

TR-I-0284

混合連続分布 HMM 移動ベクトル場  
平滑化話者適応方式の文節認識性能の評価

Evaluation of speaker adaptation method  
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

1992.10.6

概要

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れば、HMM の連結学習を用いて話者適応を行なうことができる。この観点から、移動ベクトル場平滑化を用いた混合連続分布 HMM の話者適応方式である“移動ベクトル場平滑化話者適応方式: transfer vector field smoothing model (VFS)”を先に提案した。VFS は、基本的に少量学習単語時の分布の移動問題を扱うものであり、初期モデルに不特定話者モデルを選ぶこともできる。本報告では、男性7名、女性9名の各入力話者が発声した714文節の認識実験により、初期モデルに男性および女性のモデルを用いて、入力話者の性別により初期モデルを選択した場合、および不特定話者モデルを初期モデルに用いた場合の“VFS”認識性能の比較を行なった。

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 1992

© 1992 by ATR Interpreting Telephony Research Laboratories

## 目次

1	はじめに	2
2	混合連続分布 HMM 移動ベクトル場平滑化話者適応方式	2
3	実験条件	2
4	文節認識実験	3
4.1	初期モデルが男性の場合の評価	3
4.2	初期モデルの検討	4
4.2.1	初期モデルに女性話者モデルを用いた場合	4
4.2.2	初期モデルに不特定話者モデルを用いた場合	4
5	おわりに	4
A	各認識方式の比較	10
B	初期モデルが男性モデル (MHT) の場合の認識結果	11
C	初期モデルが女性モデル (FSU) の場合の文節認識率	17
D	不特定話者モデルの文節認識率と初期モデルに不特定話者モデルを用いた話者適応の文節認識率	21
E	入力話者が MAU の場合の認識率	24

## 1 はじめに

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れば、HMM の連結学習を用いて話者適応を行なうことができる。この観点から、移動ベクトル場平滑化を用いた混合連続分布 HMM の話者適応方式である“移動ベクトル場平滑化話者適応方式: transfer vector field smoothing model (VFS)”を先に提案した [8]。VFS は、十分に学習された初期モデルの情報を取り込むことにより、不十分な学習資料しか得られない場合の再学習における 2 つの問題 1) 未学習モデルの存在、2) 学習されたモデルの推定誤差、を解くものである。話者適応には、初期モデルとして不特定話者モデルを用いる方法と、話者選択等により初期モデルを入力話者に対して選択する方法が考えられる [1, 2, 4, 5, 6, 7, 11, 12, 16]。VFS は、基本的に少量学習単語時の分布の移動問題を扱うものであり、初期モデルに不特定話者モデルを選ぶこともできる。本報告では、男性 7 名、女性 9 名の各入力話者が発声した 714 文節の認識実験により、初期モデルに男性および女性のモデルを用いて、入力話者の性別により初期モデルを選択した場合、および不特定話者モデルを初期モデルに用いた場合の“VFS”認識性能の比較を行なった。

## 2 混合連続分布 HMM 移動ベクトル場平滑化話者適応方式

本方式は、入力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”、“2. 移動ベクトルの内挿”と“3. 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。

まず、標準話者のモデルを入力話者の初期モデルとし、発話内容が既知の適応用入力音声を用いて連結学習を行なう。連結学習の結果得られたガウス分布の平均値と初期モデルのガウス分布の平均値との差分を移動ベクトルとする。本方式では、連結学習による平均値の移動を、求めた移動ベクトルに沿った移動と見なす。各平均値に対する移動ベクトルの集合を移動ベクトル場として捉え、移動ベクトルの求まっていない平均値の近傍の移動ベクトル場の局所的 / 大局的な方向をもとに、移動ベクトルを内挿する。平滑化も同様に、近傍の移動ベクトル場の方向をもとになされる [8, 9]。

## 3 実験条件

入力話者は表 1 に示した、男性 7 名、女性 9 名の計 16 名である。初期モデルは、男性 1 名 (MHT)、女性 1 名 (FSU) から作成した。この男女の話者は、従来の実験で評価してきた話者であり、今回の実験のために最適な話者を選んだわけではない。HMM の学習データには 5240 単語および音素バランス 216 単語を用いた。HMM は対角共分散行列の混合ガウス分布型である。HMM は 49 モデルあり、分布数は各音素 HMM 毎に異なる。表 2 に HMM の種類と分布数を示す。この場合のガウス分布の総計は 1383 である。無音区間の HMM の構造は 2 状態 1 ループ、その他の HMM は 4 状態 3 ループであり、各状態からのループはタイドアークとした [9, 10]。

