修 士 論 文

MLLR における回帰行列の重みづけ線形和を用いた適応法に関する研究

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

小山 岳史

2005 年 3 月

修 士 論 文

MLLRにおける回帰行列の重みづけ線形和を用いた適応法に関する研究

指導教官　党建武 教授

審査委員主査　党建武 教授
審査委員　赤木正人 教授
審査委員　小谷一孔 助教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

210037 小山 岳史

提出年月: 2005年2月

概 要

本論文では、最尤線形回帰 (MLLR) における回帰行列の線形和を用いた適応法を提案する。MLLR 法の利点は、音響モデル空間を幾つかの回帰クラスに分割し、それぞれのクラス毎に回帰行列を求め適応を行うことにより、適応データが存在しないモデルにおいても適応が可能となる点である。しかし、たとえ同じクラスに属していても、それぞれのモデルに応じて適切な回帰行列を用いて適応を行ったほうが良いのではないかと考えられる。そこで本研究では、他のクラスの回帰行列の線形和を用いることで, 同一クラス内のモデルにおいても、各々のモデルにおける他の回帰クラスの影響を考慮した回帰行列を求める手法を提案した。

目 次

第 1 章　序論 **1**
1.1　本研究の背景と目的 1
1.2　本論文の構成 1

第 2 章　隠れマルコフモデルによる音声認識と話者適応 **2**
2.1　隠れマルコフモデルによる音声認識 2
2.1.1　HMM の定義 2
2.1.2　HMM パラメータの学習 3
2.1.3　HMM による音声認識 3
2.2　隠れマルコフモデルにおける話者適応 4
2.2.1　話者適応とは 4
2.2.2　MLLR 法 5

第 3 章　回帰行列の線形和を用いた適応法 **8**
3.1　MLLR 法の利点と考えられる問題点 8
3.2　提案手法 8
3.2.1　距離関数を用いた重み係数の計算 9
3.2.2　ラグランジュの補間公式を用いた重み係数の計算 10

第 4 章　比較、評価実験 **12**
4.1　実験条件 12
4.2　実験結果 13
4.2.1　距離関数を用いた重み係数での認識結果 13
4.2.2　ラグランジュの補間公式を用いた重み係数での認識結果 14

第 5 章　考察 **15**
5.1　閾値を考慮した回帰行列の線形和による認識結果 16
5.2　線形和の項数選択と認識結果 19

第 6 章　結論 **21**
6.1　結論 21
6.2　今後の課題 21

図 目 次

3.1 回帰クラス数 2 の場合 8
3.2 通常の MLLR での適応 11
3.3 距離関数を用いた重みをつけた適応 11

5.1 proposal1-2 が適応される例 16
5.2 行列の選択基準 19

表 目 次

4.1 距離関数で重みをつけた結果 13
4.2 各回帰クラスにおける結果の平均値 13
4.3 各回帰クラスにおける結果の最大値 13
4.4 ラグランジュの補間公式で重みをつけた結果 14
4.5 各回帰クラスにおける結果の平均値 14
4.6 各回帰クラスにおける結果の最大値 14

5.1 提案法を適用するモデルの範囲を変更した結果 17
5.2 提案法を適用するモデルの範囲を変更した結果の平均値 18
5.3 提案法を適用するモデルの範囲を変更した結果の最大値 18
5.4 線形和の項数選択を行った場合の認識結果 20
5.5 線形和の項数選択を行った認識の各クラス数での平均値 20
5.6 線形和の項数選択を行った認識の各クラス数での最大値 20

第1章　序論

1.1　本研究の背景と目的

HMM を用いた音声認識において、一般的に不特定話者モデルを用いた認識は特定話者モデルを用いた場合に比べて認識性能が低い。また、ある特定の話者における認識性能が他の話者に比べて著しく低いという現象が起こる。この原因として、認識に用いている HMM のパラメータがその話者にマッチしていないことが考えられる。そこで、認識に用いる話者の音声データを使用し、HMM のパラメータを調整することにより認識性能の向上を図るアプローチが欠かせないものとなる。これは話者適応と呼ばれる手法である.

話者適応の代表的手法として、最尤線形回帰 (MLLR) がその取り扱い易さと性能の高さにより広く用いられている。MLLR では音響モデルを幾つかの回帰クラスに分割し, それぞれのクラス毎に回帰行列を求めて適応することで、適応データが存在しないモデルに関しても適応を可能としている.

しかし、回帰クラスの中心付近のモデルと境界付近のモデルなど、たとえ同じクラスに属していても違う適応を行った方が良い場合が存在すると考えられる。そこで本研究では、他のクラスの回帰行列の線形和を用いることで、同一クラス内のモデルにおいても、各々のモデルにおける他の回帰クラスの影響を考慮した回帰行列を求める手法を提案した。

1.2　本論文の構成

1 章では序論として、本研究の背景と目的を述べた。2 章では隠れマルコフモデルによる音声認識について述べる。3 章では、本研究の提案手法である、回帰行列の線形和を用いた適応法やその計算法について述べる。4 章で提案法と従来法である MLLR との性能の比較実験やその結果について述べる。5 章では、4 章での実験の結果を踏まえた上で、6 章では本研究で得られた結論を述べ、今後の課題について検討する。

第2章　隠れマルコフモデルによる音声認識と話者適応

音声認識におけるパラメータ系列のモデル化の手法として、隠れマルコフモデル (HMM: Hidden Marcov Model) が広く用いられている。本章では、HMM を用いた音声認識について述べる。

2.1　隠れマルコフモデルによる音声認識

2.1.1　HMM の定義

HMM は時系列信号の確率モデルであり、複数の定常信号源の間を遷移することで、非定常な時系列信号をモデル化している。以下にその一例を示す。

出力ベクトル $\boldsymbol{o}_t$ を出力する確率分布が $b_i(\boldsymbol{o}_t)$ であるような信号源（状態）が、状態遷移確率 $a_{ij} = P(q_t = j|q_{t-1} = i)$ で遷移するものとして定義される。ただし、i, j は状態番号とする。

音声関連の応用では、出力ベクトル $\boldsymbol{o}_t$ を、MFCC、LPC 等の音声の短時間的なスペクトルを表現する音声パラメータである。また、音声のモデル化においては、因果性を表現するため、図のように状態を横 1 列に並べたときに左方向への遷移がない（時間が逆戻りしない）left-to-right 型と呼ばれるモデルが用いられている。

