音声言語情報処理　4－12
(1994. 12. 16)

最大事後確率推定と移動ベクトル場平滑化の組合せによる逐次話者・回線適応

高橋 淳一　　嵯峨山 茂樹
NTT ヒューマンインタフェース研究所
〒238-03 神奈川県横須賀市武 1-2356

あらまし　　オンライン適応化の基本技術となる逐次適応化手法について述べる．追加学習の枠組を有するクラス内学習法である最大事後確率 (MAP) 推定法と，クラス間平滑化法である移動ベクトル場平滑化 (VFS) 法を組み合わせた逐次型 MAP/VFS 法を提案し，話者適応，話者と回線の同時適応の問題に対する本方法の逐次適応化の効果を実験的に評価した．その結果，本方法の特徴である，MAP 推定法の追加学習機能と VFS 法の適応化の加速機能により，逐次的に認識性能の向上が図れることを明らかにした．これにより，適応学習データの蓄積を必要としない word-by-word の逐次適応化実現の見通しを得た．

和文キーワード　　音声認識，HMM，話者適応，回線適応，最大事後確率推定，MAP，移動ベクトル場平滑化，VFS，逐次

Vector-Field-Smoothed Bayesian Learning for Incremental Speaker/Telephone-Line Adaptation

Jun-ichi TAKAHASHI and *Shigeki SAGAYAMA*
NTT Human Interface Labolatories
1-2356 Take Yokosuka-shi Kanagawa 238-03 Japan

Abstract　　This paper presents an incremental adaptation method which is the basic technique for on-line adaptation. The proposed method combines MAP (Maximum *a Posteriori*) estimation as intra-class training, which has an additional training scheme, with VFS (Vector Field Smoothing) method as inter-class smoothing. The recognition performance of the method was evaluated through some experiments for speaker adaptation and for simultaneous adaptation of speaker and telephone line. It was found that a fast and incremental adaptation can be achieved by the additional training scheme and smoothing technique including interpolation. As the result, fast and word-by-word incremental adaptation without pooling the adaptation traning data can be realized.

英文 key words　　speech recognition, HMM, speaker adaptation, line adaptation, Bayesian learning, maximum *a posteriori* estimation, transfer vector filed smoothing, incremental

1 はじめに

適応化は，話者の個人性変動に起因する不特定話者音声認識と特定話者音声認識の間の認識性能差，音声信号の入力・伝送系(マイク，電話機，回線など)の特性ミスマッチによる認識性能劣化を改善するための重要な技術である．音声認識システムに適応化機能を組み込む場合，ユーザと認識システム間のマン-マシン・インタフェースを良好に保ちつつ，如何に適応学習データをユーザから獲得して効率的に適応処理を行なうかが問題となる．

ユーザと音声認識システムとの対話をフィードバックし，それを適応学習データとして用いて，逐次，モデルを適応学習し，システムの認識性能を向上させる逐次適応機能は，適応学習データを事前に獲得する必要がなく，かつ，ユーザは，適応化処理を意識せずに，発声する毎により良い(認識性能の高い)音声認識機能を享受できる点から，重要かつ魅力的である．現実的なサービス応用の観点から考えると，このような適応処理では，できる限り少ない適応学習データでの適応，すなわち，高速な適応化能力が必要とされる．例えば，メニューからの項目選択によりユーザの要求を実現するサービスを想定すると，項目選択の回数は数回であり，選択は項目名の単語発声で行なわれると考えられ，このような意味から，5～10単語程度の適応学習での性能向上が目標となろう．

本報告では，上記の考え方に基づき，オンライン適応実現の基本技術となる逐次型適応方式として，最大事後確率(MAP:Maximum *a Posteriori*)推定法と移動ベクトル場平滑化(VFS:Vector Field Smoothing)法とを組み合わせた逐次型**MAP/VFS**法を提案する．MAP推定法は，事前知識の効果的な利用により，少ない適応学習データでも適応化性能の向上を図れる方式として，ここ数年精力的研究され，蓄積型MAP適応[1,2,3,4,5]，逐次型MAP適応[6,7,8]とも話者適応においてその有効性が報告されている．文献[6,7]は，MAP推定法のオンライン適応への応用に関する．本方式は，MAP推定法の特徴 — 追加学習機能 — と，高速な適応手法として提案されたクラス間平滑化手法であるVFS法[12]との効果的な組合せを特徴とする．なお，筆者らや外村らにより，蓄積した適応学習データを一括して使用する蓄積型MAP推定法とVFS法を組み合わせる方式[9,10]も提案されているが，これは従来のVFS法で用いていた最尤(ML:Maximum Likelihood)推定法をMAP推定法に置き換えたものであり，本報告で提案する逐次的な適応方式とは異なる．本方式に関しては，IEEEのワークショップIVTTA94において既に一部報告[11]したが，本報告ではその内容を含め，話者適応，話者と回線の同時適応に対する認識実験・評価を通して，本方式により，逐次的な認識性能向上，及び，4，5単語での高速適応化が実現できることを示す．

