TR-IT-0106

混合連続分布型HMMの話者適応における モデルパラメータの学習に関する一考察
Analysis of model-parameter training for speaker adaptation

山田 武志
Yamada Takeshi

外村 政啓
Tonomura Masahiro

1995.3.31

混合連続分布型HMMの話者適応において、学習を少量サンプルで行なうためには移動ベクトル場平滑化法（VFS）を用いるのが非常に有効である。しかしながら、VFSでは分布間の距離が近いものは類似した移動ベクトルを持つと仮定しており、その妥当性については議論の余地が残されている。本実習の目的は、話者適応におけるモデルパラメータの学習について分析することにより、VFSの精度をより高めるための手がかりを探すことである。

実習報告書
「混合連続分布型 HMM の話者適応における モデルパラメータの学習に関する一考察」

山田 武志

ATR 音声翻訳通信研究所第一研究室
xyamada@itl.atr.co.jp
奈良先端科学技術大学院大学情報科学研究科
takesi-y@is.aist-nara.ac.jp

1995 年 3 月 31 日

1 はじめに

混合連続分布型 HMM の話者適応において、学習を少量サンプルで行なうためには移動ベクトル場平滑化法（VFS）を用いるのが非常に有効である。しかしながら、 VFS では分布間の距離が近いものは類似した移動ベクトルを持つと仮定しており、その妥当性については議論の余地が残されている。例えば、分布間の距離が近くても大きく異なった移動ベクトルを持つ場合や、逆に分布間の距離が非常に離れていても類似した移動ベクトルを持つ場合等を考えることができ、モデルパラメータの学習に悪影響を与えている可能性があるからである。

本実習の目的は、話者適応におけるモデルパラメータの学習について分析することにより、VFS の精度をより高めるための手がかりを探すことである。

2 準備

本実習で用いた音声認識システムの仕様を表 1に示す。モデルパラメータの学習について分析することが目的なので、 VFS は使用していない。

表 1: 音声認識システムの仕様

サンプリング周波数	12 [kHz]
分析窓	20 [ms] ハミング窓
フレームシフト	5 [ms]
パラメータ	16 次 LPC ケプストラム + log パワー + 16 次 Δ ケプストラム + Δ log パワー
認識方式	5 混合対角共分散連続分布型 HMM
モデル数	状態数 200 の HMnet による環境依存 479 モデル
学習データ	男性 146 名 + 女性 139 名（各話者 50 文章）
評価データ	適応話者の SB3 タスク（279 文節）
適応方式	ML 推定
適応パラメータ	平均値ベクトルと分散、又は平均値ベクトルのみ
適応話者	男性 4 名（MAU, MMY, MSH, MTM） + 女性 3 名（FAF, FMS, FYM） SB1 タスク

学習サンプル数（文節数） 10, 20, 40, 80, 160, 320, 598 の 7 通りで話者適応を行ない、話者適応前後のモデルパラメータを用いて分析する。

参考までに、平均値ベクトルと分散を学習した場合と平均値ベクトルのみ学習した場合の音素認識率を、適応話者 MSH について図 1に示す。ただし、学習サンプル数 0 は話者適応なしを意味する。学習サンプル数が十分大きいときには平均値ベクトルと分散を学習した方が認識率が高くなり、逆に学習サンプル数が少ないときには平均値ベクトルのみ学習した方が認識率が高くなった。これは、分散を学習するためには十分大きな学習サンプル数が必要であることを意味している。図 2と図 3から、他の適応話者についても同様の傾向があることが分かる。図 2は平均値ベクトルと分散を学習した場合の音素認識率を各適応話者について示したものであり、図 3は平均値ベクトルのみ学習した場合の音素認識率を各適応話者について示したものである。