不特定話者モデルは男性 12 名が発声した 8832 単語 (736 単語  $\times$  12 名) を用いて作成した。学習に用いた話者と音声資料を表 3 に示す。HMM の構造は、全て 4 状態 3 ループである。各状態の分布数は 15 であり、ガウス分布の総計は 2205 である。

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  $\Delta$  ケプストラム、対数パワー、 $\Delta$  対数パワーの 34 次元ベクトルを用いた。

評価は、各入力話者が発声した 714 文節 (モデル会話 1~10, A, B) を用いた文節認識で行なった。ただし、話者 FAS の、モデル会話 8 の 15 番目の文章に誤発声があったため、FAS は誤発声の 1 文節を除く 713 文節で認識を行なった。(“ジョージ・オハラです” という発声を “ジョージ” と “オハラです” の 2 文節として発声している。) 表 4 に本実験に用いた文法名とタスクの複雑さを示す。文節認識には、状態に留まる継続時間制御を用いない HMM-LR 手法を用いた [10]。話者適応には、音素バランス 216 単語の先頭から 50 単語を用いた。話者適応に用いた 12 名の平均音声長は 38.2sec である。話者適応時には、各発声の前後に最大 100msec の無音区間を付加して話者適応を行なった。連結学習の結果から計算された移動ベクトルを全て使用して、VFS の内挿および平滑化を行なった。内挿および平滑化を制御するためのファジネスの値は 1.4 とした。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おり [10]、本実験では、HMM の分散を 1.2 倍にして認識を行なった。認識時には、表 2 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 4 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 4 状態への遷移確率の 1/2 の値を与えた。

## 4 文節認識実験

### 4.1 初期モデルが男性の場合の評価

初期モデルに男性話者モデルを選んだ場合の認識結果と、比較として不特定話者モデル (SID) を用いた場合およびコードブック間の対応関係を用いたヒストグラムマッピング (CBM) に基づく離散分布型 HMM の話者適応結果を表 6 に示す。離散分布型 HMM の認識には、状態毎の時間長制御を用いている。HMM の学習には Hard-VQ、認識には Fuzzy-VQ ( $k$  近傍数 6、ファジネス 1.6) を用いた。

離散系の分析条件は、ハミング窓長 21.33ms (256 ポイント)、14 次 LPC 分析、フレーム周期 9ms である。特徴量には、16 次 LPC 相関 (WLR 距離尺度ベース)、16 次  $\Delta$  ケプストラム ( $\pm 8$  フレーム幅)、正規化パワーを用いた [13, 14, 15]。表 5 に HMM-LR の実験条件を示す。参考として、各話者の認識率を表 9~43 に示す。表 37~43 には、不特定話者モデルの分散の最適化のために行なった予備実験結果も合わせて示す。

表 6 より、CBM 手法は男女間で文節認識性能に差は殆んどみられないことがわかる。VFS は、男性のみの平均文節認識率では、CBM の男性の平均認識率を 4.5% 上回っているが、女性の平均認識率は逆に 9% 低い値を示していることが分かる。このことより、VFS では初期モデルと入力話者の依存性が CBM よりも強いことが分かる。また、SID の文節認識性能は CBM と同程度の性能であることも分かる。

## 4.2 初期モデルの検討

### 4.2.1 初期モデルに女性話者モデルを用いた場合

前節の結果より、VFS は初期モデルと入力話者に依存性のあることが分かった。本節では、初期モデルに女性話者のモデルを選んだ場合の女性話者の認識結果を示す。結果を表 7 に示す。表 7 より、初期モデルが男性の場合と比較して、14.2%(68.7%→82.9%) 文節認識率が向上し、初期モデルに女性を選択した方が良いことがわかる。また同表より、初期モデルを男性から女性のモデルにした場合、話者適応を行わない場合の認識率が 51.7%(8.8%→60.5%) 向上していることが分かる。これらのことより、初期モデルには入力話者と音声の特徴の近い話者を選択することが重要であることが分かる。本実験では、単に性別で初期モデルを選択したが、音声空間における音響的構造の類似性に基づく初期モデル選択 [17] 等を行なうことにより、更に認識率が向上する可能性がある。

