博　士　論　文

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

指導教官　石塚 満 教授

東京大学大学院 情報理工学系研究科

48-47417　西村 義隆

要旨

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．

Abstract

Research in robotics is growing and many humanoid robots are being produced. The humanoid robots are remarkable media because they have many modalities and can act in the real world like humans. Humanoid presentations are effective in a real environment, because they can move and look around at the audience similar to a human presenter. Thus, in this paper, I propose a humanoid robot's presentation system that includes the speech interaction.

There are many tools to make multimodal contents using animated agents, but there are no tools for humanoid robots. Thus, I have utilized the technology developed for animated agents. MPML (Multimodal Presentation Markup Language) is a medium-level scripting language allowing many non-specialists to easily write multimodal presentations with life-like animated agents. Like HTML for making Web pages, MPML is expected to serve as a core for describing and producing multimodal contents. I have developed a new MPML version called MPML-HR (MPML for humanoid robots). Using MPML-HR it is easy to write and thus generate a set of humanoid robot behaviors.

A parallel execution mechanism of action, speech and autonomous head motions has been introduced in the presentation system to make the presentation more life-like. The system will have two threads when the sequent two commands are action and speech; these two commands are executed simultaneously. The autonomous head motions are evoked when there is only utterance and no action. The angle of the head motion is generated at random.

While the first version of MPML-HR allows everyone to make presentation contents with humanoid robots, it doesn't support interaction functions with audiences. Furthermore, although several markup languages for interactive systems have been developed, their interaction functions in presentation have been weak or insufficient. Therefore, I introduced an interaction function to the robot's presentation system. It enables to describe speech interaction in MPML-HR. This interactive function of MPML-HR has been realized so as to make use of the merit of MPML-HR, i.e., the easiness of description. In the interaction, the presentation system can deal with audiences' utterances such as requesting to repeat previous explanations or to skip some contents. In order to cope with such interruptions, MPML-HR needs to be extended so that the content designer can specify interruption utterance patterns for speech recognition and the destination point the presentation should move to for each type of interruption. To make it possible to specify the point to move to, the presentation content needs to be divided into short pieces so that the robot can move to the head of one of such short pieces. To divide the contents, the page tag in MPML-HR scripts

is used. Each part embraced by the page corresponds to one slide on the screen. Since each slide can be considered to be one topic in the presentation, this tag is suitable for indicating the point to move to when an audience interrupts.

One issue in handling audience interruptions is that there can be speech recognition errors. If the robot reacts to an erroneous speech recognition result, the presentation will move to an unintended point. Therefore two functionalities are incorporated into MPML-HR. One is to ask the audience back what he/she said or ask the audience back the recognition result is correct or incorrect when the speech recognition confidence is low. The other is to go back to the original point in the presentation when the presentation moved by a speech recognition error and to allow the audience to correct an utterance.

In the presentation, the humanoid should listen to the user's speech by using its own microphones, because it is not realistic to assume that the user always wears a headset in a daily environment. In such a humanoid, "noise" generated by its actuators becomes a crucial problem. The humanoid is basically a highly redundant system, so it includes a lot of motors as well as cooling fans for humanoid-embedded processors to realize human-like behaviors autonomously. These noises can be easily captured by the humanoid's microphones because the noise sources are closer to the microphones than the target speech source. Thus, the SNR (signal-to-noise ratio) of input speech becomes quite low (sometimes less than 0 dB). However, it is possible to estimate these noises by using information on the humanoid's motions and gestures. I have proposed a method to improve ASR (Automatic Speech Recognition) for a humanoid with motor noises by utilizing its motion/gesture information. The method consists of noise suppression and MFT-ASR (missing-feature-theory-based ASR). The proposed noise suppression technique is based on spectral subtraction, and white noise is added to blur distortion of suppression. MFT-ASR improves ASR by masking unreliable acoustic features in the input sound. The motion/gesture information is used for obtaining the unreliable acoustic features. Furthermore, I also evaluated with the acoustic model adaptation technique called MLLR (Maximum Likelihood Linear Regression). Un-supervised MLLR was used for the adaptation. I evaluated the proposed method through recognition of speech recorded by using Honda ASIMO in a room with reverberation. The noise data contained 34 kinds of noises: motor noises without motions, gesture noises, walking noises, and so on. The experimental results show that the proposed method outperforms the conventional multi-condition training technique.

The aforementioned system has been realized with Honda ASIMO, the most famous humanoid robot in the world.

目 次

図 目 次

表 目 次

第1章　序論

1.1　本論文の目的

本論文では，簡単な記述により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を作成し，ヒューマノイドロボットを用いて実現する枠組みの提案を行う．

ヒューマノイドロボットは発話やジェスチャ，移動などのモダリティを持ち，ネットワークへも接続可能なことから，プレゼンテーションのような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．しかし，ロボットの制御プログラムはプログラミングやロボットのインタフェースに関する知識なしに構築することができず，制御プログラムからコンテンツを作成するのは大きな労力を必要とする．そこで第一に，簡単にヒューマノイドロボットを用いたプレゼンテーションコンテンツを記述することのできる言語の提案を行い，システムへ実装することを目的とする．

また，プレゼンテーションではユーザからの質問が行われる場合も考えられる．人間によるプレゼンテーションでも，質疑応答があることで内容の理解が深まる．プレゼンテーションを充実したものとするためには音声インタラクションをできることが好ましい．そこで第二にプレゼンテーション中のユーザからの質問を想定し，コンテンツの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音声インタラクション機構の導入を行うことを目的とする．

音声インタラクションを行うアプリケーションでは，音声認識誤りについての考慮がなされる場合が多い．音声認識誤りが発生すると，ロボットはユーザの意図しない対応をとってしまう．そこで，システムレベルで問い返しなどを行い，音声認識誤りに対応する試みや，音声認識レベルで雑音に頑健な手法を取り入れるなどの試みがされている．本論文では第三に，ヒューマノイドロボットによるプレゼンテーション中の音声認識において特に問題となる動作音に頑健な音声認識手法を提案することを目的とする．

1.2　背景

近年，ロボットが注目を浴びており，研究・開発が盛んである．これまでロボットというと，産業用のアームロボットなど特定のタスクを対象とした機械を指していた．しかし近年，人の形をし，二足歩行可能なヒューマノイドロボットや，動物の形をしたペット型ロボットが作られている．特にヒューマノイドロボットはその姿形から，人間をサポートする仕事を行うことが期待されていると考えられる．そのため，ヒューマノイドロボットはコミュニケーション能力や知能を持ち，人間と同じような振る舞いをできることが理想である．

しかし，その実現には多くの課題がある．例えば，人同士のコミュニケーションにおいて一般的に利用されている音声をロボットが完全に認識することは現在の技術では実現できていない．また，完全な音声認識を実現できたとしても，人間と同じ対応をとるような人工知能は現在の技術では提案されていない．

ヒューマノイドロボットは，人間と同等の能力を達成できない面も存在するが，人間よりも優れた部分も存在する．その一つに，多くの情報を扱える点がある．現在開発されているロボットはネットワークに接続可能なものが多く存在し，またロボットは移動可能なコンピュータとも捉えられるため，記憶された情報を正確に提供するタスクにおいては人間に優る部分がある．また，ヒューマノイドロボットは発話，ジェスチャ，移動などのモダリティを持ち，人間に理解しやすい形で情報提供を行うことが可能である．

そこで，ヒューマノイドロボットを有効に活躍させるためには，ロボットの長所を活かしたタスクを設定することが重要である．実際に，開発されているヒューマノイドロボットは案内や受付など，情報提供に関連するタスクをこなすものが多い．本論文では，情報提供というロボットの長所に着目し，タスクにプレゼンテーションを設定する．プレゼンテーションでは，ロボットのジェスチャや発話といったモダリティを活かし，スクリーンへの資料の提示や実際の物の提示をしながら説明を行うことで，ロボットを有効に活躍させることができると考えられる．また，実空間で動作するロボットの印象は強く，このような側面においてもプレゼンテーションは有効なタスクであると考えられる．

ヒューマノイドロボットを用いてプレゼンテーションを実現する際の問題として，ロボットの制御の困難な点が挙げられる．ロボットを操作するためには専用の制御プログラムの作成が必要であり，プログラミング言語に精通しており，ロボットのインタフェースの知識がなければ作成できない．ロボットを身近なものにし，誰でも簡単に操作できるようにするためには，簡単な記述でロボットの操作ができることが望まれる．そこで，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コンテンツを簡単に記述し，実現する枠組みについて提案と実装を行う．

1.3　本論文の構成

本論文は 6 章よりなる．第 1 章「序論」では，本研究の目的と背景について示す．本論文では近年注目を浴びているヒューマノイドロボットに着目し，ヒューマノイドロボットの有効な活用分野として期待されるプレゼンテーションシステムの実現を目指す．第 2 章「ロボットの研究開発」では，本研究の背景となるコミュニケーションロボットの開発の現状について示し，ロボットの有効な活躍分野について検討する．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」ではまず，関連研究としてアニメーションエージェントを用いたコンテンツ記述言語について示す．アニメーションエージェントのコンテンツ生成に関する研究分野では，数多くのコンテンツ記述言語が開発されている．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，アニメーションエージェントを用いたコンテンツ記述言語を拡張し，ヒューマノイドロボット用のプレゼンテーション記述言語を提案する．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を自動的に動作させる自律動作機構の導入を行う．さらに，アニメーションエージェントと比較したヒューマノイドロボットによるプレゼンテーションの有効な活躍分野を検討するため，心理学的評価実験を行った結果について示す．第 4 章「インタラクション機構の導入」では，第 3 章で提案するプレゼンテーション記述言語にインタラクション機構を導入する．プレゼンテーションでは，視聴者からの質疑に答えることで，内容の理解を深めることができるため，インタラクションは重要な機能である．そこで，プレゼンテーション記述言語の容易な記述という長所を活かしつつ，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ようなインタラクション機構を導入する．また，ロボットの音声インタラクションでは，音声認識誤りが想定されるため，音声認識誤りに対する処理として，聞き返しや確認，誤りの指摘によりコンテンツを音声割り込み前の状態に戻す機構を導入する．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する．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，音声認識レベルで動作音に適応した手法の提案を行う．動作音は推定可能である点に着目し，Missing Feature Theory に基づいた手法を提案する．実験を通して提案手法の有効性を示し，リアルタイムでの実装についても示す．第 6 章「まとめ」では，本論文の結論を示し，今後の展望について述べる．

第2章　ロボットの研究開発

2.1　はじめに

これまで，ロボットというと産業用のアームロボットのように，特定のタスクを対象とし，単純な力仕事や人間ができないような細部の仕事をこなすものが多かった．しかし近年，動物型ロボットやヒューマノイドロボットが数多く開発され，知能やコミュニケーション能力を有するものが数多く存在する．

JIS 規格によると，ロボットは図 2.1 のように分類されている [3]．シーケンスロボット，プレイバックロボット，数値制御ロボットは第一世代ロボットと呼ばれ，あらかじめ定められた通りの動作を繰り返し行うことが可能である．感覚制御ロボット，適応制御ロボットは第二世代ロボットと呼ばれ，何らかの感覚情報をもち，この情報をもとに自己の行動をある程度修正する機能を持つ．学習制御ロボットは第三世代ロボットと呼ばれ，作業経験を学習し行動に反映させる．そして，人工知能によって自らの行動を判断・決定し動作するロボットは第四世代ロボットと言われている．これはまさに，ロボットの開発の歴史を表しているものと考えられる．近年，知能をもち，人間と協調するロボットの研究開発が盛んである．それらのロボットは人間や動物の形をしているものが多く，人と協調するためコミュニケーション機能を持つ．本章では，こうしたコミュニケーションロボットの機能やタスクについて示す．2.2 節では，現在開発されているコミュニケーションロボットについて示す．2.3 節では，ロボットのコミュニケーションのための機能について示す．2.3 節では，コミュニケーション能力を持つロボットの活躍分野について示す．

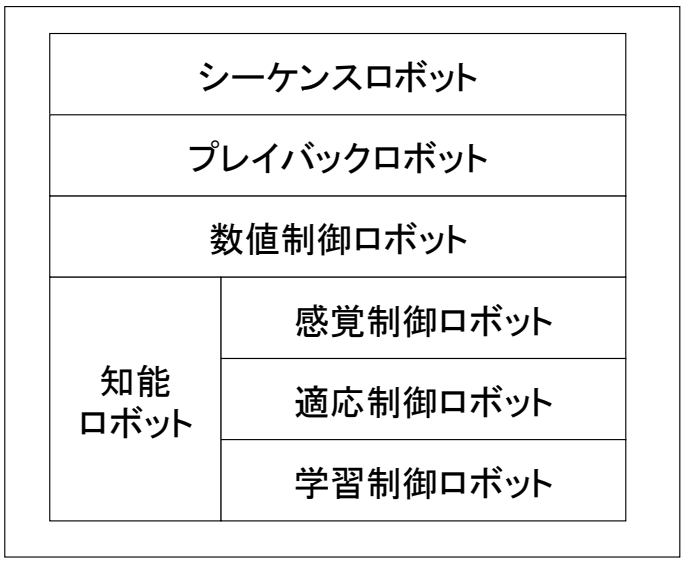


図 2.1: JIS によるロボットの分類

シーケンスロボット：あらかじめ設定された情報（順序・条件及び位置など）に従って動作の各段階を逐次進めていくロボット.
プレイバックロボット：人間がロボットを動かすことによって，順序・条件・位置及びその他の情報を教示し，その情報に従って作業を行えるロボット.
数値制御ロボット：ロボットを動かすことなく，順序・条件・位置及びその他の情報を数値・言語などにより教示し，その情報に従って作業を行えるロボット.
知能ロボット：人工知能によって行動決定できるロボット．人工知能とは認識能力・学習能力・抽象的思考能力・環境適応能力などを人工的に実現したものである.
感覚制御ロボット：感覚情報を用いて，動作の制御を行うロボット.
適応制御ロボット：適応制御機能をもつロボット．適応制御機能とは環境の変化などに応じて制御等の特性を所要の条件を満たすように変化させる制御機能をいう
学習制御ロボット：学習制御機能をもつロボット．学習制御機能とは作業経験などを反映させ，適切な作業を行う制御機能をいう

2.2　コミュニケーションロボット

これまでの産業用機械などのロボットでは，人間がロボット用のコントローラを用いて操作しなければならなかった．ところが，近年開発されているコミュニケーションロボットは，人間にとって扱いやすい音声などによるモダリティを通してロボットへの入力をすることができる．

コミュニケーションロボットは，主に動物の形をしたものと人間の形をしたものに分けられる．動物の形をしたものは，人間のペットとしての存在を意識し，名前を呼んだり触ったりすると簡単なインタラクションを行う知能を備える．人間の形をしたものは，より高度な知能を持ち，人間の発話を認識し，人間をサポートする機能を持つ．本節では，動物の形をしたロボットと人間の形をしたロボット (ヒューマノイドロボット) に分けて，開発されているものを紹介する．

2.2.1　動物型ロボット

動物型ロボットとして注目を浴びたものは，1999 年に発売されたソニーの AIBO であろう [4, 5, 6]．これまでロボットは工場などにおいて生産のために用いられるイメージが強かったが，AIBO の発売は家庭内へのロボットの進出を行った．その後，2006 年の発売完了に至るまで，次々と新たなバージョンが開発された．AIBO は音声を用いた人とのインタラクションが可能である．“おすわり”, “ふせ”, “こっちへおいで” などの単語に反応して行動をすることができ，自分の名前を覚えることも可能である．また，頭や背中を撫でると，それに反応したり，撫で方によってしかられたのか，ほめられたかなどの区別もできる．AIBO は身体表現によって応答を行うが，頭のランプで感情を示すこともできる．このように，AIBO は本物の犬のように人間の言葉を少し理解でき，人間との簡単なコミュニケーションが可能である．

人間のペットとして定番の猫の形をした NeCoRo の開発も行われている [9, 10, 11]．AIBO のボディはプラスチックで覆われており，実際の犬にはないランプを搭載していることから，動物とは一線を画するロボットを意識していると考えられる．しかし，NeCoRo は猫のような毛で全身を覆われており，外観も猫そっくりの作りとなっている．知能的な面だけではなく，外観についてもより動物に近づけていこうという意図が感じられる．NeCoRo も AIBO と同じように音声による簡単なコミュニケーションや，人が触ることによりインタラクションをすることが可能である．

同様に，本物に近い外観をもつ動物型ロボットにアザラシ型ロボット PARO[7, 8] がある．PARO は，見た目や触り心地など，デザインの面で本物に近づけるといったアプローチがなされている．例えば，触り心地，鳴き声，動作などはアザラシの生態調査に基づき設計されている．また，温度制御により，アザラシと同じ体温の体を持つ

ことも特徴と言える．PARO はこうした特徴から，人に安らぎや楽しみを与え，メンタルコミットロボットという領域で活躍している．

以上のように，動物型ロボットでは，人からの話しかけや接触を通してコミュニケーションができるものが開発されている．また，動物らしい外観や体，しぐさの追求などもなされている．

2.2.2　ヒューマノイドロボット

ヒューマノイドロボットの研究は当初，二足歩行の実現に関するものが盛んであったよう思われる．世界最初の二足歩行ロボットの開発例は 1974 年に発表された早稲田大学の WABOT-1 (WASEDA Robot No.1)[12] である．WABOT-1 は手，足，視覚，音声応答システムから構成されており，人間との簡単なコミュニケーションが可能であった．その後もヒューマノイドロボットの開発はさまざまな機関で行われ，1996 年にはホンダ技術研究所から P2 が発表された．P2 は身長 1820 mm，体重 210 kg で，胴体部にコンピュータやモータードライブ，バッテリー，無線機器などを内蔵している．そして，自在な動きや階段昇降，台車を押すなどの動作を，ワイヤレスでかつ自律動作により実現している．2000 年には，より小型化された ASIMO が完成し，注目を集めた．2004 年には走る ASIMO も発表された．二足歩行技術の研究はいまもなお続けられているが，近年では，コミュニケーション能力に関する研究も盛んに行われている．

ソニーの SDR-4X[14, 15] は豊かなコミュニケーション能力を持つ．SDR-4X には顔や音声から，誰であるか検出することができ，これを活かしたコミュニケーションを実現する．音声認識では，通常，辞書に登録した単語しか認識できないが，SDR-4X は辞書にない未知語を学習する技術も導入されている．対話技術では人や物の場所を一時的に記憶する短期記憶に加え，名前や顔を記憶する長期記憶を持つことで複雑な対話や行動を実現している．このように，ヒューマノイドロボットを用いたコミュニケーションでは，動物型ロボットを用いたコミュニケーションよりも人間に近く，高度な知能を持つものが目標とされている．

コミュニケーションロボットとして総合的な機能を備えたものの一つに ATR 知能ロボティクス研究所で開発された Robovie がある．Robovie は二つの目，車輪による移動機構，腕，体を覆う接触センサなどを持っており，コミュニケーションが可能である．Robovie を用いた人間とのコミュニケーションに関する研究が数多く行われている．小野らは，道案内タスクにおいて，人間とヒューマノイドロボットがジェスチャを同調化させることにより，空間情報が円滑に伝わることを明らかにした [18]．また，神田らはロボットが言語やジェスチャを用いて，他のロボットとコミュニケーション能力があることを示すことにより，人間は円滑にそのロボットとコミュニケーションが始められることを明らかにした [19].