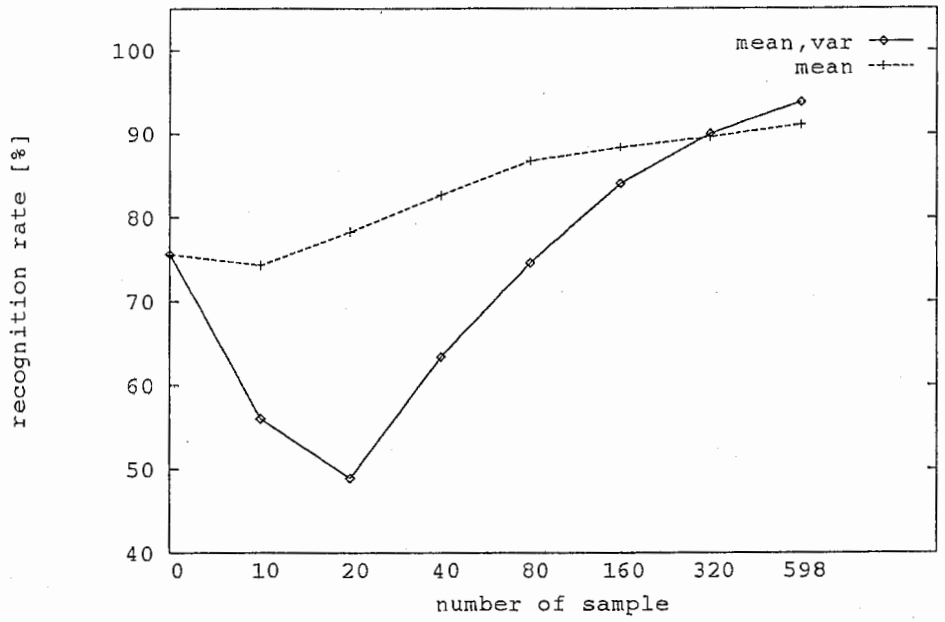

図 1: 話者適応後の音素認識率（MSH）

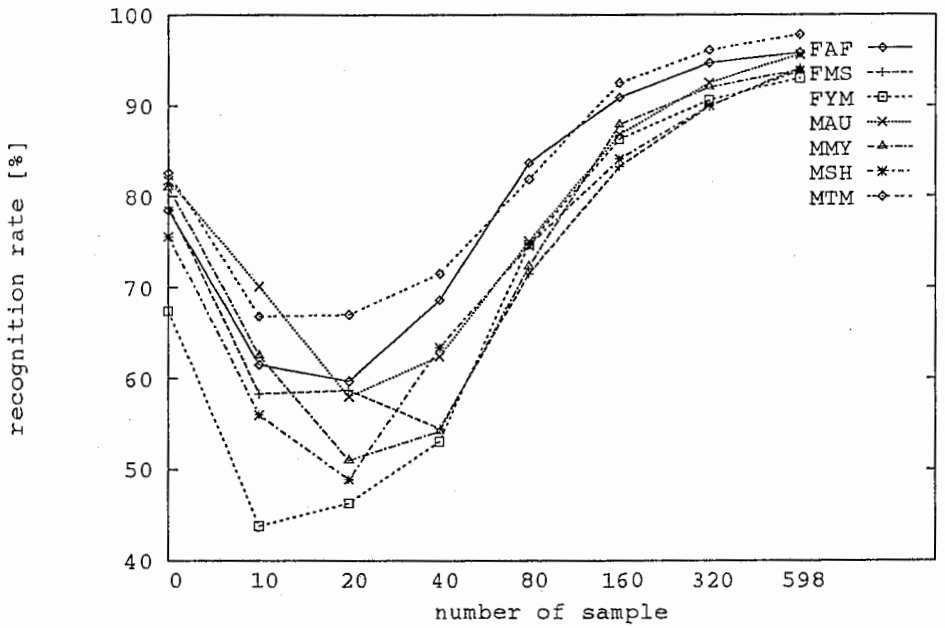

図 2: 話者適応後の音素認識率（平均値ベクトルと分散を学習した場合）

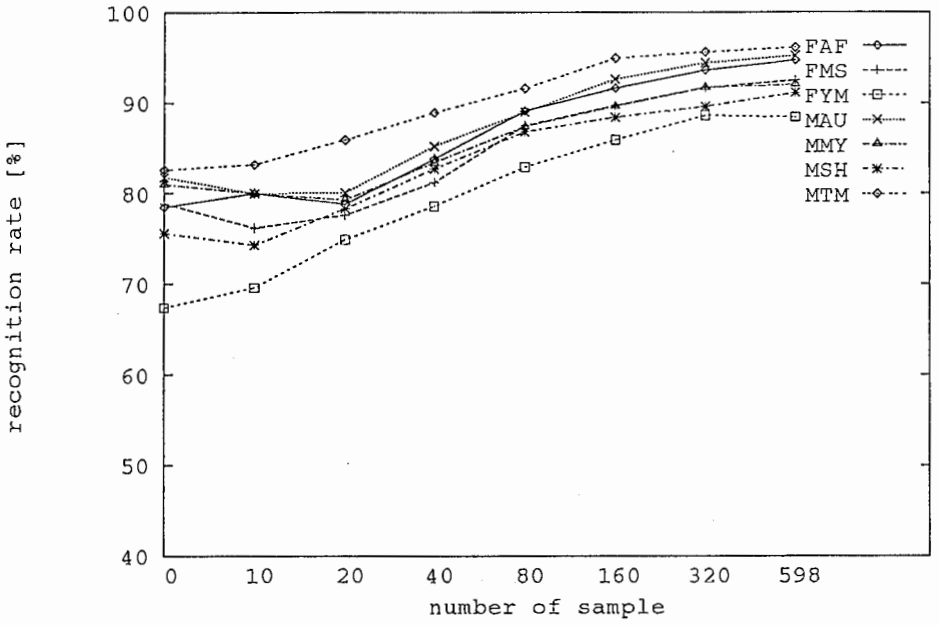

図 3: 話者適応後の音素認識率（平均値ベクトルのみを学習した場合）

3 分析

話者適応前後の平均値ベクトルμ, $\hat{\mu}$と対角共分散行列Σ, $\widehat{\Sigma}$から、平均値ベクトルの移動ベクトル $A_\mu = \hat{\mu} - \mu$ と分散の変換ベクトル $A_\Sigma = \widehat{\Sigma}\Sigma^{-1}$ （A_Σは対角要素しか持たないのでベクトルとして扱う）を求めることができる。また、A_μとA_Σの大きさを次式で定義する。

$$|A_\mu| = \sum_{n=1}^{N} a_{\mu,n}^2 \quad (\geq 0)$$

$$|A_\Sigma| = \frac{1}{N}\sum_{n=1}^{N} a_{\Sigma,n}^2 \quad (\geq 1)$$

ここで、$a_{\mu,n}, a_{\Sigma,n}$ は、各々A_μとA_Σの第n要素である。

3.1 移動ベクトルの大きさ

移動ベクトルの大きさという観点から分析を行なった。移動ベクトルの大きさと変換ベクトルの大きさを学習サンプル数別に話者 MAU について各々図4から図7に示す。横軸は各ベクトルの番号であり、縦軸は各ベクトルの大きさである。移動ベクトルの大きさは、値が大きくなるほど移動の度合が大きいことを表しており、全く移動しないとき0である。変換ベクトルの大きさは、全く移動しないとき1であり、値が0に近付くほど分散が縮小することを意味し、値が1より大きくなるほど分散が拡大することを意味する。

図4から図7より、学習サンプル数が増加するにつれて、移動する分布数も増加することが分かる。しかしながら、学習サンプル数が596のときにも依然として移動していない分布が多く