以下，第2章では，逐次型適応方式MAP/VFSの原理について説明する．第3章では，具体的な適応処理フローについて述べ，第4章では認識実験による提案方式の評価とその結果について詳述する．実験対象は，話者適応と，ここ数年音声認識応用の要求が高まっている電話音声での話者と回線の同時適応とした．実験では，主として，MAP法による逐次型適応との比較を行なった．第5章では提案方式MAP/VFSとMAPとの性能差の原因について考察し，第6章は要点をまとめる．

2 逐次型適応方式MAP/VFSの原理

最大事後確率(MAP)推定法，移動ベクトル場平滑化(VFS)法の各方式の原理を詳述し，MAP/VFS法におけるMAPとVFSの組合せ方に関する基本的な考え方について述べる．

2.1 最大事後確率(MAP)推定法

MAP推定法は，推定対象とするモデルパラメータθが事前分布$g(\theta)$に従うランダム変数であると仮定し，事前分布と実際に観測されたサンプル値の分布より得られる事後分布からの事後確率を最大にすることを規準とした推定法である(式(1))．

$$\theta_{MAP} = \underset{\theta}{\operatorname{argmax}} f(\mathbf{x} \mid \theta) g(\theta) \tag{1}$$

式(1)において，$f(\mathbf{x} \mid \theta)$は，サンプル$\mathbf{x}$の分布を表す．事前分布$g(\theta)$を一様分布(定数)とすれば，式(1)は最尤推定式に同じであり，最尤推定法はMAP推定法の特殊な場合であることがわかる．MAP推定では，事前分布の定義，及び，事前分布のパラメータ推定をどのようにするかが重要な問題である．これまでの研究により，適応化の問題においては，不特定話者モデルのパラメータの分布を先験的知識と仮定した，混合連続HMM(Hidden Markov Models)に対応したMAP推定が定式化されている[2,3,4]．平均ベクトルを推定対象のパラメータとすると，その推定式は，次式で表される．

$$\hat{\mathbf{m}}_k = \frac{\tau_k \mathbf{m}_k + \sum_{t=1}^{T} c_{kt} \mathbf{x}_t}{\tau_k + \sum_{t=1}^{T} c_{kt}} \tag{2}$$

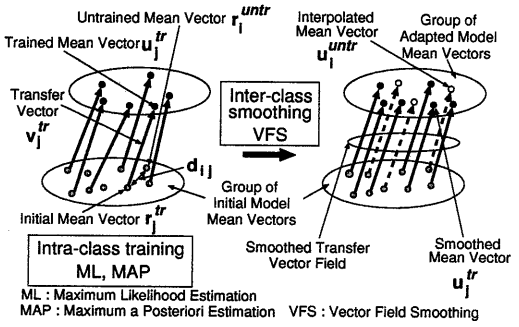

図1. 移動ベクトル場平滑化 (VFS) 法の原理

$$c_{kt} = \frac{w_k N(\mathbf{x}_t \mid \hat{\mathbf{m}}_k, \mathbf{\Sigma}_k)}{\sum_k w_k N(\mathbf{x}_t \mid \hat{\mathbf{m}}_k, \mathbf{\Sigma}_k)} \tag{3}$$

ここで，$\hat{\mathbf{m}}_k$，$\mathbf{m}_k$ は，それぞれ，MAP 推定により学習された k 番目の分布の平均ベクトル，及び，初期モデルの平均ベクトルを表す．$\mathbf{\Sigma}_k$ は，k 番目の分布の共分散行列であり，$N(\cdot \mid \mathbf{m}, \mathbf{\Sigma})$ は平均 $\mathbf{m}$, 分散 $\mathbf{\Sigma}$ の正規分布を表す．w_k は k 番目の分布の分岐確率である．τ_k は適応化パラメータであり，適応学習データのサンプル平均と初期モデルの平均ベクトルとの重み付き平均 (式 (2)) における初期モデルの平均ベクトルに与えられる重み値である．適応パラメータ τ_k は，次式から求めた [2, 4].

$$\tau_{jk} = \frac{p \sum_l c_{jkl}}{\sum_l (\mathbf{y}_{jkl} - \mathbf{m}_{jk})^t (\sum_k w_k \mathbf{\Sigma}_k)^{-1} (\mathbf{y}_{jkl} - \mathbf{m}_{jk})} \tag{4}$$