出力確率分布を単一の多次元正規分布、HMM の状態数を N としたとき、HMM のパラメータ θ は、初期状態確率 $\boldsymbol{\pi} = \{\pi_i\}_{i=1}^N$、状態遷移確率 $A = \{a_{ij}\}_{i,j=1}^N$、各状態 i での出力確率 $B = \{b_i(\cdot)_{i=1}^N\}$ により $\theta = (\boldsymbol{\pi}, A, B)$ で表される。この場合、状態が $\boldsymbol{Q} = \{q_1, q_2, \cdots, q_T\}$ と遷移し、かつ出力ベクトル系列 $\boldsymbol{O} = \{\boldsymbol{o}_1, \boldsymbol{o}_2, \cdots, \boldsymbol{o}_T\}$ を出力する確率は、

$$P(\boldsymbol{O}, \boldsymbol{Q}|\theta) = \prod_{t=1}^{T} a_{q_{t-1}q_t} b_{q_t}(\boldsymbol{o}_t) \tag{2.1}$$

で与えられる。ただし $a_{q_0 i} = \pi_i$ である。したがって、$\boldsymbol{O} = \{\boldsymbol{o}_1, \boldsymbol{o}_2, \cdots, \boldsymbol{o}_T\}$ がパラメータが θ である HMM から出力される確率は、すべての状態遷移について和をとることにより、

$$P(\boldsymbol{O}|\theta) = \sum_{\boldsymbol{Q}} P(\boldsymbol{O}, \boldsymbol{Q}|\theta) = \sum_{\boldsymbol{Q}} \prod_{t=1}^{T} a_{q_{t-1}q_t} b_{q_t}(\boldsymbol{o}_t) \tag{2.2}$$

と表すことができる。出力確率分布として M 個の多次元正規分布の重みつき和を用いる場合は、各分布の重み $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^{M}$, 平均ベクトル $\boldsymbol{\mu} = \{\boldsymbol{\mu}_i\}_{i=1}^{M}$、共分散行列 $\boldsymbol{\Sigma} = \{\Sigma_i\}_{i=1}^{M}$ が出力分布におけるパラメータとなる。

単純に式 (2.2) を計算した場合、計算量が非常に大きくなってしまうが ($O(2T \cdot N^T)$)、効率的な計算法として foward/backward アルゴリズムがある。これを用いることにより、計算量は $O(N^2T)$ に削減される。

2.1.2 HMM パラメータの学習

HMM のモデルパラメータ θ は、与えられた学習用のベクトル系列 $\boldsymbol{O}$ に対して、式 (2.2) で与えられる $P(\boldsymbol{O}|\theta)$ を最大にするように決定する。

$$\theta_{estimated} = \arg\max_{\theta} P(O|\theta) \tag{2.3}$$

このような推定法は最尤推定法と呼ばれており、式 (2.3) の最大化は、EM アルゴリズムに基づいた Baum-Welch アルゴリズムで解くことができる。これは、何らかの初期モデルから、次式で定義される補助関数 (Q 関数とも呼ばれる)

$$Q(\theta, \bar{\theta}) = \sum_{\boldsymbol{Q}} P(\boldsymbol{Q}|\boldsymbol{O}, \theta) \log P(\boldsymbol{O}, \boldsymbol{Q}|\bar{\theta}) \tag{2.4}$$

を最大化する $\bar{\theta}$ を求め、$\theta \leftarrow \hat{\theta}$ と置き換える操作を繰り返していくものである。これにより、

$$Q(\theta, \bar{\theta}) \geq Q(\theta, \theta) \Rightarrow P(\boldsymbol{O}|\bar{\theta}) \geq P(\boldsymbol{O}|\theta) \tag{2.5}$$

を示すことができる。つまり上記アルゴリズムにより $P(\boldsymbol{O}|\theta)$ の値の単調増加性が保証され、$P(\boldsymbol{O}|\theta)$ の局所的最適解を求めることができる。なお、式 (2.4) の最大化は、foward/backward アルゴリズムにより効率的に行うことができる。

2.1.3 HMM による音声認識

音声認識は、与えられた $\boldsymbol{O}$ に対して、任意の音素列（または単語列）W の中から、$P(W|\boldsymbol{O})$ を最大にする W_{max} を求めることである。

$$W_{max} = \arg\max_{W} P(W|\boldsymbol{O}) \tag{2.6}$$

上式の右辺をベイズの定理を用いて変形すると、

$$W_{max} = \arg\max_{W} \frac{P(\boldsymbol{O}|W)P(W)}{P(\boldsymbol{O})} \tag{2.7}$$

となる。ここで、$P(\boldsymbol{O})$ は入力ベクトル系列 $\boldsymbol{O}$ が観測される期待値であり、W の最大化には無関係であること、また W に対応する HMM を θ_W で与えることにより、

$$W_{max} = \arg\max_{W} P(\boldsymbol{O}|\theta_W)P(W) \tag{2.8}$$

と示すことができる。通常 θ_W は音素モデルを連結して作られる。式 (2.8) における $P(W)$ は言語モデル、$P(\boldsymbol{O}|\theta_W)$ は音響モデルと呼ばれる。

式 (2.8) において、音響モデル $P(\boldsymbol{O}|\theta_W)$ の部分は、

$$\begin{aligned} P(\boldsymbol{O}|\theta_W) &= \sum_{\boldsymbol{Q}} P(\boldsymbol{O},\boldsymbol{Q}|\theta_W) \\ &\simeq \max_{\boldsymbol{Q}} P(\boldsymbol{O},\boldsymbol{Q}|\theta_W) \end{aligned} \tag{2.9}$$

で近似する。この近似は viterbi 近似と呼ばれ、与えられたベクトル系列 $\boldsymbol{O}$ と θ に対して、$P(\boldsymbol{O},\boldsymbol{Q}|\theta_W)$ を最大にする状態系列 $\boldsymbol{Q}$ と、そのときの $P(\boldsymbol{O},\boldsymbol{Q}|\theta_W)$ の値とを動的計画法に基づいて求めるのが vitarbi アルゴリズムである。これにより、状態と音声との時間的な対応関係を求めることができる。

vitarbi 近似による式 (2.8) の最大化は、

$$W_{max} = \arg\max_{W} \max_{\boldsymbol{Q}} P(\boldsymbol{O}|\lambda_W)P(W) \tag{2.10}$$

と示すことができる。なお、式 (2.10) の最大化を vitarbi アルゴリズムで直接解くには探索空間が膨大になるので、ビームサーチ等の探索手法が用いられる。

