

TR-I-0266

混合連続分布HMMを用いた
移動ベクトル場平滑化話者適応方式

Speaker adaptation based on transfer vector field smoothing
model with continuous mixture density HMMs

大倉計美

杉山雅英

Kazumi OHKURA Masahide SUGIYAMA

嵯峨山茂樹

Shigeki SAGAYAMA

1992.6.12

概要

話者適応の問題を少量学習音声資料による HMM の再学習の問題として捉える。既に十分な学習資料を用いて学習された標準話者のモデルのもつ情報を用いることにより、不十分な学習資料で混合連続分布 HMM を学習した場合に生じる情報の不足を補い、未学習モデルの存在とモデルの推定誤差の問題を解決する“移動ベクトル場平滑化話者適応方式”について述べ、有効性を示す。

ATR 自動翻訳電話研究所
ATR Interpreting Telephony Reserch Laboratories

© (株)ATR 自動翻訳電話研究所 1992
© 1992 by ATR Interpreting Telephony Reserch Laboratories

目次

1	はじめに	2
2	連続分布型 HMM 話者適応	2
2.1	混合連続分布型 HMM 話者適応アルゴリズム	2
3	音素認識実験	5
3.1	実験条件	5
3.2	話者適応実験における平滑化の効果	6
3.3	話者適応と発話様式適応実験	7
4	文節認識実験	9
5	おわりに	11
A	適応資料中の音素出現頻度の詳細	14
B	文節認識結果の詳細	27

1 はじめに

近年、連続分布型 HMM (Hidden Markov Model) を用いた音声認識手法が盛んに研究されている。同時に、標準話者のモデルを入力話者に適応する研究も行なわれている [1]-[8]。このような話者適応手法には、不特定話者モデルをベースにした手法と、標準話者のモデルをベースにした手法が提案されている。どちらの場合も不十分な学習資料しか得られない場合、1) 学習資料に含まれない音素が学習できないという未学習モデルの存在と、2) 少数サンプルにより学習されたモデルの推定誤差、の問題が存在する。本報告では、話者適応の問題を少量学習資料による HMM の再学習の問題として捉え、先に提案されている話者適応方式 [14] における移動ベクトル場平滑化の考えを取り入れた、混合連続分布 HMM における“移動ベクトル場平滑化話者適応方式”を提案する。本手法は、既に十分な学習資料を用いて学習された標準話者のモデルのもつ情報を使用し、不十分な学習資料で混合連続分布 HMM を学習した場合に生じる情報の不足を補い、未学習モデルの存在とモデルの推定誤差の問題を解決するものである。本報告では、標準話者のモデルを入力話者に適応することで本手法の有効性を示すが、本手法は、不特定話者モデルの分布のうち入力話者に近い分布を選択し移動 (適応) する手法や、不特定話者モデルをベースに逐次的に話者適応を行なう手法、および話者選択手法との組合せ等への応用も可能であると考えられる。

一方、学習資料と認識資料の発話様式の違いがもたらす認識性能の低下に対する検討もなされている [8]-[12]。HMM の学習資料に文節発声資料を含めて学習した HMM は、文節発声とは異なる発話様式である単語等の音声資料を用いて学習された HMM よりも高い文節認識率を示すことが明らかにされている [11]。本手法では混合連続分布 HMM の連結学習を用いており、連結学習により得られるモデルは適応資料の発話様式を学習している可能性がある。このような考えから、本報告では、本手法を発話様式適応へ応用した結果も合わせて示す。以下の章は、次の様に構成される。第 2 章では、“移動ベクトル場平滑化話者適応方式”のアルゴリズムを“1. HMM の連結学習によるガウス分布の平均値の再学習”と、“移動ベクトルの内外挿による補間”および、“3. 移動ベクトルの平滑化”の 3 ステップに分けて説明する。第 3 章では、移動ベクトルに対する平滑化の有効性および、本適応方式の発話様式適応への応用とその有効性を 23 音素認識実験により示す。第 4 章では、文節認識実験により、本手法の話者適応と発話様式適応への応用と有効性を示す。

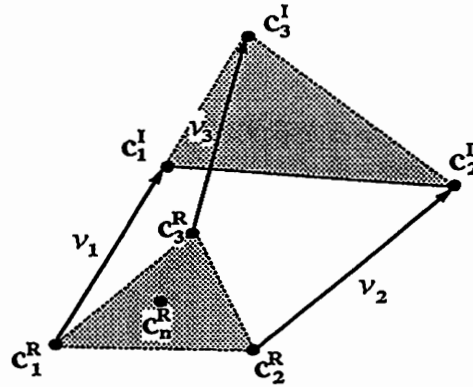
2 連続分布型 HMM 話者適応

連続分布型 HMM の場合、ガウス分布に関するパラメータ (平均値、分散)、分岐確率および遷移確率を話者に適応することができるが、このうち平均値の適応が扱い易く、かつ効果があると考えられるので、本方式では平均値のみの適応を行なう。本方式は、未知話者の適応用音声の発話内容が既知である条件下において、“1. HMM の連結学習によるガウス分布の平均値の再学習”と“2. 移動ベクトルの内外挿による補間”と“3. 移動ベクトルの平滑化”の 3 ステップより構成され、最終的に平滑化移動ベクトルを用いて話者適応モデルを求めるものである。

2.1 混合連続分布型 HMM 話者適応アルゴリズム

1. HMM の連結学習によるガウス分布の平均値の再学習

(1) Calculation of transfer vector (v_1, v_2, v_3)



(2) Interpolation & extrapolation (v_n)

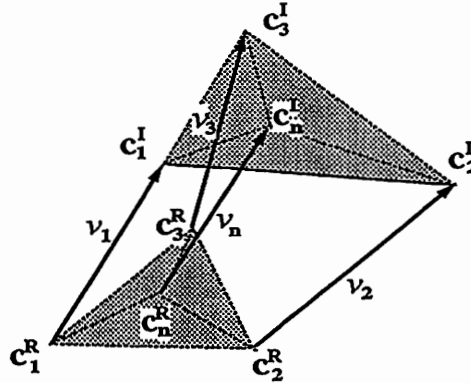


図 1: 移動ベクトルの補間の概念図

本ステップでは、未知話者の音声に含まれる音素に関して、標準話者の音素 HMM の平均値を再学習する。

- (1) 標準話者の音素 HMM を未知話者の音素 HMM の初期モデルとする。
- (2) 未知話者の入力音声の音素系列に対応するように未知話者の HMM を連結し、HMM の遷移確率、出現確率(平均、分散)および分岐確率のうち平均のみを連結学習する。

2. 移動ベクトルの内外挿による補間

本ステップは連結学習前後の HMM の平均ベクトルの差分を移動ベクトルとみなし、学習されなかった HMM の平均ベクトルの移動ベクトルを補間し平均ベクトルを移動するものである [1][2][8][13][14][15]。

(1) 移動ベクトルの計算

未知話者の全音素 HMM のガウス分布の平均ベクトルの組 ($C^I = (c_1^I, \dots, c_K^I)$, K : 全てのガウス分布の個数)のうち学習された k 番目の平均ベクトル c_k^I ($k \in K_1, K_1$: 学習音声中に存在した音素の HMM の平均ベクトルの番号の集合)と、標準話者のガウス分布の平均ベクトルの組 C^R 中で対応する c_k^R より、平均ベクトルの差分ベクトル v_k を求め、これを話者空間の移動ベクトルとする。

$$v_k = c_k^I - c_k^R \quad (k \in K_1)$$

(2) 移動ベクトルの補間

C^I のうち学習音声中に存在しなかった音素に対する HMM に属する平均ベクトル c_n^I ($n \in K_2, K_2$: 学習音声中に存在しなかった音素に対する HMM の平均ベクトルの番号の集合) を学習された k 番目 ($k \in K_1$) の移動ベクトル v_k および、 c_n^R と c_k^R 間のファジイ級関数 $\mu_{n,k}$ [18] から求めた移動ベクトル v_n を用いて c_n^I に移動する。

$$\begin{aligned} v_n &= \sum_{k \in K_1} \mu_{n,k} v_k \\ c_n^I &= c_n^R + v_n \\ \mu_{n,k} &= 1 / \left[\sum_{j \in K_1} (d_{n,k} / d_{n,j})^{1/(m-1)} \right] \end{aligned}$$