### 4.2.2 初期モデルに不特定話者モデルを用いた場合

本節では、不特定話者モデルを初期モデルにした場合を評価する。ただし、不特定話者モデルは男性が発声した音声資料のみから作成されているため、評価は男性 7 名で行なった。結果を表 8 に示す。表 8 より、平滑化を行わない場合、不特定話者モデルの認識率は未適応(不特定話者モデルをそのまま用いた場合)の認識率よりも低い認識率を示すが、平滑化を行なうことにより認識率が 9%(77.1%→86.2%) 程度向上することが分かる。この認識率は、初期モデルに男性話者モデルを用いた場合(83.2%) よりも 3% 高い認識率である。このことより、不特定話者モデルベースの話者適応においても平滑化の手法が有効であることが分かる。

## 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 (VFS) の有効性を文節認識実験により評価した。評価では、連続分布型 HMM 不特定話者モデルとコードブックマッピング (CBM) に基づく離散分布型 HMM 話者適応との比較を行ない、VFS は初期モデルと入力話者の依存関係が CBM よりも強いが、初期モデルを選択すれば CBM よりも高い認識性能を示すことが分かった。また、不特定話者モデルを初期モデルにした場合、50 単語適応で 86.2% の認識率を得た。本報告では、不特定話者モデルの分布を全て VFS により移動したが、分布の中には入力話者に対して合わない分布が存在すると考えられ、今後は不特定話者内の分布の選択と VFS を組み合わせた手法の検討を行なう予定である。また、初期モデルの選択を単なる性別ではなく音響的構造の類似性に基づく基準等により行なった場合と、不特定話者モデルベースの話者適応性能の比較を行なう必要がある。

### 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、不特定話者モデルを提供して頂いた小坂哲夫氏並びに日頃熱心に討論頂く 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] 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,” 信学技報, SP92-16, pp. 23-28 (Jun. 1992).
- [10] 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).
- [11] 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).
- [12] 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).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] 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).

[17] 宮沢, 嗟峨山: “移動ベクトル場平滑化話者適応方式における標準話者選択方式の検討,”  
音学講論 (Oct. 1992).

表 1: 入力話者

		年齢	所属
男 性	MTK	51 才	声優
	MIK	50 才	アナウンサー
	MAO	48 才	アナウンサー
	MHO	23 才	大学 (アナウンス研究会)
	MHF	20 才	大学 (一般)
	MST	33 才	一般
	MSH	22 才	一般
女 性	FAK	31 才	声優
	FKN	41 才	声優
	FNY	50 才	アナウンサー
	FFO	22 才	大学 (アナウンス研究会)
	FAS	21 才	大学 (アナウンス研究会)
	FEO	21 才	大学 (アナウンス研究会)
	FMS	21 才	大学 (アナウンス研究会)
	FRS	21 才	大学 (一般)
	FMO	22 才	大学 (一般)

表 2: HMM のガウス分布混合数

混合数	音素 HMM の種類
15	無音
12	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 gl 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。



表 3: 不特定話者モデルの学習資料の詳細

話者	M001,M002,M003,M004,M101,M102,M103,M104,M401,M402,M403,M404
音声資料	WI(520 単語) および B(音素バランス 216 単語)

表 4: 文法名とタスクの複雑性

文法名	dp10-re8.gra
音素パープレキシティ	3.6
単語パープレキシティ	41.8

表 5: HMM-LR の設定

	グローバル ビーム幅	ローカル ビーム幅	セル数	分散の倍率
離散分布 HMM (CBM)	250	32	1000	-
連続分布 HMM (VFS)	600*	64	2000	1.2
連続分布 HMM (SID)	1200	64	2000	1.2

\*: 語頭から 9 音素まではビーム幅を 600  
 10 から 13 音素まで 200  
 14 から 17 音素まで 100  
 18 音素以上 60

表 6: 入力話者男性 7 名および女性 9 名の平均認識率

初期モデル: 男性

		SID			VFS	CBM
入力話者	男性	77.1%	適応	平滑化あり	83.2% (93.9%)	78.8% (94.9%)
				平滑化なし	74.6% (92.2%)	
		未適応		47.7% (63.8%)	76.0% (93.4%)	
	女性	-	適応	平滑化あり	68.7% (82.0%)	77.7% (96.9%)
			平滑化なし	65.4% (80.1%)		
			未適応		8.8% (15.7%)	39.2% (62.6%)