2.2 隠れマルコフモデルにおける話者適応

2.2.1 話者適応とは

一般的に、不特定話者音声認識は、特定話者音声認識に比べ認識性能が低い。また、話者間で認識性能の大きな偏りがあり、誤認識のきわめて大きい、一部の話者によって全体の認証性能が決まってしまうことが知られている。このような現象は”Sheep and Goats 現象”と呼ばれている。この原因として、認識に用いられる HMM のパラメータがその話者に適していないことが挙げられる。

HMM パラメータを、ある話者の音声データから得られた情報を元に更新することを話者適応という。使用する音声データに関する教師信号の有無により、教師あり適応と教師

なし適応とに分かれる。通常、教師あり適応は、教師なし適応に比べて大きな適応効果を得られるが、実際に使用する場合、話者の音声データの教師信号（音素系列等の情報）をどのように得るかという問題がある。

話者適応の手法として良く用いられるものに、最大事後確率推定 (MAP: Maximum A Posteriori) 法 [5]、最尤線形回帰 (MLLR: Maximum Likelihood Liner Regression) 法 [6] がある。次節で、本研究のベースである MLLR 法について述べる。

2.2.2 MLLR 法

MLLR は連続分布 HMM に対して、適応データの尤度を最大にするような平均ベクトルを推定する変換行列 W を求めることである。ここで、HMM の各状態は単一ガウス出力分布を持つとし、HMM の状態 s の出力分布の平均ベクトルを $\boldsymbol{\mu}_s$ とすると平均ベクトルの推定値は次式で求られる。

$$\hat{\boldsymbol{\mu}}_s = \mathrm{W_S} \cdot \boldsymbol{\xi}_s \tag{2.11}$$

ここで、W_s は $n \times (n+1)$ の変換行列であり、$\boldsymbol{\xi}_s$ は拡張平均ベクトルである。$\boldsymbol{\xi}_s$ は次式で与える。

$$\boldsymbol{\xi}_s = [\omega, \mu_{s_1}, \cdots, \mu_{s_n}]' \tag{2.12}$$

ω はオフセット値を表し、通常 1 を用いる。

MLLR では、時刻 t、状態 s におけるベクトルが観測される確立密度関数を、正規分布 $N(\boldsymbol{\mu}_s, \Sigma_s)$ を用いて

$$b_s(\boldsymbol{o_t}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_s|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{o}_t - \boldsymbol{\mu}_s)'\Sigma_s^{-1}(\boldsymbol{o}_t - \boldsymbol{\mu}_s)\right) \tag{2.13}$$

と仮定する。変換行列 W_s は、Baum-Welch アルゴリズムで使う補助関数を利用して求める。

$$Q(\lambda, \bar{\lambda}) = constan + \sum_{\theta \in \Theta} \sum_{t=1}^{T} P(\boldsymbol{O}, \theta|\lambda) \cdot \log b_{\theta_t}(\boldsymbol{o}_t) \tag{2.14}$$

但し、λ は推定前、$\bar{\lambda}$ は推定後のモデルパラメータの集合であり、$\boldsymbol{O}$ は適応データの特徴ベクトルの系列 $\{\boldsymbol{o}_1, ..., \boldsymbol{o}_T\}$($T$ はフレーム数) を、Θ は状態の系列集合を表している。ここで、S を全状態野集合を表し、状態 s と時刻 t における与えられた観測系列 $\boldsymbol{O}$ に対する事前確率は次式のように表す。

$$\gamma_s(t) = \frac{1}{P(\boldsymbol{O}|\lambda)} \sum_{\theta \in \Theta} P(\boldsymbol{O}, \theta_t = s|\lambda) \tag{2.15}$$

式 (2.14) は次のように表せる。

$$Q(\lambda,\hat{\lambda}) = a + P(\boldsymbol{O}|\lambda)\sum_{j=1}^{S}\sum_{t=1}^{T}\gamma_j(t)\log b_j(\boldsymbol{o}_t) \tag{2.16}$$

ここで、a は定数である。また、$\log b_j(\boldsymbol{o}_t)$ を展開すると、

$$Q(\lambda,\hat{\lambda}) = a - \frac{1}{2}P(\boldsymbol{O}|\lambda)\sum_{j=1}^{S}\sum_{t=1}^{T}\gamma_j(t)[n\log(2\pi) + \log|\Sigma_j| + h(\boldsymbol{o}_t, j)] \tag{2.17}$$

ここで、$h(\boldsymbol{o}_t, j)$ 次式で与える。

$$h(\boldsymbol{o}_t, j) = (\boldsymbol{o}_t - \hat{W}_j\boldsymbol{\xi}_j)'\Sigma_j^{-1}(\boldsymbol{o}_t - \hat{W}_j\boldsymbol{\xi}_j) \tag{2.18}$$

$Q(\lambda,\hat{\lambda})$ に対して微分をとると、

$$\frac{d}{d\hat{W}_s}Q(\lambda,\hat{\lambda}) = -\frac{1}{2}P(\boldsymbol{O}|\lambda)\frac{d}{d\hat{W}_s}\sum_{j=1}^{S}\sum_{t=1}^{T}\gamma_j(t)[n\log(2\pi) + \log|\Sigma_j| + h(\boldsymbol{o}_t, j)] \tag{2.19}$$

最大値を表す $\hat{W}_s$ は、

$$\frac{d}{d\hat{W}_s}Q(\lambda,\hat{\lambda}) = 0 \tag{2.20}$$

$$P(\boldsymbol{O}|\lambda)\sum_{t=1}^{T}\gamma_s(t)\Sigma_s^{-1}[\boldsymbol{o}_t - \hat{W}_s\boldsymbol{\xi}_s]\boldsymbol{\xi}_s' = 0 \tag{2.21}$$

$$\sum_{t=1}^{T}\gamma_s(t)\Sigma_s^{-1}\boldsymbol{o}_t\boldsymbol{\xi}_s' = \sum_{t=1}^{T}\gamma_s(t)\Sigma_s^{-1}\hat{W}_s\boldsymbol{\xi}_s\boldsymbol{\xi}_s'$$

この式が W_s の導出の基本となる式である。
ここで、W_s が R 個の状態 $\{s_1, s_2, ..., s_R\}$ で共有されていると考えると

$$\sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{s_r}(t)\Sigma_{s_r}^{-1}o_t\boldsymbol{\xi}_{sr}' = \sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{s_r}(t)\Sigma_{s_r}^{-1}W_s\boldsymbol{\xi}_{s_r}\boldsymbol{\xi}_{sr}' \tag{2.22}$$