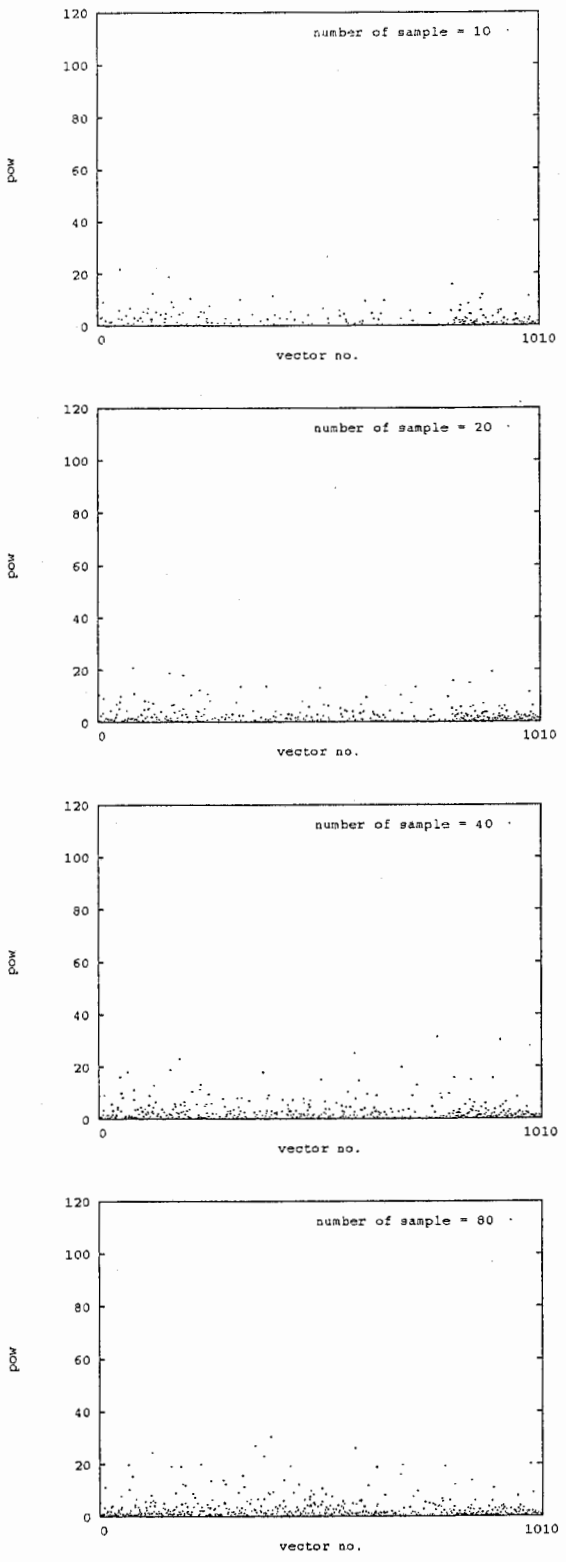

図 4: 移動ベクトルの大きさ（学習サンプル数 10, 20, 40, 80）

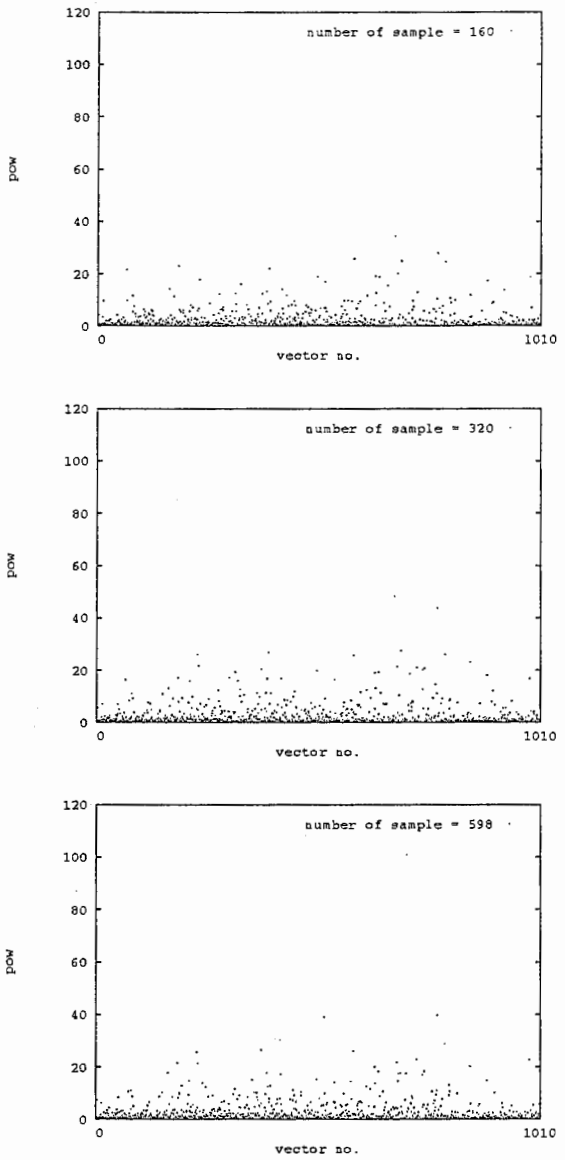

図 5: 移動ベクトルの大きさ（学習サンプル数 160, 320, 598）

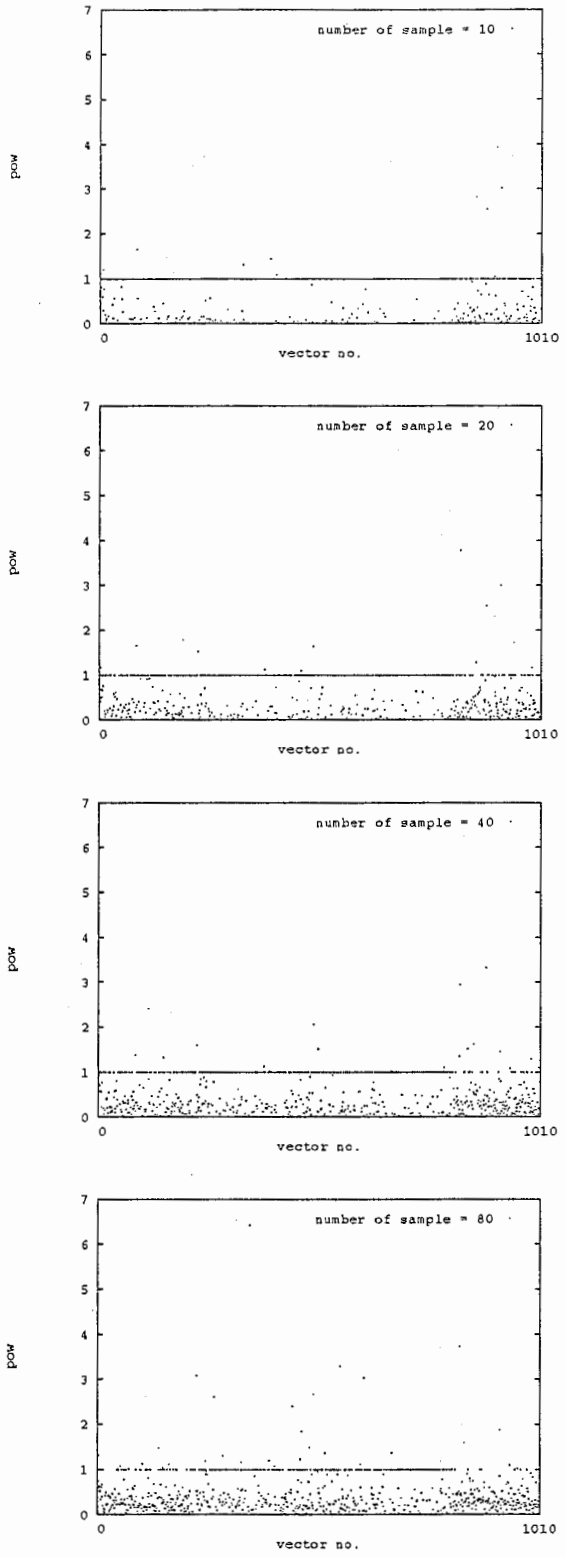

図 6: 変換ベクトルの大きさ（学習サンプル数 10, 20, 40, 80）

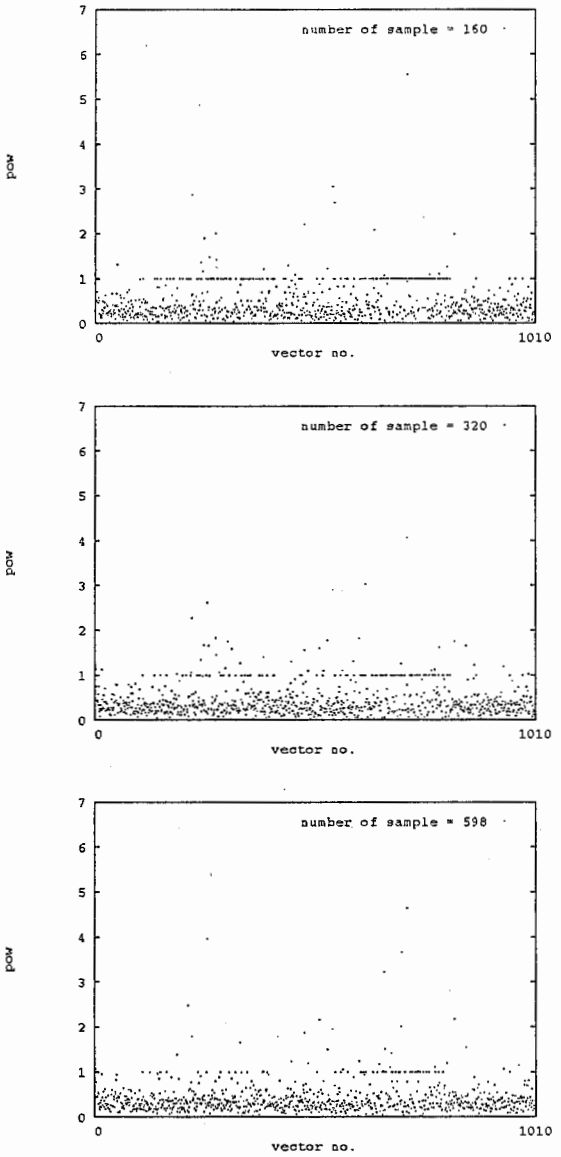

図 7: 変換ベクトルの大きさ（学習サンプル数 160, 320, 598）

存在するが、これは1分布に対して1フレーム以下の学習サンプルしか存在しないときには、話者適応プログラムが分布を移動させないからであると思われる。あるいは、移動しない分布が表す話者空間が適応話者の話者空間と類似している、すなわち学習はしているが分布が移動していないと考えることもできるが、不特定話者を特定話者に適応しようとしている以上、前者の理由の方が大きいと考えられる。

先に述べたように、VFSでは分布間の距離が近いものは類似した移動ベクトルを持つと仮定している。しかしながら、学習サンプル数が10のときに注目すると、移動ベクトルの大きさにかなりの幅があることから、VFSを行なったとき全く異なる移動ベクトルに置き換えられる可能性があると思われる。

3.2 移動ベクトルの類似性

移動ベクトルの類似性という観点から分析を行なった。特定の音素モデル間に移動ベクトルの類似性が存在するなら、それを考慮することによりVFSの精度を高くすることができると考えられるからである。以下、分析手順について述べる。