()内は、5位累積認識率

表 7: 入力話者女性 9 名の平均認識率

初期モデル	未適応	平滑化なし	平滑化あり
男性話者	8.8% (15.7%)	65.4% (80.1%)	68.7% (82.0%)
女性話者	60.5% (75.9%)	73.1% (87.8%)	82.9% (93.4%)

()内は、5位累積認識率

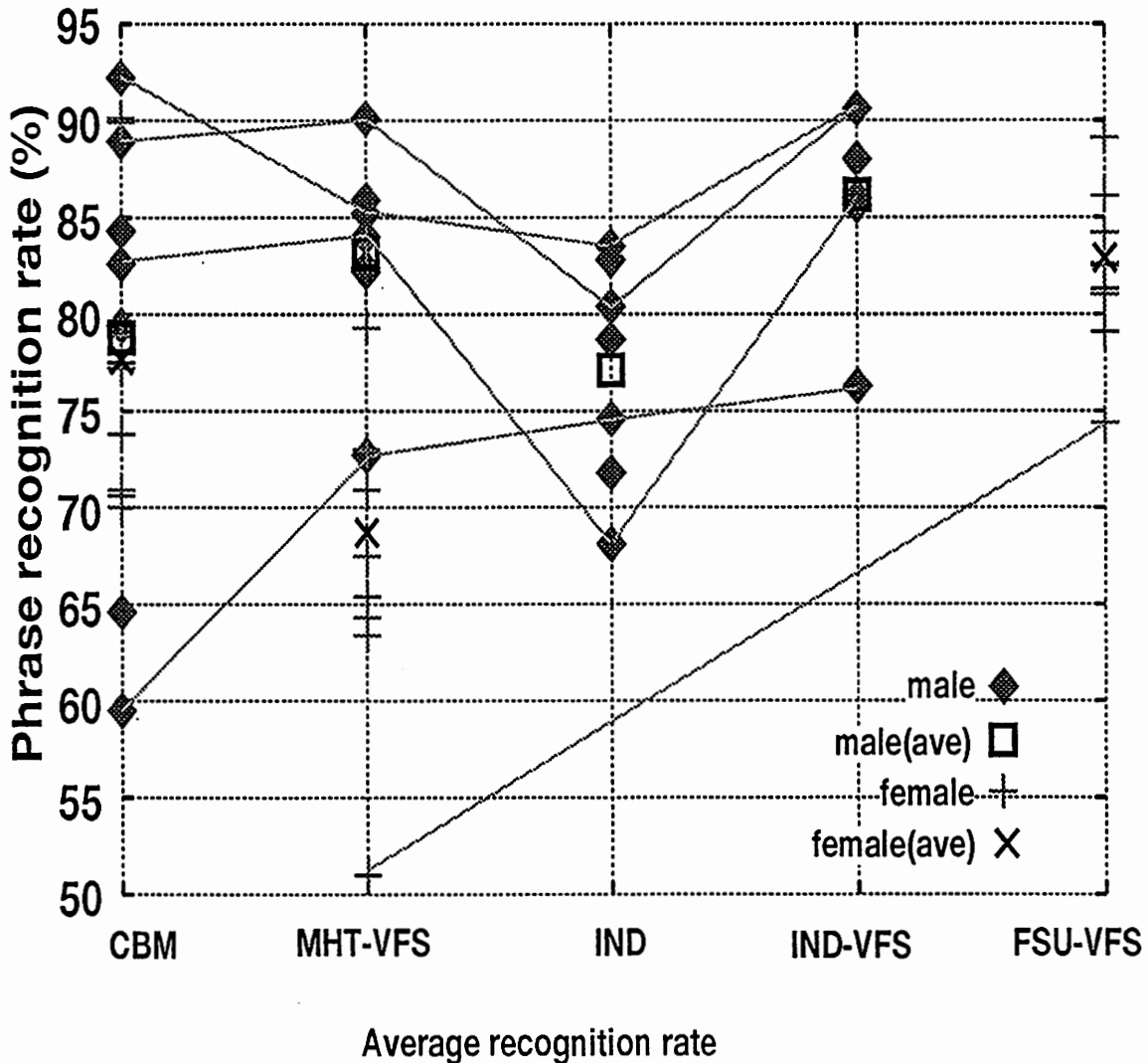
表 8: 入力話者男性 7 名の平均認識率

初期モデル	未適応	平滑化なし	平滑化あり
男性話者	47.7% (63.8%)	74.6% (92.2%)	83.2% (93.9%)
不特定話者	77.1% (93.3%)	76.3% (89.5%)	86.2% (95.1%)

