

氏 名	やま もと しゅん いち 山 本 俊 一
学位(専攻分野)	博 士 (情 報 学)
学 位 記 番 号	情 博 第 292 号
学位授与の日付	平 成 20 年 3 月 24 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研 究 科 ・ 専 攻	情 報 学 研 究 科 知 能 情 報 学 専 攻
学 位 論 文 題 目	Real-Time Robot Audition Software Based on Missing Feature Theory for Multiple Simultaneous Talkers in Real Environments (実環境における同時発話認識のためのミッシングフィーチャ理論に基づくリアルタイムロボット聴覚ソフトウェア)
論文調査委員	(主 査) 教 授 奥 乃 博 教 授 河 原 達 也 准教授 尾 形 哲 也

論 文 内 容 の 要 旨

本論文は、実環境における混合音のためのリアルタイムロボット聴覚ソフトウェアHARKに関する研究をまとめたものである。ロボット自身の耳（マイクロフォン）で聞くロボット聴覚では、接話型マイクロフォンを利用した音声認識とは異なり、混合音処理が不可欠である。本論文では、混合音処理における音源定位、音源分離、音声認識の総合的处理という課題に対するアプローチが述べられている。

第1章は序論で、ロボット聴覚を定義し、ヒューマンロボットインタラクションの観点から混合音処理について述べ、従来手法のロボット聴覚の問題点を明らかにしている。

第2章では、ロボット聴覚の研究動向、特に、混合音に対する音源定位、音源分離、音声認識の従来研究について概観し、それらの問題点を解決するために3つの課題を提示している。第1の課題は、混合音処理の統合である。第2の課題は、実時間処理である。第3の課題は、様々なロボットや様々な環境で使用可能な事前知識を最小限にした可搬性である。

第3章では、第1の課題に対して、ミッシングフィーチャ理論（MFT）に基づく音源分離と音声認識の統合を提案し、その重要課題であるミッシングフィーチャマスク自動生成について述べている。音源分離後のマルチチャネルポストフィルタで、定常雑音とチャネル間漏洩を推定し、分離音の特徴量の信頼性を評価し、マスクを生成することにより、MFTに基づいた音声認識システムによる音声認識の性能向上を図っている。さらに、分離音に白色雑音を加算することにより、事前知識を必要とせずに分離歪やチャネル間漏洩推定誤差の影響を抑え、音声認識精度のさらなる向上を達成している。HARKはこれらの機能の総体である。2種類の音源定位を実装し、音源分離・分離音認識に与える影響についても評価し、両者の有効性について述べている。最後に、3種類のロボット—SIG2, Robovie R2, Honda ASIMO—にHARKを実装し、残響特性の異なる部屋で、話者間の間隔を変え、2話者と3話者の同時発話認識実験を行い、マイクロフォンの配置やロボットの体形が異なっても、音声認識精度が向上することを確認している。

第4章では、第3の課題に対して、システム中の多数のパラメータ最適化が不可欠であることを指摘し、遺伝的アルゴリズムによる連続値パラメータの最適化について述べている。音源方向や音源数に対して、最適なパラメータ集合を管理することで、使用する環境に応じた最適なパラメータ集合が選択される機構も実現し、評価実験により認識率が向上することを確認している。

第5章では、第2の課題に対して、リアルタイムロボット聴覚ソフトウェアHARKのミドルウェアFlowDesignerによる設計と、モジュールによる実装について述べている。音声認識を除いたHARKの全モジュールは、ミドルウェアで実装し、音声認識システムとはクライアントサーバモデルで実装している。この結果、3話者同時発話認識に対して、音響信号入力終了後から分離および認識の終了までの時間が0.4秒であることを確認している。

第6章では、HARKの応用例として同時発話認識を用いた2種類の応用システムについて述べている。1つは、料理の

同時注文タスクであり、もう1つは音声のみを利用した口じゃんけんの審判タスクである。料理同時注文タスクでは、3人の話者の発話終了後1.9秒で応答が返される。それぞれの応用システムについて設計を論じ、性能評価を行っている。

第7章では、もう1つのHARKの応用例として音の可視化機能を有する録音再生システムについて述べている。“Overview first, zoom and filter, then detailed on demand.”という原理に基づいた録音再生システムのユーザインタフェースの設計について述べ、本録音再生システムを利用した議事録の録音再生実験により、有効性を示している。

第8章では本研究のまとめを行い、ロボット聴覚の研究に残された課題および今後の展開について述べている。

第9章は、結論である。

論文審査の結果の要旨

本論文は、実環境における混合音のためのリアルタイムロボット聴覚ソフトウェアHARKに関する研究をまとめたものである。得られた主な成果は次の通りである。

1. 音源分離から得られる情報を使用して、分離音の特徴量の信頼性を推定することによって、ミッシングフィーチャマスク自動生成技術を確立し、ミッシングフィーチャ理論（MFT）に基づいた音声認識システムにより、混合音に対する音声認識が向上することを示した。
2. ミドルウェア FlowDesigner により、音源定位、音源分離、ミッシングフィーチャマスク自動生成等の各種モジュールを統合し、音声認識システムとクライアントサーバで実装することにより、音声発話終了から認識結果出力までの遅延を0.4秒に抑える実時間処理を達成した。また、音源定位モジュールに複数の技法を実装し、統合手法の柔軟性を示した。
3. 複数の残響特性が異なる環境で、体形の異なる3種類のロボット、複数話者の異なる配置などで、複数話者同時発話認識実験を行い、音声認識精度の詳細な測定とGAを使用したパラメータ最適化により、マイクロフォンの相対的な配置だけが与えられれば、他の事前知識は使用せずに機能するHARKの有効性を示した。
4. 複数話者同時発話の応用として、3人の人が同時に料理を注文しても受け付けるロボットや、口によるじゃんけんの審判ロボットのプロトタイプを実装し、複数話者同時発話認識の応用の可能性を示した。
5. HARKを完成したシステムとしてまとめ上げ、様々なロボットの聴覚機能や音環境可視化システムに応用可能であるという汎用性を示した。

以上本論文は、実環境での複数話者同時発話認識を実時間で行うロボット聴覚技術を確立するとともに、複数の体形の異なるロボットに搭載し、複数話者とのインタラクションのプロトタイプを試作することにより、その効果を具体的に示しており、学術上、實際上寄与するところが少なくない。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20年2月22日実施した論文内容とそれに関連した試問の結果合格と認めた。