ここで、 $d_{n,k}$ は、 c_n^R と c_k^R の距離を表す。

上記の処理の (1)(2) を図 1 を用いて説明する。図は、全ての HMM に含まれるガウス分布の合計が 4 個である場合について示してある。連結学習により c_1^R, c_2^R, c_3^R が、 c_1^I, c_2^I, c_3^I にそれぞれ移動し、 c_n^R は学習されなかった場合を示している。この場合 c_n^I は、 c_1^R, c_2^R, c_3^R と移動ベクトル v_1, v_2, v_3 および $\mu_{n,1}, \mu_{n,2}, \mu_{n,3}$ を用いて計算される。

3. 移動ベクトルの平滑化

上述のステップで得られたモデルは、十分な適応語数が得られていない場合に推定誤差を含んでいる。このような推定誤差を含むものから求められた移動ベクトルの方向は、非連続的な動きをしていると考えられる。そこで、話者空間を移動するための移動ベクトルに連続性の拘束条件を入れ、移動ベクトルの方向性を揃える (平滑化を行なう) ことにより推定誤差の吸収を行なう。

- (1) c_k^R の k -近傍にある c_m^R と c_k^R との間のファジイ級関数 $\mu_{k,m} (k \neq m)$ を求める。
- (2) c_k^I と k -近傍にある m 番目の平均ベクトル c_m^I と c_m^R の差分ベクトル v_m を求める。
- (3) $\mu_{k,m}$ を用いて、差分ベクトル v_m に平滑化を行ない、平滑化移動ベクトル v_k^S を求める。

$$v_k^S = \sum_{m \in N(k)} \alpha_m \mu_{k,m} v_m / \sum_{m \in N(k)} \alpha_m \mu_{k,m}$$

ここで、 $N(k)$ は c_k^R の k -近傍にある平均ベクトルの番号、 α_m は v_m の信頼度を与える定数であり、 $k = m$ の場合、 $\mu_{k,m} = 1$ とする。

- (4) v_k^S と c_k^R を用いて c_k^R を未知話者へ適応する。

$$c_k^S = c_k^R + v_k^S$$

Smoothing ($v_k \rightarrow v_k^S$)

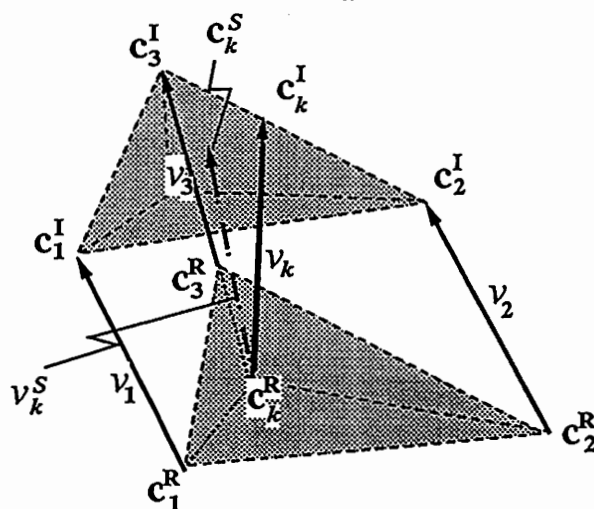


図 2: 平滑化の概念図

ここで、 c_k^S は、平滑化を行なって得られた話者適応後の音素 HMM のガウス分布の平均ベクトルである。

本実験では $\alpha_m = 1 (m \in K_1)$ 、 $\alpha_m = 0 (m \in K_2)$ とした。また、 $\mu_{k,m:(k \neq m)}$ は、 $m \in K_1$ である c_m^R 全てを用いて求めた。

上記の処理を図 2 を用いて説明する。図は、全ての HMM に含まれるガウス分布の合計が 4 個である場合について示してある。“HMM の連結学習による平均値の再学習” および、“移動ベクトルの内外挿による補間” の処理により、 $c_1^R, c_2^R, c_3^R, c_k^R$ が $c_1^I, c_2^I, c_3^I, c_k^I$ にそれぞれ移動したとする。いま、 c_k^I に対応する移動ベクトル v_k を考える。移動ベクトル v_k は、 v_1, v_2, v_3, v_k とそれぞれに対応するファジイ級関数と各移動ベクトルに対する信頼性の重み α_m により平滑化され v_k^S が計算される。

3 音素認識実験

3.1 実験条件

音声資料には ATR データベースを用いた。標準話者は男性 1 名、未知話者は男女各 1 名、分析条件は、サンプリング周波数 12kHz、ハミング窓長 20ms、16 次 LPC 分析、フレーム周期 5ms である。特徴量には、16 次 LPC ケプストラム、16 次 Δ ケプストラム、対数パワー、 Δ 対数パワーの 34 次元ベクトルを用いた。標準話者の HMM は混合連続分布型であり、各音素 HMM の分布数は音素により異なっている。表 1 に HMM の種類とガウス分布の混合数を示す。49 モデルに含まれるガウス分布数の総数は 1152 である。HMM の構造は、図 3 に示したように、母音 (長母音を含む)、撥音および無音区間 HMM は 2 状態 1 ループ、その他の HMM は 4 状態 3 ループである。標準話者の HMM の学習には 5240 単語および音素バランス 216 単語を使用した。

表 1: HMM のガウス分布混合数

混合数	音素 HMM の種類
15	無音 N a i u e o aa ii uu ee oo
14	H K1 K2 S SH TS2 b2 m n ng r t1 t2 y z
10	d2 sy zy
7	CH2 TS1 cy g1 ky w
5	b1 d1 hy
3	CH1 P1 P2 gyl ngy ny ry
2	by my py

表中の小文字表記の子音は、
 後続母音を 15ms 含めて学習した音素を表す。
 表中の数字は、
 1: 語頭の音素より作成した HMM
 2: 語中の音素より作成した HMM
 であることを表す。

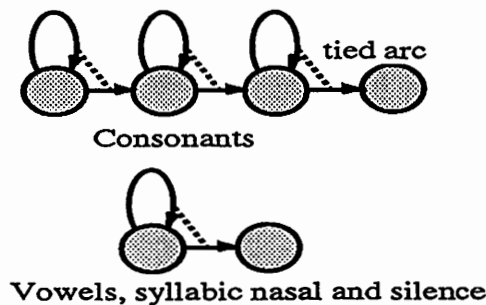
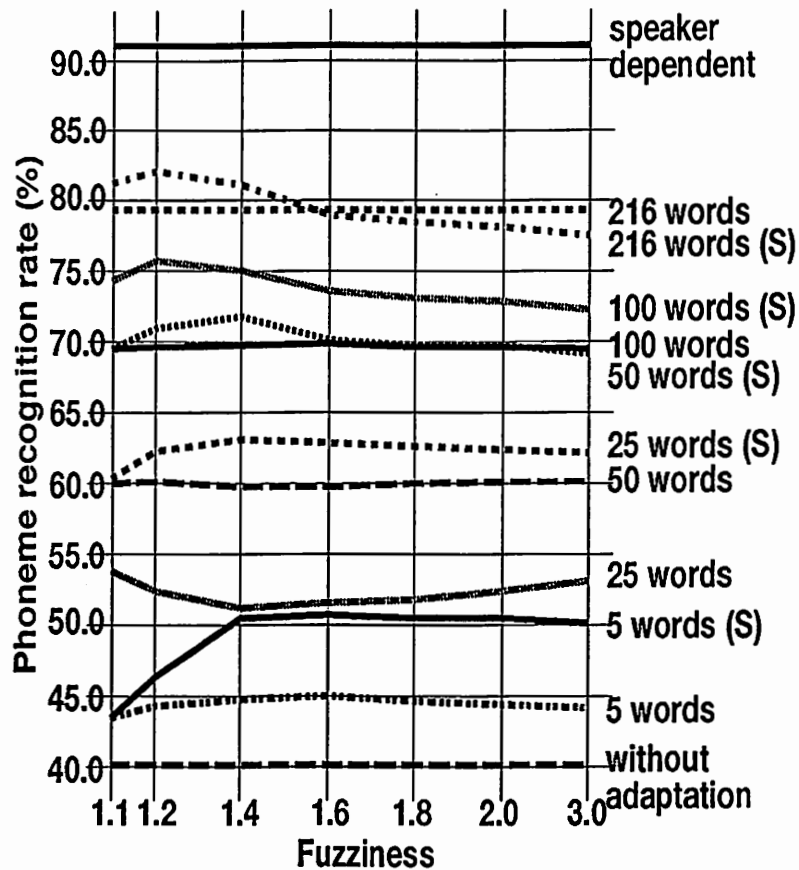


