

氏 名	岩 崎 宣 生
学 位 の 種 類	博 士 (情報工学)
学 位 記 番 号	情工博甲第295号
学位授与の日付	平成27年3月25日
学位授与の条件	学位規則第4条第1項該当
学位論文題目	リアルタイム性を考慮したスパース尺度に基づく 音源到来方向推定に関する研究
論文審査委員	主 査 教 授 井 上 勝 裕
	〃 尾 知 博
	〃 古 賀 雅 伸
	〃 五反田 博

## 学 位 論 文 内 容 の 要 旨

テレビ会議で明瞭な対話を行えるようにするため、また、音声認識システムの認識率を向上させるための前処理として、高性能な雑音除去法の開発が望まれている。雑音を抑圧する技術として、適応ビームフォーミング (ABF: Adaptive Beam Forming) により指向特性を制御する方法がある。これは、複数のマイクロホンを用いて、音の空間的な情報である到来方向 (DOA: Direction Of Arrival) を推定し、推定方向に指向特性を向けることで雑音を抑圧する方法であるが、その雑音除去能力はDOAの推定精度に左右される。そのため、ABFの概念を発展させたMUSIC (Multiple Signal Classification) 法が標準的なDOA推定手法として、レーダーやソナー、地震探査などの分野で利用されている。しかし、MUSIC法は、マイクロホン数が信号源数より少ない状況では正常に機能しない。この問題は、マイクロホン数を増やせば解決できると考えられるが、マイクロホン数の増加とともに、計算量が増加し、システムの規模が大型化するという問題が生じる。これを回避するため、近年、マイクロホン数が信号源数より少ない場合でも、信号のスパース性を利用してDOAを推定する方法が提案されている。スパース性とは、すべての時間周波数で支配的な信号は高々1個しか存在しないという性質である。このDOA推定法は、スパース性に基づく音源分離法として提案されたDUET (Degenerate Unmixing and Estimation Technique)法を発展させたものであることから、多数のデータをため込んでクラスタリングを行う必要があるが、クラスタリング等のバッチ処理を必要とする方法では、突発音などの瞬間的な音や、移動音源のDOAを推定することは難しい。

以上の観点から、本論文では、広く普及しているステレオマイクロホンの使用を前提に、スパース性に基づいてフレーム毎に音源のDOAを推定する方法を新しく提案したものである。フレーム毎にDOAが推定できれば、話者が揺れ動きながら話すときや、マイクロホン内蔵の携帯端末の対面角度が変化するときでも、話者のDOAをリアルタイムに追跡できることが期待される。

本論文の構成は以下の通りである。

第1章では、研究の背景を述べ、本論文の提案内容を概観している。

第2章では、ABF原理の流れをくむDSB (Delay and Sum Beamformer) 法とMUSIC法、DUETに基づくDOA推定法について、それぞれの根本原理やアルゴリズムを説明するとともに長所や短所を以下のように明らかにしている。

- ・ DSB法はアルゴリズムが簡単であるが推定精度は低く単一音源のときにしか機能しない。
- ・ MUSIC法はDOA推定精度が高いが、マイクロホン数が音源数より少ない状況では機能しない。
- ・ DUETに基づくDOA推定の場合、マイクロホン数が音源数より少なくても機能するが、DSB法やMUSIC法と同様にバッチ処理による方法であるため、移動音源のDOA推定に適用することは難しい。

第3章では、音声のフレーム毎 (64msec) のDOA推定に関して予備的な検討を行っている。音声のDOAに関する情報は音声区間だけに含まれ、無音区間には含まれないが、現実には、無関係な雑音成分が無音区間に重畳する。そのため、少数フレームでDOAを推定する場合、雑音の影響を受けることが懸念される。そこで、音声区間内の連続する20フレームをもとに、MUSIC法とDUETに基づくDOA推定法で推定した場合、どの程度の推定精度が得られるかをシミュレーションにより予備的に検討し、局所DOAとCramer-Rao Boundの観点から、DUET原理に基づくDOA推定法のMUSIC法に対する優位性を明らかにしている。

第4章では、以上の検討結果をもとに、フレーム毎のDOA推定法を提案している。マイクロホン収録音は、1音源フレーム、複数音源フレーム、無音源フレームに分けられる。これらのフレームにおける、局所DOAの頻度分布の形状は明確に異なり、1音源フレームでの分布は1つのピークをもつ単峰的な分布となる。そのため、頻度分布のピークを探索することにより音源のDOAを推定できる。従って、単峰的な分布では高い値をとり、双峰的な分布、平坦な分布の順に低い値をとるHoyerの尺度は、1音源フレームの選別に有効であり、この性質を利用すれば、適切な閾値を設けることで1音源フレームが選別できる。以上をもとに、当該フレームの局所DOAの頻度分布に対するHoyer尺度を計算して1音源フレームを選別した後、分布のピークを探索して、ピークとなるときの方向をDOA推定値として採択する方法を提案している。