()内は、5位累積認識率

## A 各認識方式の比較

移動ベクトル場平滑化方式 (VFS) とコードブックマッピング (CBM) および不特定話者モデル (IND) の文節認識率を図1に示す。VFSにおいては、初期モデルに男性モデル (MHT-VFS) を用いた場合、女性モデルを用いた場合 (FSU-VFS)、不特定話者モデルを用いた場合 (IND-VFS) の文節認識率を示した。図には、一部の入力話者について、各認識方式間の対応関係を示した。



図中の直線は、入力話者の各方式間における対応関係を表す。

図1: 各認識方式の比較

## B 初期モデルが男性モデル (MHT) の場合の認識結果

初期モデルに男性モデル (MHT) を用いた場合の入力話者 16 名の文節認識率を以下に示す。

表 9: 入力話者:FAK

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1 位	10.6%	82.4%	83.6%	54.2%	89.9%
~ 2 位	14.3%	87.3%	87.7%	68.2%	98.9%
~ 3 位	17.8%	88.8%	88.7%	73.7%	99.6%
~ 4 位	19.0%	89.8%	89.5%	77.7%	99.6%
~ 5 位	20.4%	90.1%	89.8%	80.5%	99.7%

表 10: 入力話者:FAS

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1 位	7.4%	58.9%	65.4%	31.8%	73.8%
~ 2 位	9.4%	69.6%	73.9%	40.8%	86.4%
~ 3 位	9.9%	74.1%	77.7%	45.9%	92.6%
~ 4 位	10.4%	76.3%	79.4%	48.7%	94.8%
~ 5 位	11.8%	77.4%	80.5%	50.6%	95.8%

表 11: 入力話者:FEO

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1 位	6.8%	63.9%	64.3%	42.4%	78.9%
~ 2 位	9.6%	70.6%	75.2%	54.8%	91.2%
~ 3 位	11.6%	76.9%	77.5%	61.2%	95.2%
~ 4 位	13.6%	78.9%	79.4%	64.6%	96.9%
~ 5 位	14.0%	80.0%	81.1%	68.2%	97.9%

表 12: 入力話者:FFO

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	15.5%	70.6%	73.0%	51.7%	77.5%
~2位	22.0%	77.7%	81.4%	65.7%	91.9%
~3位	24.9%	81.0%	83.9%	71.0%	94.8%
~4位	27.2%	82.1%	84.7%	74.5%	96.9%
~5位	28.2%	83.2%	85.4%	76.3%	97.6%

表 13: 入力話者:FKN

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	4.9%	49.2%	51.0%	33.2%	77.2%
~2位	6.2%	55.7%	57.3%	42.9%	90.6%
~3位	7.3%	60.8%	60.6%	49.2%	93.4%
~4位	8.4%	62.5%	62.5%	51.7%	95.0%
~5位	8.8%	64.1%	63.7%	54.2%	95.8%

表 14: 入力話者:FMO

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	4.9%	65.0%	70.9%	24.6%	70.9%
~2位	6.7%	72.8%	78.4%	35.2%	88.4%
~3位	7.7%	76.9%	81.5%	40.1%	92.2%
~4位	8.5%	78.6%	83.2%	44.7%	93.7%
~5位	9.2%	79.8%	84.2%	47.1%	95.1%

表 15: 入力話者:FMS

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	6.7%	55.0%	63.4%	19.0%	70.6%
~ 2位	9.2%	65.0%	72.4%	25.9%	87.1%
~ 3位	10.9%	68.6%	76.9%	31.1%	92.4%
~ 4位	12.3%	71.1%	78.4%	34.2%	95.4%
~ 5位	12.9%	73.2%	79.0%	37.8%	96.1%