ヒューマノイドロボットでは，人間とのコミュニケーションの実現に関する研究や，人間がヒューマノイドロボットと接するために，どのように設計することで円滑なコミュニケーションが達成できるかなどが研究されている．

2.3　ロボットの機能

人とのコミュニケーションを実現するためには，人間からの問いかけを認識する知覚機能と，人間への応答を行う表現機能が必要である．本節では，知覚機能と表現機能についてまとめる

2.3.1　ロボットの表現機能

身体表現

ロボットは身体を有しているため，その身体を活かした表現が可能である．こうした表現は全てのコミュニケーションロボットで可能である．例えば，動物型ロボットでは尻尾を振ったり，お座り，お手などの動作が身体表現と言えるであろう．ヒューマノイドロボットでは主にジェスチャがある．Robovie, ASIMO および SDR-4X ではジェスチャに関する動作が用意されている．

ロボットにはその他にも，移動といった表現方法がある．例えば，何かの位置を知らせるのに上手く表現できない場合，ロボット自身が移動することで相手に自分の意図を伝えることが可能である．

音声

人同士の最も基本的なコミュニケーション手段は音声によるものである．人とのコミュニケーションを実現するロボットでは，音声対話の実現は重要な課題であり，ヒューマノイドロボットではほとんどのもので搭載されている機能である．初期に開発された WABOT-1 においても音声コミュニケーション機能が備わっていた．ヒューマノイドロボットのみならず，動物型ロボットでも，鳴き声という形で音声を手段としたインタラクションをできるものが多い．

表情

ロボットを親しみやすくするためには，人間とほぼ同じ機能を持つことが必要である．しかし人間は，人間と少しだけ違う箇所があると，その違いに不気味さを感じてしまう．1970 年に森が発表した “不気味の谷” 理論 [20] において類似度と親和感についての関係が述べられている．この理論では，類似度が 100% よりも少し手前の箇所において，親和感が極度に低くなる不気味の谷があることが示されている．

人間とのコミュニケーションにおける表現手段の一つに表情があるが，表情を作り出すためには顔を人の形に近づける必要がある．しかしながら，不気味の谷理論から

親しみやすいロボットとするためには高い技術が要求される．そのため，現在開発されているロボットの中には表情に関する機能が用意されていないものも数多く存在する．

表情に関する機能を有するロボットについて紹介すると，Schulte らの顔による博物館案内ロボットは表情による感情表現が可能である [21]．MIT で開発された Kismet は音声認識，音声合成，画像認識，表情・視線の表出が可能な顔ロボットである [22, 23]．9 種類の基本顔表情が用意されており，インタラクションを行う．よりリアルな外観を追求し，表情を持つロボットにアクトロイドがある．アクトロイドは人間そっくりの外観を持っており，表情も微妙に変化する．

表情とは異なるが，視線制御可能なロボットとして，Robovie や MIT で開発された Cog[24] などがある．また，NeCoRo では目の開閉ができ，AIBO はランプの色の違いにより感情の表現を行う．

2.3.2　ロボットの知覚機能

視覚

“百聞は一見にしかず”，と言われるように，人間は視覚から得る情報が大きい．視覚は，人や環境の認識などに必要であり，ロボットにとっても重要な知覚機能であると考えられる．

コミュニケーションにおいて視覚は，個人検出 (誰であるかを理解する) や，視線制御などに用いられる．例えば，SDR-4X では，視覚からの情報により人の顔を認識し，誰であるか理解することが可能である．認識した情報を対話に反映させることで，人間にとって親しみのある対話を実現することができる．

Robovie は視線制御機能を持つ．人の顔の位置を検出し，その方向を向いてコミュニケーションを行うことで，人間にとって親和感の高い対話を実現することが可能である．また，話者の方を向いて対話するロボットの開発について小林らの報告がある [25]．横山らの報告では視線制御を用いることで発話交替が円滑になることを示している [26]．

その他の視覚情報の利用として，発話区間の検出を行い，聴覚機能の補助として用いることも検討されている [27]．

聴覚

聴覚は，人の発話を理解するのに必要な機能である．人同士のコミュニケーションでは音声を手段とするものが一般的に利用されるため，人とのコミュニケーションを実現する上で最も重要な知覚機能であると言える．聴覚の研究では，SIG がある [28]．

SIG はロバストな知覚機能を持つヒューマノイドロボットであり，マイクロホンアレーを用いて，二話者の音源分離や移動音源の抽出などが可能である．

ロボットの音声認識では，環境雑音や自己の動作雑音など，多くの雑音を含む環境での認識が必要となる．そこで，SIG のように複数マイクを備えることで環境雑音へ対応したり，AIBO のように自己の動作中は音声認識を行わず，止まってから認識を行う stop-perceive-act 法などが用いられている．

聴覚を用いて，言語情報だけでなく，ピッチや話速など，パラ言語情報を用いた話し手の好感度推定なども検討されている [29].

触覚

触覚は主に，動物型ロボットが備えている．AIBO や NeCoRo，PARO はともに体を触られたことを知覚し，その撫で方や場所などにより異なるインタラクションを行うことが可能である．また，Robovie など，ヒューマノイドロボットでも触覚を持つものが存在する．

2.4　ロボットのタスク

動物型ロボットは，癒しを目的としたものが多く，人間との単純なインタラクションや，動物らしい仕草の演出などが行われている．

一方でヒューマノイドロボットは，人間をサポートする仕事をこなすことが期待されている．ところが，ロボットによって搭載している機能に違いがあり，ロボットの持つ機能は人間の知能やコミュニケーション能力には程遠い．そこで，ロボットの長所を活かしたタスク設定がなされている．本節では，ロボットが活躍しているタスクについて示す．

おつかい

日立製作所で開発された EMIEW は，ウェイタとしての仕事をこなす．客から注文を聞き，カウンターで渡された商品を客のところまで運んでくることが可能である．コミュニケーション能力と移動能力を上手く組み合わせたタスクである．奈良先端科学技術大学院大学，和歌山大学，産業技術総合研究所の共同研究では，ヒューマノイドロボット HRP-2 にものをとってくるタスクを実装している．例えば，“新聞をもってきて” とユーザが指示すると，ロボットは新聞受けから新聞を持ってくることができる．

受付，案内

受付や案内は，さまざまなロボットで実装されている．例えば，産業技術総合研究所で開発された Jijo-2 は人の居場所や道案内が可能である [30]．また，人に道を聞きながら移動する機能なども備えている．Minerva は展示会場の案内を行うロボットである [31]．愛知万博にもさまざまなロボットが出展された．アクトロイドは人間に似せた姿形をしており，MC をこなす．また，ガードロボ i は警備ロボットの機能と案内ロボットの機能を持ち，移動機能を活かした案内が可能である．受付案内ロボ ASKA はユーザからの問いかけに対応することが可能である [32]．ASKA は人間の顔検出を行い，こちらを向いていると判断したときにのみ音声入力の受付を行う．

ネットワークを通じた情報提供

近年，ネットワークロボットが注目を浴びている．ネットワークと通じることでインターネットからの情報提供を行うことができる．ロボットサービス提供のための規格として，RSi (Robot Services Initiative) が提案されている [34]．これにより，AIBO, Ifbot, wakamaru など形も機能も異なるロボットから天気情報を提供するサービスを実現している．また，ネットワークロボットに関するプロジェクトで研究がされており

[33]，今後，人とのインタラクションを通じてインターネットから情報を提供するロボットが増えてくると考えられる．

プレゼンテーション

ロボットによる受付や案内に関するタスクは多くの試みがされているが，プレゼンテーションに関するタスクの実現は少ない．一例として，HOAP-1 を用いてプレゼンテーションを実現したものがある [36].

プレゼンテーションは，情報の提供を行うタスクである．情報の提供は，ロボットの得意とするところであり，受付，案内タスクとこの点で同質である．また，スクリーンを用いたプレゼンテーションは，ブラウザを用いて WEB ページを表示しながら情報提供を行うことも可能である．また，ロボットが移動可能である点も考慮すれば，スクリーンの中に留まらず，スクリーン外の物を対象として人間と同じようにプレゼンテーションをさせることも可能である．膨大な情報を背景としたネットワークを通じ，情報提供が将来的に盛んに行われる可能性がある点を考慮すると，プレゼンテーションはロボットの有効なタスクと捉えることができる．

2.5　本章のまとめ

本章では，コミュニケーションロボットの機能やタスクについて示した．ロボットは人間と同じように知能やコミュニケーション能力を持つことが理想であるが，その達成にはまだまだ研究開発が必要である．そこで，現在開発されているロボットを有効活用するため，ロボットの長所を活かしたタスク設計がされている．ロボットの長所として，情報提供機能，移動機能，コミュニケーション機能などがあり，これらを活かした案内や受付ロボットが数多く存在する．また，ネットワークからの情報提供も注目されており，今後，情報提供機能を活かしたロボットのタスクが多く設定されると考えられる．案内や受付と同様に，プレゼンテーションも，ロボットのモダリティや情報提供機能を活かした有効なタスクである．プレゼンテーションはスクリーンを用いることで，ネットワーク上の情報を視覚的に提供することが可能である．商品説明などにおいては，身体を活かして商品を指示しながら説明することができる．プレゼンテーションをロボットのタスクとして実現することで，ロボットの活躍分野が広がると考えられる．

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

3.1　はじめに

第 2 章で示したように，プレゼンテーションは，ヒューマノイドロボットのモダリティと情報提供能力を活かした有効な活用分野である．しかし，ロボットの操作には制御用プログラムの作成が必要であり，専門家でなければロボットを操作することは難しい．ロボットを身近に誰でも扱えるようにするためには，簡単な記述でコンテンツが作成できることが必要である．そこで，本章では，簡単な記述でヒューマノイドロボットを用いたプレゼンテーションコンテンツを作成する方法について提案を行う．3.2 節では，関連研究について示す．アニメーションエージェントを用いたコンテンツの作成に関する研究分野では，簡単な記述でコンテンツを作成するための言語について検討がなされている．それらについて紹介するとともに，特にプレゼンテーション用の記述言語である MPML について詳説する．また，ロボットの操作言語と，ロボットの評価に関する報告についても紹介する．3.3 節では，簡単な機構でヒューマノイドロボットによるプレゼンテーションを作成するための記述言語を提案するとともに，実装方法について示す．実装にあたっては，自然なプレゼンテーションを実現するため，発話と動作を同時に行うことができる機構の導入と，動作がないときに頭を動作させる自律動作の導入を行う．3.4 節では，提案するヒューマノイドロボットを用いてプレゼンテーションを実現する枠組みの第一の評価として，心理学的評価を行う．評価は，数多くのマルチモーダルコンテンツが存在するアニメーションエージェントと比較することにより行った．そして，アニメーションエージェントとヒューマノイドロボットのプレゼンテーションから受ける印象の違いについて明らかにし，ヒューマノイドロボットを用いたプレゼンテーションの有効な活用分野について検討する．3.5 節では，第二の評価として，提案する記述言語の記述容易性について検討を行う．

3.2　関連研究

スクリーン上のエージェントであるアニメーションエージェントを用いてマルチモダールコンテンツを作成するための言語がいくつか提案されている．特にプレゼンテーションを対象としたコンテンツ記述言語として，MPML (Multimodal Presentation Markup Language) がある．一方，ロボットについては，ハードウェアごとに機能が異なるため，共通の言語によりマルチモーダルコンテンツを作成するような試みは行われておらず，その前段階ともいえる，インタフェースを共通化することで，操作プログラムを効率化する枠組みが提案されている．本節では，これらの研究について示す．

また，人とのインタラクションを行うコミュニケーションロボットの評価についての報告がいくつかなされており，心理学的手法を用いて行われている．これらについても示す．

3.2.1　アニメーションエージェントを用いたマルチモーダル記述言語

コンピュータ技術の進歩により，WEB 上や TV 上などでアニメーションエージェントを用いたマルチモーダルコンテンツが数多く存在している．マルチモーダルコンテンツをゼロから構築するのは大変な作業であため，簡単な記述でコンテンツを作成するための枠組みが提案されている．

それぞれのコンテンツ記述言語にはいくつかの特徴があるが，機能の面から考えると，顔によるエージェントを用いたコンテンツを記述する APML と VHML, 身体のあるエージェントを用いたコンテンツを記述する AML, CML, STEP, TVML がある．以下，これらの記述言語について示す．

APML

APML (Affective Presentation Markup Language)[37, 38] は顔からなるエージェントを用いて表情の生成と，音声合成による出力を行うコンテンツを作成することができる．上位レベルの表情 (怒りや悲しいなど) をあらかじめ定義しておくことで，表情の生成を簡単に行うことができる．APML は MagiCster という対話システムの中で，エージェントを用いた出力部分を生成するために開発された言語である．音声合成を行うテキストを記述し，ピッチなどを調整するためのタグなどで修飾することで感情なども含めた音声合成が可能である．

VHML

VHML (Virtual Human Markup Language)[41, 42] は，トーキングヘッドと呼ばれる顔をエージェントとしたコンテンツを作成することができる．発話内容や，トーキングヘッドの感情，表情を記述することにより，豊かな顔の表情を交えたコンテンツを作成することができる．

VHML ではいくつかのサブシステムを用いてトーキングヘッドによるコンテンツを実現している．音声，表情などの感情を定義する EML (Emotion Markup Language), 音声合成エンジンによる音声出力を定義する SML (Speech Markup Language), 顔の表情や感情を含めたアニメーションを定義する FAML (Facial Animation Markup Language), 顔によるアニメーションの代わりにテキストを表示するための HTML を用いる．また，身体の表現を定義する BAML (Body Animation Markup Language), 対話によるインタラクションを定義する DMML (Dialogue Manager Markup Language) を組み込むことが可能なように設計されている．

AML, CML

AML (Avatar Markup Language)[39], CML (Character Markup Language)[40] はともにアニメーションエージェントによるマルチモーダルコンテンツを生成するための記述言語である．AML は低位レベルから詳細にエージェントの動作を記述することができ，AFML (Avatar Face Markup Language) 部と ABML (Avatar Body Markup Language) 部から構成される．AFML 部では顔や音声合成に関する記述を行い，ABML 部では身体を用いた歩行や対象指示などの動作に関する記述を行う．顔，身体それぞれの動作を記述することで，より細かなコンテンツを作成することが可能である．また，動作の開始時間に関する命令も用意されており，アニメーションエージェントの動作を詳細に管理することができる．

CML は，AML よりも上位レベルでの記述を行うことにより，簡単にコンテンツを記述することが可能である．AML が顔と身体と別々に記述するのに対し，CML では顔の記述に関するタグは用意されていない．AML と比較して詳細な記述はできないが，その代わりに簡単にコンテンツを作成できるという利点がある．

STEP

STEP (Scripting Technology for Embodied Persona)[43, 44] は，アニメーションエージェントの動作を詳細に表現することのできる記述言語である．アニメーションエージェントの記述言語は，XML に準拠しているものが多いが，STEP では Prolog ライクな記述を行う．身体の部位 (HumanRoot, skullbase, shoulder, elbow, wrist, hip, knee,

ankle) と位置 (front, back, up, down, left, right など) の組み合わせにより，エージェントの動きやポーズを定義する．この言語はエージェントの動作を詳細に記述できるが，発話命令は実装されていない．

TVML

TVML (TVprogram Making Language)[45, 46] は TV 番組の作成を目的に作られた記述言語である．XML には準拠していないが，低位レベルで記述する必要がなく，簡単なコンテンツの作成が可能である．TVML は TV 番組の制作を目的としているため，記述可能な要素はアニメーションエージェントに留まらない．例えばニュース番組ではニュースキャスターの座る椅子や机が必要であるし，演出のため植木などが置かれている光景を目にする．TVML ではこのような小道具の記述も可能である．その他に，テレビ番組制作で重要なのは照明の設定やカメラワークである．TVML ではこの点についても命令が用意されている．また，エージェントによる発話だけでなく，バックグラウンドに流すサウンドを記述することも可能である．NHK では，TVML を用いて作成したテレビ番組を放送している．

その他

その他，複数のパラメータ指定を用いることで，詳細なアニメーションエージェントによる動作や表現を記述する PAR (Parameterized Action Representation) [47, 48]，XML に準拠し，対話のためのアニメーションエージェントの動作を記述する RRL (Rich Representation Language)[49, 50]，XML に準拠し，発話 (Verbal) と発話以外 (Non verbal) の表現を含めて表現を行うことを目的とした MURML (Multimodal Utterance Representation Markup Language) [51]，文化や社会，動作，心理，意図といった人間の特性を記述して，情報の中に埋め込むための HumanML[52] などが提案されている．

3.2.2　マルチモーダルプレゼンテーション記述言語 MPML

MPML はアニメーションエージェントを用いたプレゼンテーションコンテンツを記述するための言語である．MPML の特徴には以下のものがある．

容易な記述性

MPML は，XML(Extensible Markup Language) に準拠しており，数種類のタグを用いることでコンテンツ本体を記述することができる．このため，コンピュータプログラム経験のない人でも容易にコンテンツを作成することができる．

システム非依存

MPML は中位レベルの言語であるため，システムとのインタフェースを構築することで，いかなるシステムの上でも実行可能である．そのため，MPML にはいくつかのバージョンが存在する．図 3.1 に MPML ファミリをまとめる．

MPML は，最初に開発された MPML ver. 1.0[55] から現在に至るまで使用できるエージェントやその機能によって様々なバージョンが開発されてきた．MPML ver. 2.0e[56] では，感情制御のためのタグが導入され，キャラクタの動作や音声の感情制御を容易に記述することが可能となった．MPML-VR (MPML for Virtual Reality)[57] は，VRML の技術を使用して 3 次元空間の中でキャラクタを制御し，プレゼンテーションを行わせることができる．MPML-mobile[58] では，キャラクタを使ったコンテンツを携帯電話上で実行することができる．MPML-mobile では，音声を出力する代わりにテキストが表示される．

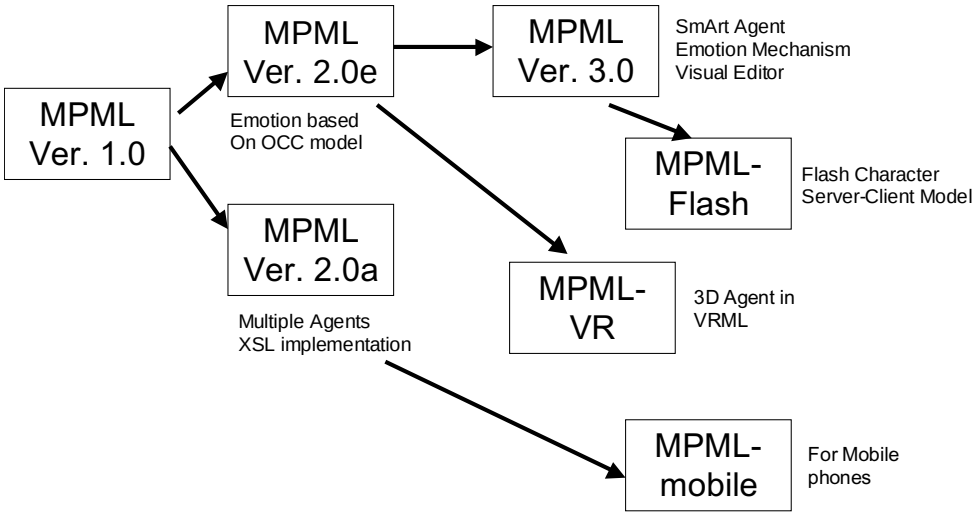


図 3.1: MPML ファミリ

豊富な制御機能

MPML では，アニメーションエージェントをプレゼンタとしてコンテンツを作成するため，ジャスチャや移動，発話などのエージェントを制御するための関数が用意されている．ジャスチャは，使用するエージェントにより異なるが，数十種類の動作が用意されているものもあり，これを使用することで充実したコンテンツを作成することが可能である．

