

帯域パワーを特徴量に加えたHMM によるワードスポッティング

5H-2

小池敏和 青木昌紀 荒井秀一 城戸健一

千葉工業大学 情報工学科

1. はじめに

HMMを用いて連続音声からキーワードをスポッティングする場合に特徴量として、ケプストラムやΔケプストラムに加えパワー成分が良く用いられている。

一方、同スポッティングにおける湧き出し単語絞り込みにおいて、全帯域パワー成分を用いなくてもわずか2帯域だけで十分絞り込めることが確認された。

そこで、本稿ではケプストラムとΔケプストラムに加え2帯域のパワー成分をHMMの特徴量に用いたワードスポッティング結果を報告する。

2. HMMとワードスポッティング法

本稿では音素単位でHMMを作成し、単語の音素記号列順に結合することにより単語HMMを構成し、連続音声からスポッティングした。次に、使用したHMMの構成とワードスポッティングの方法を示す。

2.1 出力確率の確率密度関数

本稿で用いたHMMは平均と分散からなる連続型HMMである[1]。また、各パラメータの推定には確率密度関数として、次式に示す多次元正規分布を用いた[2]。

$$b = \left(\frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \right) e^{-\frac{1}{2}(y-\mu)^t \Sigma^{-1} (y-\mu)}$$

y : 出力ベクトル μ : 平均

Σ : 共分散行列 t : 転置 -1 : 逆行列

Investigations of Key Word Spotting by
Phoneme HMM with Band Power
Toshikazu Koike Masaki Aoki Shuichi
Arai Ken'iti Kido
Chiba Institute of Technology
2-17-1 Tsudanuma Narashino Chiba

2.2 スポッティング法

単語の音素記号列順に、図1のように音素モデルでの最終状態を次の音素モデルの初期状態とすることにより音素モデルを結合して単語モデルを構成する[3][4]。

ワードスポッティングでは、まず、入力フレームに同期して連続 Viterbi アルゴリズムに基づく対数ゆう度計算を行う。次に、最適パス選択は状態経路長で正規化した累積ゆう度の中からゆう度が局所的にピークを示すものを抽出する。さらに、バックトラックにより最大出力を持つパスを候補とする[5][6]。

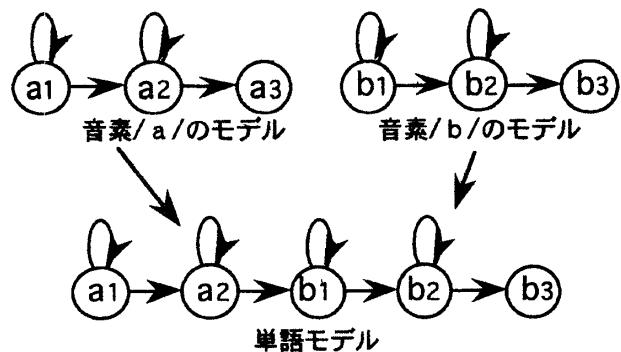


図1 HMMのモデル

3 実験

3.1 実験条件

実験条件を表1に示す。音素モデルを作成するための学習サンプルは東北大学・松下音声データベースから男女各5名の発声した212単語による音素を使用した。また、スポッティングデータはNHKニュースの男性アナウンサーが発声した音声を使用した。

学習は総音素数34音素用い、延べ50単語についてのスポッティング結果を用いて検討した。

表1 実験条件

学習データ	東北・松下データベース
認識データ	NHKニュース
サンプリング	12KHz, 16bit
フレーム長/周期	10ms/5ms
音素数	34音素
キーワード	延べ50単語
特徴パラメータ	LPCケプストラム12次 △ケプストラム12次 △パワー

3.2 湧き出し絞り込みによる帯域パワーの検討

ケプストラムと△ケプストラムと△パワー成分を特徴量とするワードスポッティングの結果からニュースデータ中に複数回発声されたキーワードを用いて、HMMと異なる分析方法（帯域パワーを用いたDPマッチング）で、湧き出し単語絞り込みを行った。その結果を図2に示す。この結果から6帯域のパワー全てを用いなくても2帯域で十分絞り込めることが示されている。

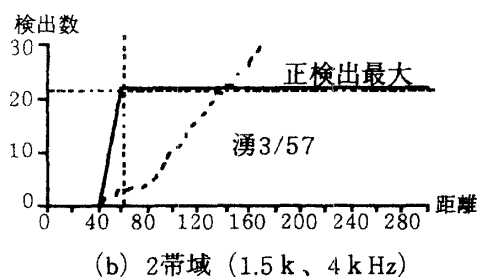
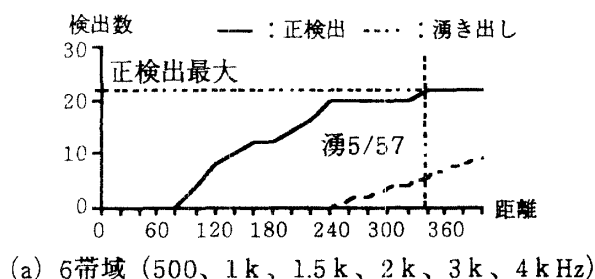


図2 湧き出し絞り込みによる帯域パワーの比較

3.3 ワードスポッティング

上述の結果から、HMMの特徴量に2帯域のパワー成分を付加した実験を行った。その結果を表2に

示す。△パワーの時と比較すると、状態数が同じ条件では、検出率はよいが湧き出し単語が多くなる結果が得られた。このことからキーワード検出には効果があるが、その反面、全帯域パワーに比べ情報量が少なくなったために他の音声部分との切り分けが難しくなったものと考えられる。

表2: キーワードスポッティング

	正検出率 (%) / 湧き出し単語数	
帯域パワー (1.5K, 4K) 状態数4	86	472
△パワー 状態数3	88	467
△パワー 状態数4	82	73

4. 今後

帯域パワーに更に処理を施し (FFTによりケプストラムを求めるなど)、比較検討を進める。

文献

- [1] 小池, 荒井, 城戸: 「多次元相関出力分布を用いた音素HMM学習に関する一検討」音響講論 1-5-5 (1996-3)
- [2] 中川 「確率モデルによる音声認識」電子情報通信学会 (1988)
- [3] 中川, 大黒, 橋本: 「構文解析駆動型日本語音声認識システム」信学 D II No.8 PP1276-1283 (1989)
- [4] 丸山, 花沢, 川端, 鹿野: 「HMM音韻学習で用いた英語音声認識」信学技報 S P 88-119 (1988)
- [5] 鈴木, 今村: 「並列プロセッサによるHMMワードスポッティング」信学技報 S P 90-35 (1990)
- [6] 今村: 「HMMによる電話音声のスポッティング」信学技報 S P 90-18 (1990)