

発話者の特徴に基づく雑音除去に関する研究

西田義人[†] 田中成典[‡] 安彦智史[†] 木本直樹[†] 足立佳哉[‡]

関西大学大学院総合情報学研究科[†] 関西大学総合情報学部[‡]

1. はじめに

近年、携帯電話などの音声通信装置が普及している。音声通信装置は、屋内に限らず、屋外や車内などの環境下においても利用されるため、音声認識に不可欠な雑音の除去に関する研究が盛んに行われている。代表的な雑音除去の既存研究として、雑音の定常性を利用した研究[1][2]や音響モデルを利用した研究[3][4]が行われている。前者では、雑音の周波数を推定することにより、空調音などの定常雑音を除去できる。しかし、周波数の推定が困難な突発音などの非定常雑音は除去できない。後者では、駅やオフィスなどにおいて周辺環境を想定した音響モデルを作成することにより、定常雑音や突発音を除去できるが、想定と異なる環境では雑音を除去できないため、適応環境が限定される。また、適応環境を限定することなく定常雑音や突発音を除去する手法[5]が提案されているが、発話者の音声と他者の音声を区別できないという問題があり、喧騒などの雑音を除去できない。そこで、本研究では、発話者の音声を特定し、発話者以外の音声を除去することにより、定常雑音や突発音のみならず、他者の音声を含む雑音も除去する雑音除去法を提案する。

2. 研究の概要

本研究では、発話者の音声と雑音が混在した音声データから雑音を除去する手法を提案する。対象とする雑音は、定常雑音、突発音と喧騒を含む雑音とする。本システム(図1)は、1) 音声抽出機能、2) 音源分離機能、3) 発話者音声抽出機能で構成される。入力データは、発話者の音声と雑音が混在した音声データとし、出力データは、雑音除去を施した音声データとする。

2.1 音声抽出機能

本機能では、入力データに含まれる定常雑音

Research for Denoising Based on Features of Speaker
[†] Yoshito Nishita, Satoshi Abiko, Naoki Kimoto
 Graduate School of Informatics, Kansai University, 2-1-1
 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan
[‡] Shigenori Tanaka, Yoshiya Adachi
 Faculty of Informatics, Kansai University, 2-1-1
 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

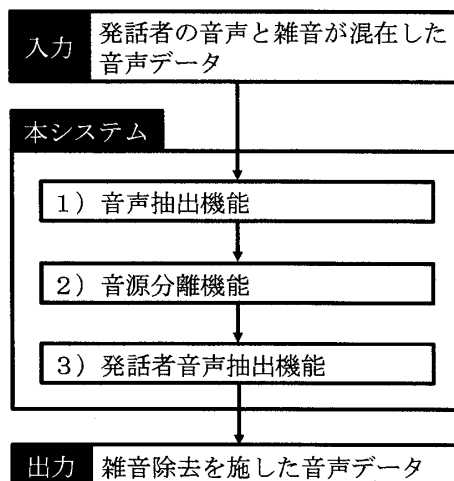


図1 システムの概要

と突発音を除去することにより、音声のみを抽出する。まず、入力データをフーリエ変換することにより、入力データの振幅スペクトルと位相スペクトルを抽出する。次に、抽出したスペクトル情報から、定常雑音と突発音の特徴を基に雑音のスペクトルを推定し、除去を行う。最後に、定常雑音と突発音のスペクトルを除去した振幅スペクトルと位相スペクトルを逆フーリエ変換することにより、入力データから音声を抽出する。

2.2 音源分離機能

本機能では、音声抽出機能で抽出した音声から、発話者の音声と他者の音声を分離する。分離には、音声の個人性を利用する。人の音声は、個人によって周波数領域などが異なるという特徴を持つ。音声の個人性は声紋認証などに利用されており、指紋や筆跡同様に個人を識別する有力な特徴の1つである。そのため、音声の個人性に基づき、発話者の音声と他者の音声を分離する。

2.3 発話者音声抽出機能

本機能では、音源分離機能で分離した音声の振幅スペクトルを比較し、発話者の音声を特定する。本研究では、発話者とマイクが近接して

いる状況を想定している。そのため、分離した音声を比較し、最も振幅スペクトルが高い音声を発話者の音声と判断する。発話者以外の音声を他者の音声とし、除去することにより、発話者の音声を抽出する。

3. システムの実証実験と考察

本システムの有用性を実証するため、入力データに含まれる雑音が、出力データにおいて除去されているかを検証した。

3.1 実証実験

実証実験では、出力データを聴取し、雑音が除去、または、十分に抑制されていると判断できた場合を成功と判断した。また、出力データの波形も表示し、目視による確認も併せて行った。