が成り立つ。
共有された再推定式を導出するには、まず、式 (2.22) を次のように書き換える。

$$\sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{s_r}(t)\Sigma_{s_r}^{-1}\boldsymbol{o}_t\boldsymbol{\xi}_{sr}' = \sum_{r=1}^{R}V^{(r)}\hat{W}_sD^{(r)} \tag{2.23}$$

ここで、

$$
\begin{aligned}
Z &= \sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{s_r}(t)\Sigma_{s_r}^{-1}\boldsymbol{o}_t\boldsymbol{\xi}'_{s_r} && (2.24)\\
V^{(r)} &= \sum_{t=1}^{T}\gamma_{s_r}(t)\Sigma_{s_r}^{-1} && (2.25)\\
D^{(r)} &= \boldsymbol{\xi}_{s_r}\boldsymbol{\xi}'_{s_r} && (2.26)\\
g_{jq}^{(i)} &= \sum_{r=1}^{R} v_{ij}^{(r)} d_{jq}^{(r)} && (2.27)
\end{aligned}
$$

とすれば、

$$
w'_i = G^{(i)-1} z'_i \tag{2.28}
$$

を計算することで、W_s が求まる。但し、w_i, z_i はそれぞれ W_s, Z の i 行目を意味し、$g_{jq}^{(i)}, v_{jq}^{(i)}, d_{jq}^{(i)}$ をはそれぞれ $G^{(i)}, V^{(i)}, D^{(i)}$ の j 行 q 列の要素を意味する。

MLLR の多回帰クラスへの拡張

適応単語が少量である場合、一つの解決方法として複数の状態を一つの回帰クラスに共有化する。すべての状態に対する適応データを利用し変換行列を求める。状態が共有化される場合式は次のように拡張することができる。

$$
\begin{aligned}
\sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{s_r}(t)\Sigma_{s_r}^{-1}\boldsymbol{o}_t\boldsymbol{\xi}'_{s_r} &= \sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{s_r}(t)\Sigma_{s_r}^{-1}\hat{W}_s\boldsymbol{\xi}_{s_r}\boldsymbol{\xi}'_{s_r}\\
\sum_{t=1}^{T}\sum_{r=1}^{R}\gamma_{s_r}(t)\Sigma_{s_r}^{-1}\boldsymbol{o}_t\boldsymbol{\xi}'_{s_r} &= \sum_{r=1}^{R} V^{(r)}\hat{W}_s D^{(r)} && (2.29)
\end{aligned}
$$

行列 G は 3 次元的な構造を持ち $(n+1 \times n+1 \times n)$。この式は次のように解く。

$$
\begin{aligned}
z_i &= \hat{w}_i G^{(i)} && (2.30)\\
\hat{w}'_i &= G^{(i)-1} z'_i && (2.31)\\
[\hat{w}_i]_{(n+1\times 1)} &= [g_{jq}]_{(n+1\times n+1)}\,[z_i]_{(n+1\times 1)} && (2.32)
\end{aligned}
$$

第3章　回帰行列の線形和を用いた適応法

3.1　MLLR法の利点と考えられる問題点

2.2.2節で述べたように、MLLR法では音響空間をクラスタリングして作られる回帰クラスごとに回帰行列を求め適応を行う。これにより、適応データが与えられなかったモデルに関しても適応が可能になる。

しかし、1つのクラス内全てのモデルを同一の回帰行列を用いて適応した場合、明らかに状態の異なる、クラスの中心のモデルと他のクラスとの境界のモデルも同一の回帰行列を用いて適応することになり、ここに問題があると考えられる。この問題を解決する方法として、回帰クラスの数を多く設定し、より精度の高い適応を行うことが挙げられるが、それに応じて必要な適応データ数も多くなってしまう。

3.2　提案手法

本研究では、従来のMLLR法に基づいた回帰行列を求めた後、その重み付き線形和をモデル毎に計算し、その行列により適応を行うという手法を提案した。このように他のクラスの回帰行列を考慮することによって、個々のモデルの状況ににより適した回帰行列を求めることが出来ると考えられる。図3.1は、回帰クラス数が2の場合における例である。

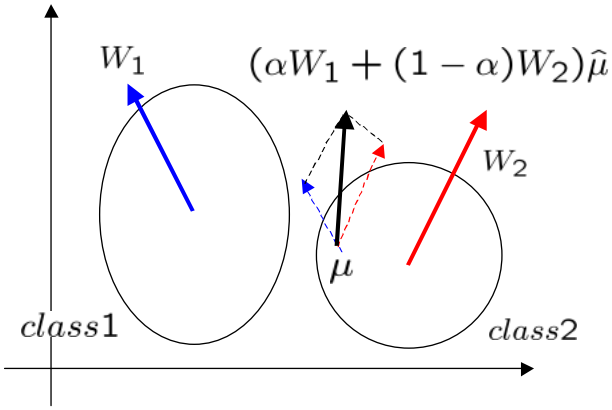

図 3.1: 回帰クラス数2の場合

あるモデルの平均ベクトルを $\boldsymbol{\mu}$、その拡張平均ベクトルを $\hat{\boldsymbol{\mu}}$ とする。class1,class2 における回帰行列をそれぞれ W_1, W_2、重み係数を α, β とすると、このモデルにおける回帰行列 W_{new} は以下の式で表される。

$$W_{new} = \alpha W_1 + \beta W_2 \tag{3.1}$$

重み係数は以下の条件を満たすように正規化する。

$$\alpha + \beta = 1 \tag{3.2}$$

この条件により、提案法が従来の MLLR 法を包含したものになる（ex. class1 に属するモデルの場合ならば、$\alpha = 1, \beta = 0$ となる）と同時に、W_{new} によって適応されたモデル $W_{new}\hat{\boldsymbol{\mu}}$ は、$W_1\hat{\boldsymbol{\mu}}$ と $W_2\hat{\boldsymbol{\mu}}$ を結ぶ直線上に存在することになる。

クラス数が N の場合も同様である。

$$W_{new} = c_1 W_1 + c_2 W_2 + \cdots + c_N W_N \tag{3.3}$$

ただし、

$$c_1 + c_2 + \cdots + c_N = 1 \tag{3.4}$$

本研究では、重み係数を求める手法として、以下の 2 つを用いた。

- クラスの中心からの距離
- ラグランジュの補間公式

以下の節で、それぞれの計算方法を説明する。

3.2.1 距離関数を用いた重み係数の計算

クラス数を N、あるモデルと i 番目のクラスの中心との距離を d_i としたとき、式 (3.3) における重み係数 c_i を以下のように定義する。

$$c_i = \frac{D_i}{\sum_{k=1}^{N} D_k} \quad \left(D_i = \frac{1}{d_i}\right) \tag{3.5}$$