ここで，p は特徴パラメータの次元数である．$\mathbf{y}_{jkl}$ はサンプル平均 $\sum_l c_{jklt}\mathbf{x}_t / c_{jkl}$ であり，c_{jkl} は $c_{jklt} = c_{jkl}$ ($\mathbf{x}_t \in C_l$ の場合)，$c_{jkl} = 0 (\mathbf{x}_t \notin C_l$ の場合) である．クラス C_l はサンプルデータ $\mathbf{x}_t$ にかかわるすべての混合要素とした．

2.2 移動ベクトル場平滑化 (VFS) 法

移動ベクトル場平滑化 (VFS) 法の原理を図1に示す．VFS 法はクラス間平滑化法の一種であり，クラス内学習が行なわれていることを前提とする．図1には，限られたデータ量の適応学習データによって初期モデル中の一部のモデルが学習され，学習できなかったモデルを VFS 法により求める過程が示されている．適応対象のモデルパラメータは平均ベクトルである．VFS 法は補間と平滑化の2種類の処理からなる．適応学習データにより学習できなかったモデルの平均ベクトルは，補間処理により推定される．また，適応学習データ量が少量であることから，学習されたモデルの平均ベクトルには推定誤差が含まれると考えられ，平滑化処理によりこの推定誤差を補正する．

図1に示された各記号は，初期モデルの平均ベクトル：$\mathbf{r}_i$，適応された (クラス内学習により学習された，または，VFS 法により適応された) 平均ベクトル: $\mathbf{u}_i$，適応された平均ベクトルと初期モデルの平均ベクトルとの差分ベクトル：$\mathbf{v}_i$ である，また，各記号に付与された添字 tr, $untr$ は，それぞれ，そのベクトルが適応学習データを用いたクラス内学習により学習されたこと，学習されなかったことを表す．

補間処理による未学習の平均ベクトルの推定式を 式 (5),(6) に示す．

$$\mathbf{v}_j^{tr} = \mathbf{u}_j^{tr} - \mathbf{r}_j^{tr} \tag{5}$$

$$\mathbf{u}_i^{untr} = \frac{\sum_{j \in G_K} w(d_{ij}) \mathbf{v}_j^{tr}}{\sum_{j \in G_K} w(d_{ij})} + \mathbf{r}_i^{untr} \tag{6}$$

平滑化処理による既学習の平均ベクトルの補正式を式 (7) に示す．

$$\mathbf{u}_i^{tr} = \frac{\sum_{j \in G_K} w(d_{ij}) \mathbf{v}_j^{tr} + \mathbf{v}_i^{tr}}{\sum_{j \in G_K} w(d_{ij}) + 1} + \mathbf{r}_i^{tr} \tag{7}$$

上記の式 (5) ～ (7) において，差分ベクトル $\mathbf{v}_i$ は移動ベクトルと呼ばれる．VFS 法では，適応化による初期モデルから適応化モデルへの移動ベクトル場は滑らかであると仮定する，d_{ij} は，初期モデルの平均ベクトル群の i 番目，j 番目の平均ベクトル間の距離 (通常はユークリッド距離) を表す．$w(d_{ij})$ は，距離 d_{ij} を変数とした重み関数であり，上記各計算では，移動ベクトルの寄与する割合に相当する重み値を与える．ここでは，重み関数としてガウス窓関数 $w(d_{ij}) = \exp(-d_{ij}^2/s)$ を用いた [13]. s は移動ベクトルの方向を調節する平滑化パラメータであり，$s = 0$ では平滑化なし，$s = \infty$ では G_K が全ベクトルであれば平行移動を表す．G_K は K 近傍の平均ベクトルの集合を表す．

2.3 MAP/VFS 法

上記 2.1 から，MAP 推定法は，モデルパラメータに内在する事前知識と新たな学習データとの統合学習を特徴とし，初期モデルの学習に用いた大量のデータを再度使用することなく，新たな学習データを用いた学習によ