感情表現

MPML では OCC モデル [60][61] で定義された感情を使用することができる．OCC モデルは，感情誘発と関連付けされる多様な要素を考慮した感情モデルとして，現在多くの感情理論の基礎となっている．OCC モデルでは，イベントの結果，エージェントの行動，対象の様子を社会における三つの感情を誘発する要素とし，それに基づいて 22 種類の感情が定義されている．MPML にはこの 22 種類の感情が用意されており，感情を指定するだけで，プレゼンタとなるキャラクタの身振り，手振りや移動，発声の仕方が変化する．感情表現豊かなプレゼンテーションが，感情をタグ付け指定するだけで容易に実現可能である．

MPML の命令セット

図 3.2 に MPML のタグ (命令) 構造を示す．これは，ver. 2.0e の木構造を示したものである．MPML にはさまざまなバージョンが存在するが，安定した動作実績があり，必要な機能がカバーされている ver. 2.0e について示す．

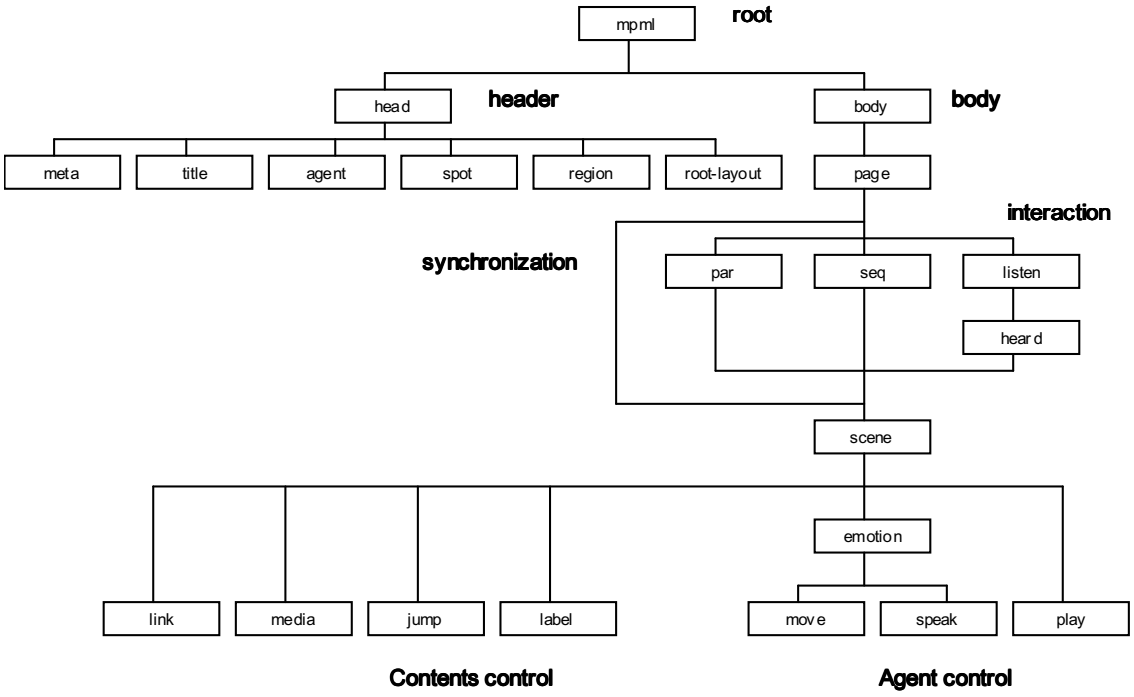


図 3.2: タグ構造 (MPML ver. 2.0e)

MPML は mpml をルートとし，ヘッダ部を記述する head とコンテンツ本体を記述する body から構成される．コンテンツ本体には，アニメーションエージェントを操作する命令の他，インタラクションのための命令や，同期のための命令なども存在する．

ヘッダ部には，コンテンツの情報を記述する meta, title や，スクリーン上のコンテンツを表示する領域を指定する region, root-layout, spot などのタグがある．また，エージェントを指定する agent により，決められたアニメーションエージェントだけではなく，自分で定義したエージェントを用いて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．

本体は，page により，ページごとにコンテンツが記述される．プレゼンテーションは，通常 PC や OHP のスライドを用いて行う．この際，スクリーンに表示されるスライド 1 枚の単位が MPML における page の概念である．MPML では，page を組み合わせることにより，コンテンツを記述する．

アニメーションエージェントの制御は，emotion, speak, play, move により行われる．speak はアニメーションエージェントに発話をさせるための命令である．音声合成器を用いて発話がされるのと同時に，吹き出しからテキストによるスクリーンへの表示も行われる．move はアニメーションエージェントを移動させるための命令である．座標はスクリーンの (x,y) により指定される．emotion はエージェントの感情を指定するタグである．emotion で囲まれた speak は音声合成の話速やピッチが調整され，その感情に合わせた音声合成がなされる．例えば，悲しい感情では，話速が遅くピッチも低く音声合成がされ，嬉しい感情では，話速が早くピッチも高く音声合成がされる．move が emotion で囲まれた場合，エージェントの移動速度が変わる．例えば，悲しい感情では移動速度が遅くなり，嬉しい感情では移動速度が速くなる．play はエージェントにジェスチャをさせるための命令である．エージェントごとに用意されているジェスチャの数は異なるが，一つのエージェントにつき数十程度が用意されている．これには例えば，バイバイなどのジェスチャがある．

エージェントの動作を同期させるための命令が par である．par で囲まれた命令は開始時点が同期して実行される．これとは逆に，明示的に同期させないようにする seq も存在する．この命令は，par の中で同期したくない命令を記述するときなどに有効である．

MPML には簡単なインタラクションをするための命令も存在する．listen が記述された箇所では，ユーザからの発話待ちを行う．音声認識文法の定義などは heard で行う．ユーザからの発話があると，jump などの命令によりコンテンツを遷移させることが可能であり，遷移先となる箇所を示す label 命令も用意されている．

3.2.3　ロボットの操作言語

ロボットは，アニメーションエージェントのようにコンテンツを簡単に作成するための言語の提案はないが，インタフェースを共通化することで開発効率を高めようとする試みが行われている．例えば，RT ミドルウェア [67, 68] は分散オブジェクト技術 CORBA を用いて分散した CPU 間を簡単に接続して開発することを目的とする．ORiN[66] はネットワークに接続された装置 (ロボット) に統合的にアクセスするためのインタフェースを目指す．ORCA[63] は分散オブジェクト技術を用いてロボット間の通信とロボット内の通信を統一的に行うことを目指す．OpenHRP[64] は CORBA を用いた分散オブジェクトシステムである．プラグインやアダプタといった機構を導入することで多数の開発者が平行して開発を進めつつ，それぞれの成果を相互に利用しやすい環境を提供する．

また，一般的なプログラミング言語，C++などの SDK を提供することにより，ロボットプログラムの開発環境を提供する試みもある．OPEN-R[62] では，OPEN-R SDK が無償で公開されており，ユーザはこれを用いることで，C++などを用いてロボットの動作を記述することができる．OPEN-R は，ペット型ロボット AIBO とヒューマノイドロボット QLIO-SDR での動作実績がある．2006 年には Microsoft 社が Microsoft Robotics Studio の発表を行った．これは，ロボットのインタフェースを統一し，Visual Studio 製品を用いてロボットのプログラミングを行うための枠組みである．

3.2.4　コミュニケーションロボットの評価

評価方法

コミュニケーションロボットの評価を行う場合，定量的評価と定性的評価がある．定量的評価では，被験者にロボットのコンテンツについて数値で評価させることで主観評価を行うことができる．客観評価を行う場合は，ロボットのコンテンツを見る被験者の視線や身体動作を測ったり，被験者にセンサをつけ，脈拍や脳波を測ったりすることで実現できる．また，ロボットから人間に何かを依頼するようなコンテンツでは，依頼が達成できたか否かにより客観評価を行うことも可能である．定性的評価では，ロボットのコンテンツを見た感想をアンケート形式で被験者に書いてもらうことで主観評価ができる．また，客観評価は，ロボットと被験者の会話などを書き出すことにより実現できる．

ロボットの評価では，定性的評価と比べると数値として表れ評価しやすいことから，定量的評価が行われることが多い．また，客観評価を行うためには，計測機械などが必要となるため，主観評価を行うことが多い．実際にロボットの評価についての報告を示すと，中田らはシャチ型ロボットを用いて親和感を演出する行動の生成方法を提

案し，SD 法を用いた評価をしている [69]．尾形らは SD 法を用いて情緒モデルを有するロボットの人間とのコミュニケーションにおける評価をしている [70]．中田らは 5 つの形容詞対を用いた評価を行っているが，神田らはさらに形容詞対を増やし，28 対の形容詞対を用いた SD 法によりロボットのインタラクションについて心理学的評価実験を行っている．また，評価においてインタラクションの静的側面だけでなく，動的側面についての考察も行っている [71, 72, 73, 74, 75]．これらの評価は印象評価のための心理学的手法である SD 法を用いて行われている．

SD 法

SD 法 (Semantic Differential Method)[76] は Osgood らによって提案された印象の評価方法であり，企業イメージの測定や商品イメージの測定などに用いられている．SD 法では，数十の形容詞対を用いて評価対象について主観評価を行い，印象の判断を行う．形容詞対は “よい－わるい” といったポジティブな形容詞とネガティブな形容詞から構成され，7 段階または 9 段階程度の尺度で評価を行うことが多い．また，心理学分野の研究から，適切な形容詞対を用いた場合には，因子分析の結果，3 つの要素 (Evaluation:評価，Potency:潜在性，Activity:活動性) に分解されることが見出されている．SD 法の分析では，複数の変数の背後に存在する因子を発見するため，因子分析が行われる．評価対象の因子ごとの得点を見ることで，印象評価をすることができる．

3.2.5　ヒューマノイドロボットによるプレゼンテーションの実現のための考察

ロボットの開発効率を高めるために，インタフェースを共通化する枠組みが提案されているが，これらは C 言語などのプログラミング言語を対象としており，プログラミングに関する知識が必要である．また，低位レベルで記述を行う必要があるため，初心者が簡単にコンテンツを作ることは難しい．ロボットについては，アニメーションエージェントにあるような，中位または上位レベルでコンテンツを作成することのできるような言語の提案はなされていない．そこで，ヒューマノイドロボットによるプレゼンテーションコンテンツを簡単に作成できるようにするためには，アニメーションエージェントで用いられているような中位または上位レベルの言語をヒューマノイドロボットにも用いることが適切であると考えられる．

アニメーションエージェント用の言語には多くのものが提案されているが，プレゼンテーションコンテンツを作成するためには，発話や動作などのモダリティを利用でき，記述もシンプルであるものが望ましい．簡単な記述が可能な言語についてみると，MPML，TVML，CML が挙げられる．TVML はテレビ番組に特化した仕様となって

いるため，プレゼンテーションへの利用には考慮が必要である．その点，CML は一般的なマルチモーダルコンテンツの作成が可能であり，MPML はプレゼンテーションに特化している．この二つを比較すると，MPML はプレゼンテーションに特化しているため，CML と比べても記述量が少なくコンテンツを作成することが可能である．また，システム非依存であるため，PC のみならず携帯電話での動作実績もあり，MPML を用いることでヒューマノイドロボットによるプレゼンテーションコンテンツを簡単に作成するための枠組みが実現できると考えられる．

コミュニケーションロボットの評価については，心理学的手法，特に SD 法を用いたものが利用されている．ロボットによるプレゼンテーションを評価する場合においても，印象を評価することで，プレゼンテーションが受け入れられているか否かを知ることができると考えられる．また，アニメーションエージェントを用いたコンテンツは TV や WEB 上など数多く存在する．これらの背景には，アニメーションエージェントによるマルチモーダルコンテンツは人々に受け入れられていることがあると考えられる．そこで，アニメーションエージェントによるプレゼンテーションとヒューマノイドロボットによるプレゼンテーションの印象を比較することで，ヒューマノイドロボットによるプレゼンテーションの位置づけを評価することができると考えられる．

3.3　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の提案

本節では，簡単にヒューマノイドロボットによるプレゼンテーションコンテンツを記述するための言語の提案と，システムへの実装について示す．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は，アニメーションエージェント用プレゼンテーション記述言語 MPML を拡張することで実現する．拡張にあたっては，エージェントが実在する空間の違いについて考慮し，新たなタグを導入する．また，ヒューマノイドロボットにホンダ ASIMO を用いることでシステムの実装を行う．さらに，自然なプレゼンテーションを実現するため，発話と動作を同時に行う機構と，自律的な動作を行う機構についても導入する．

3.3.1　MPML-HR の命令セット

本節では，MPML-HR のタグについて説明する．まず，新たに導入した `point` について示す．

①**Point**
< **point**　**x** = “　”　**y** = “　” > と記述することで，スクリーン上の (x, y) 座標を指示する．スクリーンエージェントを用いた従来の MPML では，スライドのある座標を指示したい場合，エージェントの移動命令である `move` により実現することができた．しかし，ロボットは実空間上に存在するため，移動命令とは別にスクリーンの座標を指示する命令が必要である．新たに導入した `point` を用いることで，ヒューマノイドロボットは (x, y) で指定されたスクリーン上の座標を指示するよう動作する．現段階では厳密なスクリーン上の位置を指し示すことはできないが，スクリーン上の左右，上中下 6 分割された中の 1 つの領域を右手もしくは左手によって指し示す．指定された座標がスクリーンの右半分の座標であればロボットはスクリーンの右側に移動して右手でスクリーンを指す．逆にスクリーンの左半分の座標であればロボットはスクリーンの左側に移動して左手でスクリーンを指す．以下の流れに従ってロボットは行動する．

1. スクリーンの右側または左側へ移動する
2. 体をスクリーン側へ 20 度傾ける
3. 右手または左手を上げ，上，中，下のエリアを指し示す

図 3.3 に < **point** > 命令を実行した例を示す．
次に，新たな引数を導入した `move` について示す．

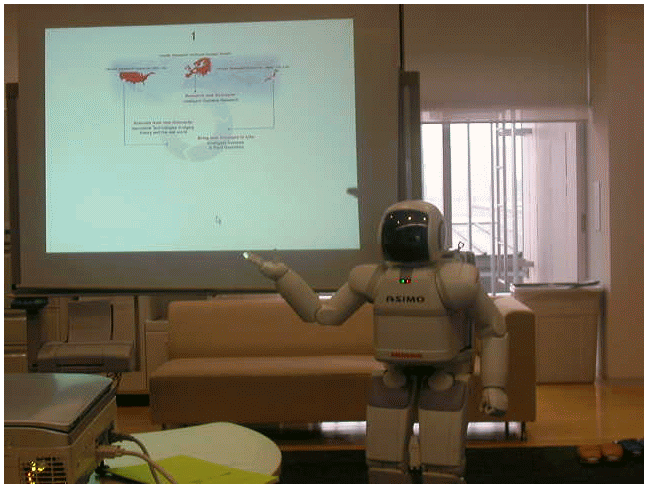


図 3.3: スクリーン指示動作

②Move

< **move**　**x** = “　”　**y** = “　”　**z** = “　” > と記述することで，ヒューマノイドロボットを (x, y) で指定された座標に移動させ，さらに z 度だけ身体を回転させることができる．MPML はアニメーションエージェントを用いるため，移動命令によりスクリーン内をエージェントが移動するが，ロボットは実空間上を移動する．そこで，新たに引数 z を導入しており，身体の回転についての記述が可能である．また，従来の MPML との互換性を高めるため，z は省略可能とし，z を省略した場合は 0 が代入される．

その他のエージェント制御命令である `play`, `speak`, `emotion` について示す．

③Play

< **play**　**act** = “　” > と記述することで **act** で指定したジェスチャをエージェントに実行させることができる．指定可能なジェスチャはエージェントごとにあらかじめ定義されているため，ユーザはジェスチャに関する細かい制御について気にすることなく，コンテンツ中にジェスチャを組み込むことができる．図 3.4 に ASIMO の動作例を示す．

④Speak

`speak` でテキストを囲むことで，ヒューマノイドロボットに発話をさせることができる．またこのタグは，`emotion` で囲むことで発話の際の感情が変化する．

⑤Emotion

`speak` と `move` は `emotion` によって囲むことができる．`emotion` は感情を含めたコンテンツ作成を容易にするために導入されたタグである．このタグを用いることで，感情に合わせて音声と移動スピードが変化する．MPML-HR では，ピッチやテンポを変化

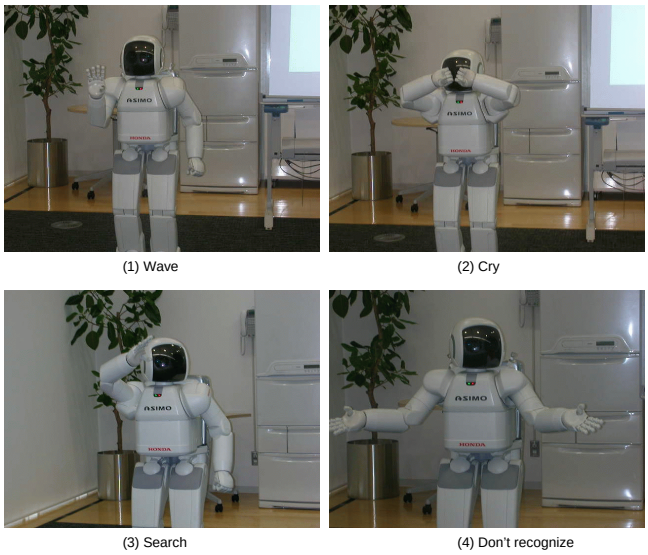


図 3.4: ASIMO の動作例

させることにより感情を含めた音声合成を可能にしているが，移動スピードの変化に関しては実装を行っていない．感情は OCC モデルに基づき happy-for, worried, anger といった 22 種類が用意されている．< **emotion　types** = “　〃 > とすることによって感情の指定を行う．

その他，コンテンツを作成するためには，ヘッダ部を定義する `head` ，本体を示す `body`，タイトルを示す `title`，エージェントを指定する `agent` およびプレゼンテーションのスライドを指定する `page` が必要であるが，これらは MPML ver. 2.0e と同様である．

3.3.2　システムの実装

図 3.5 にプレゼンテーションを実行する際のシステム構成を示す．プレゼンタにはヒューマノイドロボット ASIMO(Advanced Step in Inovative MObility) を用いた．ASIMO は Honda によって開発されている高性能二足歩行ロボットである [13].