表 16: 入力話者:FNY

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	17.2%	76.2%	79.3%	55.7%	90.1%
~ 2位	21.7%	83.1%	86.3%	68.3%	97.6%
~ 3位	23.2%	85.4%	89.9%	74.4%	98.5%
~ 4位	24.8%	87.5%	90.9%	79.4%	99.0%
~ 5位	25.9%	88.8%	91.5%	82.5%	99.0%

表 17: 入力話者:FRS

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	5.7%	67.1%	67.5%	40.3%	70.0%
~ 2位	7.3%	77.7%	77.2%	51.8%	84.5%
~ 3位	8.1%	80.8%	80.0%	58.3%	91.9%
~ 4位	9.4%	83.1%	81.8%	63.3%	94.1%
~ 5位	10.4%	83.9%	82.4%	66.2%	95.0%

表 18: 入力話者:MAO

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	56.4%	79.0%	90.1%	87.4%	88.9%
~2位	62.7%	84.0%	93.8%	95.5%	96.6%
~3位	67.4%	86.1%	95.2%	97.8%	98.2%
~4位	69.6%	87.5%	95.7%	98.0%	98.5%
~5位	71.6%	88.2%	96.1%	98.6%	98.5%

表 19: 入力話者:MHF

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	51.5%	78.0%	82.4%	62.5%	64.6%
~2位	62.5%	84.6%	90.6%	79.3%	78.7%
~3位	66.8%	87.3%	92.4%	83.1%	82.8%
~4位	69.6%	88.4%	93.7%	86.3%	85.7%
~5位	71.3%	89.1%	94.0%	88.1%	87.1%

表 20: 入力話者:MHO

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	48.9%	57.3%	72.7%	61.6%	59.5%
~2位	57.6%	67.8%	82.2%	75.8%	76.8%
~3位	61.2%	74.6%	85.4%	81.1%	83.8%
~4位	64.0%	77.9%	87.4%	84.3%	86.0%
~5位	66.0%	80.3%	88.5%	86.8%	88.0%

表 21: 入力話者:MIK

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	55.6%	79.6%	85.9%	92.4%	92.2%
~ 2位	62.9%	85.4%	91.3%	97.6%	97.5%
~ 3位	66.0%	88.5%	93.7%	98.3%	97.9%
~ 4位	68.3%	89.8%	94.4%	98.5%	98.2%
~ 5位	69.5%	91.0%	95.0%	98.5%	98.2%

表 22: 入力話者:MSH

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	57.3%	76.2%	82.2%	73.0%	79.4%
~ 2位	68.2%	83.3%	90.8%	83.1%	92.0%
~ 3位	71.4%	87.1%	92.4%	86.6%	95.1%
~ 4位	74.4%	88.7%	94.1%	88.9%	96.1%
~ 5位	75.9%	90.1%	95.1%	90.5%	97.3%

表 23: 入力話者:MST

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	44.7%	77.6%	85.2%	77.9%	84.3%
~ 2位	53.6%	83.3%	91.5%	92.0%	93.8%
~ 3位	57.0%	87.0%	93.7%	95.1%	96.1%
~ 4位	58.4%	88.1%	95.0%	97.5%	97.6%
~ 5位	61.2%	90.2%	95.5%	97.5%	98.2%



表 24: 入力話者:MTK

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	19.6%	74.4%	84.0%	76.9%	82.6%
~ 2位	23.9%	80.8%	90.8%	88.1%	91.6%
~ 3位	27.7%	83.9%	91.9%	92.0%	95.4%
~ 4位	30.3%	85.7%	92.7%	93.0%	96.2%
~ 5位	31.2%	86.3%	92.9%	93.7%	96.9%

表 25: 女性 9 名の平均認識率

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	8.8%	65.4%	68.7%	39.2%	77.7%
~ 5位	15.7%	80.1%	82.0%	62.6%	96.9%

表 26: 男性 7 名の平均認識率

	VFS			CBM	
	未適応	平滑化なし	平滑化あり	未適応	適応
1位	47.7%	74.6%	83.2%	76.0%	78.8%
~ 5位	63.8%	92.2%	93.9%	93.4%	94.9%