距離の逆数をとることで、距離が小さいクラスに大きな重みがかかる。本研究では距離関数として、以下の式で表されるバタチャリア距離を用いた。$\mu^{cl}, \mu^{m}, \Sigma^{cl}, \Sigma^{m}$ はそれぞれ回帰クラス、モデルの平均と分散である。

$$d_i = \frac{1}{8}(\mu^m - \mu^{cl})^t \left(\frac{\Sigma^m + \Sigma^{cl}}{2}\right)^{-1} (\mu^m - \mu^{cl}) + \frac{1}{2} \ln \frac{\left|(\Sigma^m + \Sigma^{cl})/2\right|}{\left|\Sigma^m\right|^{\frac{1}{2}} \left|\Sigma^{cl}\right|^{\frac{1}{2}}} \tag{3.6}$$

3.2.2 ラグランジュの補間公式を用いた重み係数の計算

ラグランジュの補間公式とは、独立変数の範囲として 1 つの区間が与えられており、その区間内である関数の値のいくつかが知られているとき、同じ区間内でその関数の別の値の近似値を求めるための公式である。

$n+1$ 個の異なる点 $x = x_0, x_1, \cdots, x_n$ に対して、任意の n 次多項式 $f(x)$ を以下のように近似する。

$$\begin{aligned} f(x) &= f(x_0)\frac{(x-x_1)(x-x_2)\cdots(x-x_n)}{(x_0-x_1)(x_0-x_2)\cdots(x_0-x_n)} \\ &+ f(x_1)\frac{(x-x_0)(x-x_2)\cdots(x-x_n)}{(x_1-x_0)(x_1-x_2)\cdots(x_1-x_n)} \\ &+ \cdots \\ &+ f(x_n)\frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\cdots(x_n-x_{n-1})} \end{aligned} \tag{3.7}$$

この $f(x)$ は $x = x_0, x_1, \cdots, x_n$ で、それぞれ $f(x_0), f(x_1), \cdots, f(x_n)$ となる。この公式において、

$x_i \Rightarrow \boldsymbol{\mu}_i^{cl}$：$i$ 番目のクラスのセントロイド
$x \Rightarrow \boldsymbol{\mu}^m$：モデルの平均ベクトル
$f(x_i) \Rightarrow W_i$：i 番目のクラスの回帰行列
$(x - x_i) \Rightarrow D(x, x_i)$：モデルとクラスおよびクラス間のバタチャリヤ距離

と置き換えることにより、本研究に適用した。

$$\begin{aligned} W_{new} &= W_0\frac{D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_1^{cl})D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_2^{cl})\cdots D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_n^{cl})}{D(\boldsymbol{\mu}_0^{cl},\boldsymbol{\mu}_1^{cl})D(\boldsymbol{\mu}_0^{cl},x\boldsymbol{\mu}_2^{cl})\cdots D(\boldsymbol{\mu}_0^{cl},\boldsymbol{\mu}_n^{cl})} \\ &+ W_1\frac{D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_0^{cl})D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_2^{cl})\cdots D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_n^{cl})}{D(\boldsymbol{\mu}_1^{cl},\boldsymbol{\mu}_0^{cl})D(\boldsymbol{\mu}_1^{cl},\boldsymbol{\mu}_2^{cl})\cdots D(\boldsymbol{\mu}_1^{cl},\boldsymbol{\mu}_n^{cl})} \\ &+ \cdots \\ &+ W_n\frac{D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_0^{cl})D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_1^{cl})\cdots D(\boldsymbol{\mu}^m,\boldsymbol{\mu}_{n-1}^{cl})}{D(\boldsymbol{\mu}_n^{cl},\boldsymbol{\mu}_0^{cl})D(\boldsymbol{\mu}_n^{cl},\boldsymbol{\mu}_1^{cl})\cdots D(\boldsymbol{\mu}_n^{cl},\boldsymbol{\mu}_{n-1}^{cl})} \end{aligned} \tag{3.8}$$

この式により重み係数を求めた後、式 (3.4) を満たすように正規化する。

提案法を用いた適応のイメージ図

二次元データでの MLLR と提案法を用いた場合の、モデルの適応例を以下に示す。図 3.2 は従来法である MLLR、図 3.3 は距離関数を用いて回帰行列に重みをつけた場合の図である。

矢印の根元が適応前のモデルを示しており、上下の図ともに同じものである。矢印の先が適応後のモデルを示している。この図より、他の回帰行列を考慮することで、モデルの移動が平滑化され、極端な適応が行われなくなることがわかる。

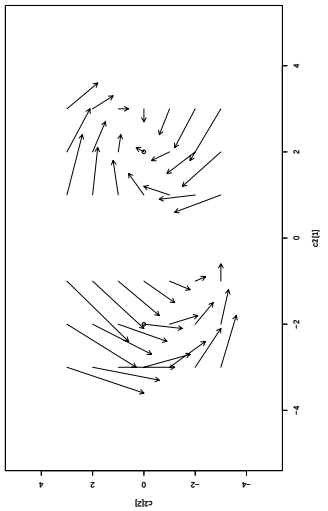

図 3.2: 通常の MLLR での適応

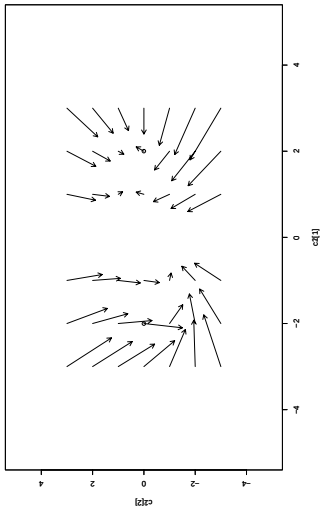

図 3.3: 距離関数を用いた重みをつけた適応

第4章　比較、評価実験

本研究で提案した回帰行列の線形和を用いた適応法と、従来法である MLLR 法の比較実験を行い、その性能を評価した。

4.1　実験条件

- 実験データ
  ATR 研究用データベース Aset の男性 10 名、女性 5 名の重要単語 5240 単語
  - 不特定話者モデル
    男性 7 名 (mht,mms,mmy,mtk,mtm,mtt,mxm)
    女性 3 名 (ffs,fkn,fms)
    計 10 名のデータにより学習
  - 評価用データ
    不特定話者モデル生成時に用いた話者以外の
    男性 3 名 (mau,mnm,msh)、女性 2 名 (faf,fyn) の偶数番目単語のうち 151 単語
  - 適応用データ
    偶数番目の単語のうち、評価用以外の単語をランダムに選択
- 音響パラメータ
  0 次を含む MFCC13 次元
- 音響分析
  サンプリング周波数 12KHz、20ms ハミング窓、フレーム周期 10ms
- HMM の構成
  Left-to-Right3 状態、各状態は 1 混合
- 回帰クラス数
  2、3、5、8、10