ASIMO は高さ 1200[mm] で重量 52[kg] である．関節の自由度は総計 26 自由度であり，足は左右それぞれ 6 自由度，腕は左右それぞれ 5 自由度である．手は 1 自由度を有しており，グウやパーを表現できる．また，500[g] の握力があり，軽い物を持つこともできる．頭は 2 自由度があり，首を左右に動かしたり頷いたりすることができる．また，1.6[km/h] で歩くことができ，踊りやお辞儀など様々な種類の動作を行うことができる．

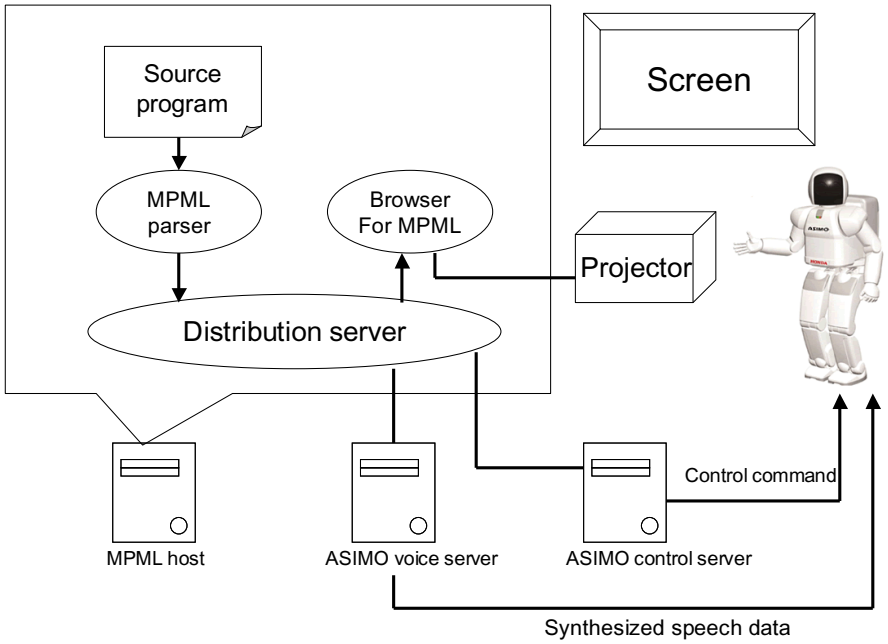


図 3.5: MPML-HR のシステム構成

ASIMO の制御は ASIMO control server からの無線通信により行う．音声出力については ASIMO voice server が制御を行う．ASIMO voice server では音声合成を行い，その音声データを ASIMO のスピーカまたは ASIMO voice server に接続されたスピーカから出力する．また，この構成図では ASIMO voice server は独立の 1 台の PC として示されているが，MPML host の中に置くことも可能である．

MPML-HR の実行は MPML host 内で行う．MPML host と ASIMO Control server および ASIMO voice server は TCP/IP によって接続されており，ネットワークを介して命令が送られる．MPML のソーススクリプトは MPML parser によって解釈され，中間記述に変換される．Distribution server ではこの命令系列を受け取り，その命令の種類によって Browser, voice server, ASIMO control server のいずれかに命令を送信する．それぞれのサーバは命令を受信するためのインタフェースがあり，Distribution server は MPML parser から受け取った中間記述をサーバ毎のインタフェースに変換して命令を送信する．命令を受信したサーバは，対応する処理を行い，Distribution server へ ACK 命令を送信する．Distribution server は ACK 命令の受信により命令の終了を検出し，次の処理を行う．

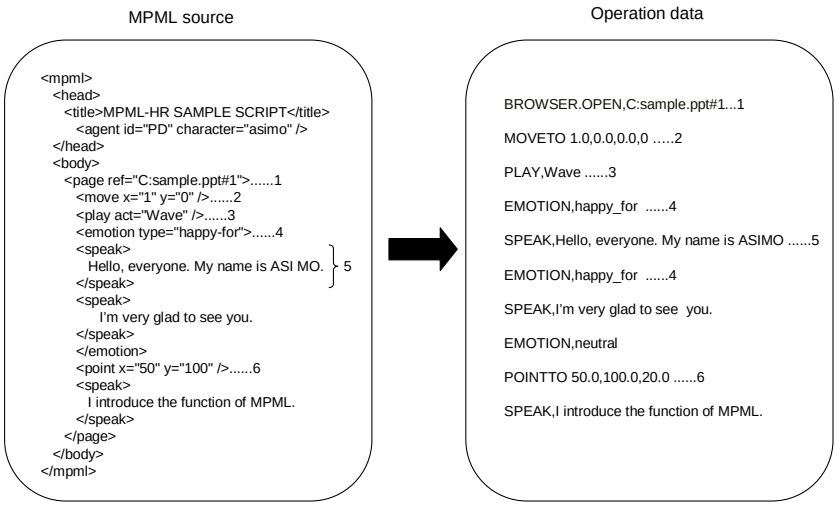


図 3.6: MPML-HR により記述したサンプルスクリプト

MPML-HR によるコンテンツ記述例

図 3.6 に MPML-HR の記述例を示す．左側は MPML-HR のソーススクリプトであり，右側は MPML parser によって変換された中間記述である．それぞれについて説明する．

1

ヒューマノイドロボットが説明を行うスライドの定義である．このページは Windows PC と繋いだ LCD のプロジェクタによってスクリーンに投影される．この例では，“sample.ppt” の 1 ページ目がスクリーンに映し出される．HTML など他の形式のファイルをスクリーンに映し出すことも可能である．

2

移動命令の記述である．ASIMO は横（x 軸）に 1[m] 移動する．移動命令では 3 番目の引数 z[deg] が記述可能であるが，この例では，二つの引数 (x，y) しか記述されていないため，体の向きは変えずに移動のみを行う．

3

ジェスチャの記述である．ASIMO は視聴者に手を振る動作を行う．手を振る動作以外にも，“Greet”，“GestureRight” といった様々な動作が用意されている．

4

感情についての記述である．それぞれの感情において声の大きさと速度が定義されている．感情タグによって囲まれた発話は指定された感情によって変化する．

5

発話の記述である．ASIMO は “Hello, everyone. My name is ASIMO.” と話す．この命

令は，“happy-for” を感情として指定した感情タグによって囲まれているため，ASIMO は嬉しそうに話す．

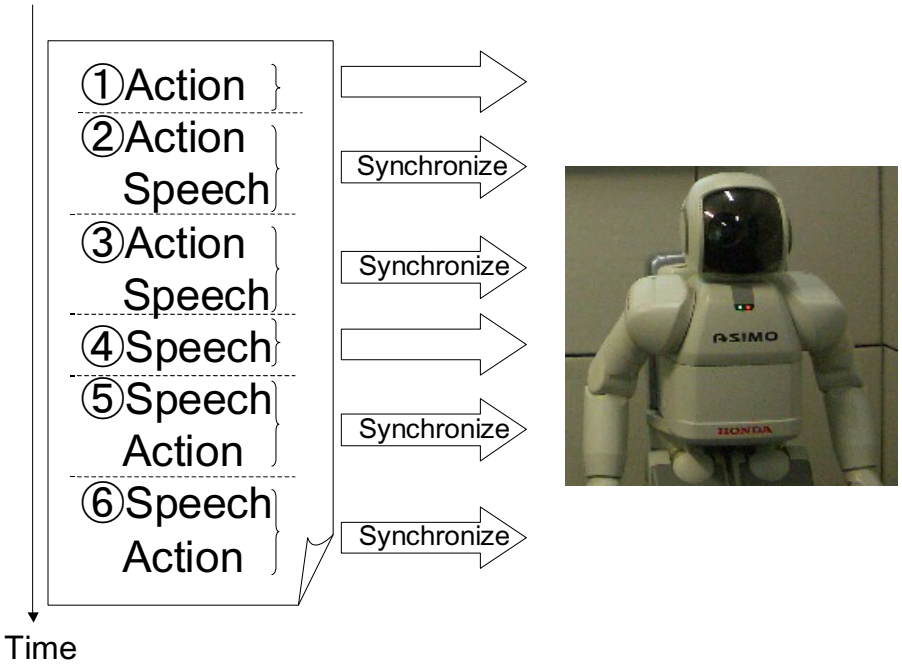
6

対象指示動作の記述である．ASIMO はスクリーンの左側に移動し，体を内側に 20 度向け，右手で案内を行う．この一連の動作は “move” と “play act = “GestureRight”” の記述に相当する．

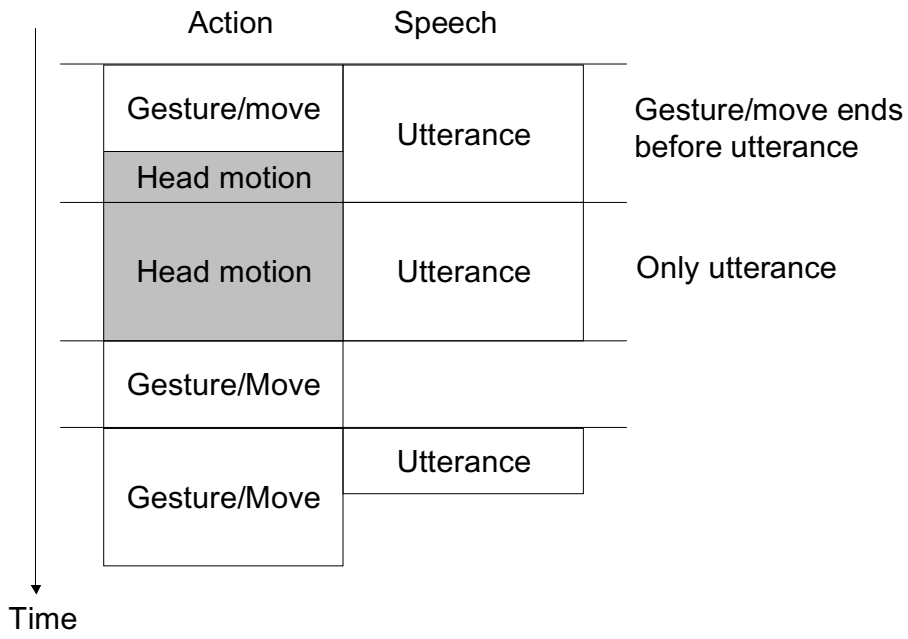
3.3.3　音声と動作の同時実行と自律的な動作

人間によるプレゼンテーションでは通常，発話と動作は同時になされる．そこで，ヒューマノイドロボットによるプレゼンテーションにおいても，人間と同じように動作と発話は並列で実行されることが必要である．しかし，MPML-HR により記述されたスクリプトを逐次実行していくと，それは発話と移動，ジェスチャが順に実行されるだけのプレゼンテーションになってしまう．このようなプレゼンテーションはとても不自然であり，視聴者に違和感を与える．そこで，動作と発話を並列で行うよう実装することとした．並列動作については，図 3.5 の Distribution server で管理を行うこととした．MPML parser からの命令を受け取る際に，一つ先の命令も読み込み，現在の命令と先の命令が同時実行可能な場合は，開始時点を同期して実行する．同時実行可能とは，現在の命令と一つ先の命令が，発話命令と動作命令 (移動命令，ジェスチャ命令または指示命令) の組み合わせである場合が該当する．図 3.7a) に同期の例を示す．①の場面では，一つ先の命令は同時実行不可能なので，単独で実行する．②の場面では，一つ先の命令は同時実行可能なので，開始時点を同期して実行する．同じように，③，⑤，⑥の場面でも同期を行い，④の場面では同期しない．

また，ロボットの自律的な動作についても実装を行った．人間のプレゼンテーションでは，プレゼンタは手や頭などの部分が動くことが多い．また，アイコンタクトにより視聴者とのコミュニケーションをとることで，よいプレゼンテーションを実現することができる．そこで，自律動作の実装では，ロボットの頭をランダムで動かすこととした．MPML では Microsoft エージェントを用いており，このエージェントの機能として動作が長期間ない場合にアイドル動作が行われる．自律的動作の実装においても，これと同じように，動作がないときに行うことがよいと考えられる．そこで，発話のみがあり，動作がない場合に自律動作を挿入することとした．この実装は図 3.5 の Distribution server で行う．図 3.7b) に自律的な動作のタイミングについて示す．発話のみがあるタイミングとして，発話命令のみが送られ，動作命令がない場合がある．このような場合は自律的な動作が挿入される．また，発話と動作は同期して実行することがあるので，開始時点で同期し，動作命令の方が先に終了してしまった場合，動作終了から発話終了までの間，自律的な動作が挿入される．



a) Timing of synchronization



b) Timing of autonomous head motion

図 3.7: 並列動作と自律動作のタイミング

3.4　ヒューマノイドロボットによるプレゼンテーションの心理学的評価

本節では，ヒューマノイドロボットによるプレゼンテーションの評価について，心理学的な見地から実験を行う．アニメーションエージェントを用いたコンテンツはテレビや Web 上などで数多く見られ，人々に受け入れられていると考えられる．そこで，アニメーションエージェントとヒューマノイドロボットをプレゼンテーションを通して比較することで実験を行う．それぞれのエージェントの印象の違いを明らかにし，ヒューマノイドロボットの有効な適用分野についての考察を行う．

3.4.1　実験条件

それぞれのエージェントを用いたプレゼンテーションを評価するため，ASIMO に MPML-HR, Peedy に MPML を用いてプレゼンテーションコンテンツを作成した．図 3.8 にそれぞれのエージェントの外見を示す．ヒューマノイドロボットには二足歩行ロボット ASIMO を用い，スクリーンエージェントには Microsoft Agents の一つである “Peedy” [77] を用いた．Peedy は鳥の形をしており，約 60 のジェスチャが用意されている．ASIMO の音声合成には NTT-IT 社製の FineVoice を用い，Peedy には Microsoft Agents で標準で使用される Microsoft Speech を用いることとした．Peedy と ASIMO の外見は異なるが，Peedy は最も一般的なキャラクタであり，コンテンツなどにもよく用いられている．Microsoft Agents の中でもジェスチャの種類が豊富であるためスクリーンエージェントとして Peedy を用いることとした．

プレゼンテーションコンテンツには天気予報を内容とするものを用いた．これは，天気予報は普段誰もが見慣れており，プレゼンテーションの内容による印象の影響を極力抑えることができるからである．コンテンツはニュースなどにおける天気予報を内容とするもので，全国の明日の予報，週間予報，最高気温，最低気温の説明を行う．また，台風の時期の予報であったため，それに対する注意なども含められている．スクリーンに表示される資料もこれらを示すものであり，エージェントはその資料の説明をするために適宜移動を行う．ジェスチャの回数と発話内容は ASIMO と Peedy 共に同一とし，プレゼンテーションの長さは共に約 5 分である．

被験者は 31 人であり，年齢は 20 代前半から 40 代半ば，男性と女性の数はほぼ同数である．この被験者は一般人の中から募集したため，例えば学生である，コンピュータに詳しい，といった偏りのある被験者とはなっていない．被験者は 3 グループに分けて評価した．グループ A は 9 人 (男性 5 人女性 4 人)，グループ B は 10 人 (男性 5 人女性 5 人)，グループ C は 12 人 (男性 5 人女性 7 人) である．グループ A では ASIMO によるプレゼンテーションを先に見せ，Peedy によるプレゼンテーションを後に見せ

Humanoid robot
ASIMO

2D character agent
PEEDY

図 3.8: 比較したプレゼンタ

た．グループ B および C ではこの順序を逆にし，Peedy を先，ASIMO を後として見せた．被験者にはそれぞれのプレゼンテーションの後すぐに評価を行ってもらった．被験者のうち 3 名は記載漏れがあったので評価から除外し，残りの 28 名について分析を行った．グループ B では実験の過程において，システムのトラブルが発生し，プレゼンテーションが途中で停止してしまった．そこで，以下の分析においては主にグループ A と C の被験者の評価をもとに分析を行った．

評価は 3 つのアンケートから成り立っている．一つ目は SD 法によるものである．30 のポジティブな形容詞とネガティブな形容詞の対を用いて 7 段階の尺度で評価する．7 段階の尺度とは例えば，良い-悪いという形容詞対であれば，非常に良い，かなり良い，やや良い，どちらでもない，やや悪い，かなり悪い，非常に悪いである．二つ目はプレゼンテーションに関する直接的な質問によるアンケートである．10 項目について 1 から 7 の 7 段階の得点をつける形式となっている．最後はプレゼンテーション全体で気づいたこと，感じたこと，よい点，悪い点などを自由記載形式で書く形式となっている．

3.4.2　実験結果

SD 法 (Semantic Differential Method)

分析の過程において，被験者によるアンケート結果を 1 から 7 までの数値に置き換えた．7 は最もポジティブな形容詞の評価 (e.g. 非常に良い) であり，1 は最もネガティブな形容詞の評価 (e.g. 非常に悪い) である．SD 法の評価については 30 形容詞対の結

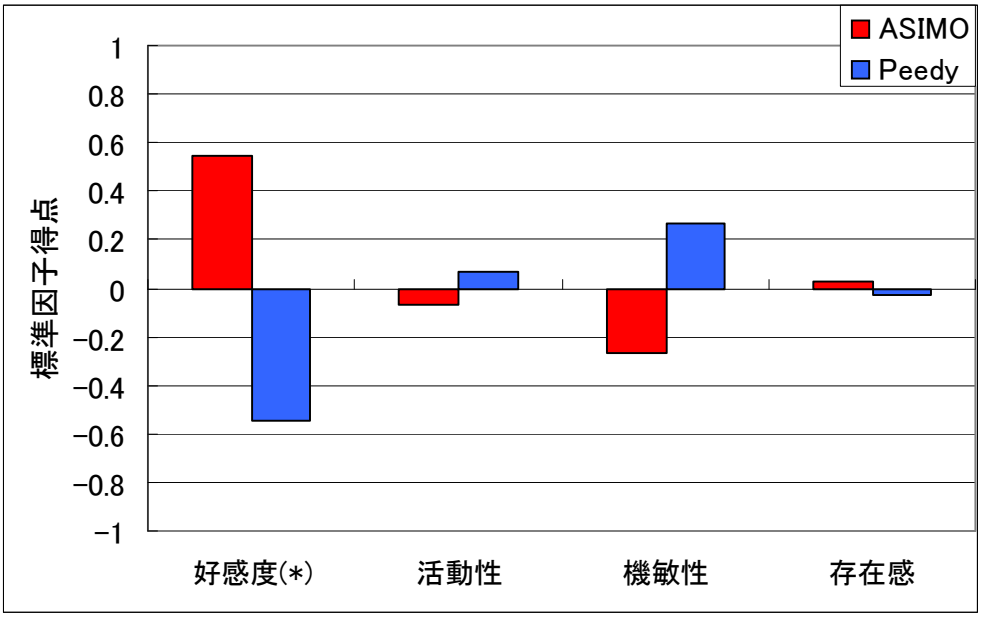
果に対して因子分析を行った．後続因子との固有値との差に基づいて 4 因子が妥当と判断し，抽出した．表 3.1 にバリマックス回転後の因子行列を示す．因子行列の各要素は因子負荷量を示し，各形容詞対と因子との相関を表す尺度である．

因子 I は ‘好きな’，‘良い’，‘かわいらしい’ および ‘感じのよい’ などの形容詞対において因子負荷量が高かったため，好感度と名づけた．因子 II は ‘動的な’，‘陽気な’ および ‘派手な’ などの形容詞対において因子負荷量が高かったため，活動性と名づけた．因子 III は ‘速い’，‘敏感な’ などの形容詞対において因子負荷量が高かったため，機敏性と名づけた．因子 IV は存在感と名づけた．表 3.1 に各形容詞対ごとの結果について示す．この結果については，プレゼンテーションの順序による被験者の人数の均衡をとることおよび，グループ B はシステムが途中で停止したことを考慮し，グループ A および C の被験者について分析している．なお，有意性の検証には t 検定を用いた．‘賢い’ という形容詞対については ASIMO の方が評価がよく，形容詞対の中で最も高い有意性が表れている．その他の形容詞対について，有意水準は異なるものの，ASIMO の方がポジティブな評価を獲得している．

図 3.9 に標準因子得点および標準因子得点の標準偏差を示す．標準因子得点とは，平均 0, 標準偏差 1 となるよう正規化したものである．因子ごとについて見ると，好感度因子においてのみ 0.01 %水準で有意な結果が表れた (*で表している)．有意性の検証には t 検定を用いたが，他の因子について有意な結果は表れなかった．この結果から ASIMO によるプレゼンテーションは視聴者に強く良い印象を与えていることが分かる．

