

## 7B-1

マルチテンプレートをを用いた対判別  
組合せ法による男女話者音素認識

北岡 教英, 河原 達也, 堂下 修司

京都大学工学部情報工学教室

## 1 はじめに

大語彙・不特定話者の音声認識のためには、音素単位の認識が望ましい。我々は、その識別クラスの対毎に特徴抽出を行い、対判別の識別結果を組み合わせる手法を提案してきた。一方、音素固有の音響的特徴には個人差があり、特に男女間で大きく、この男女差に対応するために、マルチテンプレートが有効である。

対判別では、対毎に独立した特徴空間を持ち、マルチテンプレート化により、直接的に2クラス間の分離度が改善されると考えられる。本稿では、対判別におけるマルチテンプレート化の方法を提案する。

## 2 対判別に基づく連続型 HMM による認識

連続音声の子音の調音結合に対処するため、先行母音から後続母音までの音節の音声パターンの系列をモデル化して認識を行なう。モデル化には Hidden Markov Model(HMM)を用いる。各モデル毎に、与えられたパターンの時系列の生起確率を計算し、この値が最大のものを認識結果とする。

HMMには離散型と連続型があるが、量子化を伴わない連続型を採用している。連続型 HMM では、入力として HMM の状態におけるパターンベクトルの生起確率を求める。しかし、通常の HMM では、クラス間を分離する概念がない。また、異なるクラス間では、識別に有効な特徴は一般に異なるので、全てのクラス間に共通の変数を用いるのは類似のクラス間での混同を招く。

そこで我々は、状態(クラス)の対毎に最適な特徴抽出を行ない、識別器を構築し、それに基づいて連続型 HMM の出力確率を計算する [1]。個々の識別器では2つのクラスの出力確率の比を計算し、それらを組み合わせることで正規化することにより最終的な出力確率を計算する。

## 3 男女話者への対応

## 3.1 実験条件

認識実験の試料は、男性 48 名、女性 16 名により発声された 708 単語中の子音で、学習用として男性 3088、女性 1855、評価用として男性 1317、女性 620 である。

Phoneme Recognition for Both Sexes using Pair-Wise Discrimination Method with Multiple Templates  
Norihide KITAOKA, Tatsuya KAWAHARA and Shuji DOSHITA  
Department of Information Science, Kyoto University

子音の認識カテゴリとして 27 種類を設定する。それぞれに対応する HMM は、母音定常部を含めて 5~6 状態からなる。HMM の状態の総数すなわち識別クラス数は 90 である。入力音声をフレーム毎に音響分析して、28 チャンネルスペクトラムとパワーの 29 次元ベクトルを得る。入力パターンは、着目しているフレームの前後 8 フレームから得られる 232 次元ベクトルである。

対判別組み合わせ法に用いた対の組み合わせは、全ての対の組み合わせ ( ${}_{90}C_2 = 4005$  対) から、識別に十分となるものを選択した 971 対である [2]。

## 3.2 シングルテンプレートによる認識実験とマルチテンプレートの必要性

テンプレートを単一性別の話者音声、及び男女話者音声を区別せずにまとめて学習したシステムによる認識実験を行った。その平均認識率を表 1 に示す。

表 1: シングルテンプレートによる認識率 (%)

テンプレート	試料			
	男声		女声	
	closed	open	closed	open
男性	89.3	83.1	—	54.7
女性	—	40.1	88.0	72.3
男女	80.8	76.4	78.2	68.5

単一性別による学習では、テンプレートの学習に用いた試料と認識試料の性別が異なる場合、大きく認識率が低下する。また、男女話者共用の音声テンプレートのシステムでも、単一性別のテンプレートと認識試料の性別が同じ場合に比べて closed で 9%、open で 4~7% 程度低下し、十分でない。

一般に、音素固有の音響的特徴には個人差があり、特に男女間では大きい。このため、同一の音素でも特徴空間上で男女間で距離の大きいものが存在する。そこで、対判別にマルチテンプレートを導入する。対毎にマルチテンプレート化する有効性は次の点である。

誤り傾向を反映した改善が可能

男女両性を対象とした場合、一部の音素の認識率が著しく低下している。また、それらも、互いに混同する音素が限られている。混同率の増加した音素の対を重点的に改善することで、全体の認識率の向上につながる。

あるクラス間の分離が他のクラス間に影響しない  
 多クラス一括判別と異なり、対毎に独立した特徴空間を持つので、マルチ化によりその対の特徴空間が変化しても、他の対の特徴空間は変化しない。

3.3 マルチテンプレート化の手法

マルチ化には、以下の3つの方法が考えられる。

方法1：男女共通の特徴空間を用いる方法

マルチ化対  $s_1 - s_2$  において、同一特徴空間上で4つのテンプレート  $\{s_1^m, s_2^m, s_1^f, s_2^f\}$  ( $m$ : 男性,  $f$ : 女性) を置き、混合分布から  $s_1, s_2$  のランク付け及び事後確率の計算を行う(図1)。