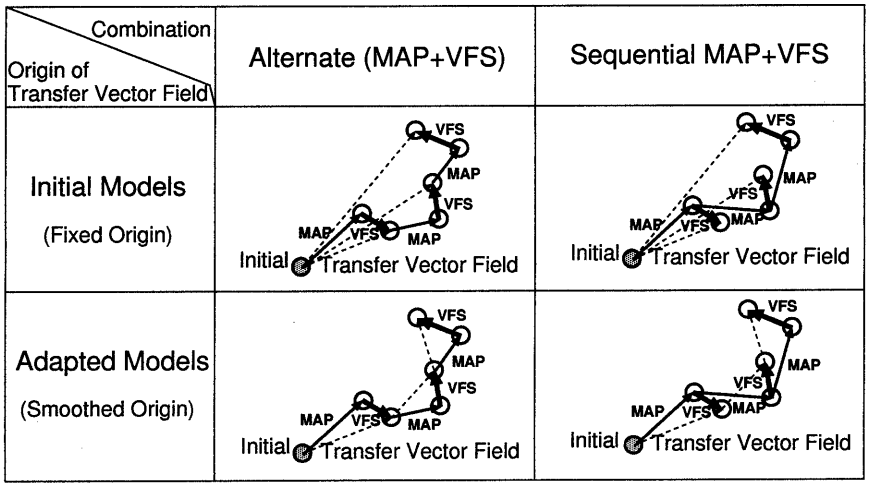

図 2. MAP と VFS の組み合わせ方

り，初期モデルの学習データと新たな学習データとを合わせた一括学習と同等の学習を実現する枠組をもつ．適応化の問題への応用では，新たな学習データを用いた学習を追加学習と見なし，逐次的なこの追加学習の繰り返しが逐次適応に相当すると考えられる．上記 2.2 から，VFS 法は，限られたデータ量の適応学習データでは学習できないモデルの補間による推定及び平滑化の処理による推定誤差の低減により，適応学習のデータ量不足を補う働きをするものと見なすことができる．

MAP 推定法，VFS 法共に，十分に学習された音響モデルを初期モデルとして用いることを前提とする点では同じであるが，学習・推定機構の枠組については異なる．MAP 推定法は，適応学習データに存在する音響モデルのみを対象とした学習法であり，クラス内学習と見なすことができる．一方，VFS 法は，適応学習データを用いた学習結果を利用し，適応学習データに存在しない音響モデルの補間推定や平滑化による既学習モデルの推定誤差の補正を行なう点から，モデル間の関係に着目した推定であり，クラス間平滑化と見なすことができる．

MAP/VFS 法は，クラス内学習とクラス間平滑化を組み合わせて互いに他方の機能を強化することにより期待できる適応の高速化，及び，MAP 推定の特徴である追加学習機能を活かした逐次的な適応機能の実現を狙った適応方式である．

3 逐次型 MAP/VFS の適応処理

3.1 MAP と VFS の組み合わせ方

検討した 4 種類の MAP と VFS の組み合わせ方を図 2 に示す．各組み合わせは，

- 逐次的な MAP 推定の処理過程における VFS の組み込み方:
  (a)Alternate(MAP+VFS).
  (b)Sequential MAP + VFS
- VFS 法の移動ベクトル場の原点のとり方:
  (c) 初期モデル (固定)
  (d) 平滑化モデル (逐次更新)

において互いに異なる．MAP 推定の処理過程への VFS の組み込み方 (a) では，MAP 推定により適応学習したモデルに基づいて，VFS 法により，未学習のモデルを補間推定かつ既学習モデルの平滑化を実行し，得られたモデルを次の MAP 推定の種モデルとして用いる．そして，このような処理を，適応学習データが得られる毎に繰り返し行なう．これに対して，(b) では，逐次的な MAP 推定の処理過程において，VFS を後処理として用いる方法である．各回の MAP 推定に用いる種モデルは，前回の MAP 推定により得られたモデルである．VFS は，各回の MAP 推定により適応学習されたモデルに基づいて実行する．VFS 法の移動ベクトル場の原点のとり方については，(c) の初期モデルを常に原点とする固定型と，(d) に示すように，原点を VFS により平滑化されたモデルに

表 1. 話者適応実験の分析条件

Method :	LPC Analysis
Sampling Freqency :	12 [kHz]
Window Length :	32 [ms] (Hamming)
Frame Shift :	8 [ms]
Preemphasis :	1-0.97z^{-1}
Feature Parameters :	16 LPC cepstrum
	16 Δ cepstrum
	Δ log-power

逐次更新する方法が考えられる．予備実験の結果から，(a) と (c) の組合せから成る方式を今回の逐次型 MAP/VFS の適応処理方式とした．

3.2 適応処理フロー

上記の (a) と (c) の組合せからなる MAP/VFS の適応処理フローでは，各適応学習データが得られる毎に，MAP 推定，VFS による補間・平滑化処理を行なう．各回の MAP 推定では，前回の VFS により得られたモデルを種モデルとして用いる．各回の VFS の処理では，各回の MAP 推定でそれぞれの適応学習データを用いて学習されたモデルをその回まで累積し，それらをその回の適応学習モデルと見なしてそれらに対応する初期モデルとの差分から移動ベクトルを求め，補間・平滑化処理を実行する．従って，各回の適応処理では，その回において得られた適応学習データのみを用いて MAP 推定を実行するが，VFS では，モデルパラメータの形式で累積されたその回までの適応学習の知識を用いた処理が実行される．これにより，適応学習データそのものを蓄積することがないので，記憶量を削減した効率的な適応処理が実現可能である．