### C 初期モデルが女性モデル (FSU) の場合の文節認識率

初期モデルに女性モデル (FSU) を用いた場合の女性入力話者 9 名の文節認識率を以下に示す。

表 27: 入力話者:FAK

	VFS		
	未適応	平滑化なし	平滑化あり
1 位	72.0%	86.1%	92.2%
~ 2 位	79.6%	91.0%	95.8%
~ 3 位	81.7%	93.1%	96.5%
~ 4 位	82.6%	93.8%	96.8%
~ 5 位	83.5%	94.0%	97.1%

表 28: 入力話者:FAS

	VFS		
	未適応	平滑化なし	平滑化あり
1 位	60.9%	66.5%	81.3%
~ 2 位	72.1%	78.3%	87.9%
~ 3 位	75.7%	83.3%	91.2%
~ 4 位	79.2%	84.4%	92.6%
~ 5 位	80.6%	84.7%	93.4%

表 29: 入力話者:FEO

	VFS		
	未適応	平滑化なし	平滑化あり
1 位	66.7%	71.1%	82.5%
~ 2 位	74.6%	81.5%	91.2%
~ 3 位	77.6%	85.2%	93.3%
~ 4 位	78.7%	86.7%	93.6%
~ 5 位	79.7%	88.0%	94.3%

表 30: 入力話者:FFO

	VFS		
	未適応	平滑化なし	平滑化あり
1 位	74.5%	76.8%	84.2%
~ 2 位	82.9%	84.2%	91.6%
~ 3 位	85.6%	87.5%	93.6%
~ 4 位	87.3%	88.5%	94.1%
~ 5 位	87.7%	89.2%	94.4%

表 31: 入力話者:FKN

	VFS		
	未適応	平滑化なし	平滑化あり
1 位	29.0%	61.6%	74.4%
~ 2 位	37.4%	70.9%	81.8%
~ 3 位	40.6%	74.8%	85.0%
~ 4 位	42.7%	77.0%	86.6%
~ 5 位	44.4%	79.6%	87.7%

表 32: 入力話者:FMO

	VFS		
	未適応	平滑化なし	平滑化あり
1 位	45.8%	72.5%	82.6%
~ 2 位	57.3%	80.4%	88.4%
~ 3 位	62.2%	82.9%	91.0%
~ 4 位	64.0%	84.5%	92.0%
~ 5 位	66.5%	85.7%	92.3%

表 33: 入力話者:FMS

	VFS		
	未適応	平滑化なし	平滑化あり
1位	56.3%	69.2%	79.1%
～2位	66.1%	79.1%	88.2%
～3位	70.9%	83.6%	89.9%
～4位	72.8%	85.7%	91.5%
～5位	73.9%	86.7%	91.9%

表 34: 入力話者:FNY

	VFS		
	未適応	平滑化なし	平滑化あり
1位	77.0%	79.6%	89.1%
～2位	85.3%	86.1%	93.8%
～3位	87.0%	88.9%	95.1%
～4位	87.7%	90.2%	95.8%
～5位	89.1%	91.5%	95.8%

表 35: 入力話者:FRS

	VFS		
	未適応	平滑化なし	平滑化あり
1位	62.7%	74.5%	81.0%
～2位	71.8%	85.6%	89.9%
～3位	75.5%	88.9%	92.0%
～4位	76.5%	90.6%	93.3%
～5位	77.6%	91.0%	93.7%

表 36: 女性 9 名の平均認識率

	VFS		
	未適応	平滑化なし	平滑化あり
1 位	60.5%	73.1%	82.9%
~ 5 位	75.9%	87.8%	93.4%

## D 不特定話者モデルの文節認識率と初期モデルに不特定話者モデルを用いた話者適応の文節認識率

不特定話者モデルを用いた場合および VFS 話者適応の初期モデルに不特定話者モデル (SID) を用いた場合の男性入力話者 7 名の文節認識率を以下に示す。表中には、不特定話者認識認識における分散の最適化のために行なった実験結果も示す。

表 37: 入力話者:MAO

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	78.4%	80.4%	80.7%	82.5%	90.6%
~ 2 位	85.7%	87.1%	87.8%	89.1%	94.5%
~ 3 位	88.7%	90.1%	90.3%	90.6%	95.4%
~ 4 位	91.3%	92.0%	92.4%	91.7%	95.7%
~ 5 位	92.2%	92.9%	92.9%	92.2%	96.1%

