

TR-I-0099

話者重畳型HMMを用いた調音様式の話者適応化

Speaker Adaptation of Articulatory Variation  
by Supplemented HMM

中村 哲, 服部浩明, 鹿野清宏

Satoshi NAKAMURA, Hiroaki HATTORI, Kiyohiro SHIKANO

1989. 8

### 概要

HMMにおける話者適応化を改善するために複数話者HMM学習を用いた話者適応化方法を提案する。話者間における発声の違いとしては、(1)静的なスペクトルの違い(2)調音様式に起因する時系列パターンの違いがある。本稿では、(2)の時系列パターンの違いを適応化するために、複数話者HMM学習による話者重畳型HMMを用いた話者適応化法を提案し有効性を示す。

ATR 自動翻訳電話研究所

ATR Interpreting Telephony Research Laboratories

© (株)ATR 自動翻訳電話研究所 1989

© 1989 by ATR Interpreting Telephony Research Laboratories

## 目次

1.	まえがき .....	1
2.	複数話者HMM学習を用いた話者適応化 .....	1
3.	実験 .....	2
	3.1 実験条件 .....	2
	3.2 スペクトル写像の精度の評価 .....	2
	3.3 話者重畳型HMM-話者適応化 .....	2
	3.4 マルチテンプレート型HMM-話者適応化 .....	2
4.	まとめ .....	3
5.	謝辞 .....	3
6.	文献 .....	3

## 1. まえがき

本稿では、HMMにおける話者適応化を改善するために複数話者HMM学習を用いた話者適応化方法を提案する。話者間における発声の違いとしては、(1)静的なスペクトルの違い(2)調音様式に起因する時系列パターンの違いがある。現在までに我々は、ベクトル量子化話者適応化法を用いて、(1)静的なスペクトルの違いについての適応化を行い有効性を示した<sup>(1,2,3,4)</sup>。本稿では、(2)の時系列パターンの違いを適応化するために、複数話者HMM学習を用いた話者適応化法を提案し有効性を示す。

## 2. 複数話者HMM学習を用いた話者適応化

異話者間の発声の時系列パターンの違いには、方言、話者固有の言い回し、調音様式の違いなどが存在するが、本稿ではこのうち音韻の調音様式に起因する時系列パターンの違いの適応化を試みる。時系列パターンの適応化としては、フレーム毎の静的なスペクトル写像の延長として時系列パターンの写像を行う方法があるが、時間長正規化や写像関数の学習に大量の学習データを必要とするという問題がある。一方、時系列パターンの違いによる誤認識は、未知話者の音声フレーム毎に標準話者の音声の系列に写像されても、標準パターンにその系列が含まれておらずに棄却されてしまうことが原因と考えられる。そこで、残された方法としては、時系列パターンをマルチテンプレート化し入力話者の時系列パターンの変動を吸収するという方法が考えられる<sup>(1,5,6)</sup>。DTWにおける時系列パターンのマルチテンプレート化としては、文献(1,6)に挙げる方法が提案されている。本稿では、基準標準話者に加えてさらに複数の標準話者を用意し、基準話者のスペクトルでありながら複数の話者から得られた種々の時系列パターンを1つのモデルの中に含むHMM(話者重畳型HMMと呼ぶ)を生成する。この方法では、マルチテンプレートとして明示的に標準パターンのテンプレート数を増やさずに種々の時系列パターンに対処することができる。次に話者重畳型HMMを推定するアルゴリズムを示す。

Step 1. 基準標準話者に加えて複数の標準話者の学習データを用意する。

Step 2. 複数の標準話者から基準標準話者へのスペクトル写像関数をベクトル量子化話者適応アルゴリズムにより求める。

Step 3. 全ての標準話者の学習データを基準標準話者のスペクトル空間に写像する。

Step 4. 全ての写像された学習データを用いて音韻毎のHMMの学習を行う。未知話者の入力音声の認識は、予め求められた未知話者から基準標準話者へのスペクトル写像を用いて基準標準話者の時系列に写像された後、複数話者HMM学習によって求められた話者重畳型HMMを用いて行われる。

### 3. 実験

音韻認識率を用いて話者重畳型HMMモデルの有効性を検討する。ここでは、静的なスペクトル写像の精度の評価、話者重畳型HMMによる話者適応化の改善、マルチテンプレート型HMMとの比較評価について述べる。

