

## 審査の結果の要旨

氏名 西村 義隆

本論文は「ロボットによるプレゼンテーション及び音声インタラクションの実現に関する研究」と題し、6章より成る。近年注目を集めているヒューマノイドロボットの一活躍分野としてプレゼンテーションに着目し、非専門家でも簡単で容易なコンテンツ記述で音声インタラクションを含むプレゼンテーションを実現する枠組みと、音声インタラクションにおいて問題となる音声認識誤りへの対応技術を開発している。

第1章「序論」では、日本を中心にして近年研究開発が盛んに進められているヒューマノイドロボットにプレゼンタの役割を果たさせ、マルチモーダルなコンテンツを実現する枠組みを実現するという、本研究の目的を述べている。

第2章「ロボットの研究開発」では、本研究の背景となるヒューマノイドロボットの開発の現状について記している。ヒューマノイドロボットには身体表現や音声などのモダリティを持つものが多く、これらのモダリティを有効に活用する受付や案内などのタスク設定が数多く行われている。しかし、プレゼンテーションをタスクとしたものは少なく、特に簡単な記述でコンテンツを作成する枠組みについての実現例はないことを述べている。

第3章「ヒューマノイドロボット用プレゼンテーション記述言語」では、ヒューマノイドロボットによるMPML-HRと名付けたプレゼンテーション記述言語の提案と実装を行っている。キャラクターエージェントを用いるコンテンツ生成に関する分野では、幾つかのコンテンツ記述言語が開発されている。このようなコンテンツ記述言語を用いることで、C言語などの一般的なプログラミング言語を用いる場合と比べ、簡単に効率よくコンテンツを作成することが可能である。しかし、ロボットの分野では簡単な機構でマルチモーダルプレゼンテーション・コンテンツを生成できるような記述言語が存在しない。そこで、キャラクターエージェントを用いるコンテンツ記述言語を拡張することで、ヒューマノイドロボット用のプレゼンテーションコンテンツ記述言語MPML-HRを提案している。また自然なプレゼンテーションの実現についても考慮し、発話と動作が同時にできる機構と、頭部を自動的に動作させる自律動作機構の導入を行っている。世界初のヒューマノイドロボットであり最も著名なAsimoの動作環境でこの記述言語MPML-HRを実装、動作させている（ホンダ研究所との共同研究による）。これにより機能を実証すると共に、キャラクターエージェントとの比較による心理学的評価を行っている。その結果、ヒューマノイドロボットによるプレゼンテーションはより印象的となり、また実空間で商品説明なども可能なため、有効性があることを示している。

第4章「インタラクション機構の導入」では、提案のプレゼンテーションコンテンツ記述言語の記述容易性という長所を活かしつつ、インタラクション機構を導入している。このインタラクションでは、コンテンツ作成者があらかじめ想定した質問に答えることができることを目的とし、説明箇所を遷移することで実現している。また音声インタラクションを含むアプリケーションで問題となる音声認識誤りのシステムレベルでの解決方法を導

入している。具体的には、音声認識結果に対する信頼度に応じて棄却・聞き返し・確認・受理を行うこととし、誤った遷移をしてしまった場合にはユーザからの指摘により遷移前の状態に戻すこととしている。インタラクションの実現のため、プレゼンテーション記述言語に新たなインタラクションに関する命令群を導入し、ロボットの記号の対話行動制御モジュールを用いた実装を行っている。そして、インタラクションを含むその他のコンテンツ記述言語との比較を行い、プレゼンテーションコンテンツについては提案の記述言語が容易な記述を実現できることを示している。

第5章「ロボットの動作音に頑健な音声認識」では、ロボットの動作音に頑健な音声認識手法を提案している。音声インタラクションを含むアプリケーションでは、音声認識誤りへの対応が重要な課題である。特にプレゼンテーションでは動作音が発生し、この雑音はマイクに近い位置から発せられるため相対的に大きく、音声認識性能を低下させる。この問題に対処すべく、動作音は推定可能である点に着目し、**Missing Feature Theory** に基づいた手法を考案している。動作音をあらかじめ収録しておき、この雑音と現在発せられている雑音を時間領域でマッチングすることにより雑音を推定し、**Missing Feature Theory**で必要となるマスクの生成を行っている。またこの手法では、従来から有効な手法として用いられているマルチコンディション学習による音響モデルを用いた手法と **Spectral Subtraction**を用いた手法とを、白色雑音の重畳を行うことで有効に組み合わせ、加えて**Missing Feature Theory**に基づいた処理を行うことで、従来手法よりも高い認識性能を得ている。最後にシステムへの組み込みを視野に入れたリアルタイムでの実装法について示している。

第6章「結論」では、本論文の成果をまとめ、ロボットによるプレゼンテーションについての展望について述べている。

以上を要するに、本論文はヒューマノイドロボットによる音声インタラクションを含むプレゼンテーションコンテンツの簡単で容易な記述による実現という課題に対し、中位レベルの記述言語を導入し、ヒューマノイドロボットに実装することで、実現性を実証的に提示している。この実現に際し、記述容易性を実現するため、ロボットの動作に応じた命令の導入と、インタラクションにおいて説明箇所への遷移による実現、さらに音声認識誤りへの対応を行うため、システムレベルでの聞き返しや誤り遷移からの復帰、音声認識レベルでのロボット動作音への適応などの点で新技術を開発している。これらはヒューマノイドロボットの新しい活用領域を拓く意味で、電子情報学上貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位論文として合格と認められる。