

ウェーブレット係数を特徴量とする音源分離

1140390 横田優佳 【福本研究室】

1 はじめに

近年、音声認識技術が向上してきている。マイクの直近で 1 人の話者が発話した音声に対しては、高精度な音声認識が可能であるが、複数人が同時に発話した音声に対する認識率は低い。複数音声の分離ができれば、ミーティング等の議事録の自動生成が可能となる。現在、音源分離技術の中で独立成分分析に基づく手法が分離精度が高いためよく用いられる [1]。しかし、この手法は音源数以上のマイクが必要となる。ミーティング等の音源数が未知であることが多い実環境において、独立成分分析に基づく分離は難しい。

一方、信号のスパース性を仮定することで音源数に関わらず 2 つのマイクで録音した音を分離することができる DUET という手法がある [2]。DUET を用いた音源分離において、マイク間の距離とサンプリング周波数がトレードオフの関係にあるという問題点がある。マイク間の距離を数 cm 以内にする必要があり、実環境での応用が難しい。

そこで本研究ではこれらの問題を考慮し、単一マイクで録音した複数話者の音声から所望の音声をウェーブレット変換を用いて抽出するシステムを提案する。

2 音声抽出システム

提案システムの処理の流れを図 1 に示す。単一マイクを用いて混合音声を取得し、観測信号を単一話者区間と複数話者区間の信号に分ける。分割した各信号に対してウェーブレット変換を適用し、係数列を取得する。単一話者区間の係数列を分離のための特徴量として用いる。係数列の相関係数を比較し、相関係数がしきい値 0 に達するまで係数列同士の差分をとる。係数列の相関係数がしきい値 0 に達したとき、差分をとった係数列を用いて信号の再構成を行う。これらの処理により所望の信号を抽出する。

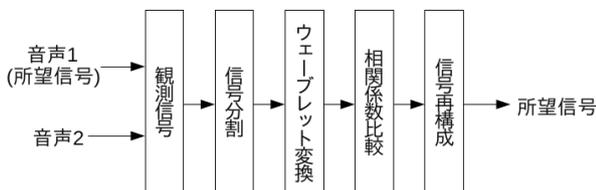


図 1 音声抽出システムの流れ

3 混合音声の分離

提案法が有効であるかを確認するために、男性と女性それぞれ 1 人ずつの混合音声に対して分離の処理を行った。前提として、録音した音声内で単一話者区間およ

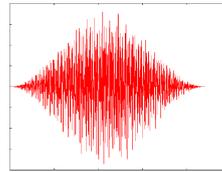


図 2 混合信号 1

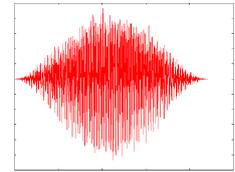


図 3 分離信号 1

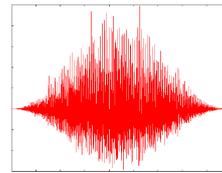


図 4 混合信号 2

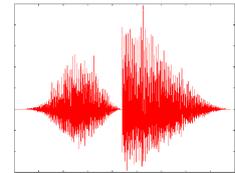


図 5 分離信号 2

び複数話者区間がわかっているものとする。シミュレーションは 2 種類の混合信号に対して行った。1 つ目の混合信号は、男性と女性がそれぞれ別に録音した音声を簡易合成した信号であり、図 2 に示す。2 つ目の混合信号は、2 人が同時に発話している音声を録音した信号であり、図 4 に示す。前者を混合信号 1、後者を混合信号 2 とする。

混合信号 1 に分離処理を行った結果を図 3、混合信号 2 の前半部分に対してのみ分離の処理を行った結果を図 5 に示す。それぞれの分離信号を分離信号 1 および分離信号 2 とする。分離信号 1 および分離信号 2 の音声を確認した結果、音声が分離されていることを確認できた。しかし、所望音声以外の音声が残っている部分があるため、音声の完全な分離はできていないことがわかった。

4 まとめ

本研究では、単一マイクで録音した複数話者の音声から所望の音声を抽出するためのシステムの提案を行った。シミュレーションの結果、完全な分離はできなかったが所望信号以外の音声を軽減することが出来た。より精度の高い分離を行うためには、適切なしきい値を検討する必要がある。

参考文献

- [1] 平井範之, 松本浩樹, 古川利博, 古屋清, “Wavelet 変換を用いたブラインド音源分離の提案,” 電子情報通信学会技術報告, Vol.104, No.456, pp.51-54, Nov.2004.
- [2] 関麻子, 渡部英二, “群遅延に基づく DUET ブラインド音源分離,” 電子情報通信学会技術研究報告, Vol.107, No.476, pp.55-60, Feb.2008.