

近傍雑音の除去による遠方発話の強調*

池上拓真[†] 西崎博光[‡] 関口芳廣[‡]

[†] 山梨大学大学院医学工学総合教育部・[‡] 医学工学総合研究部

1 はじめに

近年、雑音抑圧技術の研究が盛んである。特に複数のマイクロホンで観測した音声を用いて、音声の到来方向や高次統計量を利用する手法が多く検討されている。実環境においても、ある程度高い精度で目的音を抽出できる [1]。

しかし、音源分離に関する研究の多くは、目的の音源を観測点から 3m 以内に存在すると仮定して実験を行なっている。また、複数の音源が観測点から同程度の距離に存在する場合の問題について扱っている。実環境では目的の音源が常に観測点の近傍に存在するとは限らない。雑音源と目的の音源との位置関係も常に一定ではない。

例えば、講義、論文発表会等、聴衆が存在する環境では、録音機器は聴衆の邪魔にならない後方に設置することが多い。結果として、発話者と観測点の距離が離れ、その間に聴衆を始めとした不特定の雑音が存在する環境が仮定できる。実環境では、録音機器が必ずしも目的の音源の近くに設置できるとは限らない。

本研究では、目的の音源と観測点が 5m 離れ、かつ音源と観測点の間に非定常な雑音が存在する環境における、スペクトル減算 (Spectral Subtraction: SS) 法による雑音除去と、遅延和ビームフォーマ (Delay-and-Sum Beam Former: DSBF) [2] [3] による目的音強調を組み合わせた手法について検討した。

2 雑音抑圧手法

2.1 概要

仮定する環境の特徴は、次の 3 つである。

1. 目的の音源は観測点から遠方に位置
2. 雑音源は目的の音源と観測点の間に存在
3. 雑音は非定常

仮定した環境において、雑音除去、目的音強調の両面から目的音の高品質抽出を図った。雑音抑圧全体の処理フローを図 1 に示す。処理フローは、雑音除去を目的とする SS 法と音声強調を目的とする DSBF からなる。また、Filtering においては、SS の出力音声から DSBF に用いる入力音声の選別を行なう。

2.2 近傍雑音の除去

SS 法は、本来定常雑音を除去する手法である。周波数領域において、観測音のスペクトルから推定雑

音のスペクトルを減算することにより雑音成分を除去する。雑音を正確に推定できる環境において、効果的な手法である。しかし、推定した雑音の大きさや成分が、観測音に含まれる雑音と異なる場合、ミュージカルノイズという新たな雑音が発生する。

本研究で仮定している雑音は、時間によって特性が変化する非定常雑音である。従って、定常雑音を扱う際の推定方法では雑音を推定することができない。

そこで雑音源が観測点の近傍に位置することに着目し、独立成分分析 (Independent Component Analysis: ICA) [4] [5] による音源分離を利用する。ICA による音源分離の出力音声を観測音から減算することで、観測点の近傍に位置する雑音を除去する。実験では、音声の畳み込み問題を単純化するために、周波数領域 ICA (Frequency Domain ICA: FDICA) を用いる。観測音声と ICA の出力音声の全ての組合せにおいて減算を行なう。

2.3 目的発話の強調

観測点から離れた位置に存在する目的音を取得するために DSBF を用いる。また、SS 法で抑圧しすぎた周波数成分を相互に補い、スペクトル減算によって発生したミュージカルノイズも抑える。

スペクトル減算は、観測音と ICA の出力音声の全ての組合せにおいて行なっている。そこで、観測位置ごとに、スペクトル減算によって音質が向上したと考えられる音声を DSBF に使用した。

DSBF に用いる音声を選別するための尺度として音圧を用いる。観測音声は、雑音の付加により音圧が上昇している。雑音が理想的に除去された場合、音圧は雑音除去前と比べて小さくなる考えた。また、音圧が小さくなる場合、SS によって必要な周波数成分が失われている可能性がある。しかし、DSBF によって周波数成分を相互に補うことができるため、音圧が小さい音声の選択で良いと考えている。

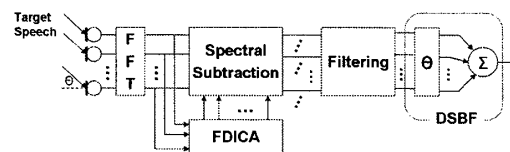


図 1 雑音抑圧処理フロー

*Remote speech enhancement based on the reduction of near noise. by Takuma IKEGAMI, Hiromitsu NISHIZAKI and Yoshihiro SEKIGUCHI (University of Yamanashi)