図 3: HMM の構造

3.2 話者適応実験における平滑化の効果

音素バランス 216 単語の先頭から 5,25,50,100,216 語を用いて話者適応を行なう。評価は文節発声資料 (SB2,SB3,SB4) より切り出した 23 音素認識で行なった。ただし、各音素最大 100 サンプルを用いた。表 2(a) に話者適応に用いた学習単語の音声長および、出現した音素の数と総数を示す。各単語の前後に 100ms の無音区間を付けて連結学習を行なった。移動ベクトルの補間時の k -近傍としては、連結学習で求められた平均ベクトル全てを使用した。ファジイ級関数を計算する場合のファジネスを、1.1~3.0 と変化させることにより平滑化の制御の効果を調べる。また、移動ベクトルの平滑化の効果を調べるために、アルゴリズムの説明で述べた“2. 移動ベクトルの内外挿による補間”で得られたモデル (c_n^1) を用いた場合と“3. 移動ベクトルの平滑化”により得られたモデル (c_n^2) を用いた場合の比較を行なう。男女の平均認識率を図 4 に示す。図 4 より、学習単語数の増加にともない、最適なファジネス値は小さくなっていくことが分かる。最適なファジネスを選んだ場合、5 単語から 216 単語全ての場合において、平滑化の制御を行なった場合の認識率は平滑化を行なわない場合の認識率を上回っており、ファジネスにより平滑化を制御する本手法の有効性が分かる。また、25 単語および 50 単語の場合



(S): 平滑化を行なった場合の認識率

図 4: 文節資料中の 23 音素平均認識率

は、表 2(a) に示したように、学習資料中に含まれる音素 HMM が 40 個近くあり、ほぼ全ての音素 HMM が学習されるが、各音素 HMM に対する学習サンプル数が少ないために十分な推定が行えない場合である。このような場合に平滑化は非常に有効な効果を示し、平滑化を行わない場合との比較において、25 単語で 9.3%(fuzziness=1.4)、50 単語で 11.7% (fuzziness=1.4) 認識率が向上している。

3.3 話者適応と発話様式適応実験

本手法は、HMM の連結学習を用いており、発話内容既知の入力を用いて適応が行なえる。前節では、単語発声資料により話者適応を行なった話者適応モデルを用いて文節から切り出した音素の認識を行なったが、本節では、次の 2 実験により、話者性と発話様式にモデルを適応できることを示す。

1. 評価用音声と同一発話様式である文節発声資料を用いた話者適応

評価音声と同一発話様式である文節発声資料を用いて話者適応を行なうことにより、話者性と発話様式にモデルを適応する。適応には、評価用文節資料とは異なる内容をもつ文節発声された文章 (SB1) の先頭から 1 ~ 11 文章を用いた。表 2(b) に学習に用いた文章中の文節数、適

表 2: 適応用音声資料長と 23 音素認識率

(a) 適応に単語発声資料を用いた場合

適応単語数	5	25	50	100	216
音声長 (sec)	3.6	18.7	37.3	75.8	162.6
認識率	50.7%	63.1%	71.8%	75.8%	82.1%
学習資料中に含まれた音素 HMM 数 (延べ音素数)	11 (32)	38 (156)	43 (319)	46 (640)	49 (1359)

学習単語中の音素の出現回数の詳細は表 6-10に示す。

(b) 適応に文節発声資料を用いた場合

適応文章数 (文節数)	1 (7)	2 (13)	3 (20)	5 (47)	11 (102)
音声長 (sec)	4.0	8.3	13.1	29.5	66.3
認識率	59.9	63.2%	68.8%	78.4%	84.5%
学習資料中に含まれた音素 HMM 数 (延べ音素数)	22 (49)	27 (100)	30 (157)	38 (334)	42 (734)

学習単語中の音素の出現回数の詳細は表 11-15に示す。

応用音声資料長および認識率を示す。表 2(a)(b) より、2 文章 (13 文節) の適応用音声長は 25 単語の約半分であるが、認識率はほぼ同等であることと、100 単語の適応用音声長とほぼ同じである 11 文章学習との認識率を比較すると、文章学習の認識率が 9% 高いことが分かる。このことより、文節発声資料を用いて適応を行なった場合、少量の音声資料で高認識率が得られていることが分かる。

2. 文節発声および連続発声した同一文章を用いた話者適応

上記の実験では、単語適応と文節適応を適応用資料の音声長により比較したが、適応に用いた単語と文節では音素コンテキストの違いがある。話者適応用の文節資料と評価用の文節資料には同一語彙が一部含まれており、文節を用いて適応を行なった場合に、発話様式の違いの他に文節の音素コンテキストを学習している可能性がある。ここでは、上記の実験に用いた文節発声文章と同一の文章を連続発声した音声 (SC1) を用いて、適応用音声資料の音素コンテキストが同一という条件下で適応を行なうことにより、発話様式の違いのみに対する適応効果を検討する。評価資料は、適応に用いた文節および連続発声とは内容の違う文章 (文節および連続発声 (SC2,SC3,SC4)) より切り出した 23 音素を用いた。表 3より、適応用資料と評価資料が同一発話様式である場合に高認識率が得られていることが分かる。この結果より、本手法を用いることにより、発話様式の違いを学習していることが分かる。

表 3: 単語・文節・連続発声を用いた話者適応
23 音素認識率

	評価資料 文節発声 (音素)	評価資料 連続発声 (音素)
適応資料: 単語発声 音声長 (100 単語): 75.8(sec) 平滑化なし	75.8% 69.9%	72.2% 65.4%
適応資料: 単語発声 音声長 (216 単語): 162.6(sec) 平滑化なし	82.1% 79.3%	76.8% 73.3%
適応資料: 文節発声 音声長 (11 文章): 66.3(sec) 平滑化なし	<u>84.5%</u> 83.6%	77.1% 74.1%
適応資料: 連続発声 音声長 (11 文章): 51.4(sec) 平滑化なし	76.6% 73.9%	<u>80.0%</u> 77.5%
特定話者認識 (HMM: 単語中 の音素で学習)	91.0%	83.5%

アンダーラインは適応資料と評価資料が
同一発話様式であることを示す
付録の表 16、表 17 に各音素認識率を示しておく

- 、 上記の 2 実験の結果は話者適応と発話様式適応が行なえたことを示している。

4 文節認識実験

本章では、HMM-LR[11]を用いた文節認識実験を通して、話者適応にバランス単語と文節発声資料を用いた場合の比較により、本手法の話者適応性能と発話様式適応性能を評価する。評価資料には、話者適応用資料とは異なる 279 文節を用いた。HMM には表 1 に示した音素 HMM と /i,u/ については無声化モデルを用いた。無声化モデルは、母音モデルに第 1 状態から第 4 状態へ NULL 遷移を付加し、遷移確率としては、第 1 状態から第 2 状態への遷移確率と NULL 遷移の遷移確率を等確率とし、第 1 状態から第 2 状態への遷移確率の 1/2 の値を与えた。単語を用いて学習した連続分布型 HMM の場合、分散に対して定数倍操作を行なうことにより文節認識率が向上することが示されている [11] が、本実験では、話者適応性能のみを検討するために分散の定数倍操作は行なわない。また、連続発声音声や任意の発話速度をもつ

表 4: 文節認識実験結果

	(a) 継続時間長 制御あり 母音, 撥音 2 状態	(b) 継続時間長 制御なし 母音, 撥音 2 状態	(c) 継続時間長 制御なし 母音, 撥音 4 状態
特定話者	87.1%	75.8%	88.3%
話者適応なし	28.7%	13.6%	21.5%
50 単語適応	70.0%	51.2%	71.5%
100 単語適応	77.4%	56.7%	81.7%
216 単語適応	82.0%	63.6%	82.6%
47 文節適応	75.1%	54.8%	72.9%
102 文節適応	79.0%	68.4%	85.1%