4.2 実験結果

4.2.1 距離関数を用いた重み係数での認識結果

適応を行った 5 話者の単語認識率の平均値を以下に示す。横軸が適応単語数、縦軸が回帰クラス数、枠内の上の数字が従来法 (MLLR)、下の数字が提案法の認識率である。

		10	15	20	30	40	60	80	100	200
2	MLLR	85.50	89.01	88.09	88.09	88.40	88.55	88.24	88.70	87.94
	proposal1	85.80	89.01	88.24	88.55	88.40	88.86	88.09	88.24	87.48
3	MLLR	85.50	88.85	87.18	89.31	88.86	89.47	89.62	89.01	89.16
	proposal1	85.80	89.31	87.79	88.70	88.09	88.70	88.24	88.40	88.40
5	MLLR	85.50	88.85	86.72	88.70	90.84	89.92	89.77	89.93	89.77
	proposal1	85.80	89.31	87.33	89.01	89.47	89.31	88.85	89.62	89.62
8	MLLR	85.50	88.85	86.72	88.70	90.84	89.77	90.53	89.01	89.92
	proposal1	85.80	89.31	87.33	89.01	89.47	89.31	89.62	88.86	88.85
10	MLLR	85.50	88.85	86.72	88.70	90.69	89.16	90.23	88.85	90.08
	proposal1	85.80	89.31	87.33	89.01	89.16	89.01	89.31	88.09	89.01

表 4.1: 距離関数で重みをつけた結果

各回帰クラス数での結果の平均値と最大値を以下に示す。なお適応単語数 0 における数値は適応前の単語認識率を示す。

ave.	0	10	15	20	30	40	60	80	100	200
MLLR	79.08	85.50	88.88	87.09	88.70	89.93	89.37	89.68	89.10	89.37
proposal1	—	85.80	89.25	87.60	88.86	88.92	89.04	88.82	88.64	88.67

表 4.2: 各回帰クラスにおける結果の平均値

max.	0	10	15	20	30	40	60	80	100	200
MLLR	79.08	85.50	89.01	88.09	89.31	90.84	89.92	90.53	89.93	90.08
proposal1	—	85.80	89.31	88.24	89.01	89.47	89.31	89.62	89.62	89.62

表 4.3: 各回帰クラスにおける結果の最大値

適応単度数が 10-30 という比較的少量の場合、5 話者の平均誤り削減率において最大 4.5%程度の効果が得られたが、それ以上の単語数の場合、逆に認識性能が劣化した。

4.2.2 ラグランジュの補間公式を用いた重み係数での認識結果

適応を行った 5 話者の単語認識率の平均値を以下に示す。横軸が適応単語数、縦軸が回帰クラス数、枠内の上の数字が従来法 (MLLR)、下の数字が提案法の認識率である。

		10	15	20	30	40	60	80	100	200
2	MLLR	85.50	89.01	88.09	88.09	88.40	88.55	88.24	88.70	87.94
	proposal2	85.95	88.55	87.79	88.40	88.55	88.24	88.40	88.55	87.94
3	MLLR	85.50	88.85	87.18	89.31	88.86	89.47	89.62	89.01	89.16
	proposal2	85.95	88.86	87.02	88.70	88.70	88.25	88.86	88.70	88.55
5	MLLR	85.50	88.85	86.72	88.70	90.84	89.92	89.77	89.93	89.77
	proposal2	85.95	88.86	86.72	89.01	88.86	88.40	87.18	87.79	87.64
8	MLLR	85.50	88.85	86.72	88.70	90.84	89.77	90.53	89.01	89.92
	proposal2	85.95	88.86	86.72	89.16	88.86	88.25	87.31	86.26	70.08
10	MLLR	85.50	88.85	86.72	88.70	90.69	89.16	90.23	88.85	90.08
	proposal2	85.95	88.86	86.72	89.16	88.86	88.55	87.48	85.04	69.16

表 4.4: ラグランジュの補間公式で重みをつけた結果

各回帰クラス数での結果の平均値と最大値を以下に示す。なお適応単語数 0 における数値は適応前の単語認識率を示す。

ave.	0	10	15	20	30	40	60	80	100	200
MLLR	79.08	85.50	88.88	87.09	88.70	89.93	89.37	89.68	89.10	89.37
proposal2	—	85.95	88.80	86.99	88.88	88.76	88.34	87.84	87.27	80.67

表 4.5: 各回帰クラスにおける結果の平均値

max.	0	10	15	20	30	40	60	80	100	200
MLLR	79.08	85.50	89.01	88.09	89.31	90.84	89.92	90.53	89.93	90.08
proposal2	—	85.95	88.86	87.79	89.16	88.86	88.55	88.86	88.70	88.55

表 4.6: 各回帰クラスにおける結果の最大値

適応単語数が 10 の場合を除いて認識性能が劣化した。また、回帰クラス数を増加させた場合、認識率がさらに低下するという結果になった。これは補間多項式が高次になるために補間関数の振動が激しくなったことが原因だと考えられる。

第5章　考察

本研究で用いた2種類の重み係数の計算法を比較すると、全体として距離関数を用いる方法において性能の向上がみられた。本章ではこの方法を元にした考察について述べる。

前章の実験により、適応単語数が比較的少量の場合は提案法の効果が見られるが、単語数がある程度以上になると、逆に性能が劣化するという結果が得られた。原因として、適応単語数が多い場合は回帰クラスが比較的頑健に求まるので、他の回帰行列を加えることが逆効果になるためであると考えられる。このことを考慮し、以下の条件を加えた追加実験を行った。

- クラスの境界付近のモデルのみ提案法を用いる。
- 線形和の項として全ての回帰行列を用いるのではなく、適切な数の行列のみを用いるようにする。

次頁より、それぞれの追加実験について述べる。

5.1 閾値を考慮した回帰行列の線形和による認識結果

3.1 節で述べたように、クラスの中心付近と周辺付近のモデルとで同一の回帰行列を用いていることが従来法での問題点として考えられる。そこで、モデルとクラスのセントロイド間の距離を考慮し、ある程度クラスの中心から離れたモデルのみに対して提案法を適応することを考えた。本研究では、モデルが属するクラスより、他のクラスとの距離の方が小さいもの、という基準を用いて判断した。