#### 3.1 実験条件

分析条件は、12kHzサンプリング、分析窓長21.3msec、フレーム周期3msec、LPC14次である。実験に用いたデータベースは、重要語5240単語と音韻バランス216単語である。話者は、未知話者が男女各1名、標準話者は未知話者とは異なる男性話者6名である。評価には有声破裂音/b,d,g/を用いる。HMMのモデルは、語頭、語中で別々のモデルとし、5240単語中の偶数番目の単語から切り出したものからHMMパラメータの学習を行い、奇数番目の単語に含まれる音韻により認識率の評価を行う。話者適応の学習には、音韻バランス216単語の前半100単語を用いる。話者適応化、HMM学習には、 $k=6$ 、ファジネス1.6のファジィベクトル量子化を用いた。

#### 3.2 スペクトル写像の精度の評価

話者適応化における静的なスペクトル写像の精度を音韻認識率を用いて評価する。方法として次の3つの比較を行う。①特定話者音声認識、②話者適応化音声認識、③未知話者の音声を標準話者の空間に写像してHMM学習、認識。この結果を表1に示す。③は①に比べて平均で6.2%の劣化であった。③は学習認識ともすべて未知話者のデータを使っているため、この劣化は静的なスペクトル写像による歪みのみに起因している。また、話者適応②は、③に加えて話者間の時系列パターンの違いを含んでいるため①に比べ平均で16.5%の劣化となっている。このことから、話者間の時系列パターンを適応化することによりさらに認識率の改善が見込めることがわかる。

#### 3.3 話者重畳型HMM-話者適応化

話者間での調音様式に起因する時系列パターンの適応化のために話者重畳型HMMを用いる。標準話者セットの中から適当な標準話者を基準標準話者とし、重畳する標準話者を1名ずつ増やして未知話者の認識率の改善を検討する。この結果を図1に示す。6名の標準話者を用いた場合82.5%の認識率であった。基準標準話者の選択により若干の傾向の違いがあるものの話者重畳型HMMを用いることで、認識率の改善が行われることがわかる。

#### 3.4 マルチテンプレート型HMM-話者適応化

マルチテンプレート型HMMによる時系列パターンの適応化を行い話者重畳型HMMとの比較を行う。次の2つのマルチテンプレート型HMMを試みる。(HMM①)基準標準話者に他の標準話者の学習データを写像して生成した話者毎のHMMをマルチテンプレートとして、基準標準話者へ写像した未知話者の音声を認識する。(HMM②)未知話者から標準話者へのスペクトルの写像を全て求め、

未知話者の音声を各標準話者へ写像して最もHMMの尤度の高いものを認識結果とする。この結果を図2,3に示す。結果としては、6名の標準話者を用いた場合HMM①の79.5%に比べてHMM②の方が平均で1.9%高い81.4%の認識率が得られた。これは、HMM①の場合、未知話者と標準話者の写像による歪と基準標準話者とその他の標準話者の間の写像による歪を受けるためである。話者重畳型HMMとの比較では、話者重畳型HMMの方がHMM②より1.1%高い認識率であった。

#### 4. まとめ

本稿では、異話者間の発声における静的なスペクトルの違いに加えて調音様式に起因する音韻毎の時系列パターンの違いの話者適応化を行う話者重畳型HMMを提案し、有声破裂音の認識による評価を行った。この結果、話者重畳HMMにより82.5%の認識率が得られ本方法の有効性が確認された。今後の課題として基準標準話者の選択などの問題が残されている。

#### 5. 謝辞

日頃ご指導頂く樽松社長、御討論頂いた音声情報処理研究室の皆様へ感謝致します。

#### 6. 文献

- (1) K.Shikano, et al., "Speaker Adaptation Through Vector Quantization" ICASSP86
- (2) 中村、鹿野, "ベクトル量子化を用いたスペクトログラムの正規化" 信学技報 SP87-17(1987-06)
- (3) 中村、鹿野, "ファジィベクトル量子化を用いたスペクトログラムの正規化" 信学技報 SP87-123(1988-02)
- (4) 中村、鹿野, "ファジィベクトル量子化に基づく話者適応化のHMM音素認識による評価" 音講論集2-P-20 1988
- (5) 鈴木、中島、石川, "ベクトル量子化による話者適応化法の連続音声認識への適用" 音講論集1-3-22 1988
- (6) S.Furui, "Unsupervised Speaker Adaptation Method Based on Hierarchical Spectral Clustering" ICASSP89

