

ノイズ環境下における音声識別の検討

study on voice recognition in noisy environments

小栗正也，三好力

Masaya Oguri, Tsutomu Miyoshi

龍谷大

Ryukoku Univ.

1. まえがき

近年、スマートデバイスやセキュリティシステムなどで音声識別技術の応用が拡大している。音声識別はハンズフリー操作による利便性の向上や業務効率化といった利点があるが、雑音や残響の影響による認識精度の低下が課題として残っている。特に環境音が入り込む野外や常にマスクをしている医療現場など、雑音・環境音が混入する環境に対応できる認識モデルの構成が求められる。

本研究では、クリーンな音声に複数の雑音や環境音を意図的に付加するデータ拡張手法に着目した。この手法を学習に取り入れ、話者認識の標準的な特徴抽出手法である i-vector[1]を用いて学習・識別を行うことでノイズを考慮した学習の有効性を検証し、その精度変化を評価することを目的とする。

2. i-vector の概要

i-vector とは音声信号から話者固有の特徴を低次元ベクトルとして抽出する方法である。GMM-UBM(Gaussian-Mixture Model-Universal Background Model)[2]に基づき、音声フレームごとの統計量を用いて発話全体を統計的に要約する特徴表現で、従来の GMM スーパーベクトルを高次元空間とみなし、その変動を表す Total Variability 空間(T 空間)を導入することで、話者差を統一的に表現できる点が特徴である。各発話はこの空間内の 1 つの低次元ベクトルとして表現され、PLDA(Probabilistic Linear Discriminant Analysis)などの判別モデルによって話者識別が行われる。本手法は計算コストが比較的安く、学習データが限定的な場合でも高い性能を発揮することから本研究ではノイズ環境下での話者識別に採用した。

3. 実験概要と結果

実験に使用する音声データは日本語話者 100 名からなる JVS コーパスを使用し、音声特徴の抽出および識別アルゴリズムには、オープンソースの音声認識ツールキット Kaldi[3]で i-vector ベースの話者識別を行う。

テストデータには何も付加していないクリーン音声と音声データにホワイトノイズ、帯域ノイズ、環境ノイズを信号対雑音比(SNR)が 10dB となるように付加したノイズ音声の 4 種類を用意し、学習データには「クリーン音声のみ」で学習したモデルと「ノイズ付加音声を含めたもの」で学習したモデルの 2 パターンを作成し、それぞれ実験を行った。

各学習条件における実験結果を表 1 に示す。

表 1 より、ホワイトノイズと帯域ノイズ環境ではノ

イズ付加音声の学習を含めることで識別率が約 10% 向上した。一方で環境ノイズ条件では識別率の向上がほとんど見られず、精度はクリーン学習時とほぼ同等であった。

表 1：各学習条件で実験を行った話者識別率

学習条件 テスト条件	クリーン音声 識別率(%)	ノイズ付加音声 識別率(%)
クリーン音声	85	83
ホワイトノイズ	75	88
帯域ノイズ	79	91
環境ノイズ	80	81

4. 考察とまとめ

実験結果からホワイトノイズ及び帯域ノイズ条件では、ノイズを含めた学習を行うことで識別率が向上したが、環境ノイズ条件では識別率の向上が見られなかった。この結果の原因として、ホワイトノイズ及び帯域ノイズは、スペクトル的に均一または限定帯域の加法的ノイズであり、学習時に同種のノイズを付加することで i-vector 抽出時の特徴のばらつきを効果的に補正できたと考えられる。一方で環境ノイズは音源入り・反射・残響などの複数の要素が混在し、スペクトル構造は非定常であるため単純なノイズ付加によるデータ拡張では環境ノイズ特有の変動に対し十分に学習できなかったため制度改善が見られなかったと考えられる。

以上より、ノイズを含めた学習データの拡張では単純なノイズに対しては有効である一方、複雑で非定常なノイズに対しては十分な効果を示さないことが明らかになった。今後の課題としては単なるデータ拡張だけでなく、複雑なノイズの構造にも応用できるより高度なノイズ除去技術や学習データの設計が必要である。

参考文献

- [1]小川哲司, 塩田さやか, “i-vector を用いた話者認識”, 日本音響学会誌 70 巻 6 号(2014), pp. 332-339
- [2] W. Campbell, D. Sturim and D. Reynolds, “Support vector machines using GMM supervectors for speaker verification,” IEEE Signal Process. Lett., 13, 308-311 (2006).
- [3] Daniel Povey et al., “The Kaldi speech recognition toolkit” in Automatic Speech Recognition and Understanding, 2011.