1. 移動ベクトル間の距離を評価尺度として、次のアルゴリズムでクラスタリングを行なう。
   (a) クラスタ中心として任意の2つの移動ベクトル$\boldsymbol{A}_i, \boldsymbol{A}_j$を選択し、全ての移動ベクトルをクラスタ$C_1, C_2$のうち距離が近い方に分類する。さらに、クラスタ$C_1, C_2$について、クラスタ中心とその他のメンバとの距離の総和S_1, S_2を計算し、$S_{all} = S_1 + S_2$とする。ただし、$S_1 = \sum_{k \in C_1} |\boldsymbol{A}_k - \boldsymbol{A}_i|^2, S_2 = \sum_{k \in C_2} |\boldsymbol{A}_k - \boldsymbol{A}_j|^2$である。
   (b) (a)を全ての$\boldsymbol{A}_i$, $\boldsymbol{A}_j$の組合せに対して計算し、S_{all}を最小とする$\boldsymbol{A}_i$, $\boldsymbol{A}_j$をクラスタC_1, C_2のクラスタ中心とする。
   (c) $S_1 \geq S_2$ならクラスタC_1を、$S_1 \leq S_2$ならクラスタC_2を分割対象とし、(a), (b)と同様の方法で2つのクラスタに分割する。これにより、3つのクラスタC_1, C_2, C_3に分割される。
   (d) クラスタC_1, C_2, C_3の各クラスタ中心により全ての移動ベクトルを分類し直し、その結果に基づいて各クラスタのクラスタ中心を選び直す。さらに、$S_{all} = S_1 + S_2 + S_3$を計算する。
   (e) S_{all}が前回の値と変化しなくなるまで(d)を繰り返す。
   (f) S_1, S_2, S_3のうち値が最大のクラスタを(a), (b)と同様の方法で2つのクラスタに分割し、(d), (e)と同様の処理を行なう。以上を、所望のクラスタ数になるまで繰り返す。
2. 同一のクラスタに分類された移動ベクトルの番号から、HMnetにおける状態番号とその状態が表す音素モデルを調べる。

以上の方法で分析を行なったが、移動ベクトルの類似性を見い出すことはできなかった。しかしながら、これは移動ベクトルの類似性が存在しないことを意味しているのはなく、むしろ次のような分析方法に起因する問題があると考えられる。

- 5混合連続分布型HMMを用いたために、対象となる分布数が増大し、かつ分析が複雑となった。

- クラスタ中心が話者間で異なるために、話者間での比較が困難となった。

上述した問題をふまえて、今後の分析方針として以下のものが望ましいと考えられる。

- 混合数が1のHMMを用いる。
- クラスタ中心をあらかじめ決めておく。例えば、全ての音素モデルから開始状態、遷移途中状態、終了状態を表す分布を一つずつ選び、その分布に対する移動ベクトルをクラスタ中心とすることが考えられる。

4 実験

3章では、VFSの精度を高めるために必要である何らかの情報を探求するという立場で、モデルパラメータの学習について分析した。本章では逆の立場から、まずモデルパラメータの学習にある特徴を仮定し、それを考慮したVFSを用いて実際に話者適応を行なうことにより、その妥当性を検証する。

ここでは、同一中心音素を持つ分布同士は非常に類似した移動ベクトルを持つという仮定をVFSに導入し、このVFSを用いて話者適応を行なった。ただし、VFSを行なうとき、同一中心音素を持つ分布が全て移動していない場合には、

1. VFSを行なわない（改良型VFS1）
2. 通常のVFSを行なう（改良型VFS2）

という2通りの方法を試みた。話者適応法を表2に示す。

表2: 話者適応法

適応方式	ML推定＋改良型VFS1、又は＋改良型VFS2
適応パラメータ	平均値ベクトルのみ
適応話者	男性4名（MAU, MMY, MSH, MTM） ＋女性3名（FAF, FMS, FYM） SB1タスク（文節数1, 3, 5, 7, 20, 47, 97 ,256）

改良型VFSを用いたときの音素認識率と通常のVFSを使用したときの音素認識率を、適応話者MAUについて図8に示す。

図8から、十分大きな学習サンプル数が与えられたときには改良型VFSを用いた方が音素認識率が高いことが分かる。従って、同一中心音素を持つ分布同士は非常に類似した移動ベクトルを持つという仮定が妥当であると考えることができる。また、学習サンプル数が少ないときの音素認識率が低いのは、同一中心音素を持つ分布がほとんど移動していないためにVFSの対象となる分布の範囲が非常に狭いからである。この問題に対しては、例えば種類（濁音、促音等）が同じ中心音素を持つ分布同士は非常に類似した移動ベクトルを持つという仮定を新たに導入して、VFSの対象となる分布の範囲を広げることが有効であると考えられる。最終的には、3章における分析を詳細に行なうことにより、VFSの精度をより一層高めることができると思われる。