この基準を元に、以下の条件で再度話者適応実験を行った。

- proposal1-2：クラスの中心から離れているモデルのみ提案法を適用。他のモデルは従来法に基づく。
- proposal1-3：proposal1-2 とは逆に、クラスの中心付近のモデルのみ提案法を適用。他のモデルは従来法に基づく。この条件は比較のために用いる。

以下の図 5.1 で例を示す。モデルが属しているクラスとの距離 (d_1) より、他のクラスとの距離 (d_2) の方が小さいモデルに提案法を用いる。

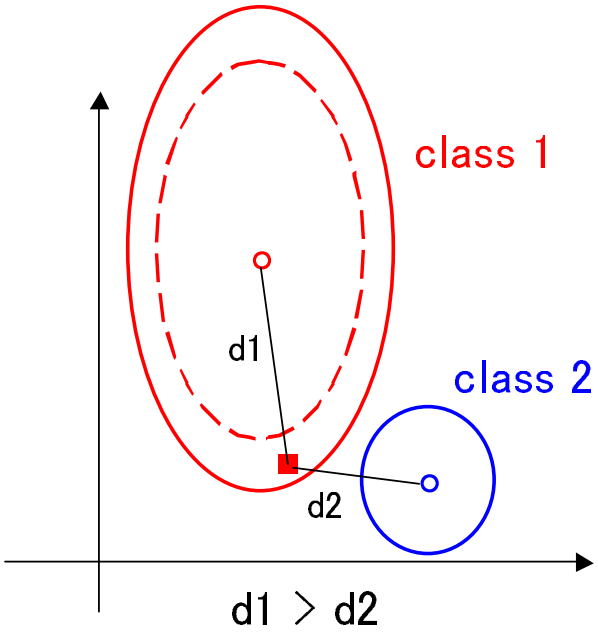

図 5.1: proposal1-2 が適応される例

他の実験条件は 4.1 節と同様。実験結果を次頁に示す。

表 5.1 は適応を行った 5 話者の単語認識率の平均値である。横軸が適応単語数、縦軸が回帰クラス数、枠内の数字は上から順に従来法 (MLLR)、提案法を全てのモデル (prop1)、クラスから離れたモデル (prop1-2)、クラスの中心付近のモデルにそれぞれ適用した場合 (prop1-3) の結果である。

class | word		10	15	20	30	40	60	80	100	200
2	MLLR	85.50	89.01	88.09	88.09	88.40	88.55	88.24	88.70	87.94
	prop1	85.80	89.01	88.24	88.55	88.40	88.86	88.09	88.24	87.48
	prop1-2	85.65	89.16	88.09	88.24	88.55	88.70	88.09	88.55	87.64
	prop1-3	85.80	88.86	88.09	88.40	88.24	88.55	88.24	88.55	87.63
3	MLLR	85.50	88.85	87.18	89.31	88.86	89.47	89.62	89.01	89.16
	prop1	85.80	89.31	87.79	88.70	88.09	88.70	88.24	88.40	88.40
	prop1-2	85.65	89.31	88.25	89.31	89.77	89.92	88.86	89.31	89.92
	prop1-3	85.80	88.86	85.80	88.86	87.94	88.40	88.40	88.55	88.40
5	MLLR	85.50	88.85	86.72	88.70	90.84	89.92	89.77	89.93	89.77
	prop1	85.80	89.31	87.33	89.01	89.47	89.31	88.85	89.62	89.62
	prop1-2	85.65	89.31	88.09	89.01	89.47	89.92	89.77	89.62	90.07
	prop1-3	85.80	88.86	85.65	89.01	89.16	89.31	88.85	89.16	89.47
8	MLLR	85.50	88.85	86.72	88.70	90.84	89.77	90.53	89.01	89.92
	prop1	85.80	89.31	87.33	89.01	89.47	89.31	89.62	88.86	88.85
	prop1-2	85.65	89.31	87.94	89.01	89.47	90.38	89.77	89.01	89.16
	prop1-3	85.80	88.86	85.65	89.01	89.16	89.92	89.77	89.16	89.01
10	MLLR	85.50	88.85	86.72	88.70	90.69	89.16	90.23	88.85	90.08
	prop1	85.80	89.31	87.33	89.01	89.16	89.01	89.31	88.09	89.01
	prop1-2	85.65	89.31	87.94	89.01	89.62	89.31	88.85	88.24	89.16
	prop1-3	85.80	88.86	85.65	89.01	89.47	89.01	89.77	88.86	88.70

表 5.1: 提案法を適用するモデルの範囲を変更した結果

表 5.2、5.3 は表 5.1 における各回帰クラス数における結果の平均値と最大値を示したものである。

ave.	10	15	20	30	40	60	80	100	200
MLLR	85.50	88.88	87.08	88.70	89.92	89.37	89.68	89.10	89.37
prop1	85.80	89.25	87.60	88.85	88.92	89.04	88.82	88.64	88.67
prop1-2	85.65	89.28	88.06	88.92	89.37	89.65	89.07	88.95	89.19
prop1-3	85.80	88.86	86.17	88.86	88.80	89.04	89.01	88.86	88.64

表 5.2: 提案法を適用するモデルの範囲を変更した結果の平均値

max.	10	15	20	30	40	60	80	100	200
MLLR	85.50	89.01	88.09	89.31	90.84	89.92	90.53	89.93	90.08
prop1	85.80	89.31	88.24	89.01	89.47	89.31	89.62	89.62	89.62
prop1-2	85.65	89.31	88.25	89.31	89.77	90.38	89.77	89.62	90.07
prop1-3	85.80	88.86	88.09	89.01	89.47	89.92	89.77	89.16	89.47

表 5.3: 提案法を適用するモデルの範囲を変更した結果の最大値

認識率の平均値、最大値ともにクラスから離れたモデルのみ提案法を用いたもの (prop 1-2) が、適応単語数が 10 である場合を除き他の 2 つに比べ性能が勝ると言う結果が得られた。この結果より、全体的な認識性能は prop1-2 の場合が一番高いと考えられる。以降の節ではこの方法をベースに検討を行う。

5.2 線形和の項数選択と認識結果

行列の線形和を計算する際、全てのクラスの回帰行列を用いるよりも, モデルに応じて適切な行列数を選択する方が良いと考えられる。本研究では、モデルからある程度離れたクラスの行列は用いないことにした。その基準として、モデルとクラスの中心までの距離を用いた。回帰クラス C_i に属するモデルに対し、C_i の中心までの距離を d_i とする。このモデルと他の回帰クラス C_j の中心までの距離 d_j が、

$$d_j \geq d_i$$

となるクラスの行列 W_j を線形和の項として用いた。図 5.2 の例では、class1 内のあるモデルに対し class2 の行列は用いるが、class3 の行列は用いないということを示している。

