

AWA 長期間収録音声コーパスを利用した時期差の分析

中原 悠佑[†] 堀内 靖雄[†] 黒岩 眞吾[†]

[†] 千葉大学 大学院融合科学研究科

1. はじめに

話者認識システムにおいて、登録時と認識時の時期差は誤棄却率を増大させる原因であり、時期差への頑健性は重要な課題である。例えば 5kHz 前後の帯域を用いることで、時期差変動に対する安定性が増すことが報告されている[1]。本稿では『AWA 長期間収録音声コーパス』[2]を利用して時期差の分析を行うとともに、この分析結果を利用した話者照合実験を行う。

2. 時期差の分析

AWA 長期間収録音声コーパスに含まれる音声データのうち、今回の分析で用いたデータを表 1 に示す。分析は、日本語 5 母音および鼻音/N/の計 6 音素について、それぞれ独立に行った。

収録日の違い、すなわち時期差によって生じる音響特徴量の変動を分析するために、まずこれらの音声データから得られる音響特徴量を、収録日によって 52 クラスに分類する。そして、このときのクラス間共分散行列を固有値分解して得られる固有ベクトルのうち、対応する固有値の大きなものを基底とする部分空間を構成することで、時期差変動の大きな部分空間を推定する。ここでクラス間共分散行列 S_B は

$$S_B = \sum_k n_k (m_k - m)(m_k - m)^T \quad (1)$$

により求められる。ただし k は収録日、 n_k および m_k はそれぞれその収録日の音声データから得られる音響特徴量の数と平均であり、 m は全ての音響特徴量の平均である。

このようにして得られた固有値最大の固有ベクトルについて、要素ごとに絶対値をとったときのスペクトルを図 1 に示す。いずれの音素についても、4kHz 前後の帯域において大きな値を持っている。

3. 話者照合実験

上述の分析によって得られた時期差変動の大きな部分空間の補空間に音響特徴量を射影することで、時期差変動の抑制された、時期差変動に対して頑健な特徴量が得られる。この特徴量を用いて話者照合実験を行った。

話者照合システムとして GMM-SVM を用いた。音声データとして、科学警察研究所が整備した『大規模話者骨導音声データベース』[3]中の通常マイク収録の音素バランス文を用いた。このデータベースには、約 3 ヶ月をおいて 2 時期分の音声データが収録されており、このうち

表 1. 分析に用いた音声データ

話者数	1 名
収録時期	1 週間ごとに 1 年間、計 52 日
収録時間帯	1 日に 3 回(朝・昼・夕)
発話内容	日本語 5 母音および鼻音/N/ (収録開始時と終了時の 2 回収録)

表 2. 時期差変動を抑制した場合の話者照合精度 (EER [%])

抑制なし	/a/	/i/	/u/	/e/	/o/	/N/
5.27	4.28	4.74	5.01	3.98	4.24	4.49

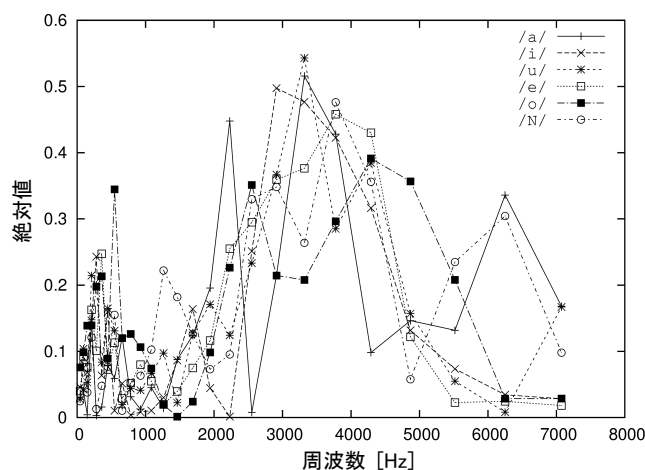


図 1. 固有値最大の固有ベクトル

1 時期目を学習に、2 時期目を評価に用いた。

話者照合実験の結果を表 2 に示す。いずれの音素を用いて部分空間を推定するかに関わらず、時期差変動の抑制によって、音素バランス文に対する照合精度が向上した。

4. 今後の課題

今後は、複数の話者を用いた分析を行う予定である。謝辞

本研究は JSPS 科研費 21300060,24650075 の助成を受けたものです。

参考文献

- [1] 早川昭二ほか, “音声の高域に含まれる個人性情報を用いた話者認識,” 音響誌, vol.51, no.11, pp.861-868, 1995.
- [2] 黒岩眞吾ほか, “AWA 長期間収録音声コーパスと時期差の分析,” 音講論(春), pp.83-86, 2012.
- [3] 蒔苗久則ほか, “大規模話者骨導音声データベースの構築と予備的な解析,” 信学技報. SP, 音声, vol.107, no.165, pp.97-102, 2007.