

論文 / 著書情報
Article / Book Information

題目(和文)	マイクロホンアレイベース目的音抽出の上限性能の解析
Title(English)	Exploring Upper-Bound Performance of Microphone-Array-Based Target Sound Extraction
著者(和文)	廣江厚夫
Author(English)	Atsuo Hiroe
出典(和文)	学位:博士(工学), 学位授与機関:東京科学大学, 報告番号:甲第268号, 授与年月日:2025年3月26日, 学位の種別:課程博士, 審査員:中臺 一博,田中 正行,川上 玲,原 精一郎,石崎 孝幸
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Institute of Science Tokyo, Report number:甲第268号, Conferred date:2025/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第		号	学位申請者氏名		廣江 厚夫	
論文審査 審査員		氏 名		職 名		氏 名	職 名
	主査	中臺 一博		教授	審査員	石崎 孝幸	准教授
	審査員	田中 正行		教授			
		川上 玲		准教授			
		原 精一郎		准教授			

論文審査の要旨（2000 字程度）

本論文は、「Exploring Upper-Bound Performance of Microphone-Array-Based Target Sound Extraction（マイクロホンアレイベース目的音抽出の上限性能の解析）」と題し、英文全6章で構成されている。 第1章「Introduction(序論)」では、研究の背景と動機について述べられている。具体的には、目的音抽出（TSE）の主要な方式としてビームフォーマー（BF）と深層学習（DL）ベース手法があることを示し、これらを組み合わせたハイブリッド方式の重要性が指摘されている。これに基づき、本論文で扱う研究の焦点をハイブリッド方式 TSE に置くことが述べられている。 第2章「Literature Review（関連研究）」では、音源分離や音声強調などの関連分野と対比を通じて TSE の定義が行われた上で、従来の TSE を線形方式と非線形方式に分類し、それぞれの利点および欠点が整理されている。また、線形方式代表例として BF、非線形方式の代表例として DL ベース TSE を挙げ、両者が相補的な関係にあるため、それらを組み合わせることで高性能な TSE が可能となることが述べられている。さらに、このような組み合わせとして、DL ベース TSE の出力を BF の参照信号として利用するカスケード構成によるハイブリッド方式が有望であると主張している。加えて、上述のハイブリッド方式に対して、構築と解析の2つのフェーズに分けて課題を整理し、本論文の貢献は、あらゆる DL ベース TSE と組み合わせ可能な新しい BF を提供し、TSE の性能と有用性を向上すること、マスクベース BF を用いて最適な TSE システムを設計するための指針を提供することであることが述べられている。 第3章「Construction phase: developing a novel BF that can accept any DL-based TSE output as a reference（構築フェーズ: 参照信号として、任意の深層学習ベースの TSE の出力が許容できる新規 BF の開発）」では、上述のハイブリッド方式 TSE を構築する際の課題である任意の DL ベース TSE の出力を参照信号として受け付け可能であること（課題 1）、その参照信号よりも高性能の目的音抽出を達成すること（課題 2）を解決するため、similarity-and-independence-aware beamformer（SIBF）が提案されている。SIBF は、DL ベース TSE の出力を振幅スペクトログラムに統一することで課題 1 の解決を、また、参照信号との依存性と音源間の独立性の両方を考慮した確率モデル（音源モデル）に基づき音源抽出を行うことで課題 2 の解決を図る手法であり、SIBF の出力を DL ベース TSE に反復入力することでさらなる性能向上を図る手法も同時に提案されている。これらの提案手法に関して、CHiME-4 データセットを用い、3 種類の音源モデルを用いた実験による有効性の実証が行われている。加えて、SIBF の理論的解析を行い、SIBF がマスクベース BF と数学的に同一の式に帰着することを示す一方、両者のマスクの解釈の違いが明確化されている。この解析により、独立成分分析とマスクベース BF が共通の枠組みで理解できることが理論的に示され、次章で扱うマスクベース BF の上限性能の解析が SIBF の性能向上のためには不可欠であることが導かれている。 第4章「Analysis phase: exploring peak extraction performance for each mask-based BF and corresponding optimal mask（解析フェーズ: 各マスクベース BF とその最適マスクに対するピーク抽出性能解析）」では、構築フェーズで得られた知見をもとに、マスクベース BF の上限性能の解析が行われている。具体的には、(1) どのマスクベース BF が最高性能を達成するか、(2) 最高性能に対応する最適マスクは共通か、(3) マスク値に対する制約は性能低下につながるか、という3つの問いに対して、マスクベース BF の全バリエーション（9 通り）を統一的な枠組みで比較し、理論的および実験的な検討が行われている。また、この過程で、マスクベース BF の性能向上のため、周波数ごとの振幅や位相を調整する後処理として「マスクベーススケーリング」も提案されている。上述の問いに対し、従来は、(1) BF ごとに最高性能が異なり、(2) 最適マスクは共通であり、(3) マスク値を 0 以上 1 以下に制限すると性能が低下すると考えられていたが、CHiME-4 データセットを用いた実験の結果、(1) 9 通りの BF バリエーションはいずれも同一の最高性能を持つこと、(2) 最適マスクは BF のバリエーションごとに異なること、(3) フィルター推定用のマスクは 0 以上 1 以下に、スケーリング用のマスクは非負かつ平均を 1 に制限しても、上限性能に到達可能であることを示している。これらの結果は、従来の想定とは正反対の新規発見であり、マスク値の制約が抽出性能に与える影響についての理解を深める成果であ
--

ると考えられる。

第5章「Discussion (考察)」では、前章の解析結果がSIBFおよびマスクベースBFの設計・適用に与える影響について考察が行われている。特に、9通りのBFバリエーションが同一の最高性能を持つことの理論的背景や、SIBFの性能向上に向けた新たな方向性について詳細な議論が行われている。

第6章「Conclusion (結論)」では、本論文の研究成果が総括され、今後の課題が整理されている。

以上を要するに、本論文は、TSEにおけるBFとDLベース手法のハイブリッドフレームワークを提案し、その構築と上限性能の解析を行ったものである。特に、新規BFであるSIBFの提案と、それを支える理論的解析は、TSEの性能向上に資する重要な成果である。さらに、従来の想定を覆す形でマスクベースBFの性能に関する新たな知見が得られた点は、本論文の大きな貢献といえる。本研究の成果は、TSEのみならず、音源分離やロボット聴覚、さらには音声処理全般に対して理論的・実用的な示唆を与えるものであり、博士(工学)の学位論文として十分な価値を有すると認められる。

注意:「論文審査の要旨及び審査員」は、東京科学大学リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。