

氏 名	なか たに とも ひろ 中 谷 智 広
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	論 情 博 第 33 号
学位授与の日付	平成 14 年 3 月 25 日
学位授与の要件	学 位 規 則 第 4 条 第 2 項 該 当
学位論文題目	Computational Auditory Scene Analysis Based on Residue-driven Architecture and Its Application to Mixed Speech Recognition (残差駆動型アーキテクチャに基づく音環境理解システムの構築と混合音認識への適用)
論文調査委員	(主 査) 教 授 山 本 裕 教 授 船 越 満 明 教 授 奥 乃 博

論 文 内 容 の 要 旨

本論文は実環境における音の認識技術に関する理論と応用についての研究をまとめたものである。特に、複数の音が同時に発生する状況下で計算機が個々の音を逐次的に分離して認識するための計算機構について研究を行った。このような計算機の聴覚能力は、自律ロボットの耳や実環境での音声認識システムを構築するために不可欠の技術である。

知的な聴覚機構について、心理物理学の分野では人の聴覚について古くから研究が行われてきた。人間は例えば混雑した会場でも選択的に音を聞き分けることができる優れた聴覚能力を有している。これらの成果の一部は、Bregman による「Auditory Scene Analysis (ASA)」などにまとめられている。これに対し、近年になって、計算機科学の分野においても、人間のように実環境で知的な聴覚能力を有する計算機構の実現を目指し「Computational Auditory Scene Analysis (CASA)」の研究がはじめられた。この分野では、混合音から様々な手がかりを用いて一貫性を持った音の流れである音響ストリームを分離し、個々の音を理解する枠組みを構築することが課題である。

本論文では、CASA の実現に向けて、特に、

1. 音響ストリーム分離のための汎用の計算モデルとして残差駆動型アーキテクチャの提案
2. 調波構造、音源方向、時間・周波数特性の手がかりを統合した、頑健な音響ストリーム分離法の構築
3. 音響ストリーム分離の混合音声への適用と、分離音声の認識評価
4. 音オントロジーを用いた音響ストリーム分離システムの統合法の構築と、背景音楽と音声を分離するシステムへの適用に関する研究成果をまとめたものであり、以下の6章からなる。

まず、第1章の序論でCASAの実現に向けた歴史的背景を述べるとともに、従来研究として、音源分離法、統合システムアーキテクチャ、雑音下での音声認識技術をまとめ、さらに、本論文の構成について概観している。続く第2章では、本論文で採用しているシステム構成法の基本となるマルチエージェントに基づく音響ストリーム分離法のコンセプトを説明するとともに、実際に、調波構造を用いた音響ストリーム分離システムの構成に適用して、マルチエージェントシステムの有効性を確認している。

第3章では、音響ストリーム分離のための汎用の計算モデルとした、第2章で説明したマルチエージェントの考えを更に一般化した残差駆動型アーキテクチャ(RDA)を提案している。RDAは、音源数未知の混ざった音から個々の音源を逐次的に分離するための計算モデルである。各音源の次の状態を逐次的に予測する処理モジュール(追跡エージェント)を動的に音源に割り当て、予測と実際の入力との誤差(残差)をもとにシステムの状態を更新する。本章では、さらに、RDAを用いて、音の特徴として調波構造、時間・周波数構造、および音源方向の手がかりを組み合わせた音響ストリーム分離法を構成している。従来の音源分離では、各手がかりを個別に用いてきたため、音源の状態に応じて分離性能が制限されていた。これに対し、各手がかりの弱点を補い合うことで、よりロバストな音源分離を実現している。従来困難であった、バイノーラル録音された混合音(2CH)から、同時に3つまでの調波構造ストリームが適切に分離されることを示している。

さらに第4章では、音響ストリーム分離の応用として聖徳太子コンピュータのプロトタイプを構築している。聖徳太子コ

ンピュータとは、複数の音声を同時に認識するコンピュータであり、従来の音声認識では実現できなかった技術である。調波構造ストリーム分離法をもとに音声ストリームの分離法を構成するとともに、分離音の品質をスペクトル歪、および音声認識の両面から評価している。分離によりスペクトル歪、音声認識率がともに改善されることを示している。