表 3.1: 因子行列と形容詞対ごとの得点

形容詞対		バリマックス回転後の因子行列					平均得点および標準偏差		
		因子 I	因子 II	因子 III	因子 IV	共通性	ASIMO	Peedy	有意水準
好きな	嫌いな	**-0.78675**	0.31936	-0.12703	0.08302	0.74400	**5.56**(1.117)	4.72(1.044)	**5%**
良い	悪い	**-0.78656**	0.51389	-0.08222	0.03956	0.89109	**5.44**(1.117)	4.44(1.212)	**5%**
かわいらしい	にくらしい	**-0.75635**	0.32510	-0.03326	-0.26540	0.74930	**5.94**(0.911)	4.67(1.247)	**1%**
感じのよい	感じのわるい	**-0.73486**	0.41393	-0.24401	-0.15967	0.79639	**5.22**(1.083)	4.22(1.133)	**5%**
興味深い	退屈な	**-0.73164**	0.35448	-0.15871	0.22901	0.73859	**5.83**(1.302)	4.44(1.461)	**5%**
優れている	劣っている	**-0.72644**	0.40691	-0.20623	0.07246	0.74107	**5.33**(1.000)	3.83(1.067)	**1%**
賢い	愚かな	**-0.71394**	0.09645	0.02644	0.23454	0.57472	**5.22**(1.030)	3.94(1.079)	**0.1%**
充実した	空虚な	**-0.68114**	0.37493	-0.10571	-0.00182	0.61570	**4.89**(1.197)	3.61(1.061)	**1%**
思いやりのある	わがままな	**-0.61707**	0.07231	-0.31953	-0.40729	0.65399	**4.44**(1.012)	3.78(1.030)	N.D.
愉快な	不愉快な	**-0.58344**	0.46376	-0.20457	-0.10475	0.60829	**5.39**(1.208)	4.67(1.291)	N.D.
はっきりした	ぼんやりした	**-0.57532**	-0.00422	-0.05391	-0.02164	0.33438	4.06(1.615)	4.06(1.471)	N.D.
暖かい	冷たい	**-0.55837**	0.44342	-0.19611	-0.24736	0.60805	**4.56**(1.212)	4.11(1.197)	N.D.
面白い	つまらない	**-0.53243**	0.51019	0.08630	0.02465	0.55183	**5.50**(1.258)	4.38(1.208)	**5%**
穏やかな	激しい	**-0.52698**	0.20534	-0.09974	-0.24078	0.38780	4.89(1.410)	**5.00**(1.106)	N.D.
人間的な	機械的な	**-0.52209**	0.42453	-0.23427	-0.10680	0.51909	**3.94**(1.580)	3.33(1.291)	N.D.
親しみやすい	親しみにくい	**-0.51876**	0.39978	-0.17959	-0.07094	0.46622	**4.56**(2.034)	4.44(1.343)	N.D.
分かりやすい	わかりにくい	**-0.51790**	0.47188	-0.02801	0.07356	0.49709	4.67(1.155)	**4.72**(1.283)	N.D.
動的な	静的な	-0.21460	**0.77320**	0.13463	0.14000	0.68161	**5.06**(1.129)	4.78(1.133)	N.D.
陽気な	陰気な	-0.30707	**0.75108**	-0.24454	-0.12478	0.73379	**4.78**(0.975)	4.56(1.343)	N.D.
派手な	地味な	-0.11537	**0.71180**	-0.06622	0.33309	0.63531	**4.16**(1.014)	3.94(1.268)	N.D.
明るい	暗い	-0.28257	**0.70948**	-0.22059	-0.24463	0.69171	4.56(1.212)	**4.89**(1.370)	N.D.
うちとけた	堅苦しい	-0.41596	**0.60230**	-0.09894	0.03290	0.54666	4.72(1.660)	**5.11**(1.149)	N.D.
生き生きした	生気のない	-0.52460	**0.55996**	-0.15679	-0.02109	0.61379	**4.67**(1.333)	3.61(1.533)	N.D.
積極的な	消極的な	-0.27835	**0.50695**	-0.21912	0.21875	0.43034	**4.61**(1.061)	4.44(1.165)	N.D.
やさしい	こわい	-0.26043	**0.44639**	-0.33811	-0.32290	0.48567	**4.77**(1.397)	4.72(1.145)	N.D.
速い	遅い	-0.03381	0.09386	**-0.75667**	0.10904	0.59440	**3.63**(1.086)	3.61(1.112)	N.D.
すばやい	のろい	-0.18258	0.00038	**-0.62056**	0.34929	0.54044	3.28(1.407)	**3.83**(1.213)	N.D.
敏感な	鈍感な	-0.22581	0.40607	**-0.51234**	-0.04253	0.48018	3.50(0.764)	3.50(1.067)	N.D.
強気な	弱気な	0.04296	0.12938	-0.11519	**0.66086**	0.46858	4.22(0.916)	**4.61**(1.112)	N.D.
複雑な	単純な	-0.01391	-0.00324	-0.09987	**0.54571**	0.30798	**3.67**(1.202)	3.44(1.165)	N.D.



因子得点の標準偏差	好感度	活動性	機敏性	存在感
ASIMO	0.845	0.891	0.916	0.843
Peedy	0.834	1.094	1.010	1.135

図 3.9: 標準因子得点

直接的な質問を用いた評価

表 3.2 に直接的な質問を用いた評価の結果について示す．SD 法の評価と同様，グループ A およびグループ C の被験者を用いた．平均得点，標準偏差および t 検定による有意性の評価について示している．ASIMO の方がよい評価を得た項目は “興味をもてたか”, “テンポがよかったか”, “好印象だったか”, “人間に近いか”, “全体” の 5 項目であった．“興味” と “印象” については 5%水準で有意な結果となった．

Peedy の方がよい評価を得た項目は “集中できたか”, “感情表現が適切か”, “身振りが適切か”, “ポインティングが適切か” の 5 項目であった．“身振り” と “ポインティング” のについては 5%水準で有意な結果となった．

表 3.2: 直接的な質問の結果

	Peedy	ASIMO	有意水準
分かりやすさ	**4.85** (1.492)	4.35 (1.682)	N.D.
集中の程度	**4.95** (1.284)	4.35 (2.104)	N.D.
興味の程度	4.75 (1.479)	**5.70** (1.308)	**5%**
テンポのよさ	3.70 (1.187)	**4.35** (1.621)	N.D.
好印象だったか	4.55 (1.627)	**5.55** (1.161)	**5%**
感情表現の適切さ	**3.70** (1.926)	3.60 (1.497)	N.D.
身振りの適切さ	**4.25** (1.757)	3.00 (1.483)	**5%**
ポインティング	**4.80** (1.435)	4.05 (1.359)	**5%**
人間らしさ	3.20 (1.778)	**3.85** (1.682)	N.D.
全体として	4.30 (1.520)	**4.8** (1.503)	N.D.

自由記載アンケート

表 3.3 に自由記載形式のアンケートにおいて主に挙げられた意見について示す．この結果については，様々な意見を取り入れるため，全ての参加者 (グループ A, B および C) のデータを基にした．結果についてみると，音声合成の品質について不満の声が多く，ジェスチャに関する指摘も多かった．

ロボットの感情表現については意見が二つに分かれた．一つの意見は，ASIMO には顔の表情がないので感情表現を表すためにはジェスチャをより大げさにした方がよいというものであった．これとは逆に，もう一つの意見はロボットのプレゼンテーションにおいて感情表現は不要であるというものであった．

また，ASIMO について，ASIMO に興味があるため注目がいってしまい，プレゼンテーションに集中できなかったという意見が多く見受けられた．さらに，ASIMO のポインティングはスクリーンエージェントと比べて劣っているという意見も見られた．

表 3.3: 自由記載アンケートの結果

	割合
ASIMO に関心がいき プレゼンテーションに集中できない	61.3%
音声合成の品質について	83.9%
ジェスチャについて	80.6%
感情表現について	32.3%

グループ分けによる分析

被験者の中には ASIMO をよい評価とする者と，逆に Peedy をよい評価とする者がいた．そこで，直接的な質問を用いた 10 項目の評価の得点を合計し，ASIMO をよい評価とした被験者のグループと Peedy をよい評価とした被験者のグループの 2 つに分割して分析を行った．その得点の分布状況を図 3.10 に示す．Peedy をよい評価としたグループの得点の分布は ASIMO のグループよりも幅広く分散している．ASIMO をよい評価としたグループの人数と Peedy をよい評価としたグループの人数は同数であった．

さらに，SD 法を用いて，それぞれのグループの傾向を見ることとした．図 3.11 に因子分析の結果を示す．ASIMO のグループについては，好感度因子について圧倒的に ASIMO の方をよい評価としている．Peedy のグループでは，機敏性因子と存在感因子で Peedy の方をよい評価としている．Peedy のグループにおける好感度因子について見ると，平均的に ASIMO の方が評価が高い．好感度因子について ASIMO をよいとする評価は ASIMO のグループのような有意な結果とはならなかったが，ASIMO のプレゼンテーションを通して良い印象を受けていると考えられる．

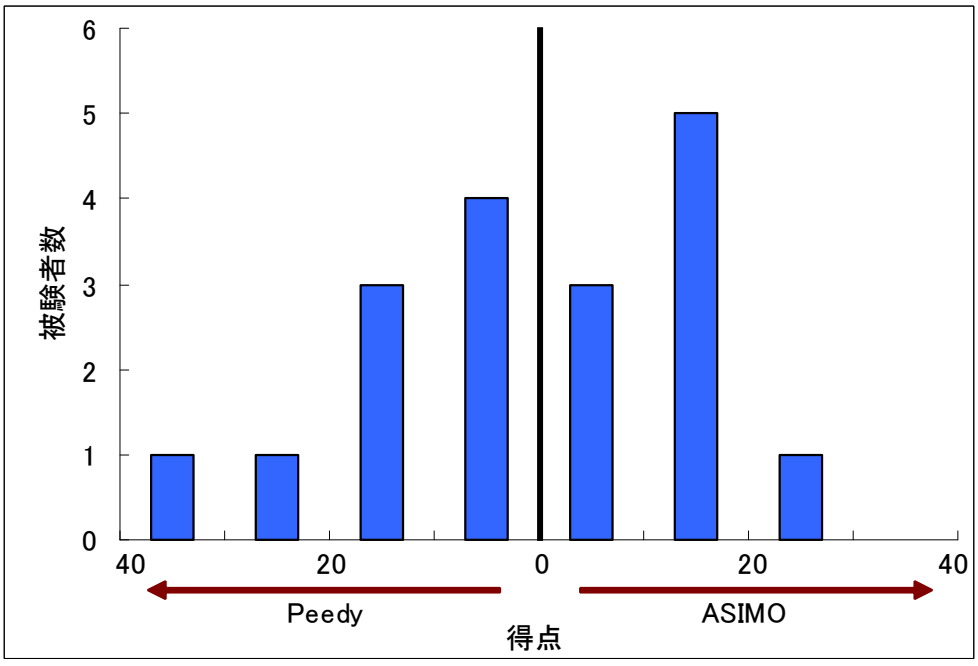
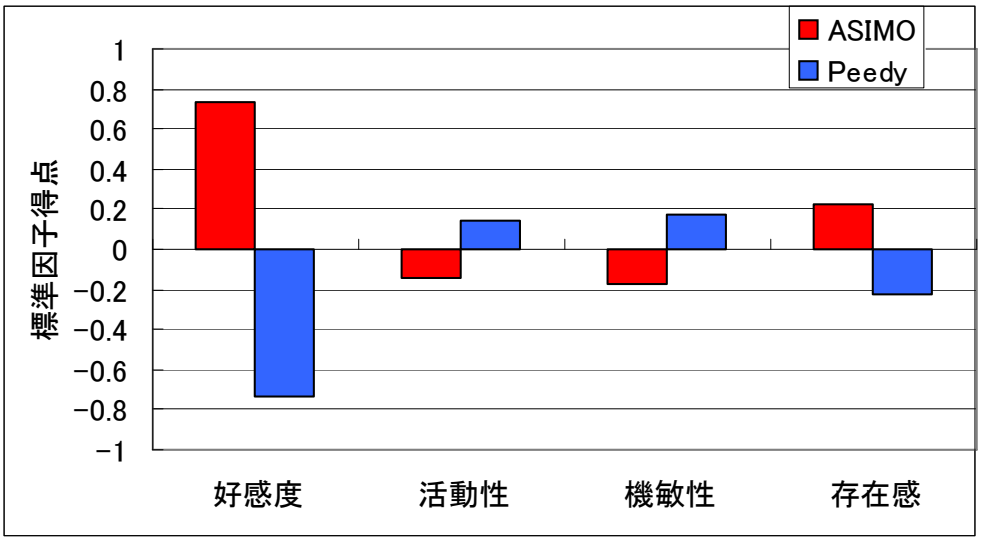


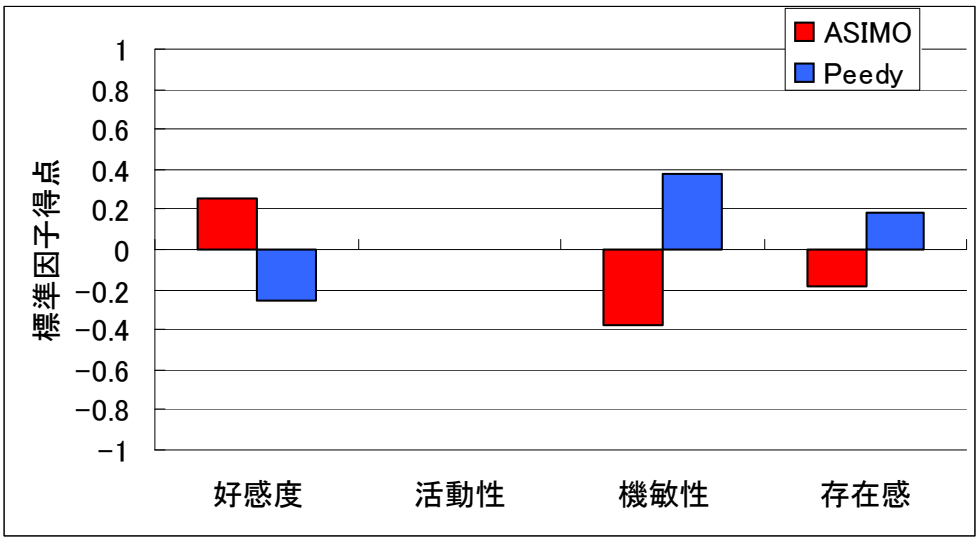
図 3.10: 得点の分布

(a) ASIMO をよい評価としたグループ



因子得点の標準偏差	好感度	活動性	機敏性	存在感
ASIMO	0.729	0.593	0.972	0.799
Peedy	0.613	1.269	0.996	1.121

(b) Peedy をよい評価としたグループ



因子得点の標準偏差	好感度	活動性	機敏性	存在感
ASIMO	1.027	1.080	0.804	0.877
Peedy	0.900	0.914	1.037	1.079

図 3.11: グループごとの標準因子得点

3.4.3　心理評価実験に対する考察

ヒューマノイドロボットが与える印象

被験者がヒューマノイドロボットのプレゼンテーションから受ける印象は以下のようにまとめられる．表 3.1 に示されるように，ASIMO は ‘好きな’, ‘良い’, ‘感じのよい’, ‘優れている’, ‘充実した’ といった形容詞対で高い評価を得ている．また，図 3.9 においては好感度についてよい評価を得ている．同様の傾向は表 3.2 の “興味”, “印象”, “全体” についても表れている．図 3.11 に示されるように Peedy を良く評価したグループにおいても ASIMO の好感度は Peedy よりも高く評価されており，このような傾向は一般的なものであると考えられる．これらの項目は，キャラクタ自体の優位性や印象の良さを示しており，プレゼンテーションにおいては重要な要素と考えられる．したがって，ASIMO はキャラクタ自体の印象がよく，コンテンツのプレゼンタとして有効なエージェントであると考えられる．

一方で，表 3.1 における ‘明るい’, ‘うちとけた’, ‘強気な’ といった形容詞対では評価が悪い．今回スクリーンエージェントとして評価対象とした Peedy はとても明るく陽気なキャラクタであり，性格がジェスチャなどにも表れている．そのような影響が実験結果に現れたと考えている．ASIMO はこれらの形容詞対について中間的な評価である 4 点以上を獲得しており，キャラクタの性格による相対的なものであり，絶対的な悪い評価ではないと考えられる．

適切なプレゼンテーションを行うためのエージェントの動作

ASIMO は実空間上に身体を持っており，動作により視聴者に与える印象が強いと考えられる．そのため，よいプレゼンテーションを行うためには，動作の選択に関する配慮が必要と思われる．ASIMO は表 3.2 に示される “身振りの適切さ” や “感情表現の適切さ” といった項目で Peedy に劣る結果となっている．ASIMO は強い印象を被験者に与えるため，適切でない表現が存在すると，Peedy と比べて悪い評価につながりやすいと考えられる．本評価実験では，ある程度の動作を含めたコンテンツを目指したため，余計な動作や適切ではない感情表現が含まれていた．このような影響により ASIMO の評価が悪くなったと考えられる．

感情表現については，3.4.2 節で示したように，ASIMO の表現をより大げさにした方がよいというものと，大げさにするのは避けるべきという意見があった．ASIMO の感情を視聴者に分かりやすくさせる，という側面では大げさにした方がよいと考えられるが，天気予報のプレゼンテーションという側面を重視すれば，大げさな感情表現は視聴者に違和感を与えかねず，避けるべきである．このような理由から，被験者の意見が分かれたものと考えている．

速さに関する項目では以下のような結果が得られた．表 3.1 における ‘すばやい’ という項目では ASIMO は悪い評価を得ている．ロボットの移動速度はスクリーン上のエージェントが動く速さと比べると実際に遅い．しかし，表 3.2 における “テンポ” において ASIMO はよい評価を得ている．この評価については，ジェスチャも考慮に含まれており，その影響が結果を導いていると考えられる．ASIMO と Peedy のジェスチャの時間にはそれほど大きな差は無いが，Peedy の方が無意味に長い時間を要するジェスチャが含まれていたように思える．

以上より，ヒューマノイドロボットのプレゼンテーションでは，スクリーンエージェントのプレゼンテーションと比べて，内容に合わない過度な表現を特に注意して避けるべきと考えられる．また，移動速度が遅いことから，ヒューマノイドロボットのプレゼンテーションでは必要最低限の移動命令を含めたコンテンツを作るべきと考えられる．

スクリーン上の資料を用いたプレゼンテーション

Peedy は表 3.1 における ‘わかりやすい’ や表 3.2 における “理解” や “ポインティング” といった項目でよい評価を得ている．今回対象としたプレゼンテーションはスクリーン上の資料を用いたものである．Peedy はスクリーン上に存在するため，より資料に近い位置に存在し，適切に指し示すことができる．このような要因が結果を導いていると考えられる．

スクリーンエージェントとヒューマノイドロボットをアドバイザとしたタスクの評価について，スクリーン上で完結するタスクについてはスクリーンエージェントの方が評価がよいがスクリーン上だけでは完結しないタスクについてはヒューマノイドロボットの方が評価がよいという報告がある [79]．この報告と本評価結果は同様の傾向があると考えられる．スクリーンエージェントを用いてスクリーン上の資料を用いたプレゼンテーションを行った場合，視聴者ははスクリーンのみに注目すればよい．しかし，ヒューマノイドロボットを用いたプレゼンテーションでは見ている人はロボットとスクリーンの両方に注目しなければならない．こうした距離的な差異も結果に影響しているものと考えられる．評価結果と先行研究を合わせると，スクリーン上の資料を用いたプレゼンテーションについてはスクリーンエージェントの方が正確に資料を指し示すことができ，評価も高く有効である．逆に，スクリーン外の資料を用いた場合，スクリーンエージェントでは適切なポインティングができず，ヒューマノイドロボットの方が有効であると考えられる．

ヒューマノイドロボットによるプレゼンテーションの適用分野

ヒューマノイドロボットは，アニメーションエージェントと比べて印象がよいという長所をもつ一方で，スクリーン上の資料を用いた説明はアニメーションエージェントと比べると評価が悪い．そこで，ヒューマノイドロボットによるプレゼンテーションの有効な適用分野として，第一に，展示場などにおけるプレゼンテーションが考えられる．このような場面では，講義のような内容の詳細を知るのではなく，インパクトや強い印象が重要である．したがって，印象のよいヒューマノイドロボットがプレゼンタとなることで，視聴者の興味をひくプレゼンテーションを行うことができると考えられる．

第二に，スクリーン上の資料のみでなく，実空間上に存在する三次元の商品等を説明する場合が考えられる．商品と同じ実空間上に存在するロボットがプレゼンテーションを行うことで，適切な指示が可能であり，視聴者に分かりやすい説明を達成することが可能である．

第三に，スクリーン上の資料を用いたプレゼンテーションであっても，アニメーションエージェントを補助的に参加させることが考えられる．スクリーン上の説明や指示はアニメーションエージェントに任せ，その他の説明はロボットが担当することで，長所を活かしたプレゼンテーションとすることが可能である．また，プレゼンタが複数いることで，論点について議論することもでき，プレゼンタが一人の場合と比べて分かりやすいコンテンツを作ることが可能である．アニメーションエージェントとヒューマノイドロボットを協調させたプレゼンテーションについては，MPML-HR ver. 2.0 として開発を行った (図 3.12).