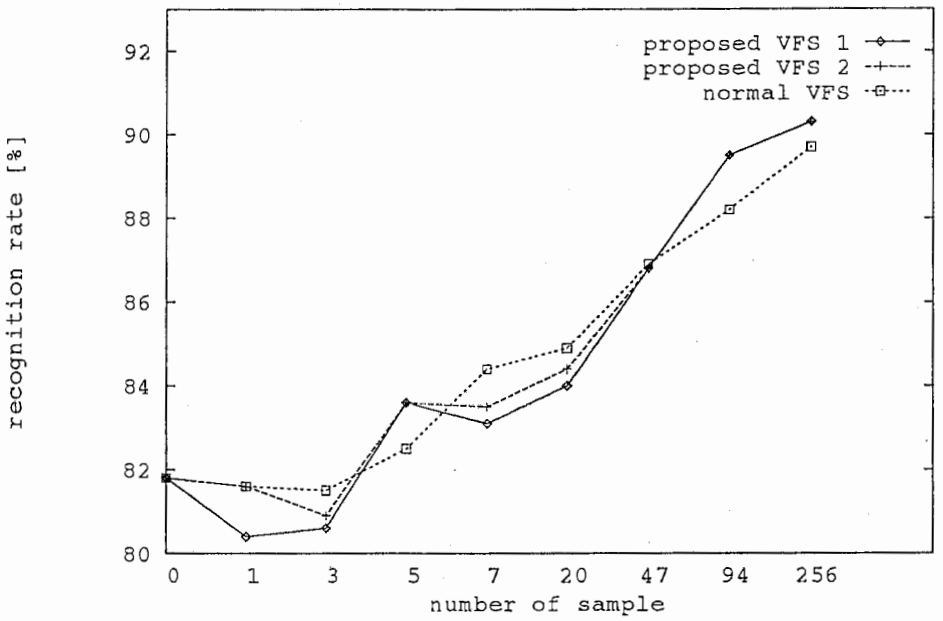

図 8: 話者適応後の音素認識率（MAU）

5 感想

本実習で行なったような分析は、様々な研究分野において必要性が高いと思われるにも関わらず、今まで取り組んだ経験が皆無だった。この経験を今後の研究に生かしたいと考えている。

6 謝辞

快適な実習環境を提供して頂いたATR音声翻訳通信研究所第一研究室の研究員諸氏に感謝の意を表します。また、本実習の機会を与えて下さったATR音声翻訳通信研究所の山崎泰弘社長、並びにATR音声翻訳通信研究所第一研究室の匂坂芳典室長に深く感謝致します。

参考文献

[1] ATR編, " 自動翻訳電話", オーム社, (1994)

[2] 中村 哲, " 音声認識における話者適応", 信学技法, SP94-3, (1994-05)

付録

A　プログラム解説

extract_var

- 機能：SSS-ToolKit で作成した HMnet ファイルから共分散行列の対角要素ベクトルを抽出する。
- 使用法：extract_var [-io]
  - -i 入力ファイル名（HMnet ファイル名）
  - -o 出力ファイル名（デフォルトは標準出力）

extract_mean

- 機能：SSS-ToolKit で作成した HMnet ファイルから平均値ベクトルを抽出する。
- 使用法：extract_mean [-io]
  - -i 入力ファイル名（HMnet ファイル名）
  - -o 出力ファイル名（デフォルトは標準出力）

analysis_var2

- 機能：混合連続HMMの多次元ガウス分布において、適応前の共分散行列と適応後の共分散行列から変換行列を求め、各変換行列の対角要素ベクトルによりクラスタリングする。
- 備考：クラスタリングに k-means アルゴリズムの改良版を使用している。
- 使用法：analysis_var2 [-n12o]
  - -n クラスタリング数
  - -1 入力ファイル名（適応前の分散ファイル名）
  - -2 入力ファイル名（適応後の分散ファイル名）
  - -o 出力ファイル名（デフォルトは標準出力）

analysis_mean2

- 機能：混合連続HMMの多次元ガウス分布において、適応前の平均値ベクトルと適応後の平均値ベクトルから移動ベクトルを求め、クラスタリングする。
- 備考：クラスタリングに k-means アルゴリズムの改良版を使用している。
- 使用法：analysis_mean2 [-n12o]
  - -n クラスタリング数

