

どんな研究

音に溢れた環境において機械が人の声を正しく認識するためには、聞きたい人の声のみを取り出す音声強調技術が必要です。本研究では、**移動しながら話す状況でも、人の移動を捉え、聞きたい人の声を聞き取り続ける**ビームフォーミングの技術を紹介します。

どこが凄い

ビームフォーミングでは音源や雑音の到来方向の情報（空間情報）が必要です。本研究では時間変化する空間情報の推定問題を移動に応じて発話を分割する問題と捉え、**深層学習を用いて解く手法を提案することにより、目的音源が移動しても高精度な空間情報の推定が可能となりました。**

めざす未来

現在でも、比較的静かな状況では機械は人の声を高精度に認識できるようになりましたが、さらなる技術発展により、雑音が多い、複数人が自由に話す、歩き回りながら話す、といった**どのような状況の中でも、人と機械が自然にやり取りできる未来**をめざします。

ビームフォーミングによる音声強調

■ ビームフォーミングとは

- 複数のマイク(マイクアレイ)を利用
- 目的音方向の音を強調



- 雑音を含む実環境における音声処理の基盤技術
→ スマートスピーカーなど商用デバイスで広く利用
→ 近年の音声認識チャレンジなどでも必須の技術
- 高精度な音声強調には空間情報の推定が重要
(空間情報：目的音や雑音の到来方向に関する情報)

課題

■ 音源移動への対応

- 目的音源が移動すると空間情報が変化
→ 時間変化する空間情報の推定は困難

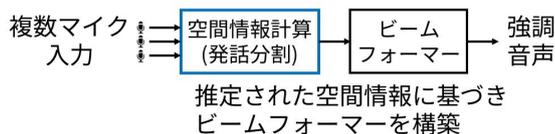
移動音源に対するビームフォーミング

■ 目標

- 話者が移動していても、**高精度な音声強調**を可能にしたい

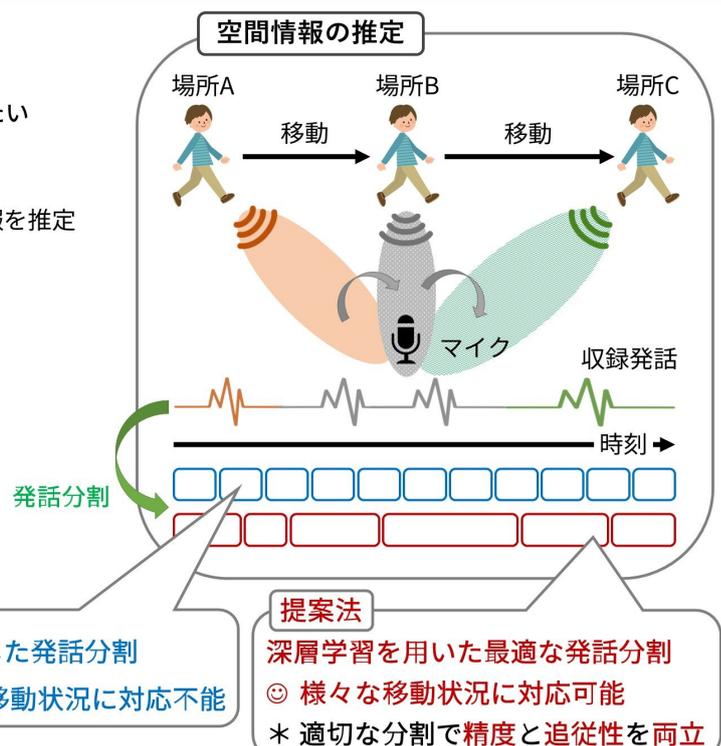
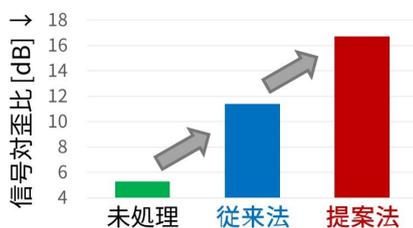
■ 提案手法

- 時間変化する空間情報の最適な推定方法を提案
→ 移動方向や移動速度に応じて発話を分割し、空間情報を推定
- 「最適な分割」により「空間情報の最適な計算」を実現



- 発話分割に深層学習の**注意機構モデル**を応用
→ 時刻毎の空間情報の**類似度**に基づき、**発話を分割**

■ 移動音源に対する音声強調実験



関連文献

[1] T. Ochiai, M. Delcroix, T. Nakatani, S. Araki, "Mask-Based Neural Beamforming for Moving Speakers With Self-Attention-Based Tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 31, pp. 835-848, 2023.

連絡先

落合 翼 (Tsubasa Ochiai) メディア情報研究部 信号処理研究グループ