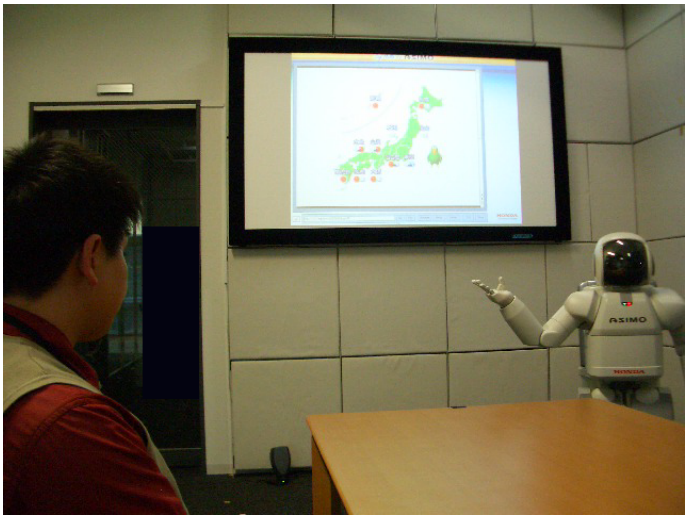


図 3.12: ロボットとアニメーションエージェントを用いたプレゼンテーション

3.5　提案する言語の記述容易性に関する考察

提案する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 MPML-HR は，容易な記述でロボットによるプレゼンテーションを実現する．そこで，この言語を用いることにより，容易な記述が実現できることについて検討を行う．

提案するヒューマノイドロボットのプレゼンテーションを実現するためのプログラム群を図 3.13 に示す．それぞれのプログラムは，TCP/IP 通信により接続されており，通信プログラムを必要とする．また，ヒューマノイドロボットによるプレゼンテーションを実現する上での本質的部分は，命令の解釈および，担当プログラム (音声合成プログラム，ロボットの制御プログラム) への送信であり，実装したシステムでは Distribution server がこの役割を担う．

Distribution Server は，C++により，2800 行程度の命令が記述されている．プログラムの流れを示すと，初めに記述されたコンテンツの命令系列を読み込み，それを担当するサーバに送信する．このとき，発話と動作の同期や，自律的動作をセットすることが行われる．命令送信後は担当サーバからの ACK 信号を待ち，ACK 信号受信により終了を検出し，次の命令を送信する．Distribution server はシステムの中核を担い，コンテンツを順に実行し管理する役割も果たしている．ユーザが，MPML-HR なしにヒューマノイドロボットによるプレゼンテーションを実現しようとする場合，Distribution Server に書かれている内容に相当するプログラムを構築しなければ，命令を順に実行し，発話，ロボットによる動作，スクリーンへの資料の表示，自律的動作を含むコンテンツを実現できない．

ヒューマノイドロボットによるプレゼンテーションを MPML-HR により実現する場合，例えばジェスチャ1 つを生成するのには，1 行記述すればよい．移動は 1 行の記述で実現可能であり，発話は `speak` で発話テキストを囲むだけで実現可能である．そして，これらを順に記述することで，一つのプレゼンテーションを作り上げることが可能である．しかし，MPML-HR を用いない場合，C++などの言語を用いて，Distribution server に相当するプログラムを記述しなければ，プレゼンテーションを実現できないため，一般的な低位レベルの言語での実現と比べると，MPML-HR では容易な記述でコンテンツを構築できることが分かる．

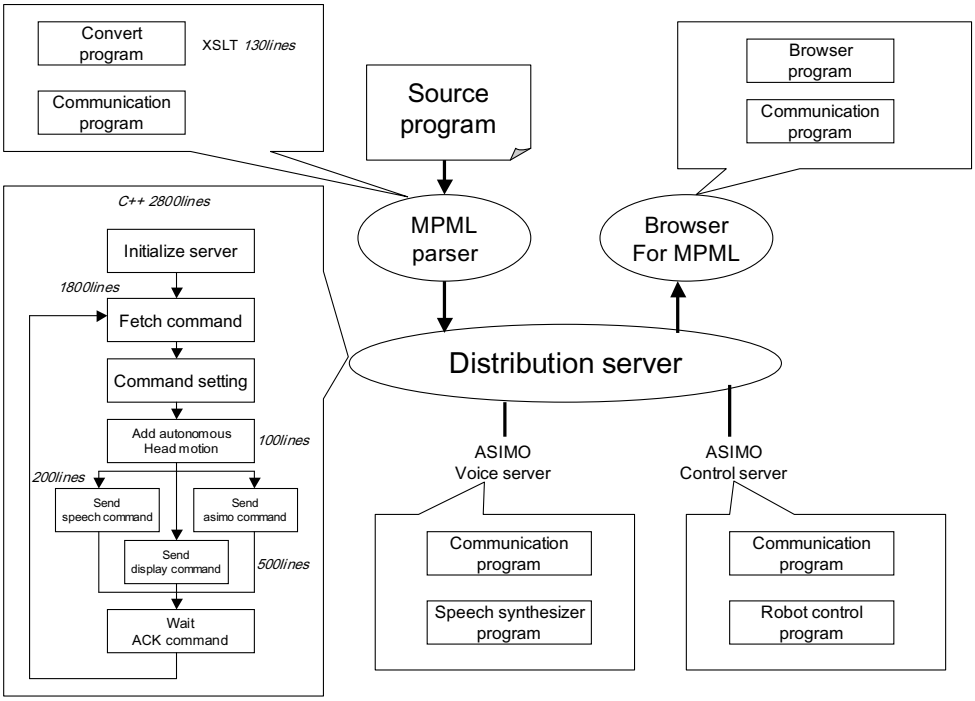


図 3.13: MPML-HR のプログラム構成

3.6　本章のまとめ

本章では，ヒューマノイドロボットによるプレゼンテーションを簡単な記述で実現するための言語について提案を行い，システムへの実装を行った．その際に，アニメーションエージェントを用いたプレゼンテーションコンテンツを記述するMPMLをヒューマノイドロボット用に拡張することで実現した．アニメーションエージェントとヒューマノイドロボットは実在する空間が異なるため，移動については身体の角度に関する引数を追加し，スクリーンの指示については新たな命令を導入した．また，自律動作や並列動作のための機構を導入することで，ヒューマノイドロボットによるプレゼンテーションの自然さを向上させた．

コミュニケーションロボットの評価については，SD法を用いたものが数多く用いられているため，アニメーションエージェントとヒューマノイドロボットをプレゼンタとしたコンテンツについて，それぞれを比較する形で心理学的評価実験を行った．そして，両者の違いを明らかにし，ロボットによるプレゼンテーションの適用分野について考察を行った．ヒューマノイドロボットはスクリーンの指示を正確にできないため，スクリーン上の資料の説明についてはアニメーションエージェントよりも評価が劣る．しかし，アニメーションエージェントと協調させることでこの問題を解決でき，また，スクリーンの外にある商品説明などはロボットの方が適切に説明できるという長所を持つ．そこで，強い印象を与えることに意義のある展示場などでのプレゼンテーションや，実空間上の商品説明などはロボットによるプレゼンテーションの有効な場面と考えられる．

さらに，MPML-HRを用いることで簡単にコンテンツが作成できることについても考察を行った．MPML-HRのような中位レベルの言語を用いない場合，インタフェースを意識しながらC言語などの低位レベルの言語によりプログラムを組まなくてはならない．このようなプログラミング言語は習得に時間がかかり，記述量も多くなりがちである．しかし，MPML-HRのような中位レベルの言語を利用することで，数十行程度の記述でコンテンツを作成することが可能となった．

第4章　インタラクション機構の導入

4.1　はじめに

第3章では，ヒューマノイドロボットを用いたプレゼンテーションシステムの実現について示した．このプレゼンテーションシステムはロボットからの一方的な説明を行うもので，ユーザからの問いかけに応答することはできない．人によるプレゼンテーションについて考えてみると，分からない箇所についての質問が発せられ，プレゼンタがそれに答えることにより内容の理解が深まる．ロボットによるプレゼンテーションにおいても音声インタラクションができることは重要な要素であると考えられる．そこで本章では，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ようなインタラクション機構を導入する．アニメーションエージェントを対象としたMPMLに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ようなインタラクションに関する命令が用意されていた．しかしこれは，プレゼンテーション中の特定の箇所でのみ音声割り込みを受け付け，その他の場所ではインタラクションをすることができなかった．視聴者からの音声割り込みはいつ発生するか分からず，特定の箇所のみでなく，プレゼンテーション中のどの場所においてもインタラクションできるような枠組みが必要であると考えられる．本章で導入するインタラクション機構はこの点を考慮し，プレゼンテーションの説明中いつでも音声割り込みを受け付けることができるように実装を行う．4.2節では，インタラクションコンテンツの実現に関する関連研究について示す．4.3節では，第3章で提案したMPML-HRのコンテンツ記述容易性という長所を活かしたインタラクション機能の導入を提案する．インタラクション実現のために新たな命令を導入し，音声認識誤りへの対応についても考慮する．システムの実装では，ロボット用の対話行動制御モジュールを利用する．4.4節では，提案するインタラクション機能を含むプレゼンテーションコンテンツを記述することのできる言語の評価として，先行研究として提案されているコンテンツ記述可能と記述量についての比較を行う．

4.2　関連研究

音声インタラクションが可能な案内システムやチケット予約システムなどは，音声対話システムとして数多く提案されている．また，インタラクションを含むコンテンツを作成するための記述言語もいくつか提案されている．本節ではこれら関連研究について紹介するとともに，ロボットの音声対話を実現するエキスパートモデルについて示す.

4.2.1　音声対話システム

人との音声コミュニケーションによりタスクを実行する音声対話システムがいくつか開発されている．その代表的なものとして，TOSBURG-II[80] がある．これは，ハンバーガーショップでの注文の受付をタスクとしたものであり，種類や数量を音声により入力することで注文タスクを達成する．音声対話システムでは，音声の誤認識が発生すると，ユーザの意図しない対応をとるため，この考慮が一つの課題である．TOSBURG-II では，雑音を含んだ音声信号を用いて音響モデルを学習することで雑音免疫学習という雑音に頑健な音声認識を行う．また，視覚的に注文内容を表示することで，ユーザへの確認を行っている．杉山らによる音声対話システムでは，音声認識の誤認識の問題を解決するため，信頼度の低い発話に対して問い返しを行うことで解決することを試みている [91]．このように，音声認識誤りに対処するため，音声認識レベル，対話システムレベルで音声認識誤りを避けるような試みがなされている．

これらは音声対話システムの実現例であるが，音声対話システムを実現するためのツールの開発も行われている [81, 84, 85]．特に，GALAXY を利用したアプリケーションはさまざまなものが開発されており，天気予報を行う JUPTER[82] やケンブリッジ市内案内をタスクとした VOYAGER，飛行機の座席情報を示す PEGASUS[83] などが存在する．

これらのツールを用いて作成した音声対話システムは，あるタスクを実行するシステムの構築を行うために有効な枠組みではあるが，一度システムを構築すると，その内容の変更は容易ではない．そこで，システムが読み込むためのコンテンツを記述する言語の提案が行われており，VoiceXML[86] や XISL (Extensible Interaction Scenario Language)[87] などがある．VoiceXML は電話での音声インタラクションを想定した記述言語であり，音声とプッシュボタンを入力とし，音声を用いた出力を行うインタラクションの記述が可能である．これを用いて音声対話システムを構築することで，内容の変更を簡単に実現することが可能である．しかし，エージェントには音声以外のモダリティがあるため，これらを利用するためには拡張が必要である．XISL は擬人化音声対話エージェント基本ソフトウェア開発プロジェクト (Galatea Project) で開発され

た記述言語であり，音声，マウス，キーボードによる入力が可能であり，音声合成，ブラウザ，アニメーションエージェントを用いた出力を行う．また，Galatea Toolkit[88] は VoiceXML や XISL を用いて音声対話システムを構築するための枠組みである．実際にこのツールを用いたアニメーションエージェントによる音声対話システムが構築されている．

また，チャットボット (Chatterbot) と呼ばれるおしゃべりエージェントの対話ルールを記述するための言語である AIML (Artificial Intelligence Markup Language)[89] や XbotML (A Markup Language for Human Computer Interaction via Chatterbots)[90] などが開発されている．しかし，現段階の実装ではテキストベースでのインタラクションを想定しており，音声やその他のモダリティを用いた応答については拡張されていない．

4.2.2 RIME (Robot Intelligence based on Multiple Experts)

ロボットを対象とした音声対話システムを実現するための枠組みとして RIME [1]がある．現状の対話システムは，タスクを限定せずに人間のようにあらゆる分野のインタラクションを行うことは難しい．そこで，1 つのシステムは，タスクに応じてドメインを限定し，そのシステムを切り替えていくことであらゆるタスクに対応するアプローチがとられている．RIME は，タスクに依存したエキスパートを複数構築することが可能であり，これを切り替えることでさまざまな分野の対話についてインタラクションを行うことが可能である．

図 4.1 に RIME のシステム構成を示す．RIME で用いるエキスパートはサブタスクごとに用意する必要がある．ロボットがサブタスクを実行する際には，対応するエキスパートは行動の決定を行う．ユーザからの発話が受け取られると，エキスパートがアクティベートされ，音声認識結果から文意を理解する．エキスパートはオブジェクト指向の枠組みのなかの一種のオブジェクトとして実現されており，対話を実現するための各エキスパートは内部状態を持ち，さまざまな情報を保持する．

システムが対話を実現するために，下位のモジュールが用意されており，エキスパートはインタフェースを通してこれらのモジュールを用いることで，ロボットの行動の実現やユーザからの発話の受付を行う．モジュールは 3 つのものがある．*understanding* モジュールは音声認識結果をエキスパートに送信し，最適なエキスパートをアクティベートする．*action selection* モジュールは選択されたエキスパートに対し，次の動作の要求を行う．*task planning* モジュールはロボットが動作している間，どのエキスパートをアクティベートするか決定する．これら 3 つのモジュールは発話割り込みを扱うために並列で動作する．

[1]RIME は文献 [92] では MEDBP と呼ばれていた．

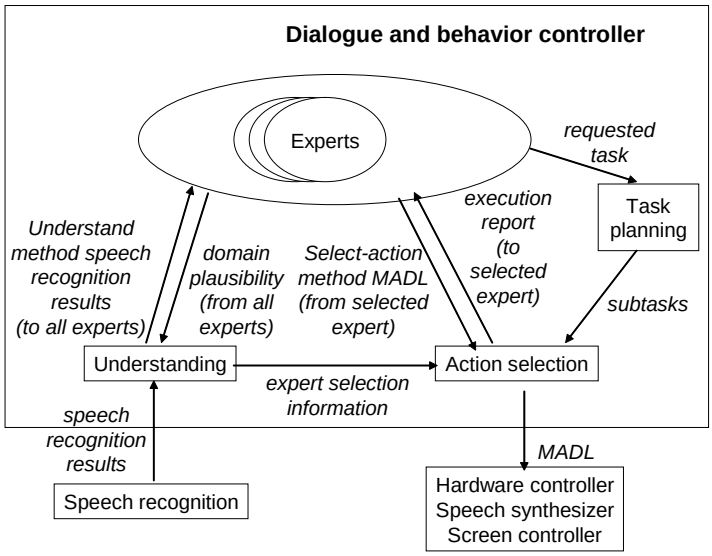


図 4.1: RIME のシステム構成

4.2.3　インタラクション機構導入についての考察

MPML-HR の長所は，誰でも簡単に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ができる点である．したがって，インタラクション機構の導入に際しても，簡単な記述を活かすことが重要である．音声対話システムを構築するためのツールが提供されているが，これらを用いるよりも自らコンテンツを作成できる VoiceXML や XISL などを利用することが好ましい．しかし，VoiceXML はロボットのモダリティを活かすためには拡張する必要がある．また，多機能な XISL を用いることでインタラクション機構の導入が可能であるが，プレゼンテーションに特化した言語ではないため，記述量は多くなりがちである．そこで，MPML-HR の長所を活かしつつインタラクション機能を導入するためには，MPML-HR にインタラクションに関する記述を追加することが好ましいと考えられる．

また，実装にあたっては，RIME を利用することが有効であると考えられる．ロボットによるプレゼンテーションは，ロボットの活躍分野の一つと考えられるが，プレゼンテーションタスクに限らず，さまざまなタスクを実現できることが望ましい．RIME を利用することで，プレゼンテーション以外のタスクの実現も容易となり，拡張性が増すものと考えられる．

4.3　インタラクション機構の実現

本節では，MPML-HR へインタラクション機構の導入を行い，MPML-HR ver. 3.0 とする．提案するシステムは，プレゼンテーション中にユーザから発せられる簡単な質問に答えることを目標とする．この機構を，プレゼンテーションの説明箇所を遷移することで実現し，そのための命令を MPML-HR へ追加する．また，インタラクションでは音声入力を受け付けるため，音声認識結果に対する信頼度を利用することで，音声認識誤りに対する考慮も行う．

実装にあたってはロボットの対話用エキスパートモデル RIME に MPML-HR 用エキスパートを追加することで実現する．RIME にはロボットの行動を制御するモジュールとユーザからの発話を理解するモジュールが用意されているため，これを用いることでインタラクションを含むプレゼンテーションの実現が容易となる．さらに，RIME を用いることで，他のタスクと切り替えながらプレゼンテーションコンテンツを実現することも可能となり，将来的な拡張を視野に入れると，ロボットによるプレゼンテーションが活かされる分野も広がると考えられる．

4.3.1　MPML-HR のインタラクション機能の拡張

説明箇所の遷移機構

プレゼンテーションにおいて視聴者からの音声割り込みとして想定されるものに，説明の省略，前に説明した箇所の再度の説明，あらかじめ想定していた質問がある．これらの要求に応えることができれば，ロボットによるプレゼンテーションはより充実したものになると考えられる．

