論文の内容の要旨

論文題目　　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

氏名　　西村 義隆

近年，ロボットが注目を浴びており，研究・開発が盛んに行われている．特に，ヒューマノイドロボットはその姿形から，人間と同じような仕事をこなすことが期待されていると考えられる．ロボットはさまざまなモダリティを有しており，ネットワークにも接続可能なことから，情報提供を行うタスクは有効な活躍分野であると考えられる．実際にヒューマノイドロボットは受付や案内などの分野で数多く活躍している．同様に情報の提供を行うタスクとして，プレゼンテーションがある．スクリーンを用いたプレゼンテーションは，音声による情報提供のみならず，資料を交えた情報提供も可能である．本論文ではヒューマノイドロボットによるプレゼンテーションが，ロボットによるタスクの有効な一分野と捉え，これを実現する方法を提案することを目的とする．

第一に，簡単な記述によりコンテンツを作成できる機構について検討した．ロボットの操作を行うためには，通常，アセンブラなどを用いて作られる制御プログラムが必要である．制御プログラムは複雑であり，専門家でないとプログラムを作ることは難しい．しかし，ロボットを身近なものとしていくためには，簡単な機構で操作できることが望ましい．スクリーンエージェントを用いたコンテンツの生成に関する研究領域では，記述言語を用いて簡単にマルチモーダルコンテンツを作成するための試みが行われている．ロボットでも，スクリーンエージェントと同じような記述言語を用いることで，プレゼンテーションコンテンツを作成することが可能であると考えられる．そこで，スクリーンエージェントを用いたマルチモーダルプレゼンテーションコンテンツを作成するための MPML (Multimodal Presentation Markup Language)をヒューマノイドロボット用に拡張し，MPML-HR (MPML for Humanoid Robots)とすることで，簡単な記述でロボットによるプレゼンテーションコンテンツを作成することを実現した．拡張にあたっては，存在する空間の違いを吸収することに留意した．具体的には，移動命令については，位置を表す二次元の座標の他，ロボットの体の方向を表す引数を加えた．スクリーン上の座標を指し示す対象指示命令はスクリーンエージェントでは移動命令によって実現可能であるが，ロボットでは新たな命令を用意することで対応した．また，ロボットは動作がなく，発話のみがあると不自然であるため，発話と動作を同時に実行できる機構を実現し，発話命令のみがあり，動作命令がないときは，首を自動的に動かすことにより不自然さを解消した．

第二に，音声インタラクションを含むプレゼンテーション機構について検討した．人間によるプレゼンテーションでは，質疑応答を行うことで不明な点を解消し，理解が深まることから，ヒューマノイドロボットによるプレゼンテーションでもインタラクション機能があることが望ましい．コンテンツの記述容易性という MPML-HR の長所を活かしたインタラクション機能の導入ができるよう検討を行った．MPML にはインタラクションに関する機能が用意されているが，音声入力を受け付ける箇所は特定の部分に限られている．MPML の他にもインタラクションコンテンツを作成する記述言語はいくつか提案されてい

るが，例えば，Voice XML は音声によるインタラクションのみを想定しており，XISL ではプレゼンテーションに特化していないため記述量が増える．そこで，既存の記述言語を用いず，MPML-HR を拡張することでインタラクション機能を導入することとした．

プレゼンテーションにおける音声インタラクションでは，聞き逃した箇所の再度の説明や，既に知っている話題の省略，分かりにくい箇所の質問などが予想される．これらの要求に対応すべく，説明箇所の遷移を用いることでインタラクションを実現した．再度の説明では，前の説明ポイントに戻り，説明の省略は省略対象の後の説明ポイントに遷移することで実現できる．また，分かりにくい箇所の説明はコンテンツ作成段階で想定質問コンテンツを構築しておくことで実現可能である. MPML-HR では, ページという概念があり，プレゼンテーションにおけるスライド１枚がコンテンツ記述における１ページに相当する．プレゼンテーションでは通常，スライドごとに１つの話題があるため，説明箇所の遷移を行うにはページ単位で行うことがよいと考えられる．そこで，ページの先頭への遷移により音声インタラクションを実現した．コンテンツの作成段階では，音声認識を受け付ける認識文法や認識を受け付けた際の遷移先の記述を行う．コンテンツの実行段階では，音声認識エンジンを常に走らせておき，音声認識が行われるとシステムへの割り込みが行われ，説明箇所が遷移する．