文節認識結果の詳細は表 25- 表 26 に示す。

音声認識への展開を考え継続時間長制御を行わない場合の検討も行なう。ビーム幅は、特定話者認識において最大 600 で認識率が飽和することを確認しており、本実験では最大 600 にした。最大 600 とは、語頭から 9 音素まではビーム幅を 600 とし、それ以降ビーム幅を順次 60 まで減らしている [11]。継続時間長は、単語から求めた値を文節用に変換した [16][17]。

表 4(a)(b) より、継続時間長制御を行わない場合は、文節認識率が低下してしまうことが分かる。これは、母音および撥音の HMM の構造が 2 状態の場合、継続時間長制御を用いないと挿入誤りが起こるためである。(viterbi アライメントとラベルに基づく正確な検討ではないが、参考として付録の表 24 に誤認識の 1 例を示しておく。) 挿入誤りを防ぐために母音および撥音の HMM の構造を 4 状態 3 ループにした場合 [11] の認識率を表 4(c) に示す。母音および撥音を 4 状態 3 ループにした場合、学習サンプル数との関係より、ガウス分布の混合数は 15 では多過ぎると考えられるため 12 とした。この場合の分布の総数は 1383 である。表 4(c) に示したように、母音および撥音を 4 状態 3 ループにした場合は 47 文節を用いた場合を除き、継続時間長制御を行なった場合 (表 4(a)) よりも高認識率を示している。これは、母音および撥音を 4 状態 3 ループにした結果、挿入誤りが減少したことと、表 5 に示したように 23 音素認識率が母音および撥音を 4 状態 3 ループにすることにより向上したことが理由の一つであると考えられる。(参考資料として、母音および撥音の状態数による各音素の認識率を表 18 および表 19 に示しておく。また母音および撥音が 4 状態 3 ループの条件で話者適応を行なった場合の、各音素の認識率を付録の表 20- 表 23 に示しておく。表より、平滑化を行なうことにより適応サンプル数の少ない子音の音素認識率が向上することが分かる。)

216 単語学習と 102 文節学習を比較すると、継続時間長制御を行わない場合 (表 4(b)(c)) は 102 文節学習における文節認識率が 216 単語学習の場合の認識率を上回っているが、継続時間長制御を行なった場合 (表 4(a)) は、逆の傾向を示している。この傾向は、男女共同様に示している。表 5 に示した 102 文節学習を用いた場合の 23 音素認識率をみると、状態数に依らず 216 単語学習よりも文節学習の方が高い認識率を示していることより、発話様式適応を行なう場合、単語から求めた値を文節用に変換した継続時間長制御は悪影響をもたらすようである。

表 5: 母音および撥音の HMM の状態数による比較
23 音素認識率

	母音, 撥音 2 状態	母音, 撥音 4 状態
特定話者	91.0%	92.1%
話者適応なし	40.1%	47.3%
50 単語適応	71.8%(60.1%)	74.5%(66.1%)
100 単語適応	75.8%(69.9%)	78.3%(72.9%)
216 単語適応	82.1%(79.3%)	84.7%(81.1%)
47 文節適応	78.4%(73.5%)	81.5%(78.6%)
102 文節適応	84.5%(83.6%)	87.4%(85.6%)

() 内は、平滑化を行わない場合の認識率
付録の表 18、表 19 に各音素の認識率を示しておく。

継続時間長制御に関する検討のため、特定話者用の音素継続時間長制御を用いた場合の話者適応を用いた文節認識実験を行なった。付録の表 27 および表 28 に結果を示す。この場合、女性の文節認識率は適応に単語および文節を用いたどちらの場合も、認識率が 3% 程度向上していることが分かる。このことより標準話者用の音素継続時間長制御を使用することは、文節認識に対して悪影響を及ぼすことがわかるが、この点については今後より詳細に検討していく必要がある。

以上の結果より、1 分程度の文節資料を用いて話者性と発話様式への適応を行なうことにより、特定話者の認識性能に近い文節認識性能を実現できることが示された。

5 おわりに

本報告では、混合連続分布 HMM を用いた移動ベクトル場平滑化話者適応方式の有効性を 23 音素認識実験と文節認識実験により示した。23 音素認識実験により、平滑化手法は、適応音声資料中にほぼ全ての音素 HMM が存在するが、各音素 HMM に対する学習サンプル数が少ないために十分な推定が行えない場合に非常に有効な手法であることが明らかになった。同時に、本方式が発話内容既知の入力に対して話者適応が行なえる利点を生かし、発話様式適応に応用できることも示した。また、文節認識実験では、音素継続時間長制御の有無の比較と発話様式適応の検討を行なった。結果として、継続時間長制御を行わずに 85.1% の文節認識率を得られ、本手法を用いた発話様式適応は有効であることが明らかになった。また、発話様式適応を行なう場合、継続時間長制御は何らかの悪影響をもたらすことがわかったが、この点については今後検討していく必要がある。本報告では、混合連続分布 HMM のパラメータのうち、最も適応効果の大きいと考えられるガウス分布の平均値のみの再学習に基づく適応を行なったが、連結学習時に分岐確率、分散、および遷移確率を学習すれば、より精密な適応モデルを作成することができると考えられる。今後は、適応文章数の検討と適応単語 / 文章数の削減、および平均値以外のパラメータの適応による認識性能の向上の検討と、不特定話者ベースの話者適応への応用の検討を行なう予定である。不特定話者モデルをベースとした話者適応への応用では、不特定のモデルは多数の話者から作られており、音声空間がかなり広がっていることが考えられ、不特定話者モデルの平均値をすべて移動した場合は、認識性能の低下をま

ねく恐れがある。このような不特定モデルを本手法を用いて話者適応する場合は、不特定モデルの空間をモデルを学習した話者の部分空間に分けておき、個々の部分空間(話者空間)を適応する方法を用いることが良いと考えられる。

謝辞

研究の機会を与えて頂いた、ATR 自動翻訳電話研究所榎松明社長に感謝いたします。また、有益な助言をいただいた服部浩明氏(現在: 日本電気中央研究所勤務)、山口耕市氏、並びに ATR の皆様に感謝いたします。

参考文献

- [1] 平田, 中川: “連続出力分布型 HMM における話者適応化の日本語音韻認識による評価,” 信学技報, SP90-16, pp.57-64 (Jun. 1990).
- [2] 中藤, 松本: “連続分布 HMM の教師なし話者適応化の検討,” 信学技報, SP90-67, pp.79-86 (Dec. 1990).
- [3] 水田, 中島: “混合連続分布音素片 HMM を用いた不特定話者単語認識方式における最適識別学習法の検討,” 信学技報, SP91-58, pp.21-28 (Sep. 1991).
- [4] 松岡, 鹿野: “混合ガウス分布不特定話者 HMM をベースとした重み係数による話者適応化法,” 音学講論, 1-1-6, pp.11-12 (Mar. 1992).
- [5] 岩崎, 中島: “混合連続分布音素片 HMM における話者適応化法の検討,” 音学講論, 1-5-21, pp.45-46 (Mar. 1991).
- [6] 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. on SP, Vol.39, pp.806-814 (Apr.1991).
- [7] F. Kubala, R. Schwartz and C. Barry: “Speaker Adaptation from a Speaker-Independent Training Corpus,” proc. of ICASSP90, pp.137-140 (Apr. 1990).
- [8] 大倉, 杉山, 嵯峨山: “混合連続分布 HMM 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-Q-17, pp.191-192 (Mar. 1992).
- [9] 松岡, 鹿野: “HMM 音素モデルの発話変動への適応,” 音学講論, 3-P-15, pp.177-178 (Mar. 1991).
- [10] 花沢, 中島: “発話様式依存型 HMM による音声認識方式の検討,” 音学講論, 2-1-1, pp.51-52 (Mar. 1992).
- [11] 山口, 嵯峨山: “混合連続分布型 HMM を用いた HMM-LR 連続音声認識,” 音学講論, 1-P-5, pp.113-114 (Mar. 1992).
- [12] 加藤, 杉山: “多入力素子をもつニューラルネットワークを用いた連続音声認識,” 音学講論, 3-1-1, pp.71-72 (Mar. 1992).
- [13] K. Shinoda, K. Iso, and T. Watanabe, “Speaker Adaptation For Demi-Syllable Based Continuous Density HMM,” Proc. of ICASSP91, S13.7, pp. 857-860 (May, 1991).
- [14] 服部, 嵯峨山: “少量学習データを用いたコードブックマッピングによる話者適応化,” 音学講論, 1-5-23, pp.49-50 (Mar. 1991).

