

022048 FBANK を用いた話者選択型不特定話者音声認識

計算機工学講座 C 松浦 祥悟

1 はじめに

先行研究において、複数の話者の音声を 1 つの HMM に学習し、FBANK を用いた不特定話者音声認識の認識率は 85.77% である。そして、認識率の向上が課題とされている [1]。

そこで本研究は、「複数の特定話者の HMM を選択的に用いる」という話者選択型の不特定話者音声認識を試みる。具体的には、話者選択に 2 つの方法を試み、従来手法と認識率を比較する。

2 話者選択型不特定話者音声認識

本認識方式では、あらかじめ特定話者と同様に各話者に対する HMM を作成する。認識する際は、話者に適切な HMM を選択する話者選択を行い、選択した HMM を用いて認識する。話者選択は 2 つの方法が考えられる。

- 1) 教師ありの話者選択：規定の単語を発話し、認識率で HMM を選択する方法
- 2) 教師なしの話者選択：任意の単語を発話し、尤度で HMM を選択する方法

3 評価実験

3.1 学習データと評価データ

データベースとして ATR 単語発話データベース Aset(1 話者につき 5,240 単語)の男女各 10 名を使用し、奇数番号の音声を学習データ、偶数番号の音声を評価データとする。

3.2 実験方法

3.2.1 教師ありの話者選択

以下に教師ありの話者選択の実験の手順を示す。

1. 認識する話者以外の 9 名の話者の音声から各 HMM を作成する。
2. 認識する話者の奇数番号の音声を各話者の HMM を利用して認識する。
3. 2 において、認識率が最も高い話者を選択する。
4. 選択した話者の HMM を利用して偶数番号の音声を認識し、認識率を求める。

3.2.2 教師なしの話者選択

以下に教師なしの話者選択の実験の手順を示す。

1. 認識する話者以外の 9 名の話者の音声から各 HMM を作成する。
2. 奇数番号の全ての単語認識における尤度の総和を求める。
3. 尤度の総和が最も高い話者を選択する。
4. 選択した話者の HMM を使用して偶数番号の音声を認識し、認識率を求める。

3.3 実験条件

本実験では、認識には HTK[2] を使用して実験を行う。特徴パラメータには FBANK、音素 HMM の共分散行列には Diagonal-covariance および、Full-covariance(以下、それぞれ Diagonal, Full と略記)を使用する。また、HMM は半連続型 HMM を使用する。その他の実験条件を表 1 に示す。

表 1: 実験条件

基本周波数	16kHz	
分析窓	Hamming 窓	
1 分析窓長	25ms	
フレーム周期	10ms	
音響モデル	3 ループ 4 状態・半連続分布型	
stream 数	3	
FBANK	24 次 FBANK+	24 次 FBANK+
特徴ベクトル	対数パワー + 対数パワー (計 50 次)	
Diagonal 混合分布数	FBANK 256 対数パワー、	FBANK 256 対数パワー 16
Full 混合分布数	FBANK 128 対数パワー、	FBANK 128 対数パワー 8

表 2: 実験結果 (平均の認識率)

	教師あり	教師なし
Diagonal	79.21%	77.00%
Full	73.18%	71.12%

3.4 実験結果

教師ありおよび教師なしの話者選択の実験結果を表 2 に示す。実験の結果、いずれの場合も Diagonal が認識率が高く、教師ありの話者選択で 79.21%、教師なしの話者選択で 77.00% となった。

従来の不特定話者音声認識による認識率は 85.77% である [1]、本提案の手法は、従来の手法と比較すると認識率は低かった。今後話者適応の手法などを利用して、精度を上げることを検討したい。

4 考察

4.1 教師ありと教師なしの話者選択の比較

教師なしの話者選択の結果は、教師ありの話者選択と比べいずれの場合も認識率は低くなった。しかし差は 2% 程度であり、大きな差はなかった。

4.2 発話数の減少

実験の結果、尤度が高い単語は数単語であった。従って尤度が高い数単語を選択するだけで、同じ認識率が得られると思われる。

5 おわりに

本研究では、話者選択型の方式で不特定話者音声認識を行った。しかし、従来の手法と比較すると認識率が低い。今後は話者適応の手法を用いて認識率を向上させていく予定である。

参考文献

- [1] 堀田, 村上, 池原: “不特定話者における同音異義語音声認識”, 日本音響学会春季研究発表会, 2006. (発表予定)
- [2] Seve Young, et al.: HTK Ver3.2 Reference manual, Cambridge University, (2002).