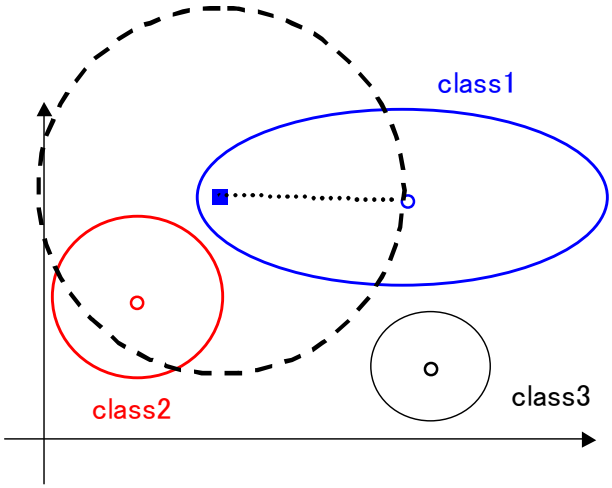

図 5.2: 行列の選択基準

次頁で線形和の項数選択を行った場合の認識結果について述べる。

表 5.4 は適応を行った 5 話者の単語認識率の平均値である。横軸が適応単語数、縦軸が回帰クラス数、枠内の数字は上から順に従来法 (MLLR)、線形和に用いる行列を距離基準で選択した場合の結果 (prop4) である。

		10	20	30	40	60	80	100	200
2	MLLR	85.50	88.09	88.09	88.40	88.55	88.24	88.70	87.94
	prop4	85.96	88.24	88.40	88.40	88.55	88.40	88.85	87.63
3	MLLR	85.50	87.18	89.31	88.86	89.47	89.62	89.01	89.16
	prop4	85.96	87.33	89.31	89.92	89.46	89.31	88.86	89.16
5	MLLR	85.50	86.72	88.70	90.84	89.92	89.77	89.93	89.77
	prop4	85.96	87.31	89.31	90.75	90.23	90.08	89.92	89.62
8	MLLR	85.50	86.72	88.70	90.84	89.77	90.53	89.01	89.92
	prop4	85.96	87.31	89.31	90.69	90.08	90.69	89.16	90.38
10	MLLR	85.50	86.72	88.70	90.69	89.16	90.23	88.85	90.08
	prop4	85.96	87.33	89.31	90.69	89.31	89.62	89.31	90.08

表 5.4: 線形和の項数選択を行った場合の認識結果

表 5.5、5.6 は表 5.4 における各回帰クラス数における結果の平均値と最大値を示したものである。

ave.	10	20	30	40	60	80	100	200
MLLR	85.50	87.08	88.70	89.92	89.37	89.68	89.10	89.37
prop4	85.96	87.51	89.13	90.09	89.53	89.62	89.22	89.37

表 5.5: 線形和の項数選択を行った認識の各クラス数での平均値

max.	10	20	30	40	60	80	100	200
MLLR	85.50	88.09	89.31	90.84	89.92	90.53	89.93	90.08
prop4	85.96	88.24	89.31	90.75	90.23	90.69	89.92	90.38

表 5.6: 線形和の項数選択を行った認識の各クラス数での最大値

単語数が多い場合にも、従来法と同程度の認識率が得られるようになった。

第6章　結論

6.1　結論

本研究で提案した回帰行列の重み付き線形和を用いた適応法は、認識率の上昇は僅かではあったが、MLLR の回帰クラス周辺のモデルにおける適応に改善の余地があるという可能性を示すものとなった。また重み係数、提案法を適用するモデルの選択や線形和に用いる行列の個数を適切に決定することにより、認識率の向上も期待できると考えられる。

6.2　今後の課題

MLLR の回帰クラス数に加え、線形和をとる行列数という 2 つのパラメータの決定法の指針を検討する必要がある。また、話者ごとにモデルの分布等の細かい分析を行うことにより、認識率が上昇した話者とそうでない話者との相違点を見つけることができれば、認識性能の向上につながる可能性もある。

また、本研究では重み係数を決定する関数をあらかじめ決めておく手法をとったが、重み係数も適応データから学習する方法も考えられる。例として、

- 回帰行列を求めた後、もう一度 EM アルゴリズムを用いて重み係数を求める。
- 適応データがあるクラスに属する確率を決め、それを用いて学習を行う。

などの方法が考えられる。

謝辞

本研究を進めるにあたって、全般的な御指導を頂いた党建武教授に心から感謝致します。また、中井満助手には、本研究に対する専門的かつ適切な御意見を頂き誠に感謝致しております。

さらに、本研究室のメンバーならびに OB の皆様には、公私にわたり大変お世話になりました。誠に簡単ではありますが、ここに厚くお礼申し上げると共に論文の結びにしたいと思います。

参考文献

[1] 中川聖一, 確率モデルによる音声認識, 電子情報通信学会, 1998

[2] 鹿野清宏, 伊藤克亘 著, 河原達也, 武田一哉, 山本幹雄 編著, 音声認識システム, オーム社, 2001

[3] L.R.Labiner,B.H Juang 著, 古井貞熙 監訳, 音声認識の基礎 (上)(下), NTT アドバンストテクノロジ, 1995

[4] 徳田恵一, 隠れマルコフモデルによる音声認識と音声合成, IPSJ Magazine Vol.45 No.10, pp.1005-1011, 2004.10

[5] J.-L. Gauvain and C.H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," IEEE Trans. on Speech and Audio Processing, vol. 2, no. 2, pp. 291–298, Apr. 1994.

[6] C. J. Legatter and P. C. Woodland, Maximum lilelihood liner regression for speech adaptation of continuous-density hidden Marcov models, Computer Speech and language, vol.9, pp.171-185, 1995

[7] 篠田浩一, 確率モデルによる音声認識のための話者適応技術, 電子情報通信学会論文誌 Vol.J87-D-II, No.2, pp.371-386, 2004

[8] 安藤彰男 著, リアルタイム音声認識, 電子情報通信学会, 2003

[9] 篠田浩一, 篠崎隆宏, 統計的手法を用いた音声モデリングの高度化とその音声認識への応用, IPSJ Magazine Vol.45 No.10, pp.1012-1019, 2004.10

[10] 日本数学会, 岩波数学辞典 第３版, 岩波書店