第5章では、異なる音響処理システムを統合するための新しい方法論として音オントロジを用いる方法を提案している。様々な性質を持った音を同時に扱うことは、実環境で音を処理するCASAにとって重要な課題である。オントロジとは、異なるシステム間で知識を共有するための共通語彙として、人工知能分野で利用されている技術である。異なる種類の音の関係を表現する知識源として音オントロジを構築し、音声と音楽を同時に扱えるプロトタイプシステムを構築している。背景音楽つきナレーションから各構成音が分離できることを実験により示している。

最後に第6章で本論文の成果をまとめ、将来の研究課題を検討している。

論文審査の結果の要旨

本論文は、実環境における音の認識技術に関する理論と応用についての研究をまとめたものである。実環境における音声認識や自律ロボットの耳の構築などにおいて、雑音等を含む環境下での音声認識の必要性は非常に高い。これに対し、現在の音声認識技術は、雑音の少ない計算機環境では利用することはできても、複数の人間が同時に発話し得る実環境では著しく認識率が低下するという問題があった。これを解決するためには、雑音に対してロバストな音声認識技術の開発が必要であるが、実環境で起こりえる非定常な雑音を現在のHMM (Hidden Markov Model) をベースとした音声認識で扱うのは非常に困難である。このため、本研究では、音声認識システムのフロントエンドとなる音源分離システムを研究開発し、混合音から目的音声だけを抽出して音声認識性能を向上させるというアプローチをとっている。得られた主な成果は次の通りである。

1. 一般の環境では、未知数の音源が独立に発生するため、それらを処理するためには、音源数未知の混ざった音から個々の音源を逐次的に分離する必要がある。従来、このような状況に対応できる音源分離モデルは提案されていなかった。本研究では、これに対処する新しい計算モデルとして、残差駆動型アーキテクチャ (Residue-Driven Architecture, RDA) を提案した。各音源の次の状態を逐次的に予測する処理モジュール (追跡エージェント) を動的に音源に割り当て、予測と実際の入力との誤差 (残差) をもとにシステムの状態を更新する。本研究では、様々な音源分離問題に実際にRDAを適用することで、提案法の拡張性および汎用性を示した。

2. 音源分離は、音の特徴として調波構造、時間・周波数構造、および音源方向などの手がかりを用いて行われるが、各手がかりは万能ではなく音源の状態に応じて分離性能が制限されている。これに対し、本研究では新しい音源分離法としてこれらの手がかりを組み合わせる方法を用いる方法を構築した。各手がかりの弱点を補い合うことで、よりロバストな音源分離を実現した。さらに、従来困難であった、バイノーラル録音された混合音 (2CH) から、同時に3つまでの調波構造ストリームが適切に分離されることを示した。

3. 音源分離の応用として複数の音声を同時に認識するシステムのプロトタイプを作成した。音声のうち調波構造をもった有声音区間を2の分離法で分離すると共に、無声音区間を有声音の分離の残差で補うことで音声分離を実現した。音源分離により混合音の音声認識率が改善されることを示した。

4. 実環境では、音声以外にも音楽など一般の音が発生する可能性があるため異なる種類の音を扱う音源分離法を統合する必要がある。この新しい方法論として音オントロジを用いる方法を提案した。音オントロジとは、異なる音のクラスとその関係を階層構造を用いて表したネットワークである。音オントロジに基づき、入力音の種類が識別され、異なる分離処理を適応的に割り当てる枠組みを構築した。音声と音楽を同時に扱えるプロトタイプシステムを構築し、背景音楽つきナレーションから各構成音が分離できることを実験により示した。

以上述べるように、本研究における成果は、実環境における音声認識の基礎研究として重要な役割を果たす。混合音を扱う前処理としての音源分離の手法と音声認識への応用を明らかにするものであり、実環境における音声認識に大きく寄与するものである。

よって、本論文は博士 (情報学) の学位論文として価値あるものと認める。また、平成14年2月14日、論文内容とそれに

関連した事項について試問を行った結果、合格と認めた。