験用に発声した．

5.2 実験結果

まず，データ量に応じてモデルの複雑さがどのように変化していくかを調べた．図5は，1名の話者において，データ量に応じて適応化により使用されるノード数（適応化ベクトル数）がどのように変化するかを示した図である．ここで，DTはAMCC-DTによる結果，MDLはAMCC-MDLによる結果を表す．DTにおける（ ）内の数字はデータ量しきい値である．DT，MDLともにデータ量が増加するにつれ，ノード数が増え，モデルの複雑度が増加していることがわかる．DTでは，ノード数の増え方はデータ量しきい値が大きいほど小さい．また，木構造が異なるため単純な比較はできないものの，MDLでは，DTにおいてデータ量しきい値100〜200に対応する複雑さをもつモデルが選択されていることがわかる．

図6に，AMCC-DT，および，AMCC-MDLの認識実験結果を示す．DTにおける（ ）内の数字はデータ量しきい値を表す．ここで参照実験（REF）としてスペクトル内挿写像[6]と事前確率として自然共役分布

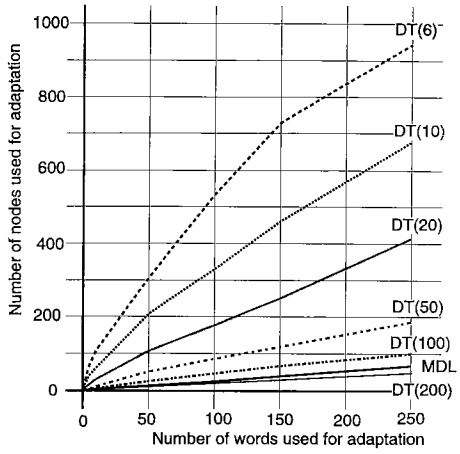

図5 適応化に用いたノード数
Fig. 5 Number of nodes used for adaptation.

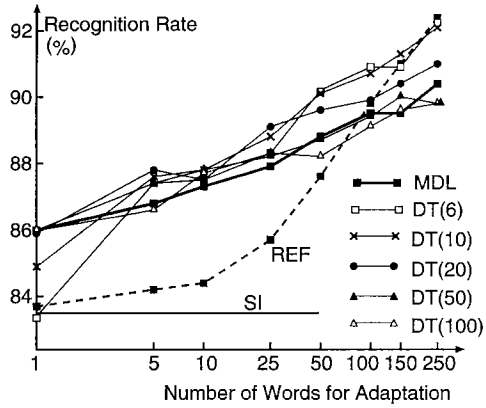

図6 提案法の認識率
Fig. 6 Recognition accuracies of proposed method.

を仮定した平均ベクトルの事後確率推定[3]の技術を用いた場合の認識結果を示す．これは成分分布間で適応化ベクトルを全く共有しない手法である．また，不特定話者認識実験（SI）の結果も示す．

まず，AMCC-DTの結果を見ると，データ量しきい値の最適値は10あるいは20であり，適応化用発声50単語のとき，不特定話者認識に比べ40%ほどエラーが削減されている．ほとんどの単語数で参照実験の結果を上回っている．データ量しきい値が小さい場合には，データ量が少ないときに認識性能が低下する傾向が，大きいときには，データ量が大きいときに認識性能が飽和する傾向が見られる．

また，AMCC-MDLも良好な性能を示している．認

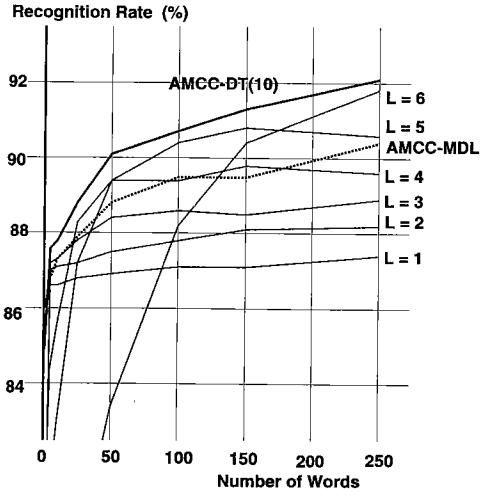

図7 階層を固定する方法の認識率
Fig. 7 Recognition accuracies of level-fixed method.