Table 1. Recognition Rates(Standard:MHT,MNM) /b,d,g/

Input Speaker	① Dependent	② Adaptation		③ Adaptation	
		MHT	MNM	MHT	MNM
MAU	93.9	79.8	73.6	89.9	87.7
FSU	93.0	79.4	75.0	87.0	84.4
Spectral Pattern	Input Speaker	Adapt : Input→Standard Speaker		Adapt : Input→Standard Speaker	
HMM Speaker Variation	Input Speaker	Standard Speaker		Input Speaker	

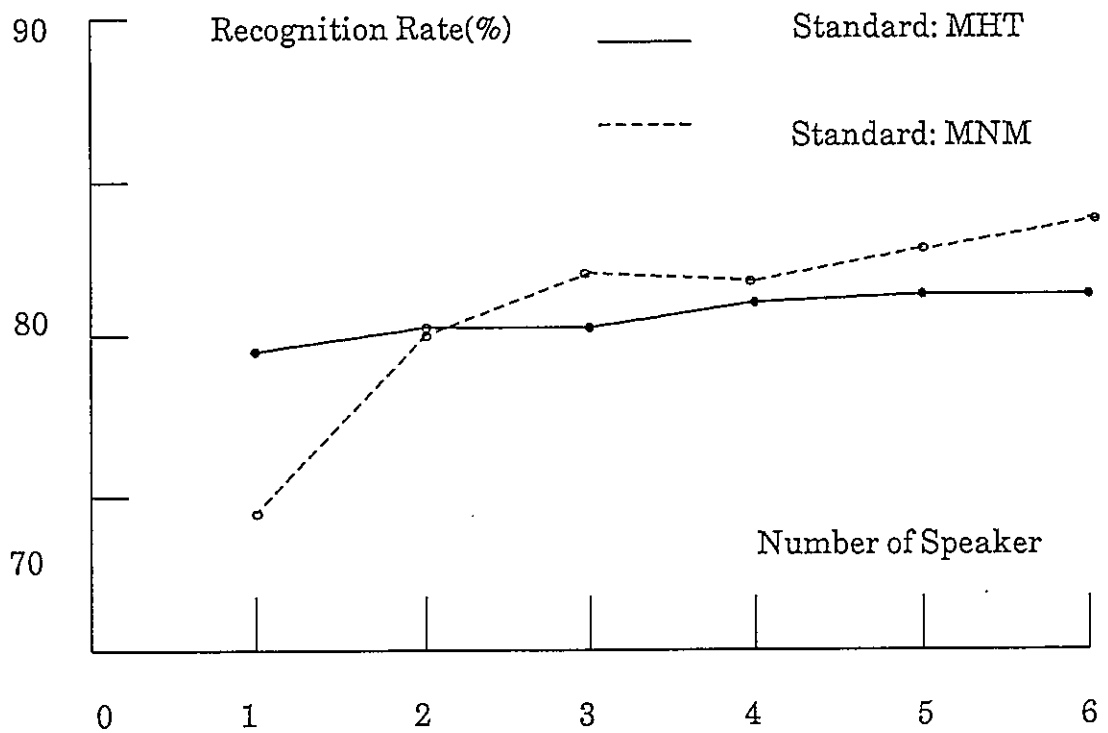


Fig. 1. Improvements by Supplemented HMM /b,d,g/

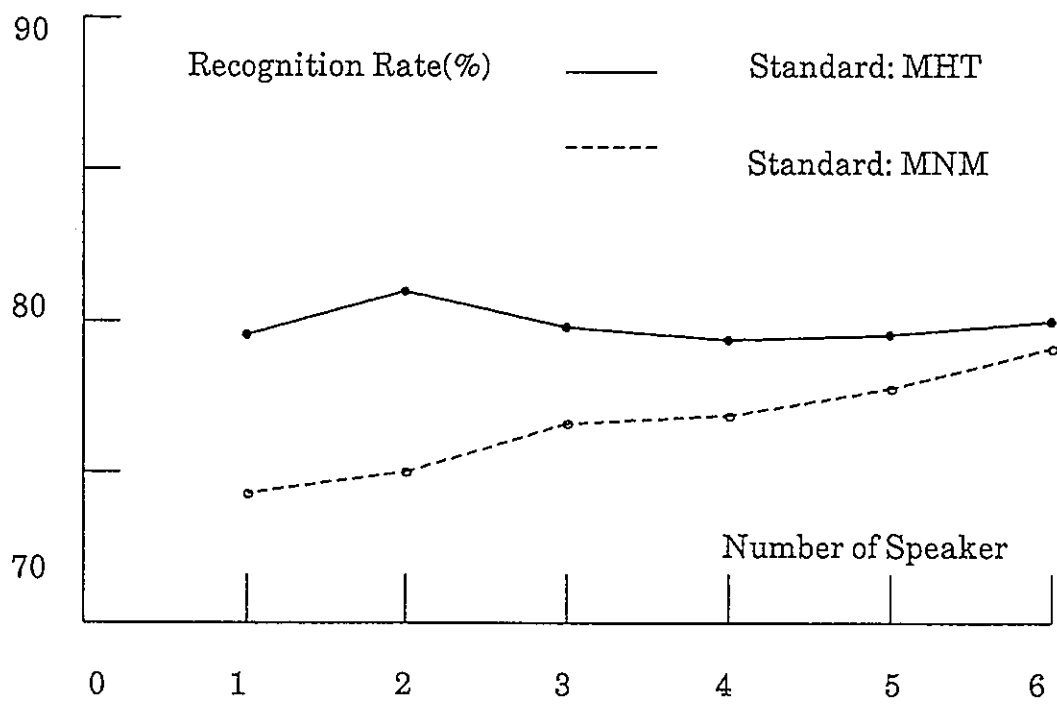


Fig. 2. Improvements by Multi-templete HMM① /b,d,g/



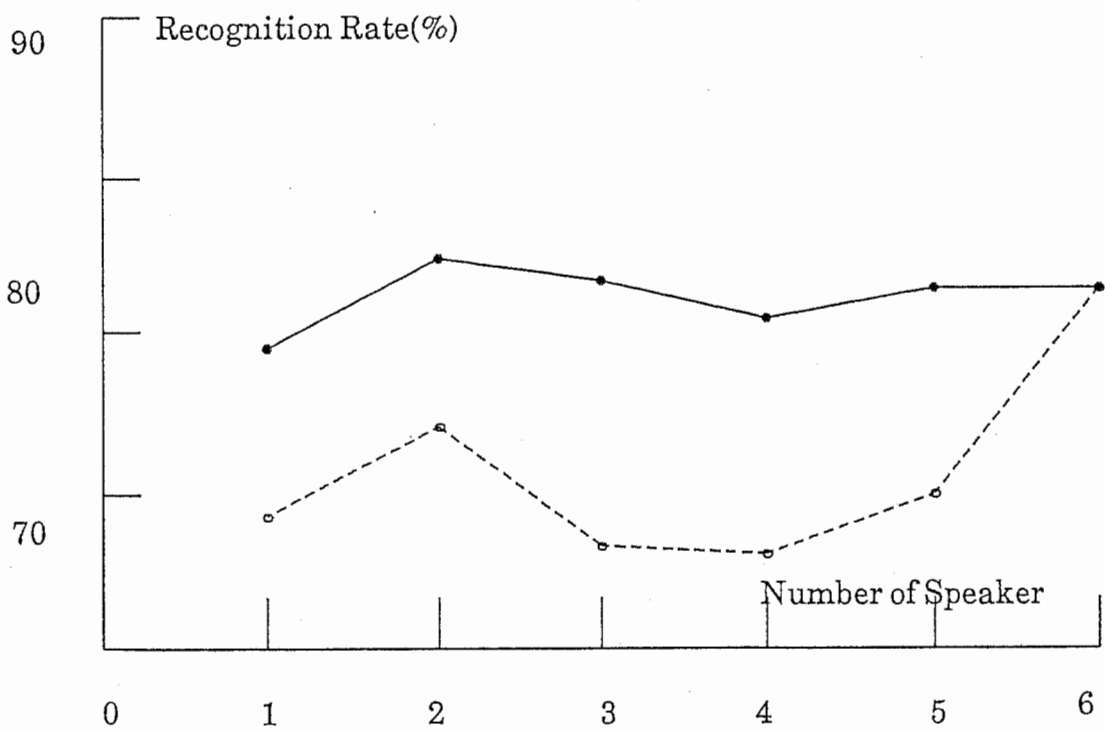


Fig. 3. Improvements by Multi-templete HMM② /b,d,g/