第5章では、実環境下でシミュレーションを行い、上述の提案法の有効性を検証している。まず、単一音源を対象にしたシミュレーションを行った結果、提案法はブロードサイド方向(マイクロホンの正面方向)から $\pm 30^\circ$  以内なら、DOA推定誤差が $2^\circ$  未満に収まることを確認した。次に、目的音源1個、妨害音源2個の3音源環境下に適用した場合、目的音源の方向がブロードサイド方向から $\pm 30^\circ$  の範囲、残響時間が200msec以下、SN比が15dB以上であれば、目的音源のDOAに対する推定誤差は概ね $3^\circ$  未満に収まることが確認され、提案法の有効性が検証できた。

第6章では、本研究で得られた結果について総括するとともに、今後の課題についてまとめている。

## 学位論文審査の結果の要旨

アレイセンサを用いた雑音抑圧原理として、適応ビームフォーミング (ABF: Adaptive Beam Forming) がよく知られている。ABFでは、まず目的音源の到来方向 (DOA: Direction Of Arrival) を推定し、その DOA推定方向にアレイセンサの指向特性を向けて空間的なフィルタを形成することで、雑音（目的信号源以外の信号源からの伝搬波）を抑制している。この場合、ABFの雑音除去能力はDOAの推定精度に大きく左右されることから、これまでDOAを高精度に推定する方法の開発が精力的に進められてきている。

本論文は、広く普及しているステレオマイクロホンの使用を前提に目的音源のDOAをリアルタイムに推定する観点から、個々の時間周波数点では支配的な信号は高々1個しか存在しないとする音のスパース性を利用して、フレーム毎に目的音源のDOAを推定する方法に関して検討を行い新推定法の構築に関する研究をとりまとめたものである。

本研究において得られた成果を要約すれば次の通りとなる。

1. ABFの流れをくむDSB (Delay and Sum Beamformer) 法やMUSIC (Multiple Signal Classification) 法によるDOA推定について、その基本原理やアルゴリズムを説明して、DSB法は簡単なアルゴリズムで実行できるが推定精度が低く単一音源のときしか適用できないことや、MUSIC法はそれに比べてDOA推定精度が高く複数音源のときでも適用できるが音源数が既知でマイクロホン数より少ない状況でしか機能しないことを明らかにしている。
2. スパース性を利用したDUET (Degenerate Unmixing and Estimation Technique)原理に基づくDOA推定について、基本概念やアルゴリズムを説明し、この場合、マイクロホン数は音源数より少なくても機能するが、すべての時間周波数点で局所DOAを求めてクラスタリングする必要があることから、十分な精度をもってDOAをリアルタイムに推定することは難しいことを明らかにしている。
3. 音声は音声区間と無音区間を繰り返す断続波であることから、現実には雑音の重畳する無音区間は除いて、音声区間内の20フレームのみを用いてDOAを推定した場合、MUSIC法による推定値はDUET原理による推定値に比べて大きくばらつくことを示すとともに、その原因をCramer-Rao Boundの観点から明らかにしている。
4. 複数音源のマイクロホン収録音について、1音源フレーム、複数音源フレーム、無音源フレームにおける局所DOAの頻度分布を調べ、その分布形状が音源数に依存して大きく異なり、1音源フレームでは1つのピークをもつ単峰的な分布になることを明らかにするとともに、上述の3つのタイプのフレームをHoyer尺度に基づいて選別する方法を考案し、その有効性を検証している。

5. 以上の検討結果をもとに、各フレームで局所DOAの頻度分布のピークを探索し、ピークとなるときの方位をDOA推定値として採択する方法を提案し、リアルタイムアルゴリズムとして具体化している。また、提案法の有効性を検証するため、目的音源と2つの妨害音源からなる環境下でシミュレーションを行い、目的音源の方向がブロードサイド方向から $\pm 30^\circ$ の範囲、残響時間が200msec以下、SN比が15dB以上であれば、目的音源のDOAに対する推定誤差は概ね $3^\circ$ 未満に収まることを確認している。

以上のことから、本論文では、スパース性に基づいてフレーム毎に音源のDOAを推定する方法を新しく提案するとともに、その有効性をシミュレーションにより確認しており、今後、多くの分野において更なる発展が期待でき、情報工学の分野に寄与するところ大きい。

本論文に関し、調査委員から、シミュレーション条件や、Cramer-Rao Boundの理論値とシミュレーション結果との関係、ホイヤー尺度の応用範囲、提案手法の新規性と他分野への応用可能性などについて質問がなされたが、いずれも著者から明確な回答が得られた。また、公聴会においても、多数の出席者があり、種々の質問がなされたが、いずれも著者の説明によって質問者の理解が得られた。

以上により、論文調査及び最終試験の結果に基づき、審査委員会において慎重に審査した結果、本論文が、博士（情報工学）の学位に十分値するものであると判断した。