氏　名	金　鉉燉（キム　ヒョンドン）

（論文内容の要旨）

本論文は、ヒューマノイドロボットが耳の位置にある２本のマイクロフォンを使用し、ロボットの動きを使用して音を聞き分けるアクティブオーディションに関する研究をまとめたものである。２本という人の耳の数と同じ最小限のマイクロフォンで聞き分ける機能は、ステレオ入力装置が標準装備になっていることから、コストの極めてよい手法であるものの、正面からの音に対して前後の曖昧性が残るので、ロボットの頭部を動かすことにより、その曖昧性を解消する必要がある。本論文では、混合音処理における曖昧性の解消のために、挙動との統合及び視聴覚情報統合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、ロボット聴覚及びアクティブオーディションを定義し、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、ロボット聴覚におけるアクティブオーディションの課題として、事前知識最少の音源定位法の開発という低レベルでの取り組みと人と正対してインタラクションを行うことを可能にする高レベルでの取り組みの必要性を述べている。

第２章では、人の音源定位機構について文献調査、特に、３次元空間定位に関する人のアクティブオーディションについての従来研究を概観し、そこで得られた知見を基に、前節で述べたロボット聴覚におけるバイノーラルアクティブオーディションへの２つのレベルからの取り組みが妥当であることを述べている。

第３章では、ロボット聴覚の研究動向、特に、ディジタル信号処理の観点から音源定位、音源分離、発話区間検出（VAD, Voice Activity Detection）、及び音声認識について従来研究を概観し、２本のマイクロフォンを使ったロボット聴覚の具体的な研究課題として４点を提示している。第１の課題は、ロボットが人と正対してインタラクションを取るために必要な音源定位である。第２の課題は、前後問題の曖昧性解消のためのロボットの頭の動かし方の設計である。第３の課題は、雑音源がある時に正対した話者の発話区間検出である。第４の課題は、顔検出と音源定位との統合による移動話者追跡法である。

第４章では、第１の課題に対して、事前情報や事前学習の不要な白色化相互相関法（CSP, Cross-power Spectrum Phase Analysis）に、EMアルゴリズムを援用して２本のマイクロフォンによる音源定位を考案している。CSP法は、単一音源に対しては有効であることが知られているものの、複数音源では曖昧性が生じ、音源定位の性能が劣化する。これに対して、単一フレーム内では優位な音源は１つであるという時間領域でのスパースネスを仮定し、単一フレーム内で得られた定位情報を基にEMアルゴリズムで接続することにより、複数音源の定位手法を確立している。評価実験により、同時に発話する２名の話者の定位精度が向上することを確認している。

第５章では、第２の課題である２本のマイクロフォンでは判定が難しい音源の前後問題に対して、マイクロフォンが装着された耳介を有するロボット頭部を斜め下と左右に動かすことにより解消する方法について述べている。また、３次元音源定位生成システムを作成し、グラフィカルユーザインタフェースを用いて自由な軌跡で合成音が作成できる複数移動話者追跡のベンチマークデータ生成法を開発し、移動話者追跡実験の再現性を保証する方法を提案している。ベンチマークによる評価の結果、１０度下に首をかしげながら高々１０度左右に頭部を動かせば、同じ左右の動きよりも周辺部分で前後問題の解消が大幅に改善することを示している。また、前章で述べた音源定位システムの定位性能が保証される移動速度の範囲も示している。

第６章では、正対した話者の発話区間検出という第３の課題に対して、複素スペクトル円心法（CSCC法, Complex Spectrum Circle Centroid）を用いて、雑音が一方向から到来するという仮定の下で雑音除去を行い、GMM（Gaussian Mixture Model）によって発話区間を検出する方法について述べている。さらに、SNR最大化ビームフォーマ（Maximum SNR beamformer）による分離を行い、音声認識システムの性能（単語認識精度）が最大１７％向上することを確認している。

第７章では、視聴覚情報統合による移動話者追跡精度の向上という第４の課題に対して、OpenCVによる正面顔の検出と肌色クラスタリングによる横顔の検出を組み合わせた頑健な顔検出方法を開発し、第３章で述べた音源定位と統合することにより、音だけの追跡では7.8度あったエラーが4.9度に削減できることを示している。

第８章では本研究のまとめを行い、２本のマイクロフォンを使ったアクティブオーディション研究で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。

氏　名	金　鉉燉（キム　ヒョンドン）

（論文審査の結果の要旨）

本論文は、実環境において２本のマイクロフォンを使って聞き分けるバイノーラルアクティブオーディションによるロボット聴覚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. ２本のマイクロフォンによる音源定位では不可避な前後判定の曖昧性を解消するために、マイクロフォンが設置されたロボット頭部を動かすアクティブオーディションを開発し、単純な左右の動きよりは、斜め下への動きの方が、比較的小さな動きで曖昧性解消が可能なことを実証し、人に関して従来から知られていた認知科学的知見がロボットにも適用可能であることを示した。
2. ２本のマイクロフォンを使用した単一音源に対して有効な白色化相互相関法（CSP法）による音源定位を基に、同一時間フレーム内では優位な音源が１つであるという仮定の下で、EMアルゴリズムにより複数音源の定位が可能であることを示した。さらに、移動話者の音響データを参照データ付きで作成する技術を開発し、複雑な軌跡の移動に対しても、再現性のある複数話者追跡実験を可能としている。
3. OpenCVと肌色クラスタリングとを組み合せた頑健な顔検出を設計し、様々な方向の顔検出を可能にし、CSP法による音源定位と統合することにより、ロバストな複数話者追跡システムを開発している。この結果、ロボットが常に話者に正対することが可能となり、人とロボットとの新しいインタラクションの可能性を示した。
4. 話者に正対するシステムのために、雑音が正面以外の単一方向から到来するという仮定の下で、複素スペクトル円心法を使用した発話区間検出法を開発し、さらに、SNR最大化ビームフォーマによる音源分離と音声認識に適用し、評価実験により分離音の信号雑音比及び音声認識率が向上することを示した。

以上本論文は、実環境でコモディティハードウェアと言ってもよいステレオ入力装置だけを使用し、かつ、事前知識を一切使用しないバイノーラルアクティブオーディションの基礎技術を開発し、様々な環境での使用が想定されるロボットに適した手法であると考えられる。さらに、複数話者の音源定位や視聴覚情報統合による複数話者追跡などの技術を提案しており、様々な環境での評価を行い、その有効性を具体的に示しており、学術上、実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8月27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。なお，本論文調査に当たっては東京工業大学大学院情報理工学研究科客員准教授中臺一博氏にその一部を依頼した。