表 38: 入力話者:MHF

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	71.3%	71.8%	70.2%	77.3%	86.0%
~ 2 位	81.9%	81.7%	81.2%	86.4%	91.7%
~ 3 位	86.6%	86.0%	86.3%	88.1%	93.0%
~ 4 位	90.2%	89.6%	89.5%	89.4%	94.1%
~ 5 位	92.4%	92.2%	91.0%	90.5%	95.0%

表 39: 入力話者:MHO

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	75.5%	74.6%	72.3%	59.4%	76.3%
~ 2 位	87.3%	86.4%	84.5%	71.1%	86.0%
~ 3 位	90.6%	90.2%	88.5%	77.5%	89.6%
~ 4 位	92.4%	91.6%	91.0%	80.1%	91.0%
~ 5 位	93.7%	92.4%	91.9%	81.7%	91.7%

表 40: 入力話者:MIK

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	82.8%	83.5%	83.6%	84.2%	90.6%
~ 2 位	92.0%	92.6%	92.6%	90.2%	95.4%
~ 3 位	94.5%	94.5%	94.7%	91.9%	95.8%
~ 4 位	95.7%	96.2%	96.5%	92.9%	95.9%
~ 5 位	96.2%	96.6%	96.9%	93.4%	96.1%

表 41: 入力話者:MSH

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	77.9%	78.7%	79.6%	79.1%	86.0%
~ 2 位	85.9%	86.4%	86.4%	84.3%	91.0%
~ 3 位	88.7%	88.5%	89.2%	86.4%	93.1%
~ 4 位	90.6%	91.3%	91.5%	87.5%	94.5%
~ 5 位	91.7%	92.2%	92.4%	88.8%	95.7%

表 42: 入力話者:MST

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	84.2%	82.8%	82.4%	78.4%	88.0%
~ 2 位	93.4%	92.9%	93.3%	86.8%	94.0%
~ 3 位	95.7%	95.5%	95.7%	89.6%	96.2%
~ 4 位	96.5%	96.8%	96.8%	90.6%	96.8%
~ 5 位	97.1%	97.1%	97.2%	91.6%	96.8%

## E 入力話者が MAU の場合の認識率

HMM-LR の性能評価等で用いられてきた男性話者 MAU の文節認識結果を以下に示す。表には、初期モデルに MHT の HMM を用いた場合の話者適応 (MHT-VFS)、特定話者認識 (SD)、不特定話者認識 (SID)、および不特定話者モデルからの話者適応 (SID-VFS) 実験結果を示した。特定話者認識用 HMM は、5240 単語および音素バランス 216 単語を用いて学習した。

また、表 46 に、従来評価してきた 279 文節 (SB3) のタスクと、今回評価した 714 文節 (モデル会話 1~10, A, B: MODEL) タスクとの複雑さの比較を示す。

表 45: 入力話者:MAU

	MHT-VFS	SD	SID			SID-VFS
分散の倍率	1.2	1.4	1.0	1.2	1.4	1.2
1 位	80.4%	88.9%	83.5%	83.1%	83.6%	83.5%
~ 2 位	89.5%	94.8%	93.3%	93.8%	93.1%	92.6%
~ 3 位	92.0%	96.9%	94.5%	95.2%	95.1%	93.7%
~ 4 位	92.7%	97.5%	95.7%	95.7%	95.7%	94.4%
~ 5 位	93.6%	97.8%	95.9%	96.6%	96.2%	95.0%

表 46: タスクの複雑性の比較

	MODEL	SB3
文法名	dp10-re8.gra	sp2.gra
音素パープレキシティ	3.6	5.9
単語パープレキシティ	41.8	95.6
ルール数	2052	1407



表 43: 入力話者:MTK

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	68.1%	68.1%	67.4%	73.2%	85.6%
~ 2 位	81.1%	81.1%	81.1%	82.1%	91.3%
~ 3 位	85.0%	85.4%	85.9%	85.6%	93.1%
~ 4 位	87.7%	87.5%	87.5%	87.4%	93.8%
~ 5 位	88.5%	88.8%	88.4%	88.2%	94.0%

表 44: 男性 7 名の平均認識率

	SID			VFS	
	分散 1.0 倍	分散 1.2 倍	分散 1.4 倍	平滑化なし	平滑化あり
1 位	76.8%	77.1%	76.6%	76.3%	86.2%
~ 5 位	93.1%	93.3%	93.0%	89.5%	95.1%