4 認識実験

話者適応 [14]，話者と回線の同時適応 [11] に対して，逐次型 MAP/VFS の性能評価を行なった．いずれも，適応パラメータは平均ベクトルのみである．

4.1 話者適応の実験とその結果

不特定話者モデルをベースとして，MAP/VFS 法と MAP 法による逐次話者適応化を行ない，その認識性能を評価した．分析条件を表 1 に示す．特徴量は，16 次 LPC ケプストラム，16 次Δケプストラム，Δ対数パワーの 33 次元ベクトルである．

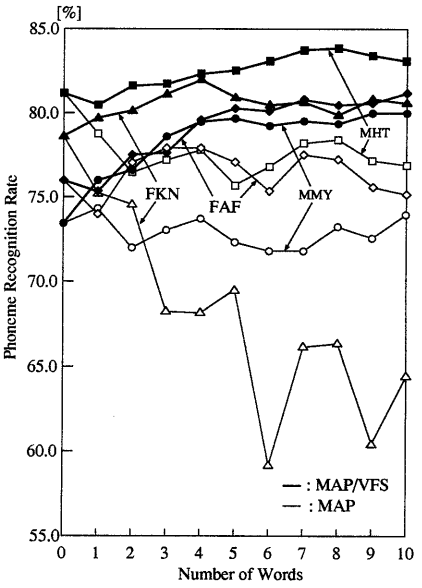

図 3. 音素認識における適応学習特性 (話者適応)

表 2. 話者適応における音素認識性能比較

Speaker	Adaptation Method	Number of Words 0	5	10
MHT	MAP/VFS	[%] 81.2 (94.9)	[%] 82.6 (95.3)	[%] 83.1 (95.5)
	MAP		75.7 (89.8)	76.8 (89.8)
MMY	MAP/VFS	[%] 73.3 (89.7)	[%] 79.7 (94.4)	[%] 80.0 (94.5)
	MAP		72.3 (88.0)	73.9 (88.4)
FAF	MAP/VFS	[%] 76.0 (90.9)	[%] 80.2 (93.0)	[%] 81.2 (93.5)
	MAP		77.1 (90.7)	75.2 (87.7)
FKN	MAP/VFS	[%] 78.7 (92.5)	[%] 80.9 (94.2)	[%] 80.6 (93.9)
	MAP		69.5 (86.7)	64.4 (81.3)
Average Error Reduction Rate	MAP/VFS		[%] 14.8	[%] 16.5
	MAP		-17.9	-22.9

( ) : Recognition rate up to the third rank

不特定話者モデルは，日本音響学会データベース 9600 文 (男性 30 名，女性 34 名) を用いた連結学習により求めた．モデルは，3 状態 4 混合の left-to-right 型の混合連続分布の音素ベースの環境独立型 HMM である．分散は対角成分のみを用いた．音素モデルは 52 種類である．

連母音などを除く 35 種類の基本的な音素に対する音素認識実験を行なった．評価には，ATR5240 単語のうち同音異義語をマージした 4852 単語を用い，うち 10 単語は適応学習データに，残りの 4842 単語を音素認識に用いた．評価に用いたテストデータの話者は，男女各 2 名 (男性：MHT, MMY, 女性：FAF, FKN) である．尚，VFS 法での平滑化パラメータ s の値は，すべての実験に

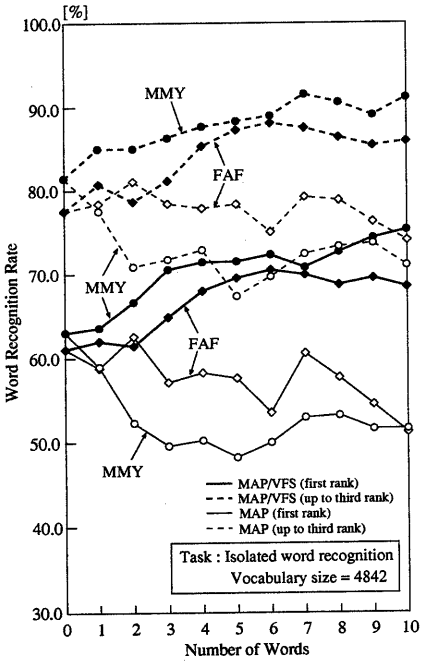

図 4. 大語彙単語認識における適応学習特性 (話者適応)

表 3. 話者適応における大語彙単語認識性能比較

Speaker	Adaptation Method	Number of Words 0	5	10
MMY	MAP/VFS	[%] 63.0 (81.4)	[%] 71.6 (88.2)	[%] 75.2 (91.0)
	MAP		48.2 (67.4)	51.6 (71.0)
FAF	MAP/VFS	[%] 61.0 (77.6)	[%] 69.6 (87.2)	[%] 68.4 (85.8)
	MAP		57.6 (78.2)	51.2 (74.0)
Average Error Reduction Rate	MAP/VFS		[%] 22.6	[%] 26.0
	MAP		-24.4	-28.0

Task : Isolated word recognition (Vocabulary size = 4842)
( ) : Recognition rate up to the third rank

おいて 10.0 とした.