- -1 入力ファイル名（適応前の平均値ベクトルファイル名）
- -2 入力ファイル名（適応後の平均値ベクトルファイル名）
- -o 出力ファイル名（デフォルトは標準出力）

calc_dist_var

- 機能：混合連続HMMの多次元ガウス分布において、適応前の共分散行列と適応後の共分散行列から変換行列を求め、変換行列の対角要素ベクトルの正規化した大きさとその平均と分散を計算する。
- 使用法：calc_dist_var [-12o]
  - -1 入力ファイル名（適応前の分散ファイル名）
  - -2 入力ファイル名（適応後の分散ファイル名）
  - -o 出力ファイル名（デフォルトは標準出力）

calc_dist_mean

- 機能：混合連続HMMの多次元ガウス分布において、適応前の平均値ベクトルと適応後の平均値ベクトルから移動ベクトルを求め、移動ベクトルの大きさとその平均と分散を計算する。
- 使用法：calc_dist_mean [-12o]
  - -1 入力ファイル名（適応前の平均値ベクトルファイル名）
  - -2 入力ファイル名（適応後の平均値ベクトルファイル名）
  - -o 出力ファイル名（デフォルトは標準出力）

analysis_phone

- 機能：analysis_var2, analysis_mean2 により作成したクラスタリングファイルにおいて、分布番号を状態番号または音素名に変換する。
- 使用法：analysis_phone [-a12o]
  - -a 状態番号、先行・中心・後続音素名を出力（省略すれば中心音素名のみ）
  - -1 入力ファイル名（HMnet ファイル名）
  - -2 入力ファイル名（クラスタリングファイル名）
  - -o 出力ファイル名（デフォルトは標準出力）

compare_vector_var

- 機能：混合連続HMMの多次元ガウス分布において、適応前の共分散行列と適応後の共分散行列から変換行列を求め、各変換行列の対角要素ベクトル間の距離を求める。
- 使用法：compare_vector_var [-i1234567o]

- -i 入力ファイル名（適応前の分散ファイル名）
- -1 ･･･ -7 入力ファイル名（適応後の分散ファイル名）
- -o 出力ファイル名（デフォルトは標準出力）

compare_vector_mean

- 機能：混合連続HMMの多次元ガウス分布において、適応前の平均値ベクトルと適応後の平均値ベクトルから移動ベクトルを求め、移動ベクトル間の距離を求める。
- 使用法：compare_vector_mean [-i1234567o]
  - -i 入力ファイル名（適応前の平均値ベクトルファイル名）
  - -1 ･･･ -7 入力ファイル名（適応後の平均値ベクトルファイル名）
  - -o 出力ファイル名（デフォルトは標準出力）

付録

B　ディレクトリ構造とファイルの解説

bin　分析のために作成したプログラムを格納

doc　実習報告書と図版関係のファイルを格納

- report.tex ：　実習報告書の LaTeX ファイル
- FIGURE ：図版関係のファイルを格納
  - dist_mean ：　移動ベクトルの大きさに関連するファイルを格納
  - dist_var ：　変換ベクトルの大きさに関連するファイルを格納
  - recrate ：　音素認識率に関連するファイルを格納

lib　汎用Cライブラリのソースファイルとオブジェクトファイルを格納

src　分析のために作成したプログラムのソースファイルと分析データを格納

- EX_01 ：　話者適応結果と分析データを格納
  - MAP ：MAP 推定による話者適応結果を格納
  - ML ：ML 推定による話者適応結果と分析結果を格納
    - HMnet ：　話者適応前の HMnet ファイル
    - HMnet.mean ：　extract_mean で作成したファイル
    - HMnet.var ：　extract_var で作成したファイル
    - ???/DIST ：移動ベクトルの大きさを格納
    - ???/CLUST100 ：　クラスタリング結果を格納
  - ML2 ：　ML 推定による話者適応結果を格納
  - ML3 ：　ML 推定 + 改良型 VFS1 による話者適応結果を格納
  - ML4 ：　ML 推定 + 改良型 VFS2 による話者適応結果を格納
  - ML5 ：　ML 推定による話者適応結果と分析結果を格納
    - HMnet_male ：　話者適応前の HMnet ファイル
    - HMnet_male.mean ：　extract_mean で作成したファイル
    - HMnet_male.var ：　extract_var で作成したファイル
    - HMnet_female ：　話者適応前の HMnet ファイル
    - HMnet_female.mean ：　extract_mean で作成したファイル
    - HMnet_female.var ：　extract_var で作成したファイル
    - ???/DIST ：移動ベクトルの大きさを格納
    - ???/CLUST100 ：　クラスタリング結果を格納

- SSS SSS-ToolKit 関連のファイルを格納
  - forward_backward.c を一部修正
  - 改良 VFS1 (mean_smoothing.c、Exe.adapt_HMnet)
  - 改良 VFS2 (mean_smoothing2.c、Exe.adapt_HMnet2)