- [15] 白木, 誉田: “セグメント符合化における話者適応化,” 信学技報, SP87-67, pp.49-56 (Oct. 1987).
- [16] 花沢, 川端, 鹿野: “Hidden Markov モデルによる日本語有声破裂音の認識,” 音響学会誌, Vol.45, No.10, pp.776-783 (Oct. 1989).
- [17] 川端, 花沢, 鹿野: “HMM 音韻認識に基づくワードスポッティング,” 音学講論, 3-P-5, pp.237-238 (Mar. 1988).
- [18] 中村, 花沢, 鹿野: “ベクトル量子化話者適応アルゴリズムの HMM 音韻認識による評価,” 信学技報, SP88-106, pp.1-8 (Dec. 1988).

A 適応資料中の音素出現頻度の詳細

表 6: バランス単語 5 単語中の音素の出現頻度

S	SH	H	z	TS2	t1
0	2	0	0	0	0
t2	K1	K2	b2	ng	m
0	0	1	0	0	2
n	r	y	d2	sy	zy
0	2	3	0	0	0
CH2	TS1	g1	w	cy	ky
0	0	1	0	0	0
b1	d1	hy	CH1	P1	P2
0	0	0	0	0	0
gy1	ngy	ny	ry	py	by
0	0	0	0	0	0
my	N	a	i	u	e
0	0	4	8	2	0
o	aa	ii	uu	eei	oou
6	0	1	0	0	0
Q					
11					

表 7: バランス単語 25 単語中の音素の出現頻度

S	SH	H	z	TS2	t1
2	2	0	1	2	2
t2	K1	K2	b2	ng	m
1	0	6	2	3	6
n	r	y	d2	sy	zy
1	5	4	1	2	4
CH2	TS1	g1	w	cy	ky
2	0	1	2	3	1
b1	d1	hy	CH1	P1	P2
3	2	1	0	1	0
gy1	ngy	ny	ry	py	by
0	0	0	1	0	1
my	N	a	i	u	e
0	7	22	18	14	5
o	aa	ii	uu	eei	oou
16	0	1	3	1	6
Q					
51					

表 8: バランス単語 50 単語中の音素の出現頻度

S	SH	H	z	TS2	t1
6	3	6	3	2	3
t2	K1	K2	b2	ng	m
6	1	10	4	4	12
n	r	y	d2	sy	zy
7	15	6	3	5	5
CH2	TS1	g1	w	cy	ky
5	0	1	4	5	1
b1	d1	hy	CH1	P1	P2
3	3	2	1	1	1
gy1	ngy	ny	ry	py	by
0	0	0	1	0	2
my	N	a	i	u	e
0	13	47	33	26	19
o	aa	ii	uu	eei	oou
28	3	2	5	3	8
Q					
101					

表 9: バランス単語 100 単語中の音素の出現頻度

S	SH	H	z	TS2	t1
14	5	11	10	4	4
t2	K1	K2	b2	ng	m
13	7	22	11	11	19
n	r	y	d2	sy	zy
12	28	9	8	11	9
CH2	TS1	g1	w	cy	ky
10	0	6	6	3	8
b1	d1	hy	CH1	P1	P2
3	3	2	2	1	4
gy1	ngy	ny	ry	py	by
4	2	1	3	0	2
my	N	a	i	u	e
0	24	77	56	61	41
o	aa	ii	uu	eei	oou
61	5	6	13	8	16
Q					
204					

表 10: バランス単語 216 単語中の音素の出現頻度

S	SH	H	z	TS2	t1
33	9	23	26	6	13
t2	K1	K2	b2	ng	m
18	26	51	18	19	36
n	r	y	d2	sy	zy
29	61	17	12	16	14
CH2	TS1	g1	w	cy	ky
9	2	8	13	12	12
b1	d1	hy	CH1	P1	P2
9	10	11	1	7	7
gy1	ngy	ny	ry	py	by
4	6	10	11	8	10
my	N	a	i	u	e
10	53	160	109	126	80
o	aa	ii	uu	eei	oou
108	8	9	38	13	54
Q					
446					

表 11: SB1: 先頭 1 文章中の音素の出現頻度

S	SH	H	z	TS2	t1
2	2	0	1	0	0
t2	K1	K2	b2	ng	m
0	3	3	0	1	3
n	r	y	d2	sy	zy
0	1	1	1	0	0
CH2	TS1	g1	w	cy	ky
1	1	0	1	0	1
b1	d1	hy	CH1	P1	P2
0	1	0	0	0	0
gy1	ngy	ny	ry	py	by
0	0	0	0	0	0
my	N	a	i	u	e
0	1	5	7	5	2
o	aa	ii	uu	eei	oou
5	0	0	1	0	0
Q					
14					

表 12: SB1: 先頭 2 文章中の音素の出現頻度

S	SH	H	z	TS2	t1
5	3	0	1	0	1
t2	K1	K2	b2	ng	m
2	5	5	1	3	4
n	r	y	d2	sy	zy
2	3	1	2	0	0
CH2	TS1	g1	w	cy	ky
2	1	0	1	0	1
b1	d1	hy	CH1	P1	P2
0	1	0	0	0	0
gy1	ngy	ny	ry	py	by
0	0	0	0	0	0
my	N	a	i	u	e
0	4	10	14	9	3
o	aa	ii	uu	eei	oou
13	0	0	1	0	2
Q					
26					

表 13: SB1: 先頭 3 文章中の音素の出現頻度

S	SH	H	z	TS2	t1
8	5	0	2	0	2
t2	K1	K2	b2	ng	m
5	6	7	1	3	5
n	r	y	d2	sy	zy
5	5	3	5	0	0
CH2	TS1	g1	w	cy	ky
2	2	0	1	1	1
b1	d1	hy	CH1	P1	P2
0	1	0	0	0	0
gy1	ngy	ny	ry	py	by
0	0	0	0	0	0
my	N	a	i	u	e
0	5	13	18	13	9
o	aa	ii	uu	eei	oou
22	0	1	1	0	4
Q					
41					

表 14: SB1: 先頭 5 文章中の音素の出現頻度

S	SH	H	z	TS2	t1
11	8	2	3	0	5
t2	K1	K2	b2	ng	m
13	13	13	1	9	11
n	r	y	d2	sy	zy
13	11	3	10	1	0
CH2	TS1	g1	w	cy	ky
2	2	2	4	1	1
b1	d1	hy	CH1	P1	P2
1	2	0	0	2	0
gy1	ngy	ny	ry	py	by
0	0	0	0	1	0
my	N	a	i	u	e
0	12	29	37	19	24
o	aa	ii	uu	eei	oou
46	1	2	1	2	9
Q					
101					

表 15: SB1: 先頭 11 文章中の音素の出現頻度

S	SH	H	z	TS2	t1
29	13	9	7	6	9
t2	K1	K2	b2	ng	m
23	24	28	4	15	22
n	r	y	d2	sy	zy
33	22	9	22	3	4
CH2	TS1	g1	w	cy	ky
4	3	5	13	1	5
b1	d1	hy	CH1	P1	P2
2	2	1	0	3	0
gy1	ngy	ny	ry	py	by
0	0	0	1	3	0
my	N	a	i	u	e
0	24	83	74	51	46
o	aa	ii	uu	eei	oou
80	1	4	7	5	21
Q					
217					