図 3 に音素認識の適応学習特性を示す. MAP/VFS 法では, 学習特性は全体に右上がりで, 特に 5 単語までは認識率が急激に増加する傾向を示している. これに対し, MAP 法では. 初期モデル (不特定話者モデル) の認識率に比べて低下するかまたはその近辺を変動する傾向を示した. 5, 10 単語の適応化での音素認識性能比較を表 2 に示す. MAP/VFS 法では, 平均 14.8%, 16.5% 誤認識が削減された. MAP 法では, それぞれ 17.9%, 22.9% 誤認識が増加した.

音素認識の評価で用いたテストデータを用いて大語彙単語認識 (語彙数 =4842) を行なった. 実験には, 4842 単語のデータからランダムに選んだ 500 単語のデータを用いた. 評価に用いたテストデータの話者は, 男女各 1 名 (男性 : MMY, 女性 : FAF) である. 図 4 に大語彙単語認識における適応学習特性を示し, 表 3 に 5, 10 単語適応での認識性能比較を示す. 適応学習特性は, 音素認識の場合と同様全体に右上がりの特性で, 4, 5 単語まで認識率が急激に増加する傾向を示している. 認識性能に関しては, MAP/VFS 法では, それぞれ, 平均 22.6%,26.0% 単語誤認識率が削減された. MAP 法では, 24.4%,28.0% 単語誤認識率が増加した. この結果から, 逐次型 MAP/VFS 法では, VFS 法により MAP 法での話者適応が加速されているものと考えられる.

表 4. 話者と回線の同時適応実験の分析条件

Method	Selective LPC (SLP) Analysis
Bandwidth	300 ~ 3400 [Hz] (Telephone)
Sampling Freqency	8 [kHz]
Window Length	32 [ms] (Hamming)
Frame Shift	8 [ms]
Preemphasis	$1\text{-}0.97z^{-1}$
Feature Parameters	12 LPC cepstrum
	12 Δ cepstrum
	Δ log-power

4.2 話者と回線の同時適応の実験とその結果

実際的な応用として重要な電話音声認識に対して, 提案方式の効果を評価した. 実験は, 電話回線の周波数特性, 電話機のマイクロホン特性, 話者の個人性の変動の同時適応化である. これは, 次に示す, 回線の周波数特性に対する補正がケプストラム空間での加算処理で実現できるという理論的な事実に, ケプストラム空間での線形変換である VFS 法の原理が適合している点に基づいている.

$$\text{時間領域} : y(t) = h(\tau) * x(t) \tag{8}$$

$$\text{ケプストラム領域} : C^Y = C^H + C^X \tag{9}$$

ここで, $h(\tau)$ は回線特性, $x(t)$ は電話機から入力される音声信号, $y(t)$ は電話回線からの出力信号であり音声認識システムの入力に相当する. C^X, C^Y, C^H はそれぞれ $x(t)$, $y(t)$, $h(\tau)$ に対するケプストラムである.

実験は, 不特定話者モデルをベースとした, 1 単語単位の適応学習による逐次型の話者と回線の同時適応であり, MAP/VFS 法と MAP 法に対して認識性能を比較・評価した. 分析条件を表 4 に示す. 電話帯域を扱う観点から, 選択予測分析 (SLP:Selective LPC) を用いた. サ

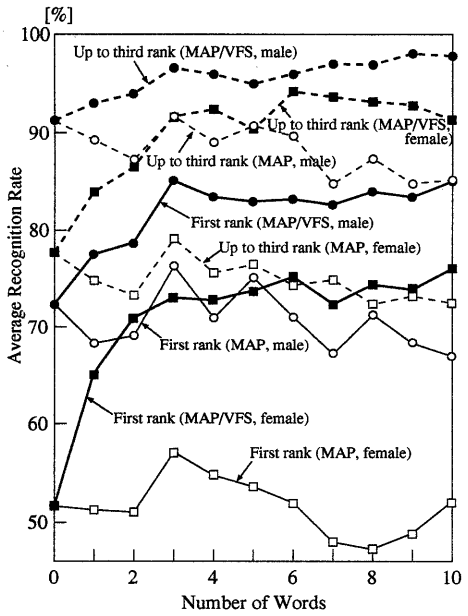

図 5. 単語認識における適応学習特性
(話者と回線の同時適応)