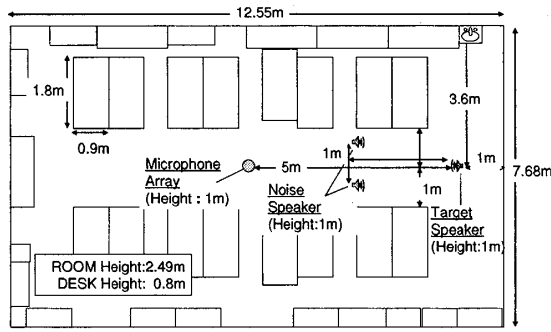


図 2 観測環境

3 遠方の目的音を対象とした雑音抑圧実験

3.1 実験条件

大学構内の教室で観測した音声を用いて提案手法の評価を行なった。音声を観測した環境を図 2 に示す。観測に用いたマイクロホンアレイは、8 素子円形アレイである。円の直径は 0.22m である。

目的音として、新聞記事読み上げ音声コーパス (JNAS) から男性 5 名、合計 20 発話を使用した。観測時は、全ての発話を連結し、5 分の音声として使用した。雑音源として、JNAS の男性の音声を使用した。ただし、目的音とは、発話者、発話内容が異なる音声を選択した。観測は日時を変えて 3 回行った。入力信号対雑音比 (Signal to Noise Ratio: SNR) は、約 0dB である。音響分析の条件を表 1 に示す。

評価尺度として、雑音抑圧度 (Noise Reduction Rate: NRR) を用いた。NRR は、雑音抑圧前後の SNR の差である。雑音抑圧の効果が大きいほど、NRR は高くなる。雑音抑圧実験は、観測音から発話区間のみを切り出した発話単位と非発話区間を含む 5 分間の観測音全体のそれぞれで行なった。

3.2 実験結果

発話単位で雑音抑圧を行なった結果を図 3 に示す。観測音全体で雑音抑圧を行なった結果を図 4 に示す。比較を目的として、観測音に対して SS のみ行った場合の NRR、観測音に対して DSBF のみ行った際の NRR も同時に示す。

表 1 音響分析条件

サンプリング周波数	16kHz
量子化ビット数	16bits
窓関数	ハミング窓
窓幅	32ms(512 点)
フレームシフト	5.3ms(85 点)

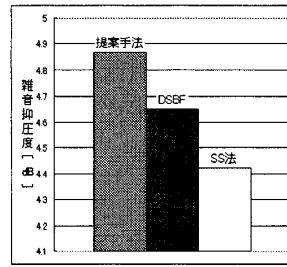


図 3 発話単位での雑音抑圧結果

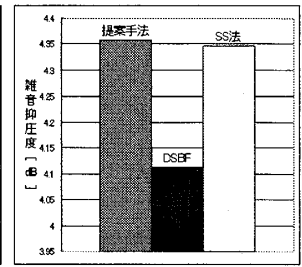


図 4 観測音全体の雑音抑圧結果

3.3 考察

発話単位での実験では、提案手法、DSBF のみの NRR が高い。遠隔の発話に対してフォーカスを向けることで、目的の発話を取得できていると考えられる。それに対して、発話単位の雑音抑圧において、SS 法のみを行なった場合の NRR は低い。今回の実験では、雑音として音声を用いている。音声は非定常な信号であり、時間とともに周波数成分が変化する。非定常雑音の除去は、定常雑音を除去する場合より、ミュージカルノイズが発生しやすい。従って、観測音声に含まれる雑音と推定した雑音のひずみがミュージカルノイズとして発生したと考えられる。

観測音全体での雑音抑圧では、提案手法、SS 法のみ NRR が高い。目的の発話を含まず雑音のみが観測される非発話区間では、DSBF の効果が得られなかったと考えられる。DSBF は、目的の音声を強調することで相対的に雑音を抑圧しているためである。提案手法では、SS 法による雑音除去を行うことで、非発話区間に関しても雑音を抑えることができています。発話単位、観測音全体のいずれにおいても、提案手法が安定して音質を改善できている。

4 おわりに

本稿では、近傍雑音の除去、音声強調を組み合わせることにより、観測点から 5m 離れた目的音の高 SNR 取得について述べた。複数の手法を組み合わせることにより、4dB 程度の NRR を得ることができた。しかし、DSBF に用いる音声の選択によっては、さらに SNR の改善が期待できる。他の雑音抑圧手法についても、さらに検討する必要がある。

参考文献

- 高橋祐, 大迫慶一, 猿渡洋, 鹿野清宏: プラインド空間的サブトラクションアレイによる駅環境音声認識, 日本音響学会講演論文集, 1-7-18, pp.653-654, 2007.9
- 浅野太, 速水悟, 松井俊浩: 遅延とアレイを用いた音源方向同定雑音抑圧システムの試作とその音声認識への応用, 電子情報通信学会, 信学技報, Vol.96, No.437, pp.23-30, 1996.12
- 田中雅史: マイクロホンアレイ処理技術, 電子情報通信学会, 信学技報, Vol.95, No.318, pp.1-8, 1995.10
- 村田昇: 入門独立成分分析, 東京電機大学出版局, pp.28-34, 2004.7
- 牧野昭二, 荒木章子, 向井良, 澤田宏: 独立成分分析に基づくプラインド音源分離, 電子情報通信学会, 信学技報, Vol.103, No.129, pp.17-24, 2003.6