音声インタラクションでは，音声認識結果に誤りが発生すると予期せぬ対応を行い，意図しない内容となる場合がある．この問題に対応すべく，音声認識誤りに頑健な手法についても検討を行った．音声認識結果に対する信頼度を用いることで，信頼度が低いものは棄却することとした．しかし，信頼度が低いもの全てを棄却してしまうと，何か発話しても無視してしまい，インタラクションが不可能になる．そこで，聞き返しや確認を行う動作を導入した．聞き返しでは，ロボットから再度の発話を要求する．これは，発話があったことは認識しているが，認識結果の信頼性が低い場合に有効である．確認では，認識結果が正しいか，はい，いいえの二者択一の答えが得られるような問い返しを行う．これは，受理するには信頼性に欠けるが，第一候補の認識結果である確率が高い場合に有効である．これらの音声認識誤りへの対応に加え，誤認識により誤った箇所へ説明が遷移した場合には，遷移後一定期間内はユーザからの指摘により，もとの説明箇所へ戻る機構を実装することで，音声認識誤りに頑健なインタラクションを実現した．

第三に，音声インタラクションに必要な音声認識性能を向上させることについて検討を行った．ロボットによる音声認識では，さまざまな雑音の混入により認識性能が低下することが問題である．これを解決するため，接話マイクを用いた音声認識が行われるが，ユーザにとっては煩わしく，ロボット自身のマイクで行うことが理想である．ロボットに混入する雑音には，他の音源から出る雑音，部屋の残響，ロボット自身が発する動作雑音がある．特にプレゼンテーションタスクではロボットの動作が多い．動作音はロボットのマイクに近い位置から発せられるため相対的に雑音レベルが大きく，認識性能に大きな影響を与える．そこで，ロボットの動作音に頑健な音声認識手法についての検討を行った．

ロボットの動作音には定常的な雑音成分と非定常的な雑音成分がある．定常的な雑音に対しては，スペクトル領域において推定雑音を減算する SS(Spectral Subtraction)と音声に雑音を重畳したデータを用いて音響モデルを学習するマルチコンディション学習による音響モデルを用いることで効果があると考えられる．しかし，雑音の大きな環境では，SS により SNR (Signal to Noise Ratio)は向上するものの，歪みが発生する．この歪みが認識性能の低下を引き起こす．また，SS は雑音を除去する処理であるのに対し，マルチコンディ

ション学習による音響モデルは雑音を含んだ音声を学習させている．これを単純に組み合わせると認識性能が低下する．そこで，SS による歪みを抑えるため白色雑音の重畳を行い，SS と白色雑音重畳後の音声データを用いて音響モデルを学習させることでこの問題を解決した．白色雑音の重畳は SNR を低下させ，一見認識性能が低下するようにも思えるが，SS による引き残し成分を平坦化することで認識性能が向上した．

ロボットの非定常成分への適応には，雑音に埋もれた信頼性の低い周波数帯域をマスクし，その帯域情報の認識結果への寄与を小さくすることで認識性能を向上させる MFT (Missing Feature Theory)を用いることについて検討した．MFT では，マスクの生成をいかに行うかが重要な問題であり，マスクの推定には，雑音の推定が必要である．ロボットは自己の動作情報を取得することが可能であり，同じ動作であればほぼ同じ動作音が出力される．つまり，動作情報を用いることで動作音の推定は可能である．そこで，動作音を推定することで MFT を有効に活用することができると考え，非定常成分への適応に用いた．雑音の推定には，あらかじめ収録した雑音と入力雑音を時間領域で一致させることで推定を行った．時間領域での一致の際には，入力信号の動作雑音以外が混入している領域を，振幅の大きさから推定し，この領域を除いてマッチングした．この手法を用いることで，音声などの動作音以外の音を含む入力信号を用いても適切な雑音推定を行うことができた．実験の結果，提案手法を用いることで，従来から有効とさ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よりも高い認識性能を達成した．また，教師なし MLLR との組み合わせにおいても提案手法の有効性を確認した．

提案するプレゼンテーションシステムを実現するため，二足歩行ロボットとして知名度の高いホンダ ASIMO を用いて実装を行った．