ンプリング周波数は 8[kHz] である．特徴量は，話者適応実験の場合と同じ種類を用いたが，ケプストラム，Δケプストラムの次数は 12 とした．不特定話者モデルは，話者適応の場合と同様，日本音響学会音声データベースの 9600 文を用いた．ただし，電話音声を模擬するため，高品質音声データを電話帯域 (300[Hz] ～ 3.4[kHz]) で分析した．モデルは 3 状態 4 混合の left-to-right 型の混合連続 HMM で，音素ベースの環境独立型モデルである．評価用データは，実電話回線経由で収録した実電話音声である．使用した電話機の送話器はカーボンマイク，発声内容は単語で，語彙は電子協提案の 100 都市名である．適応学習データは，評価用データ収録時に一緒に収録し，語彙は ATR 音素バランス 216 単語である．この中から任意に選択した 10 単語を適応学習データとして用いた．評価用データの話者は男女各 3 名である．尚，VFS 法の平滑化パラメータ s はすべての実験において 10.0 とした．

話者と回線の同時適応における単語認識の適応学習特性を図 5 に示す．図 5 では，男女各 3 名の平均認識率をプロットした．実線は第 1 位の認識率を，破線は第 3 位までの累積認識率を表す．話者適応の場合と同様，MAP/VFS 法では，特性は全体に右上がりであり，3, 4 単語までは認識率が著しく増加する傾向を示す．これ

表 5. 話者と回線の同時適応における単語認識性能比較

Speaker	Adaptation Method	Number of Training Words		
		0	5	10
Male	ML/VFS	[%]	[%] 80.7 (94.3)	[%] 81.3 (95.0)
	MAP(p)		79.7 (95.7)	76.3 (94.3)
	MAP(p)/VFS	72.3 (91.3)	83.3 (95.3)	83.7 (96.0)
	MAP(s)		75.7 (90.7)	66.7 (85.0)
	MAP/VFS(s)	*94.3 (100.0)	82.7 (95.0)	85.0 (97.7)
Female	ML/VFS	[%]	[%] 75.7 (89.7)	[%] 75.0 (92.7)
	MAP(p)		61.7 (83.7)	66.0 (85.3)
	MAP(p)/VFS	51.7 (77.7)	73.0 (89.7)	75.3 (92.0)
	MAP(s)		53.7 (76.3)	52.0 (72.3)
	MAP/VFS(s)	*88.0 (97.7)	73.7 (90.3)	76.0 (91.3)
Average Error Reduction Rate	ML/VFS		[%] 40.0	[%] 40.4
	MAP(p)		23.7	22.0
	MAP(p)/VFS		41.9	45.1
	MAP(s)		8.2	-9.8
	MAP/VFS(s)		41.5	48.1

* : Recognition rate for high-quality data (but telephone bandwidth)
( ) : Recognition rate up to third rank
(p) : pooled, (s) : sequential

に対し，MAP 法では，初期モデル (不特定話者モデル) の認識率付近を変動する傾向にある．5，10 単語適応での単語認識性能比較を表 5 に示す．表 5 では，VFS 法の回線適応に対する効果をも調べる目的で，提案法及び従来の VFS 法 (ML/VFS) を含めた 5 種類の方法に対して評価を行なった．各方法の性能を誤認識削減率で評価すると，5 単語適応に対して，ML/VFS : 40.0%，MAP (pooled) : 23.7%，MAP(pooled)/VFS : 41.9%，MAP(sequential) : 8.2%，本報告で提案する逐次型 MAP/VFS(s) : 41.5% であった．10 単語適応の場合は，それぞれ，40.4%，22.0%，45.1%，-9.8%，48.1% であった．この結果から，逐次型 MAP/VFS 法では，VFS 法により MAP 法での話者と回線の同時適応が加速されているものと考えられる．ここで，pooled，sequential は，それぞれ，蓄積型，逐次型を意味する．

5 考察

前章の実験結果に基づき，逐次型の MAP/VFS 法と MAP 法の性能差の原因について考察する．MAP 推定は，先にも述べたように，クラス内学習であるため，適応学習データに存在する音素に対応するモデルしか適応学習することができない．従って，MAP 法により逐次的に適応を行なっても，限られたデータ量の適応学習データの条件では，一部のモデルのみ学習され，学習できないモデルがかなり存在すると考えられる．これに対し，提案する MAP/VFS 法では，クラス間平滑化方式である VFS 法の補間処理により未学習モデルが推定できるので，未学習モデルに対してあたかも適応学習データが存在し，それを用いた適応学習が行なわれたような効果