そこで，インタラクションを実現するため，説明箇所を遷移するための命令である jump タグと do タグを導入することとした．しかし，これらの遷移を実現するためには，遷移先についての記述が必要である．そのため，コンテンツをいくつかのパートに分け，各パートの先頭に遷移できるようにする．コンテンツを分けるために，MPML-HR の page タグを用いた．page タグはスクリーンの 1 つのスライドごとに作られる．プレゼンテーションでは 1 つのスライドが 1 つの話題に対応すると考えられるため，視聴者からの割り込みに対して遷移するのに適した単位である．jump タグまたは do タグのアトリビュート page を指定することで，説明箇所の遷移を実現する．

また，認識文法についても定義する必要がある．認識文法の定義は grammar タグのアトリビュート recgram により行うこととした．さらに，認識文法の有効範囲を指定するための recog タグを導入し，アトリビュート page により対象ページを記述する．なお，recog タグによって指定されない文法はコンテンツの全ての範囲で有効なものとする．

音声認識誤りへの対応

視聴者からの割り込みの実現においては音声認識誤りの考慮が必要である．音声認識誤りが起こり，正解とは異なる発話割り込みと解釈してしまうと，プレゼンテーションは視聴者の意図しない場所に遷移する．このようなことが頻繁に起これば，使い勝手が悪くなり，インタラクション機能を導入する利益が失われてしまう．そこで，音声認識誤りにより，ロボットによるプレゼンテーションが誤った箇所に遷移した場合，視聴者からの指摘によりもとの状態に戻す機構について導入した．新たに導入したタグは goback である．

また，音声認識誤りを予防するための処理についても導入を行った．音声認識結果に対する信頼度が低い場合に，発話者に再度の発話を依頼する (聞き返し)，あるいはその解釈内容が正しいか Yes または No による回答を要求 (確認) するものである．発話割り込みにより認識された音声認識結果はその信頼度の大きさに応じて，棄却，聞き返し，確認，受理のいずれかがなされる．信頼度が低かった場合，誤認識である可能性が高いため，発話内容は棄却され，ロボットは何ら動作を行わない．信頼度が高かった場合には受理となる．受理された場合，ロボットは jump, do, goback で指定される動作を行う．聞き返し，確認は棄却と受理の間の信頼度であった場合になされる．その中でも信頼度が低い場合は聞き返し，信頼度が高い場合は確認となる．聞き返しでは，ロボットが “何かおっしゃいましたか？” と聞き返すことにより，発話者に対して再度の発話を要求する．何らかの発話割り込みが入った可能性が高いが，何と発話したかまでは正しく音声認識できていないような場合に有効である．確認では，音声認識内容に対して，その発話内容を行ったかどうかの聞き返しを行う．例えば，発話者が “明日の天気を説明して” といった場合，ロボットは “明日の天気を説明しますか？” と問い返す．確認の際は，常に Yes か No の答えが可能な形で問い返しを行うため，受け付け可能な認識文法が沢山ある場合と比べると二者択一の音声認識は正しくなされる可能性が高い．したがって，発話内容を正しく検出できている可能性が高い場合，聞き返しと比較して有効である．

棄却，聞き返し，確認，受理の境界を決定する閾値はあらかじめシステムで定義されている．しかし，より柔軟な割り込み機構を実現するため，この閾値を変更することも可能である．変更には，confidence-thresholds を用いて行う．棄却と聞き返しの境界には reject，聞き返しと確認の境界には ask-back，確認と受理の境界には accept 属性を指定することにより閾値の変更がなされる．

図 4.2 に示すように，reject, ask-back, accept の任意の値を同じにすることにより，聞き返し，確認を行わないよう設定することも可能である．

また，確認を行う際には，grammar タグの属性として confirm を記述することが必要である．この記述は，確認の際にロボットが発話する内容を定義するものである．この confirm 属性が省略された場合には，確認は行われず，本来確認が行われるべき範

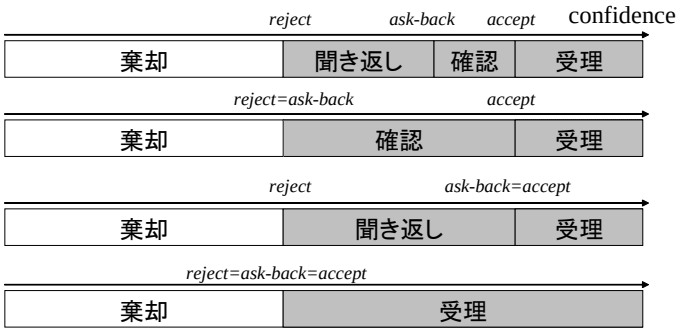


図 4.2: 信頼度の閾値設定例

囲の信頼度であった場合，聞き返し (聞き返しが行われない閾値設定では棄却) が行われる．閾値を変化させることで，さまざまな利用が可能である．例えば，認識文法が少ない場合には正しく認識できる可能性が高いため，聞き返しを省略することも有効である．また，コンテンツの最初の段階で，終了地点に遷移するような発話は想定し難いため，発話される可能性に応じて個別に閾値の設定を行うことも可能である．

ブラウザへのテキスト表示

エージェントを用いたアプリケーションでは，視覚を利用し，システムが現在理解している内容や発話内容を表示するものが数多く存在する．MPML でも，エージェントの発話は吹き出しによって画面に表示される．そこで，MPML-HR でも，ロボットによる発話内容と，認識可能な文をスクリーンに表示することとした．ロボットの発話内容を表示することで，聴覚で聞き逃した箇所も視覚で補うことができ，プレゼンテーションの理解が深まる．また，ロボットが認識可能な文を表示することで，ロボットとの対話を円滑に行うことが期待できる．

図 4.3 にブラウザの表示例を示す．ブラウザ中央にはスライドが表示され，ブラウザ右側に認識可能な文を，ブラウザ下側にはロボットの発話内容を表示する．発話内容については，発話命令の記述により自動的に表示されるが，認識可能な文についてはユーザが定義する必要がある．これは，`grammar` タグのアトリビュート `view` を用いて行う．

MPML-HR ver. 3.0 の命令セット

図 4.4 にインタラクション機構を含む MPML-HR ver.3.0 のタグ構造をまとめる．新たに導入したタグについて，以下に示す．

図 4.3: プレゼンテーション用ブラウザ

①Interaction
インタラクションに関する記述は，このタグで囲まれた範囲において行う．

②Recog
< **recog　page** = “　” > のように音声インタラクションの有効範囲をページ単位で指定する．このタグで有効範囲を指定しなかった場合，コンテンツの全ての範囲で有効な命令となる．

③Grammar
< **grammar　recgram** = “　”　**confirm** = “　”　**view** = “　” > のように記述する．アトリビュート `recgram` により，文法の定義を行う．文法の定義は，例えば

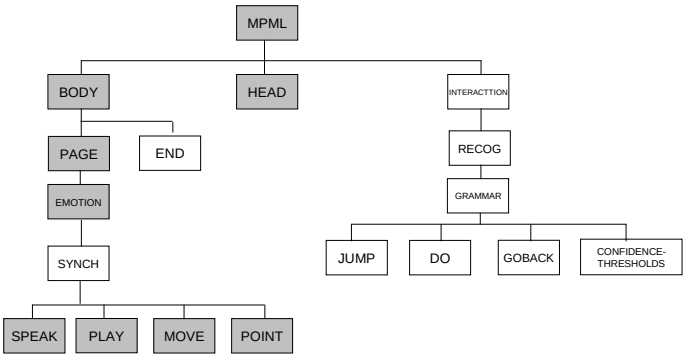


図 4.4: タグ構造 (MPML-HR ver. 3.0)

(前 | 2 枚目) のスライドを [もう一度] 説明して [ください]
のように，カギカッコ ([) で省略の定義を行うことが可能であり，丸カッコ (() で選択の定義を行うことが可能である．また，アトリビュート confirm により確認の際の発話を，アトリビュート view によりブラウザへ認識可能な文のテキスト表示を行う．

④Jump

< **jump　page** = “　″ > のようにページを指定することにより，コンテンツの状態を遷移させることが可能である．このタグで指定した場合，遷移先のページの終了後，もとのページには戻ってこない．

⑤Do

< **do　page** = “　″ > のようにページを指定することにより，コンテンツの状態を遷移させることが可能である．このタグで指定した場合，遷移先のページの終了後，もとのページに戻ってくる．この点が jump とは異なる．

⑥Goback

do や jump の代わりに定義しておくことで，説明箇所の遷移後一定期間，視聴者からの指摘に応じてコンテンツの状態を戻す．

⑦Confidence-thresholds

< **confidence − thresholds　reject** = “　″　**ask − back** = “　″　**accept** = “　″/ >
のように定義することで，図 4.2 に示す棄却と聞き返しの閾値，聞き返しと確認の閾値および確認と受理の閾値を設定することができる．

⑧End

本編のコンテンツとインタラクション用のコンテンツを切り分けるために用いる．本編のコンテンツの終了箇所にこのタグを記述することで，それ以降のコンテンツはインタラクション専用となる．つまり，このタグより後に記述されたコンテンツは jump または **do** によって指定されない限り実行されない．

⑨Synch

発話や動作の同期を行う．同期可能な命令は，speak と play, move または point である．

サンプルスクリプト

```
<?xml version="1.0" encoding="UTF-8"?>

<mpml>
<body>
 <page ref="./slide.ppt#1" id="明日の天気">
  <synch>
   <speak>明日の東京は晴れでしょう</speak>
   <play act="victory" />
  </synch>
 </page>
 <page ref="./slide.ppt#2" id="明後日の天気">                                    (1)
  <speak>明後日の東京は雨でしょう</speak>
 </page>
 <end />
 <page ref="./slide.ppt#3" id="昨日の天気">
  <speak>昨日の東京は曇り空でしたね</speak>
 </page>
</body>
<interaction>
 <grammar recgram="[それは] ちがうよ"
          confirm="遷移前に戻りますか？" view="違うよ"> ……..(3)
  <goback />……(4)
 </grammar>
 <grammar recgram="きのう [のてんき]"
         confirm="昨日の天気ですか？" view="昨日の天気">
  <do page="昨日の天気" />……..(5)
 </grammar>
 <grammar recgram="あさってときのう [のてんき]"
         confirm="明後日と昨日の天気ですか？" view="明後日と昨日の天気">
  <do page="明後日の天気" />                                                    (2)
  <do page="昨日の天気" />
 </grammar>
 <recog page="明後日の天気"> ……..……(6)
   <grammar recgram="あした [のてんき]"
            confirm="明日の天気ですか？" view="明日の天気">
     <confidence-thresholds reject="0.6"  ask-back="0.8" accept="0.9" />.....(7)
     <jump page="明日の天気"  /> ……..(8)
   </grammar>
 </recog>
</interaction>
</mpml>
```

図 4.5: MPML-HR ver. 3.0 により記述したサンプルスクリプト

図 4.5 に MPML-HR ver. 3.0 のサンプルスクリプトを示す．スクリプトは二つの部分からなる．一つは body で囲まれた範囲であり (図 4.5 (1))，プレゼンテーションコンテンツ本体を記述する．もう一つは interaction で囲まれた範囲であり (図 4.5 (2))，インタラクションに関する記述をする．grammar タグ (図 4.5 (3)) は認識文法を記述する．発話割り込みが認識文法の一つと一致すると，ロボットは grammar タグの内部に記述された内容を実行する．recog タグの内部には grammar タグを記述することができ (図 4.5(6))，これらの認識文法を受け付ける範囲を page 属性によって指定することができる．つまり，ページ範囲の指定された認識文法はその範囲内でのみ有効であり，recog タグによって囲まれていない認識文法 (図 4.5(3)) はプレゼンテーション全体で有効である．jump タグ (図 4.5(8)) と do タグ (図 4.5(5)) は発話割り込みを検出した際の実行内容を決定する．これらはどちらも，あるページに遷移するための命令である．また，複数の jump タグまたは do タグを記述することにより，一つの割り込みから複数ページを実行することも可能である．複数ページを指定する際には，grammar 内には jump のみまたは do のみで記述することが原則であるが，誤って jump と do を混在して記述した際には，最後に記述したものが優先される．goback タグ (図 4.5(4)) は，認識誤りによりプレゼンテーションコンテンツが間違った場所に遷移してしまった場合に，もとのコンテンツに戻るための命令である．この場合も，grammar タグを用いて，goback を受け付ける際の認識文法の記述を行う．grammar タグ内部には音声認識の棄却，聞き返し，確認および受理の閾値を変えるために confidence-thresholds タグ (図 4.5(7)) を記述することができる．

grammar タグでは，view 属性の記述により，受け付け可能な発話内容のスクリーンへの表示を行うことができる．MPML-HR のプレゼンテーションでは，画面にスライドが表示されるが，右サイドに受け付け可能な発話割り込みを，画面下にはロボットの発話内容を表示するスペースが設けられている (図 4.3 参照)．grammar タグで view 属性が記述された場合にはこの内容が右サイドに表示されるが，省略された場合には認識文法である recgram の内容が表示される．また，confirm 属性の記述により確認の際の発話を定義する．

図 4.6 に図 4.5 の MPML-HR ver. 3.0 サンプルスクリプトに関するインタラクション例を示す．

Robot: 明日の東京は…
User:　昨日の天気は？
Robot: はい，　(Go to “明後日の天気”)
Robot: 明後日の東京は …
User:　違うよ
Robot: 失礼しました，もとの説明に戻ります
Robot: 明日の東京は… (Go back to “明日の天気”)
User:　昨日の天気は？
Robot: はい，　(Go to “昨日の天気”)
Robot: 昨日の東京は曇り空でしたね
Robot: もとの説明に戻ります
Robot: 明日の東京は，… (Return to “明日の天気”)
…
Robot: 明後日の東京は…
User:　明日の天気
Robot: 何か言いましたか？ (聞き返し)
User:　明日の天気
Robot: 明日の天気ですか？ (確認)
User:　はい
Robot: はい，(Jump to “明日の天気”)
Robot: 明日の東京は晴れでしょう

図 4.6: プレゼンテーション中の対話例

4.3.2　マルチエキスパートモデルに基づく実装

システム概要

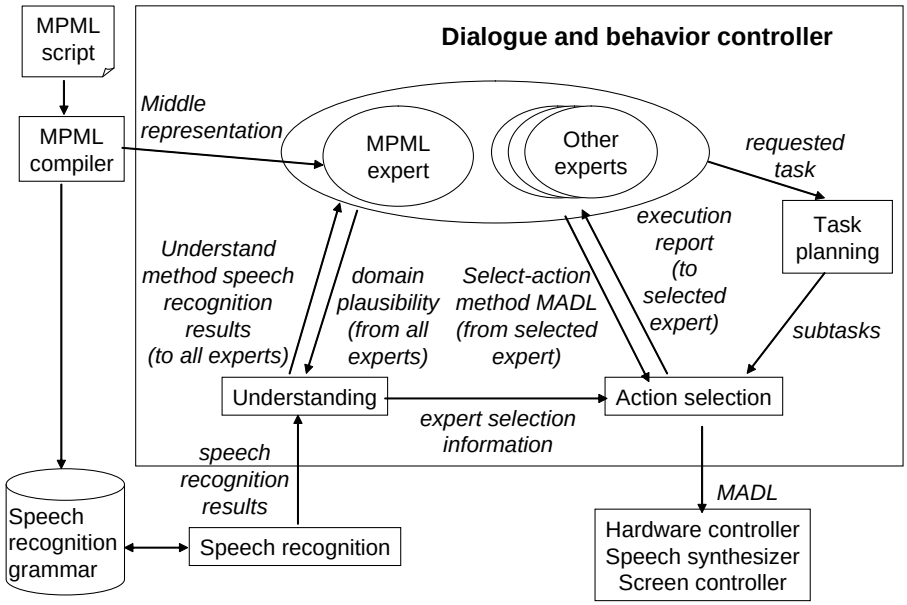


図 4.7: MPML-HR ver. 3.0 のシステム構成

図 4.7 に MPML-HR ver. 3.0 のシステム構成を示す．MPML-HR ver. 3.0 は対話ロボットの記号レベルの対話行動制御モジュールによって実現されている．対話・行動制御サブシステムはユーザからの発話を理解し，最適な次の行動を決定した後に動作を行う．本システムでは，出力形式の一単位を MADL (Multimodal Action Description Language) として表現する．この MADL には発話テキスト (e.g. “こんにちは ASIMO”)，身体を用いた動作 (e.g. “お辞儀” や “太郎に近づく”)，スライドの表示などが含まれる．そして，MADL の実行時には，全ての動作が同期して開始される．対話・行動制御サブシステムは，音声認識エンジン，音声合成器，ASIMO を制御するハードウェアコントローラと接続されている．

MPML のプレゼンテーションを実現するため，RIME に MPML 用のエキスパートを開発した．(以後，本節では MPML-HR ver. 3.0 の代わりに MPML と表記する．) 記述された MPML スクリプトは MPML コンパイラによって MPML エキスパートで用いる中間言語と音声認識用文法に変換される．コンテンツ実行の際，MPML エキスパートは初期化処理を行う．初期化処理では，MADL の動作系列と割り込みリストがスタックに積まれる．そして，一つずつ MADL を *action selection* モジュールに送信す

ることでコンテンツを実行する．その際，割り込みリストのスタックから 1 つポップし，有効な音声認識文法をセットする．ユーザからの音声割り込みが発せられた場合には，*understanding* モジュールを通して MPML エキスパートに音声認識結果と信頼度が送られる．MPML エキスパートは音声認識結果を受け取った際には信頼度に応じて棄却，聞き返し，確認，受理のいずれかを決定し，処理に応じてスタックの積み替えを行う．以上の流れにより，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．

MPML コンパイラ

MPML コンパイラは MPML スクリプトを中間記述に変換する．中間言語は複数の `block` 要素から構成され，一つの `block` 要素は MPML スクリプトの `body` 中の `page` に対応する．本編のコンテンツに加え，`jump` タグや `do` タグで指定されたページに対しても `block` が生成されるため，一つの `page` からは，最大で三つのの `block` が生成される．図 4.8 に図 4.5 の MPML スクリプトをコンパイルした中間記述の例を示す．

```
<block id="1" next="2">
 <interruption-handle recgram="gram_0_1" goto="goback." confirm="遷移前に戻りますか？"></interruption-handle>
 <interruption-handle recgram="gram_0_2" goto="3_r." confirm="昨日の天気ですか？"></interruption-handle>
 <interruption-handle recgram="gram_0_3" goto="2_r.3_r." confirm="明後日と昨日の天気ですか？"></interruption-handle>
 <madl>
  <show-slide uri="./slide.ppt#1" />
  <show-text view="side">昨日の天気¥n明後日と昨日の天気¥n</show-text>
 </madl>
 <madl>
  <utterance>明日の東京は晴れでしょう</utterance>
  <show-text view="bottom">明日の東京は晴れでしょう</show-text>
  <play name="victory" />
 </madl>
</block>
```

図 4.8: 中間記述例

`handle-interruption` 要素は認識文法ごとに生成される．この中には，文法 ID，遷移先 ID および確認発話の内容が定義されている．遷移先 ID は複数の記述がなされるよう設計されており，遷移先 ID の境界はピリオド (.) によって表される．

`madl` 要素は一つの MADL を表し，対話や行動の制御を表現する．`show-slide` 要素はスクリーン上に指定されたスライドを表示する命令である．また，`show-text` 要素は指定されたテキストをスクリーンに表示する命令である．テキストを表示する位置は，受け付け可能な認識文法を表示するスライド右側 (side) と，ロボットの発話を表示するスライド下側 (bottom) の二種類が存在する．`utterance` 要素，`play` 要素，`go-to`