表 16: 文節 (SB1)・文章 (SC1) を適応に用いた場合の認識率
 文節資料 (SB2,SB3,SB4) 中の 23 音素認識率

適応には先頭の 11 文章を使用

	MAU 入力 サンプル数 Total 1770	MAU 母音 撥音 2 状態 文節 適応 f=1.4	MAU 母音 撥音 2 状態 文章 適応 f=1.4	FSU 入力 サンプル数 Total 1706	FSU 母音 撥音 2 状態 文節 適応 f=1.4	FSU 母音 撥音 2 状態 文章 適応 f=1.4
S	100	100.0%	99.0%	100	99.0%	97.0%
SH	87	98.9%	71.3%	83	94.0%	80.7%
H	54	72.2%	68.5%	56	82.1%	69.6%
z	30	100.0%	46.7%	32	53.1%	43.8%
CH	27	92.6%	92.6%	28	92.9%	92.9%
TS	42	45.2%	78.6%	42	71.4%	78.6%
P	2	100.0%	100.0%	2	0.0%	50.0%
t	100	82.0%	76.0%	100	87.0%	81.0%
K	100	96.0%	97.0%	100	100.0%	100.0%
b	32	75.0%	75.0%	27	59.3%	33.3%
d	100	91.0%	50.0%	100	79.0%	49.0%
g	86	66.3%	47.7%	79	43.0%	21.5%
m	100	85.0%	84.0%	100	96.0%	90.0%
n	100	78.0%	55.0%	100	85.0%	88.0%
N	84	92.9%	82.1%	57	71.9%	49.1%
r	100	74.0%	58.0%	100	70.0%	68.0%
w	60	73.3%	63.3%	44	59.1%	84.1%
y	66	77.3%	47.0%	56	66.1%	42.9%
a	100	100.0%	100.0%	100	100.0%	100.0%
i	100	82.0%	87.0%	100	98.0%	88.0%
u	100	85.0%	81.0%	100	94.0%	94.0%
e	100	82.0%	80.0%	100	94.0%	89.0%
o	100	90.0%	91.0%	100	84.0%	87.0%
平均		84.3% 84.8%	75.3% 75.4%		77.3% 84.2%	72.9% 77.7%

平均の下段は、

(正解サンプル数 / 総入力サンプル数) とした場合の平均
 表中の f はファジネスの値を示す

表 17: 文節 (SB1)・文章 (SC1) を適応に用いた場合の認識率
文章資料 (SC2,SC3,SC4) 中の 23 音素認識率

適応には先頭の 11 文章を使用

	MAU 入力 サンプル数 Total 1764	MAU 母音 撥音 2 状態 文節 適応 f=1.4	MAU 母音 撥音 2 状態 文章 適応 f=1.2	FSU 入力 サンプル数 Total 1699	FSU 母音 撥音 2 状態 文節 適応 f=1.4	FSU 母音 撥音 2 状態 文章 適応 f=1.2
S	100	100.0%	100.0%	100	88.0%	90.0%
SH	86	100.0%	97.7%	83	96.4%	97.6%
H	55	58.2%	63.6%	56	71.4%	76.8%
z	30	96.7%	76.7%	32	56.3%	50.0%
CH	27	88.9%	77.8%	28	92.9%	100.0%
TS	41	51.2%	80.5%	42	83.3%	81.0%
P	2	0.0%	0.0%	2	0.0%	0.0%
t	100	76.0%	86.0%	100	77.0%	84.0%
K	100	87.0%	93.0%	100	96.0%	99.0%
b	32	62.5%	68.8%	28	46.4%	53.6%
d	100	82.0%	85.0%	100	66.0%	57.0%
g	87	47.1%	59.8%	81	50.6%	40.7%
m	100	75.0%	83.0%	100	95.0%	96.0%
n	100	69.0%	78.0%	100	71.0%	85.0%
N	87	97.7%	82.8%	53	83.0%	77.4%
r	100	66.0%	74.0%	100	70.0%	81.0%
w	60	68.3%	63.3%	47	78.7%	83.0%
y	57	33.3%	45.6%	47	36.2%	72.3%
a	100	100.0%	98.0%	100	97.0%	96.0%
i	100	50.0%	59.0%	100	90.0%	85.0%
u	100	83.0%	76.0%	100	73.0%	78.0%
e	100	63.0%	70.0%	100	86.0%	90.0%
o	100	75.0%	75.0%	100	82.0%	80.0%
平均		70.9% 75.1%	73.6% 78.4%		73.3% 79.0%	76.2% 81.5%

平均の下段は、
(正解サンプル数 / 総入力サンプル数) とした場合の平均
表中の f はファジネスの値を示す

表 18: 状態数による特定話者の認識率
 文節資料 (SB2,SB3,SB4) 中の 23 音素認識率

	MAU 入力 サンプル数 Total 1770	MAU 母音 撥音 2 状態	MAU 母音 撥音 4 状態	FSU 入力 サンプル数 Total 1706	FSU 母音 撥音 2 状態	FSU 母音 撥音 4 状態
S	100	98.0 %	98.0%	100	97.0 %	97.0%
SH	87	100.0%	100.0%	83	98.8 %	98.8%
H	54	75.9 %	75.9%	56	91.1 %	92.9%
z	30	96.7 %	100.0%	32	93.8 %	100.0%
CH	27	92.6 %	92.6%	28	96.4 %	96.4%
TS	42	92.9 %	92.9%	42	100.0%	100.0%
P	2	50.0 %	50.0%	2	100.0%	100.0%
t	100	90.0 %	98.0%	100	94.0 %	96.0%
K	100	98.0 %	99.0%	100	100.0%	100.0%
b	32	93.8 %	93.8%	27	92.6 %	96.3%
d	100	89.0 %	90.0%	100	76.0 %	76.0%
g	86	90.7 %	94.2%	79	78.5 %	78.5%
m	100	62.0 %	62.0%	100	98.0 %	100.0%
n	100	72.0 %	75.0%	100	99.0 %	99.0%
N	84	91.7 %	96.4%	57	86.0 %	86.0%
r	100	93.0 %	93.0%	100	99.0 %	100.0%
w	60	61.7 %	58.3%	44	100.0%	100.0%
y	66	87.9 %	86.4%	56	98.2 %	98.2%
a	100	100.0%	100.0%	100	100.0%	100.0%
i	100	76.0 %	83.0%	100	93.0 %	94.0%
u	100	92.0 %	91.0%	100	98.0 %	98.0%
e	100	83.0 %	91.0%	100	98.0 %	99.0%
o	100	88.0 %	85.0%	100	95.0 %	96.0%
平均		85.9 % 87.2 %	87.2% 88.8%		94.9 % 94.7 %	95.7% 95.4%

平均の下段は、
 (正解サンプル数 / 総入力サンプル数) とした場合の平均

表 19: 状態数による話者適応 (11 文章:SB1) を用いた場合の認識率
 文節資料 (SB2,SB3,SB4) 中の 23 音素認識率

	MAU	MAU	FSU	FSU
	母音 撥音 2 状態	母音 撥音 4 状態	母音 撥音 2 状態	母音 撥音 4 状態
	f=1.4	f=1.2	f=1.4	f=1.2
S	100.0%	100.0%	99.0%	97.0%
SH	98.9%	98.9%	94.0%	92.8%
H	72.2%	74.1%	82.1%	82.1%
z	100.0%	93.3%	53.1%	62.5%
CH	92.6%	88.9%	92.9%	96.4%
TS	45.2%	45.2%	71.4%	78.6%
P	100.0%	50.0%	0.0%	0.0%
t	82.0%	91.0%	87.0%	90.0%
K	96.0%	100.0%	100.0%	100.0%
b	75.0%	75.0%	59.3%	63.0%
d	91.0%	90.0%	79.0%	76.0%
g	66.3%	70.9%	43.0%	43.0%
m	85.0%	85.0%	96.0%	97.0%
n	78.0%	92.0%	85.0%	95.0%
N	92.9%	90.5%	71.9%	77.2%
r	74.0%	82.0%	70.0%	96.0%
w	73.3%	90.0%	59.1%	75.0%
y	77.3%	75.8%	66.1%	85.7%
a	100.0%	100.0%	100.0%	100.0%
i	82.0%	86.0%	98.0%	93.0%
u	85.0%	83.0%	94.0%	94.0%
e	82.0%	81.0%	94.0%	95.0%
o	90.0%	84.0%	84.0%	90.0%
平均	84.3%	83.8%	77.3%	81.7%
	84.8%	86.8%	84.2%	88.0%