3.2 結果と考察

入力データの波形 (図 2) と出力データの波形 (図 3) を目視で比較した結果、入力データ中に含まれる雑音をある程度除去した音声データを出力できていることが確認できた。しかし、本システムでは、入力データに含まれる他者の音声除去できていない区間が発生した。また、出力データを聴取したところ、発話者の音声に歪みが生じた区間が発生した。他者の音声の除去が正確に行われなかった原因として、他者の音声の周波数推定が不十分であったために雑音を除去しきれなかったと考えられる。また、発話者音声についても、他者の音声と同様に周波数推定が不十分であったため、発話者の音声を雑音と誤認識し除去したために歪みが生じたと考えられる。

4. おわりに

本研究では、発話者の音声を特定することにより、定常雑音や突発音のみならず、喧騒を含む雑音も除去できる雑音除去法を提案した。実証実験の結果、入力データ中に含まれる定常雑音や突発音、喧騒を含む雑音の除去が確認できたことから、本研究の有用性を示すことができた。しかし、他者の音声と発話者音声の周波数推定が不完全であり、他者の音声の除去と発話者音声の抽出については、より精度を高める必要がある。今後の課題として、他者の音声と発話者音声の周波数を推定する方法を改善し、提案手法の精度向上を目指す。また、本システムでは、発話者とマイクが近接した状況を想定しているが、発話者とマイクが離れていても実現可能な手法についても検討する。

参考文献

[1] 野村行弘, 斗澤秀亮, 呂建明, 関屋大雄, 谷萩隆

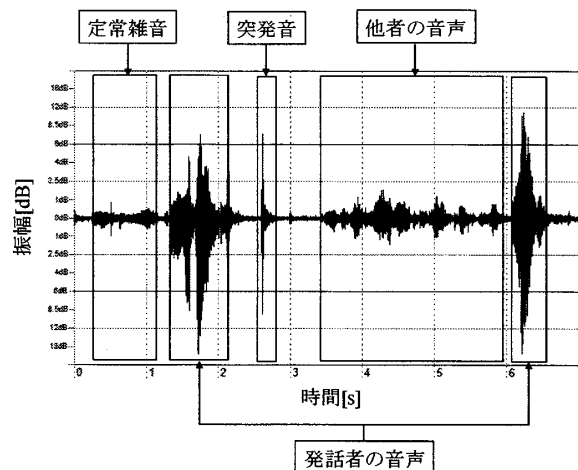


図 2 入力データの波形

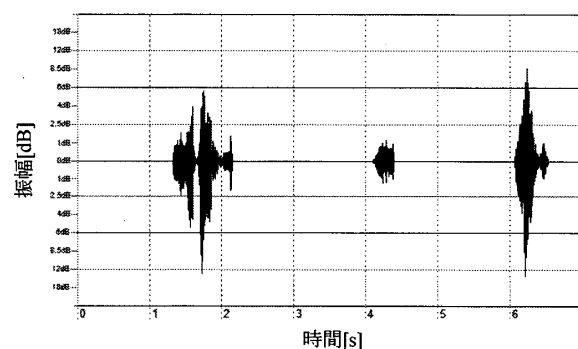


図 3 出力データの波形

詞: モフォロジー処理を用いたスペクトルサブトラクションにおけるミュージカルノイズ除去, 電子情報通信学会論文誌, 電子情報通信学会, Vol.J89-D, No.5, pp.991-1000, 2006.5.

- [2] Gülzow, T., Ludwig, T. and Heute, U. : Spectral-Subtraction Speech Enhancement in Multirate Systems with and without Non-uniform and Adaptive Bandwidths, *Signal Processing*, Vol.83, No.8, pp.1613-1631, 2003.8.
- [3] 伊田政樹, 中村哲: 雑音 GMM の適応化と SN 比別マルチパスモデルを用いた HMM 合成による高速な雑音環境適応化, 電子情報通信学会論文誌, 電子情報通信学会, Vol.J86-D-II, No.2, pp.195-203, 2003.2.
- [4] Flores, J. and Young, S. : Continuous Speech Recognition in Noise Using Spectral Subtraction and HMM Adaptation, *IEEE Acoustics Speech and Signal Processing*, IEEE, Vol.1, No.10, pp.409-412, 1994.4.
- [5] 釜森勇樹, 川村新, 飯國洋二: 位相推定を用いたゼロ位相信号による白色雑音およびインパルス雑音の除去, 応用音響研究会研究報告, 電子情報通信学会, Vol.108, No.69, pp.33-38, 2008.5.