識性能はAMCC-DTに比べるとやや低いものの，例えば，50単語を適応化に用いた場合，誤り率は不特定話者認識に比べ32%減少している．

図7には，AMCC-DTで用いた木構造において，木構造の階層を一つ定め，すべての分布に対しその階層のノードの適応化ベクトルを用いる場合の認識率を示す．$L=1$が，ルートノードの適応化ベクトルをすべての成分分布に対して用いた場合の認識率を表し，$L=6$が成分分布間で適応化ベクトルを共有しなかった場合の認識率を表す．ここでAMCC-DTのデータ量しきい値は10である．階層を固定した場合，階層が下がるにつれ，Coarce ModelからFine Modelへと変化していくことがわかる．AMCC-DTは階層を固定する方法に比べ，すべての単語数で性能が上回っている．すなわち，与えられたデータ量に対し，最適な階層を選択できていることがわかる．

5.3 考察・議論

提案したAMCC-DT，AMCC-MDLがともに良好な性能をもつことが実験を通して確認できた．

MDL基準を用いる方法（AMCC-MDL）はデータ量しきい値を用いる方法（AMCC-DT）に比べ，性能がわずかながら低い．選択されたモデルの適応化ベクトル数を比較すると，MDLの方がDTに比べ小さく，実際の最適なモデルよりもより複雑度の小さいモデルが選択されている．

AMCC-DTでは，最適なデータ量しきい値を実験を通して求める必要があり，実用においてはその点が障害となり得る．一方，AMCC-MDLではそのような調節パラメータを必要としないという利点があり，実験環境の変動に対しより頑健な手法ということができる．

6. むすび

データ量に対するモデル複雑度制御を用いた話者適応化法を提案した．あらかじめ階層構造の形で用意されたモデル集合の中から，最適なモデルを選択し適応化に使用する．半音節単位を用いた日本語大語い認識実験を行い有効性を確認した．適応化用単語が50単語のときに誤認識を最大40%削減することができた．

更に，本論文では，モデル選択の基準としてデータ量しきい値を用いる方法と，MDL基準を用いる手法との比較検討を行った．これまで，MDL基準は，人工的な数値シミュレーション実験ではその有効性が確認されているものの，実際のデータに対し応用した例は少なかった．今回の評価で，MDL基準の有効性が実証できたと考えられるが，今回提案した話者適応化の枠組みでは，最適なモデルより複雑度の小さいモデルが選択される傾向があり，その対策が課題として残っている．

今後は，階層構造の構成方法を工夫すると共に，話者のみならず，他のさまざまな環境条件の相違に対しても提案法を適用していきたい．

文献

[1] R. Schwartz, Y.L. Chow, and F. Kubala, "Rapid speaker adaptation using a probabilistic spectral mapping," Proc. ICASSP87, pp.633–636, Dallas, 1987.
[2] T. Kosaka, "Tree-structured speaker clustering for fast speaker adaptation," Proc. ICASSP94, pp.I-245–248, Aderaide, 1994.
[3] C.-H.Lee, C.-H.Lin, and B.-H.Juang, "A study on speaker adaptation of continuous density HMM parameters," Proc. ICASSP90, pp.145–148, Albuquerque, 1990.
[4] 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），vol.J75-D-II, no.8, pp.1281–1289, 1992.
[5] K. Ohkura, M. Sugiyama, and S. Sagayama, "Speaker adaptation based on transfer vector field smoothing with continuous mixture density HMMs," Proc. ICSLP92, pp.369–372, Alberta, 1992.
[6] 篠田浩一，磯 健一，渡辺隆夫，"音声認識のためのスペクトル内挿を用いた話者適応化，" 信学論（A），vol.J77-A, no.2, pp.120–127, 1994.
[7] T. Watanabe, K. Shinoda, K. Takagi, and E. Yamada, "Speech recognition using tree-structured probability density function," Proc. of ICSLP94, pp.223–226, 1994.
[8] V. Digalakis and L. Neumeyer, "Speaker adaptation using combined transformation and bayesian methods,"

Proc. ICASSP95, pp.680–683, Detroit, 1995.
[9] J. Rissanen, "Universal coding, information, prediction, and estimation," IEEE Trans. IT, vol.30, 1984.
[10] H. Li and N. Abe, "Generalizing case frames using a thesaurus and the MDL principle," Proc. of Recent Advances in Natural Language Processing, Bulgaria, pp.239–248, 1995.

（平成8年5月7日受付，7月26日再受付）

篠田　浩一　（正員）

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．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　（正員）

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，音声情報処理の研究開発に従事．昭和58～59マサチューセッツ工科大客員研究員．現在，情報メディア研究所勤務．日本音響学会，IEEE各会員．工博．