平均の下段は、
 (正解サンプル数 / 総入力サンプル数) とした場合の平均
 表中の f はファジネスの値を示す

表 20: 先頭 50 単語中の音素の出現頻度と認識率
 文節資料 (SB2,SB3,SB4) 中の 23 音素認識率

母音・撥音:4 状態 3 ループ

	適応 サン プル数	MAU 平滑化 なし	MAU 平滑化 あり	FSU 平滑化 なし	FSU 平滑化 あり
S	6	97.0 %	100.0%	96.0%	99.0 %
SH	3	66.7 %	83.9 %	57.8%	65.1 %
H	6	44.4 %	50.0 %	19.6%	39.3 %
z	3	36.7 %	83.3 %	15.6%	46.9 %
CH	1+5	100.0 %	100.0 %	89.3%	100.0%
TS	0+2	38.1 %	42.9 %	7.1%	14.3 %
P	1+1	0.0 %	0.0 %	0.0%	0.0 %
t	3+6	82.0 %	89.0 %	45.0%	57.0 %
K	1+10	94.0 %	86.0 %	91.0%	96.0 %
b	3+4	68.8 %	84.4 %	18.5%	44.4 %
d	3+3	26.0 %	54.0 %	5.0%	22.0 %
g	1+4	19.8 %	48.8 %	26.6%	35.4 %
m	12	65.0 %	77.0 %	74.0%	77.0 %
n	7	27.0 %	47.0 %	70.0%	72.0 %
N	13	91.7 %	91.7 %	80.7%	77.2 %
r	15	78.0 %	87.0 %	93.0%	94.0 %
w	4	18.3 %	55.0 %	13.6%	50.0 %
y	6	19.7 %	42.4 %	10.7%	37.5 %
a	47	100.0 %	100.0 %	100.0%	100.0 %
i	33	87.0 %	87.0 %	98.0%	95.0 %
u	26	85.0 %	83.0 %	78.0%	81.0 %
e	19	85.0 %	84.0 %	96.0%	97.0 %
o	28	85.0 %	87.0 %	86.0%	89.0 %
平均		61.5 % 67.1 %	72.3 % 76.7 %	55.3 % 65.0 %	64.7 % 72.2 %

平均の下段は、
 (正解サンプル数 / 総入力サンプル数) とした場合の平均
 イタリック文字は、
 平滑化により認識率の向上したもの
 適応サンプル数の欄の “+” は、
 語頭音素 + 語中音素の各個数
 を示す

表 21: 先頭 100 単語中の音素の出現頻度と認識率
 文節資料 (SB2,SB3,SB4) 中の 23 音素認識率

母音・撥音:4 状態 3 ループ

	適応 サン プル数	MAU 平滑化 なし	MAU 平滑化 あり	FSU 平滑化 なし	FSU 平滑化 あり
S	14	99.0 %	99.0 %	98.0 %	97.0 %
SH	5	65.5 %	88.5 %	61.4 %	73.5 %
H	11	53.7 %	64.8 %	44.6 %	51.8 %
z	10	90.0 %	90.0 %	46.9 %	46.9 %
CH	2+10	92.6 %	100.0%	82.1 %	96.4 %
TS	0+4	57.1 %	73.8 %	31.0 %	52.4 %
P	1+4	0.0 %	0.0 %	0.0 %	50.0 %
t	4+13	84.0 %	89.0 %	54.0 %	64.0 %
K	7+22	97.0 %	96.0 %	100.0 %	100.0 %
b	3+11	75.0 %	75.0 %	29.6 %	25.9 %
d	3+8	55.0 %	70.0 %	10.0 %	20.0 %
g	6+11	41.9 %	58.1 %	38.0 %	43.0 %
m	19	75.0 %	80.0 %	75.0 %	76.0 %
n	12	29.0 %	59.0 %	86.0 %	88.0 %
N	24	91.7 %	90.5 %	70.2 %	71.9 %
r	28	93.0 %	86.0 %	96.0 %	91.0 %
w	6	18.3 %	60.0 %	43.2 %	68.2 %
y	9	27.3 %	40.9 %	12.5 %	48.2 %
a	77	100.0 %	100.0 %	100.0 %	100.0 %
i	56	91.0 %	86.0 %	95.0 %	93.0 %
u	61	88.0 %	86.0 %	92.0 %	87.0 %
e	41	90.0 %	82.0 %	95.0 %	97.0 %
o	61	84.0 %	87.0 %	87.0 %	88.0 %
平均		69.5 % 74.2 %	76.6 % 80.8 %	62.9 % 71.5 %	70.8 % 75.9 %

平均の下段は、
 (正解サンプル数 / 総入力サンプル数) とした場合の平均
 イタリック文字は、
 平滑化により認識率の向上したもの
 適応サンプル数の欄の “+” は、
 語頭音素 + 語中音素の各個数
 を示す

表 22: 先頭 216 単語中の音素の出現頻度と認識率
 文節資料 (SB2,SB3,SB4) 中の 23 音素認識率

母音・撥音:4 状態 3 ループ

	適応 サン プル数	MAU 平滑化 なし	MAU 平滑化 あり	FSU 平滑化 なし	FSU 平滑化 あり
S	33	100.0 %	100.0 %	98.0 %	97.0 %
SH	9	79.3 %	95.4%	61.4 %	84.3%
H	23	94.4 %	88.9 %	92.9 %	94.6%
z	26	93.3 %	93.3 %	53.1 %	53.1 %
CH	1+9	92.6 %	88.9 %	85.7 %	100.0%
TS	2+6	38.1 %	66.7%	33.3 %	66.7%
P	7+7	0.0 %	50.0%	50.0 %	50.0 %
t	13+18	91.0 %	97.0%	61.0 %	73.0%
K	26+51	96.0 %	95.0 %	100.0 %	100.0 %
b	9+18	71.9 %	90.6%	44.4 %	33.3%
d	10+12	73.0 %	80.0%	31.0 %	54.0%
g	8+19	53.5 %	66.3%	38.0 %	43.0%
m	36	88.0 %	90.0%	90.0 %	83.0 %
n	29	64.0 %	74.0%	93.0 %	92.0 %
N	53	91.7 %	94.0%	84.2 %	77.2 %
r	61	93.0 %	90.0 %	98.0 %	95.0 %
w	13	31.7 %	65.0%	77.3 %	88.6%
y	17	54.5 %	57.6%	64.3 %	66.1%
a	160	100.0 %	100.0 %	100.0 %	100.0 %
i	109	91.0 %	87.0 %	96.0 %	94.0 %
u	126	84.0 %	81.0 %	93.0 %	94.0 %
e	80	96.0 %	86.0 %	96.0 %	97.0%
o	108	89.0 %	87.0 %	90.0 %	86.0 %
平均		76.8 % 82.2 %	83.6 % 85.9 %	75.2 % 80.0 %	79.2 % 83.5 %

平均の下段は、
 (正解サンプル数 / 総入力サンプル数) とした場合の平均
 イタリック文字は、
 平滑化により認識率の向上したもの
 適応サンプル数の欄の “+” は、
 語頭音素 + 語中音素の各個数
 を示す

表 23: SB1 の先頭 11 文章 (102 文節) 中の音素の出現頻度と認識率
 文節資料 (SB2,SB3,SB4) 中の 23 音素認識率

母音・撥音:4 状態 3 ループ

	適応 サン プル数	MAU 平滑化 なし	MAU 平滑化 あり	FSU 平滑化 なし	FSU 平滑化 あり
S	29	100.0 %	100.0 %	100.0 %	97.0 %
SH	13	98.9 %	98.9 %	97.6 %	92.8 %
H	9	63.0 %	74.1 %	75.0 %	82.1%
z	7	86.7 %	93.3 %	59.4 %	62.5%
CH	0+4	74.1 %	88.9 %	89.3 %	96.4%
TS	3+6	23.8 %	45.2 %	64.3 %	78.6%
P	0+3	50.0 %	50.0 %	0.0 %	0.0 %
t	9+23	90.0 %	91.0 %	83.0 %	90.0%
K	24+28	95.0 %	100.0%	99.0 %	100.0%
b	2+4	65.6 %	75.0 %	48.1 %	63.0%
d	2+22	92.0 %	90.0 %	84.0 %	76.0 %
g	5+15	62.8 %	70.9 %	43.0 %	43.0 %
m	22	73.0 %	85.0 %	98.0 %	97.0 %
n	33	95.0 %	92.0 %	94.0 %	95.0%
N	24	92.9 %	90.5 %	77.2 %	77.2 %
r	22	70.0 %	82.0 %	96.0 %	96.0 %
w	13	58.3 %	90.0 %	88.6 %	75.0 %
y	9	77.3 %	75.8 %	80.4 %	85.7%
a	83	100.0 %	100.0 %	100.0 %	100.0 %
i	74	89.0 %	86.0 %	97.0 %	93.0 %
u	51	82.0 %	83.0 %	97.0 %	94.0 %
e	46	88.0 %	81.0 %	96.0 %	95.0 %
o	80	85.0 %	84.0 %	86.0 %	90.0%
平均		78.8 % 83.3 %	83.8 % 86.8 %	80.6 % 87.9 %	81.7% 88.0%