が得られているものと考えられる．また，平滑化処理により，少量データの適応学習に対する推定誤差の補正も行なわれるので，逐次，すべてのモデルが学習あるいは推定され，適応効果が増幅されているものと考えられる．上記の理由から，MAP/VFS 法が MAP 法に比べて適応速度が速いものと考えられる．

また，本実験では，MAP 推定の適応化パラメータ τ を式 (4) により各要素分布毎に調整したが，ロバスト性を高める観点からすべての要素分布に対して共通の値を用いる方が良いという報告がある [4]．また，その値は，使用するモデルや適応学習データ量などに応じて最良の値を実験的に求めておくことが必要である．

6 まとめ

オンライン適応の基本技術となる逐次型話者・回線適応方式として，最大事後確率 (MAP) 推定法と移動ベクトル場平滑化 (VFS) 法を組み合わせた MAP/VFS 法を提案し，その原理と実験・評価について述べた．MAP 推定法を用いた逐次的な学習処理過程に VFS 法の補間・平滑化処理を組み込むことにより，限られたデータ量の適応学習データでも．逐次すべてのモデルが適応学習され，高速な適応が実現できることを示した．また，適応学習データの蓄積を必要としない方式であるため，記憶量を節約した効率的な適応処理が可能である．

今後は，平均ベクトル以外の他のモデルパラメータの適応手法や，適応学習データの選択について検討する予定である．

謝辞 日頃御指導いただく，ヒューマンインタフェース研究所音声情報研究部北脇部長，音声処理方式研究グループ管村リーダをはじめグループの方々に感謝いたします．また，MAP 推定法に関してアドバイス頂いたヒューマンインタフェース研究所 古井特別研究室 松岡主任研究員に感謝します．

参考文献

[1] C. -H. Lee, C. H. Lin, and B. H. Juang, "A Study on Speaker Adaptation of the Parameters of Continuous Density Hidden Markov Models," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 39, No. 4, pp.806-814, April 1991.

[2] J. -L. Gauvain and C. -H. Lee, "Bayesian Learning of Gaussian Mixture Densities for Hidden Markov Models," *Proc. DARPA Speech and Natural Language Workshop*, pp. 272-277, Arden House, Feb. 1991.

[3] J. -L. Gauvain and C. -H. Lee, "Maximum *a Posteriori* Estimation for Multivariate Gaussian Mixture Observations of Markov Chains," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 2, No. 2, pp. 291-298, Apr. 1994.

[4] C. -H. Lee and J. -L. Gauvain, "Speaker Adaptation Based on MAP Estimation of HMM Parameters," *Proc. ICASSP93*, pp. II-558-561, Apr. 1993.

[5] 中川，越川，"最大事後確率推定法を用いた連続出力確率分布型 HMM の適応化，" 日本音響学会論文誌 Vol. 49, No. 10, pp.721-728, 1993.

[6] 松岡，C. -H. Lee，"最大事後確率推定法 (MAP 推定法) によるオンライン話者適応化，" 信学技報 SP93-133, pp.39-46, 1994.

[7] T. Matsuoka and C. -H. Lee, "A Study of On-line Bayesian Adaptation for HMM-based Speech Recognition," *Proc. EUROSPEECH93*, pp.815-818,1993.

[8] Y. Turumi and S. Nakagawa, "An Unsupervised Speaker Adaptation Method for Continuous Parameter HMM by Maximum A Posteriori Probability Esimation," *Proc. ICSLP94*, pp.431-434, Sept. 1994.

[9] J. Takahashi and S. Sagayama, "Telephone Line Characteristic Adaptation Using Vector Field Smoothing Technique," *Proc. ICSLP94*, pp.991-994, Sept. 1994.

[10] 外村，小坂，松永，"最大事後確率推定法を用いた移動ベクトル場平滑化話者適応方式，" 音学講論 2-8-20, pp.77-78, 1994.

[11] J. Takahashi and S. Sagayama, "Fast Telephone Channel Adaptation Based on Vector Field Smoothing Technique," *Proc. of 2nd IEEE Workshop on Interactive Voice Technology for Telecommunications Applications (IVTTA94)*, pp.97-100, Sept. 1994.

[12] 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, Oct. 1992.

[13] J. Takami and S. Sagayama, "A Speaker Adaptation Technique for Hidden Markov Networks," *Technical Report of IEICE*, SP93-50, pp. 9-16, Aug. 1993.

[14] 高橋，嵯峨山，"最大事後確率推定と移動ベクトル場平滑化の組合せによる高速話者適応，" 音学講論 2-8-19, pp.75-76, 1994.