方法2：対毎に性別判定を行う方法

マルチ化対  $s_1 - s_2$  において、男女それぞれの特徴空間で  $s_1^m - s_2^m, s_1^f - s_2^f$  それぞれのランク付け、事後確率の計算及びその差の計算を行う。差の大きい方の性別を  $n$  とするとき、 $s_1^n - s_2^n$  の結果を対  $s_1 - s_2$  の結果とする(図2)。

方法3：入力系列に対して性別判定を行う方法

入力パターンに対して話者が男性と仮定した場合と女性と仮定した場合の識別を対判別組合せ法により行う。その際、マルチ化対では男女それぞれの特徴空間での識別結果を、シングル対では共通の識別結果を用いる。これを入力系列全体に対して行って、事後確率の系列を男女別に生成した後、系列から性別判定が行い易い母音の区間を抽出して性別判定を行って、男女いずれかの事後確率の系列を選択する(図3)。

マルチ化対選択方法

効率の低下を防ぐために、マルチ化が必要な対のみをマルチ化することを考える。その際には、男女音声で学習したシングルテンプレートによる各対の判別分析の混同率  $\alpha_1$  とマルチテンプレートによる混同率  $\alpha_2$  を求め、 $\alpha_2 - \alpha_1 < \theta$  ( $\theta$ : 閾値) となる対をシングルテンプレートとし、その他の対をマルチテンプレートとする。

表 2: マルチ化による認識率 (%)

識別方式	$\theta$ , マルチ対の数	試料			
		男声		女声	
		closed	open	closed	open
方法 1	全対,971	85.3	79.8	83.4	70.6
	0.02,348	84.1	79.3	81.6	69.5
方法 2	全対,971	70.8	66.5	78.0	67.1
方法 3	全対,971	87.8	83.7	88.0	72.6
	0.025,348	86.3	81.2	85.0	73.5

3.4 各手法を用いた認識実験

以上の方法を用いた認識実験の結果を表 2 に示す。

同一空間に4テンプレートを置く方法では、シングルと比較して数%の向上が見られた。対毎に性別判定を行う方法では、結果はむしろ低下した。これは、ある対で対象とするクラスに正解がない場合に正確に性別判定が行えないことに起因すると思われる。系列全体を用いて性別判定を行う方法では、男性用・女性用の識別器及び母音区間抽出による性別判定の精度の高さにより、単一性別用のシステムとほぼ同等の認識率を得た。

4 おわりに

本研究では、対判別組合せ法による音声パターンの識別で、男女による音響的な差異に対処するためにマルチテンプレートを導入する手法を検討した。そして、男女別の識別結果を出力する識別器と母音による性別判定器を組み合わせることで、単一性別用のシステムと同等の精度の識別が行なえることがわかった。

参考文献

- [1] Tatsuya Kawahara and Shuji Doshita. HMM based on Pair-Wise Bayes Classifiers. IEEE-ICASSP, 1992
- [2] 小川亨, 河原達也, 堂下修司. 2群対判別法における対の組合せの選択手順. 情報処理学会第41回全国大会, 7M-1, pp.235, 1990.

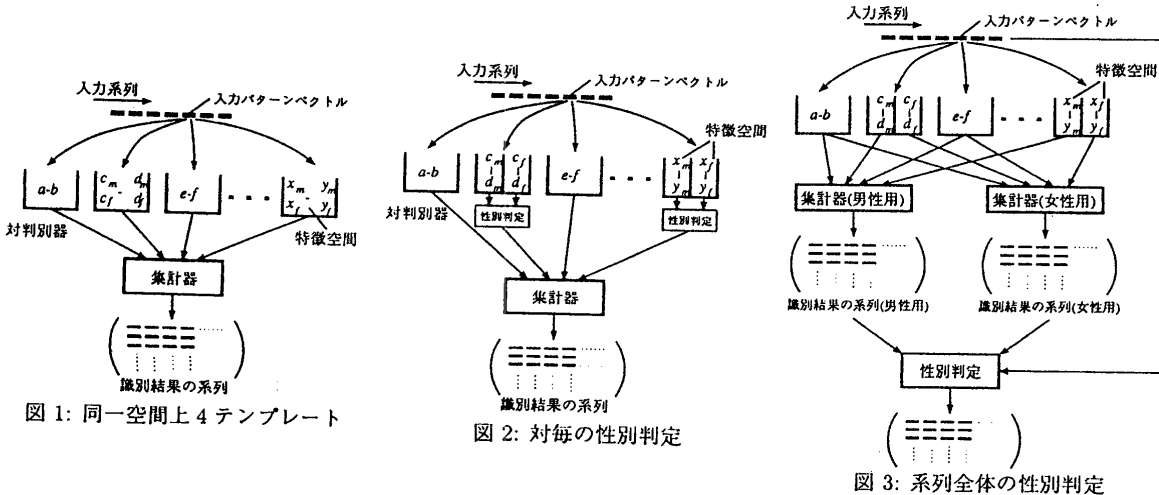


図 1: 同一空間上4テンプレート

図 2: 対毎の性別判定

図 3: 系列全体の性別判定