

| | | |
|---------|--------------------------|-------------|
| 氏名(本籍) | やまもと きよし (茨城県) | 山本 潔 |
| 学位の種類 | 博士(工学) | |
| 学位記番号 | 博甲第4150号 | |
| 学位授与年月日 | 平成18年7月25日 | |
| 学位授与の要件 | 学位規則第4条第1項該当 | |
| 審査研究科 | システム情報工学研究科 | |
| 学位論文題目 | 実環境下におけるロバスト音声インタフェースの研究 | |
| 主査 | 筑波大学教授 | 工学博士 北脇信彦 |
| 副査 | 筑波大学教授 | 工学博士 椎名毅 |
| 副査 | 筑波大学教授 | 工学博士 水谷孝一 |
| 副査 | 筑波大学助教授 | 博士(工学) 山田武志 |
| 副査 | 筑波大学(連携)助教授(産業技術総合研究所) | 工学博士 浅野太 |

論文の内容の要旨

パソコンやテレビなどを操作するインタフェースは、キーボードやマウス、リモコンなどが主流であるが、これらは利用時に手がふさがってしまったり、ボタンの数が多く操作が煩雑であったりするといった問題点がある。本研究では、人間にとってより自然で違和感のないインタフェースとして、実環境下で音声を用いて機器の操作や指示を行うロバスト(頑健)な音声インタフェースを開発することを目的とする。

音声インタフェースにより機器を操作する場合、口元にマイクロホンを設置し、高いSN比で音声を収音できることは少ない。多くの場合、機器と人間の間に距離がある、反射や残響がある、音声以外の雑音があるなどの理由により、収音時のSN比が低下し音声認識が困難になる。このような状況下で良好な音声認識率を確保するために、本研究では、はじめに要素技術の課題として、発話区間検出法、音源分離法、モデル適応法について検討した。つぎに、これらの技術を統合し、さらにカメラから得られる画像情報を用いることで、実環境での実用にも耐えうるロバスト(頑健)な音声インタフェースを構築した。

発話区間検出法の検討では、従来の入力信号のエネルギー又は音響情報に基づく方法では検出精度が悪いため、Support Vector Machine (SVM) 及び Support Vector Regression (SVR) を用いる方法を提案し、有効性を示した。提案法は音源数や音源のパワー差に関する情報を反映している固有値分布と、SVM や SVR で求めた分離平面や回帰直線を用いることで音声区間検出を実現している。

音源分離に関しては、マイクロホンアレイ出力の音響情報のほかに、カメラを用いた人間の位置情報も利用する方法を提案した。音響情報は ML (Maximum Likelihood) 法を用いることにより音源分離している。音響情報と画像情報をベイジアンネットワークで統合処理することによって分離精度の向上を図った。さらに、部屋の反射や残響などの環境変動に対するモデル適応について検討し、ロバスト性の向上を図った。

最後に、提案した音声インタフェースをリアルタイムシステムとしてロボット (HRP-2) に実装し、有効性を検証した。

審 査 の 結 果 の 要 旨

本論文は、人間にとってより身近な音声を媒体として機器の操作や指示をおこなう音声インタフェースについて検討している。マイクロホンとスピーカが離れて設置される実環境下において良好な音声認識率を確保するため、本研究では、発話区間検出法、音源分離法、モデル適応法について、新しい処理手法を提案した。さらに、提案手法をロボットシステムに実装し、実環境での実用にも耐えうるロバスト（頑健）な音声インタフェースであることを実証した点は、評価できる。

本研究では、話者が発話中は話者、雑音源、マイクロホンアレイの位置が変化しないことを条件としている。歩行中のロボットへの実装などを考慮して、今後は、移動中の音声インタフェースを検討することが望まれる。

よって、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。