平均の下段は、
 (正解サンプル数 / 総入力サンプル数) とした場合の平均
 イタリック文字は、
 平滑化により認識率の向上したもの
 適応サンプル数の欄の “+” は、
 語頭音素 + 語中音素の各個数
 を示す

B 文節認識結果の詳細

表 24: 継続時間長を行なわない場合の認識誤りの1例
母音・撥音が2状態数の場合

話者	誤りの種類	例	
男性	"i" に挟まれた "g" の脱落 "y" の挿入 "e" の挿入	kaigi → kaini	1
		tetsuzukio → tetsuzuki-y-o	7
		shoteino → shotei-e-no	1
		imasu → i-e-masu	1
	語頭・語尾の挿入	itadakimasu → oi-itadakimasu	1
		imasu → yu-imasu	1
	その他の挿入誤り	soudesuka → o-soidesuka	1
		cyoukousuru → cyoukou-saserereru	1
		yoNma.Nendesu → yo-miageru-Ndesu	1
女性	"i" に挟まれた "g" の脱落 "y" の挿入 "e" の挿入	kaigi → kaini	7
		tetsuzukio → tetsuzuki-y-o	7
		youshino → youshi-e-no	1
		sureba → soereba	1
	語頭・語尾の挿入	tsuuchino → tsuuchi-e-no	1
		moushikomi → ano-moushikomi	2
	その他の挿入誤り	cyoukousuru → cyoukou-sasererareru	1

表 25: 文節認識実験結果 (MAU)

	(a) 継続時間長 制御あり 母音, 撥音 (2 状態)	(b) 継続時間長 制御なし 母音, 撥音 (2 状態)	(c) 継続時間長 制御なし 母音, 撥音 (4 状態)
特定話者			
1 位	83.9%	71.7%	85.0%
~2 位	93.6%	81.0%	95.0%
~3 位	96.1%	83.2%	96.4%
~4 位	98.2%	84.2%	97.1%
~5 位	98.6%	86.4%	97.5%
話者適応なし			
1 位	41.2%	23.7%	37.6%
~2 位	48.0%	26.2%	45.2%
~3 位	53.1%	28.7%	46.6%
~4 位	55.9%	30.5%	48.8%
~5 位	60.2%	32.3%	50.5%
50 単語適応	f=1.4	f=1.4	f=1.2
1 位	73.5%	48.0%	72.4%
~2 位	86.7%	59.1%	81.4%
~3 位	90.7%	62.7%	85.0%
~4 位	93.2%	64.5%	87.1%
~5 位	94.3%	65.2%	88.9%
100 単語適応	f=1.2	f=1.2	f=1.2
1 位	80.7%	55.2%	82.1%
~2 位	93.2%	67.0%	90.0%
~3 位	95.7%	70.3%	91.8%
~4 位	96.1%	72.8%	92.8%
~5 位	96.8%	75.6%	93.6%
216 単語適応	f=1.2	f=1.2	f=1.2
1 位	81.7%	63.4%	83.2%
~2 位	93.6%	71.7%	92.1%
~3 位	95.7%	76.3%	94.3%
~4 位	96.4%	77.8%	94.6%
~5 位	97.5%	79.2%	95.3%
47 文節適応	f=1.4	f=1.4	f=1.4
1 位	75.6%	53.8%	74.6%
~2 位	85.7%	62.0%	83.5%
~3 位	89.3%	67.0%	85.7%
~4 位	90.7%	67.7%	87.1%
~5 位	91.8%	68.5%	97.5%
102 文節適応	f=1.4	f=1.4	f=1.2
1 位	78.9%	67.4%	85.0%
~2 位	90.7%	73.1%	90.7%
~3 位	93.2%	75.6%	91.0%
~4 位	94.6%	77.1%	92.8%
~5 位	95.7%	77.4%	93.6%

表中の f はファジネスの値を示す

表 26: 文節認識実験結果 (FSU)

	(a) 継続時間長 制御あり 母音, 撥音 (2 状態)	(b) 継続時間長 制御なし 母音, 撥音 (2 状態)	(c) 継続時間長 制御なし 母音, 撥音 (4 状態)
特定話者			
1 位	90.3%	79.9%	91.7%
~2 位	97.5%	85.6%	99.3%
~3 位	98.9%	88.9%	99.6%
~4 位	99.3%	90.3%	99.6%
~5 位	99.6%	90.7%	100.0%
話者適応なし			
1 位	16.2%	3.6%	13.7%
~2 位	23.4%	6.8%	19.8%
~3 位	27.3%	8.3%	21.9%
~4 位	29.1%	10.4%	24.8%
~5 位	33.1%	11.5%	27.3%
50 単語適応	f=1.4	f=1.4	f=1.2
1 位	66.6%	54.3%	70.5%
~2 位	79.1%	62.6%	83.5%
~3 位	82.7%	66.6%	86.3%
~4 位	84.9%	69.4%	89.2%
~5 位	86.3%	70.1%	90.7%
100 単語適応	f=1.2	f=1.2	f=1.2
1 位	74.1%	58.3%	81.3%
~2 位	83.8%	64.4%	90.3%
~3 位	87.1%	68.0%	91.4%
~4 位	88.9%	70.1%	93.5%
~5 位	90.7%	72.7%	93.5%
216 単語適応	f=1.2	f=1.2	f=1.2
1 位	82.4%	63.7%	82.0%
~2 位	91.0%	71.9%	92.1%
~3 位	93.2%	77.0%	93.9%
~4 位	93.9%	78.8%	93.9%
~5 位	93.9%	80.2%	93.9%
47 文節適応	f=1.4	f=1.4	f=1.4
1 位	74.5%	55.8%	71.2%
~2 位	84.5%	64.8%	80.6%
~3 位	87.4%	67.3%	84.9%
~4 位	89.6%	70.5%	85.6%
~5 位	90.3%	70.9%	88.5%
102 文節適応	f=1.4	f=1.4	f=1.2
1 位	79.1%	69.4%	85.3%
~2 位	87.8%	78.4%	93.2%
~3 位	89.9%	80.9%	95.7%
~4 位	91.4%	83.8%	96.0%
~5 位	91.7%	84.9%	96.8%

表中の f はファジネスの値を示す

表 27: 継続時間長制御の検討
文節認識実験結果 (MAU)

	音素継続時間長制御に MAU の推定値を使用	音素継続時間長制御に MHT の推定値を使用
216 単語適応	f=1.2	f=1.2
1 位	82.1%	81.7%
~2 位	94.3%	93.6%
~3 位	96.8%	95.7%
~4 位	97.1%	96.4%
~5 位	97.1%	97.5%
102 文節適応	f=1.4	f=1.4
1 位	78.1%	78.9%
~2 位	88.9%	90.7%
~3 位	92.1%	93.2%
~4 位	93.2%	94.6%
~5 位	94.6%	95.7%

表中の f はファジネスの値を示す

表 28: 継続時間長制御の検討
文節認識実験結果 (FSU)

	音素継続時間長制御に FSU の推定値を使用	音素継続時間長制御に MHT の推定値を使用
216 単語適応	f=1.2	f=1.2
1 位	85.3%	82.4%
~2 位	93.5%	91.0%
~3 位	96.4%	93.2%
~4 位	97.5%	93.9%
~5 位	97.5%	93.9%
102 文節適応	f=1.4	f=1.4
1 位	82.8%	79.1%
~2 位	91.0%	87.8%
~3 位	92.8%	89.9%
~4 位	94.2%	91.4%
~5 位	95.3%	91.7%

表中の f はファジネスの値を示す