要素はそれぞれ，発話，ジェスチャ，特定の位置への移動を表す．MPML での感情表現は，MADL の `emotion` 要素によって記述される．

中間記述への変換時には，文法リストファイルと文法記述ファイルが生成される．文法リストファイルは，認識可能な文法のリストが記述され，これを用いて認識文法が生成される．

MPML エキスパート

本来，MPML エキスパートはプレゼンテーションごとに作られる．しかし，コンテンツの内容以外の枠組みは共通であるため，MPML エキスパートはメソッドを共有し，中間記述と認識文法のみが異なるよう構成されている．MPML-HR はオブジェクト指向を用いることで，同じクラスとして表現されている．MPML エキスパートのクラスについて以下説明する．

initialize メソッドは中間記述を MADL スタックと割り込みリスト (文法 ID と遷移先から構成される) スタックに積むことを行う．

select-action メソッドは以下の処理を行う．まず，MADL スタックと割り込みリストスタックから一つずつポップする．ポップされた割り込みリストを基に，アクティブな文法 ID リストをセットする．そして，ロボットの動作を行うため，MADL を返す．

understand メソッドは音声認識結果に対する信頼度の値が閾値を超えているかチェックする．閾値は 4.3.1 節で示した reject, ask-back, accept の 3 つがある．信頼度が reject 以下であれば，発話割り込みによる動作は何も行われない．信頼度が reject より大きく，ask-back より小さい場合には，聞き返しを行うため，聞き返し用の発話内容の MADL をスタックの一番上にプッシュする．信頼度が ask-back より大きく，accept より小さい場合には，確認を行うため，確認用の発話内容の MADL をスタックの一番上にプッシュする．なお，確認の際には，認識結果が正しいか否か二者択一の発話を想定するため，その 2 つの認識文法のみを受け付けるよう，アクティブな文法 ID のセットを行う．信頼度が accept より大きな場合には，その発話内容により jump, do または goback の遷移が行われる．

図 4.9: ASIMO によるインタラクション

4.4　提案する記述言語の他の言語との比較

本節では，提案する MPML-HR と同様に，エージェントを用いたジェスチャや発話によるインタラクションを記述することのできる XISL との記述量での比較を行う．MPML-HR と XISL には違いがあるため，両者の共通する機能を用いて比較を行う．違いとして，入力モダリティについて MPML-HR は音声のみを受け付けるのに対し，XISL は音声，マウス，キーボードによる入力が可能である．また，出力モダリティはほぼ共通するが，MPML-HR がエージェントをロボットとするのに対し，XISL ではアニメーションエージェントである．また，MPML-HR では音声認識結果の信頼度を利用し，聞き返しやもとのコンテンツに戻る命令があるのに対し，XISL では誤認識に対する対応命令は用意されていない．

```
<mpml>
 <body>
  <page ref="./slide.ppt#1" id="first">
   <synch>
   <speak>Tomorrow, it will be fine in Tokyo.</speak>
   <play act="victory" />
   </synch>
  </page>
  <page ref="./slide.ppt#2" id="second">
   <speak>The day after tomorrow, it will rain in Tokyo.</speak>
  </page>
  <page ref="./slide.ppt#3" id="last">
   <speak>Yesterday, it was cloudy in Tokyo.</speak>
  </page>
 </body>
 <interaction>
  <grammar recgram="[How  was] yesterday">
   <do page="last" />
  </grammar>
  <recog page="second">
    <grammar recgram="[Please] (explain | talk) the first slide ">
      <jump page="first" />
   </grammar>
  </recog>
 </interaction>
 </mpml>
```

図 4.10: MPML-HR により記述したサンプルコンテンツ

```
<body>
 <dialog id="first" comb="seq" repeat="1" scope="document">
  <begin>
   <output type="mmi-browser" event="open">
    <![CDATA[
     <param name="window-name">Tomorrow_page</param>
     <param name="uri"> slide1.xml</param>
    ]]>
   </output>
   <par_output>
     <output type="agent" event="speech">
       <![CDATA[
       <param name="name">ASIMO</param>
       <param name="text">Tomorrow, it will be fine in Tokyo.</param>
       ]]>
     </output>
     <output type="agent" event="balloon">
       <![CDATA[
       <param name="name">ASIMO</param>
       <param name="text">Tomorrow, it will be fine in Tokyo.</param>
       ]]>
     </output>
     <output type="agent" event="action">
       <![CDATA[
       <param name="name">ASIMO</param>
       <param name="action">Hello</param>
       ]]>
     </output>
   </par_output>
  </begin>
  <exchange>
   <operation target="slide.xml">
    <input type="speech" event="recognize" target="slide1.grxml"match="/grammar/rule[@id=yesterday]" />
   </operation>
   <action>
    <call next = "#Yesterday_page" />
   </action>
  </exchange>
 </dialog>
 <dialog id="first" comb="seq" repeat="1" scope="document">
  <begin>
   <output type="mmi-browser" event="open">
    <![CDATA[
     <param name="window-name">AfterTomorrow_page</param>
     <param name="uri"> slide2.xml</param>
    ]]>
   </output>
   <par_output>
     <output type="agent" event="speech">
       <![CDATA[
       <param name="name">ASIMO</param>
       <param name="text">The day after tomorrow, it will rain in Tokyo.</param>
       ]]>
     </output>
     <output type="agent" event="balloon">
       <![CDATA[
       <param name="name">ASIMO</param>
       <param name="text">The day after tomorrow, it will rain in Tokyo.</param>
       ]]>
     </output>
   </par_output>
  </begin>
```

図 4.11: XISL により記述したサンプルコンテンツ 1

```
(続き)
<alt_exchange>
    <exchange>
     <operation  target="slide.xml">
       <input  type="speech" event="recognize" target="slide1.grxml"match="/grammar/rule[@id=yesterday]" />
     </operation>
     <action>
       <call   next = "#Yesterday_page" />
     </action>
    </exchange>
    <exchange>
     <operation  target="slide2.xml">
       <input  type="speech" event="recognize" target="slide2.xml"match="/grammar/rule[@id=tomorrow]" />
     </operation>
     <action>
       <goto   next = "#Tomorrow_page" />
     </action>
    </exchange>
  </alt_exchange>
 </dialog>
 <dialog  id="first" comb="seq" repeat="1" scope="document">
  <begin>
   <output  type="mmi-browser"  event="open">
    <![CDATA[
     <param  name="window-name">Yesterday_page</param>
     <param  name="uri"> slide3.xml</param>
    ]]>
   </output>
   <par_output>
     <output  type="agent"  event="speech">
       <![CDATA[
       <param  name="name">ASIMO</param>
       <param name="text">Yesterday, it was cloudy in Tokyo.</param>
       ]]>
     </output>
     <output  type="agent"  event="balloon">
       <![CDATA[
       <param  name="name">ASIMO</param>
       <param name="text">Yesterday, it was cloudy in Tokyo.</param>
       ]]>
     </output>
   </par_output>
  </begin>
  <exchange>
   <operation  target="slide.xml">
     <input  type="speech" event="recognize" target="slide1.grxml"match="/grammar/rule[@id=yesterday]" />
   </operation>
   <action>
     <call   next = "#Yesterday_page" />
   </action>
  </exchange>
 </dialog>
</body>
```

図 4.12: XISL により記述したサンプルコンテンツ 2

	MPML-HR	XISL
Tag	16	60
Attribute	12	76

表 4.1: MPML-HR と XISL の比較

図 4.10 に MPML-HR により記述したサンプルコンテンツを，図 4.11 および図 4.12 に XISL で記述した同様のサンプルコンテンツを示す．各言語により記述されたコンテンツのタグの数とアトリビュートの数をまとめると表 4.1 のようになる．MPML-HR では，同様のコンテンツを 4 分の 1 以下の記述で書くことが可能である．

MPML-HR が少ない記述でコンテンツを作成できる理由として，出力をエージェントによる動作と発話に限定し，エージェントの発話内容の画面への表示は自動的に行われるのに対し，XISL ではアニメーションエージェントだけではなく，発話のみや，テキストによる出力をすることも想定されているため，その種類の出力を行うかなどの記述が余計に必要である．また，XISL が各コンテンツのページごとにインタラクションを記述するのに対し，MPML-HR ではインタラクションをコンテンツとは別に記述するため，全てのページで有効な記述をすることができ，このような場合に記述量を削減することが可能である．

XISL は，対話システムにおいて，例えば何らかの販売コンテンツを構築することが可能である．客からの注文を受け付け，最終的に金額を算出して音声で応答するようなコンテンツを記述することができる．そのために，コンテンツの遷移時に変数を渡すことができる構造となっている．一方で MPML-HR はあらかじめユーザが記述したコンテンツに遷移することのみが可能であり，注文に応じて金額を計算するなどの処理はできない．しかし，プレゼンテーションというタスクに限定すれば MPML-HR のような機能があればほとんどのコンテンツが記述可能であり，容易な記述性という点においては MPML-HR の方がよいと考えられる．

また，アニメーションエージェントを用いたコンテンツと比べると，ロボットは自己のマイクを用いると周囲の雑音が混入する可能性が高く，音声認識誤りへの対応は重要な課題である．そのような機能について検討を行い，音声認識誤りの予防や，誤って遷移した際にもとに戻る機構を用意した MPML-HR ver. 3.0 はロボットの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記述では有効な言語であると考えられる．

4.5　本章のまとめ

本章ではインタラクション機能を有するロボットのプレゼンテーション記述言語 MPML-HR ver. 3.0 の設計と実装を行った．MPML-HR ver. 3.0 を用いることにより，視聴者からの説明の省略，再度の説明，想定される内容の説明を行うインタラクションを専門的な知識なしに作成することが可能となった．また，音声インタラクションで問題となる音声認識の誤認識に対処するため，音声認識結果の信頼度を用いて棄却，聞き返し，確認を行うこととした．さらに，誤った認識により意図しない箇所へ説明が遷移してしまった場合には，ユーザからの指摘によりもとの説明に戻る機構を導入した．

MPML-HR ver. 3.0 の構築には，マルチエキスパートに基づくロボット対話システムコントローラが役立つと考え，これをベースにした実装を行った．MPML-HR エキスパートを構築することで，将来的に他のタスクを対象とした対話システムと結合することも可能である．

また，提案するインタラクション機能の拡張はロボットを対象として実装を行ったが，アニメーションエージェントをプレゼンタとした同様の機構を実装することも可能である．

第5章　ロボットの動作音に頑健な音声認識

5.1　はじめに

第4章では，プレゼンテーションシステムへのインタラクション機能の導入について示した．音声インタラクションでは，音声認識の誤認識の問題があり，これを解決するため，システムレベルでの対応を導入した．しかし，雑音のある環境では，システムレベルでの対応のみならず，音声認識レベルでの対応もしなければ不十分な場合が存在する．具体的には，プレゼンテーションの中にはジェスチャや移動が含まれるため，動作中に音声認識を行わなければならない場合がある．しかし，ロボットが音声認識を行う場合，動作音が雑音となりSNR (Signal to Noise Ratio) が低下するという問題がある．ロボットは動作を行うために，多くのモータ，CPU, これらを冷却するためのファンを搭載している．そして，定常時にはモータ音やファン音が，動作中には手足の動作に伴うモータ音が発せられる．これらの音はマイクに近いところから発せられることもあり，場合によっては雑音が目的の音声よりも大きくなることもある．さらに，位置の変化等によりマイクに混入するモータ音やファン音も変化するため，SNRは不規則に変化する．これらの理由により，ロボットの動作中における音声認識は難しい問題である．人・ロボットコミュニケーションの研究では，高雑音下での音声認識を避けるため，ロボット自身のマイクを用いずに接話マイクによる音声認識が行われている [97]．しかし，常に接話マイクを用いることは利用者にとって煩わしく，ロボット自身のマイクで音声認識を行うことは重要であると考えられる．また，発話が検出された際には動作を止めることによって雑音を軽減する方法 (stop-perceive-act 法) もあり，Sony のペットロボット AIBO はこの方法を採用している．しかし，発話の度に動作を停止させることは円滑なプレゼンテーションの妨げともなりかねず，動作中における音声認識を実現することは重要な課題であると考えられる．

そこで，本章では，ロボットの動作中に音声認識を行うことを目指し，ロボットの動作音に頑健な音声認識手法の検討を行う．5.2 節では，雑音に頑健な音声認識に関する関連研究について示す．5.3 節では，動作音に頑健な音声認識手法を提案する．動作音は推定可能である点に着目し，Missing Feature Theory に基づく手法を提案する．提案

手法では，有効な手法として雑音環境下の音声認識に利用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る手法，Spectral Subtraction を用いる手法および音響モデルを適応する手法を Missing Feature Theory に基づく手法と効果的な組み合わることで，動作音への適応を行う．5.4 節では，提案手法の有効性を示すため，従来手法との比較実験を行い，その結果について示す．5.5 節では，提案する手法を第 4 章で示したプレゼンテーションシステムに組み込むことを視野に入れ，リアルタイムでの実装について示す．

5.2　関連研究

音声認識では，雑音への頑健性を向上させることが大きな課題であるため，さまざまな手法が提案されている．まず音声認識システムの概要について示した後に，前処理，音声特徴量，正規化，デコーダ内部での処理，音響モデルに分け，雑音への頑健性を向上させるための技術について触れる．最後にロボットの動作音への適応に有効な音声認識手法の考察について示す．

5.2.1　音声認識システム概要

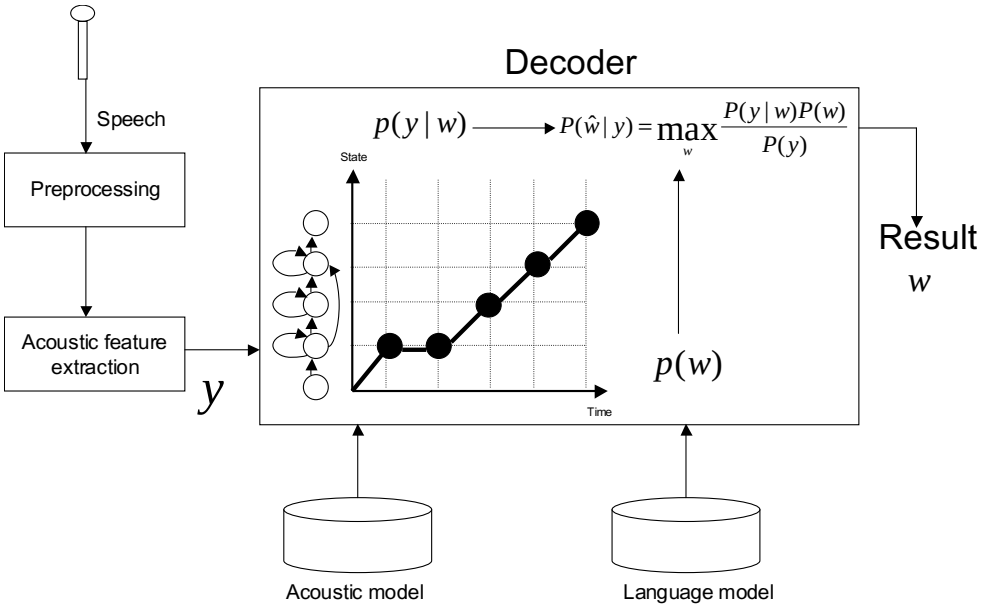


図 5.1: 音声認識システム

音声認識システムは，図 5.1 に示す流れに従って行われる．マイクから入力された音声は，前処理がなされる．主に信号処理技術を用い，雑音を除去することが行われる．カーナビシステムや，ロボット自身のマイクによる音声認識では対象とする音声以外の雑音が大きいため必須の処理であるが，接話マイクを用いた音声認識では混入する雑音が小さいためなされないこともある．

次に，雑音が除去された音声信号から音声特徴量が抽出される．音声特徴量とは，音声の波形から音の特徴を抽出する処理である．人間の聴覚では周波数領域の分析により音の知覚がなされており，コンピュータによる音声認識でも周波数領域での分析が

なされるのが一般的である．人間が発する声は唇や顎，舌などの形によって音が決定される．これらの形は調音フィルタの役割を果たし，肺から発せられた空気がフィルタリングされることである種の音となる．音は調音フィルタによって決定されるが，調音フィルタの形を表すのが，スペクトル包絡である．よって，音声特徴量の抽出では，FFT(Fast Fourier Transform) の後，DCT(Discrete Cosine Transform) を行い，ケプストラム領域に変換することでスペクトル包絡を抽出する．

得られた音声特徴量から，単語または文を予測する．音声特徴量 y, 単語または文 w とすると，確率的に最も高い $P(w|y)$ が認識結果として得られる単語または文になる．しかし，$P(w|y)$ を直接求めることは困難であるので，ベイズ規則を用いて $P(w|y) \approx \max_w \frac{P(y|w)P(w)}{P(y)}$ を求めることとなる．ここで，$P(y)$ は w とは無関係なので，$\max_w P(y|w)P(w)$ を求めることが問題となる．

$P(y|w)$ を求めるために統計的モデルが必要であが，次のようなモデルが一般的に用いられる．音声認識では，統計的モデルをコンパクトに表現できることから，ベイジアンネットワーク (Bayesian network) の一種である HMM(Hidden Marcov Model) を用いる．HMM は時系列信号の確率モデルであり，複数の定常信号源の間を遷移することで，非定常な時系列信号をモデル化する．このモデルを音声特徴量の系列に適用することで $P(y|w)$ を得る．HMM は図 5.2 に示すように，初期状態と終了状態を含め 5 状態 (または 3 状態) を持ち，各状態ごとに遷移確率と出力確率を持つ．出力確率には，混合正規分布を用いた確率密度関数を用い，尤度として定義する．正規分布は平均と分散により定義されるが，複数の正規分布を組み合わた混合正規分布を用いることで複雑な関数も表現することができる．混合数には，16 混合や 32 混合の正規分布がよく用いられる．そして，混合正規分布を用いた確率密度関数を用いることで，音声特徴量 y を出力する出力確率を求めることができる．

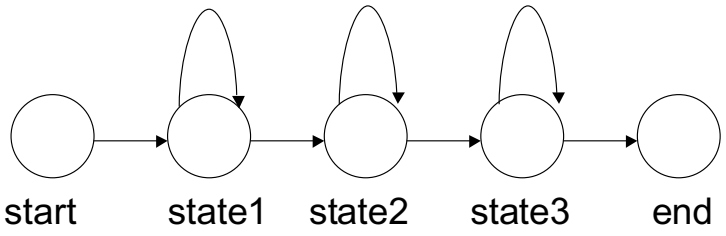


図 5.2: Hidden Marcov Model

HMM を用いた確率計算では，ビタビアルゴリズム (Viterbi algorithm) を利用することで効率的に $P(y|w)$ の計算を行う．なお，HMM の学習時にはバームウェルチのアルゴリズム (Baum-Weltch algorithm) を用いることで効率的な計算を行う．

HMM は a, i などの音素ごとにモデルを持つが，前後の音素により音声の特徴が変わることから，a-i+u, e-i+o などのように前後を含めた音素のモデルなどが利用される．a, i などの単一の音素をモノフォン，a-i+u, e-i+o などの前後を含めた音素をトライフォンと呼ぶ．音声認識ではトライフォンモデルが用いられることが多いが，音素の数は膨大であるため，類似の音素によりグループ化することで効率的な学習を行っている．

次に，$P(w)$ を求める過程について示す．この過程では言語モデルと呼ばれる統計的モデルが用いられる．これは，単語の連続のしやすさをモデル化したものである．文法的に連続し得ない単語列は，ここで排除し，認識結果に反映することができる．単語認識や，チケット予約システムなどの，少ない種類の文章を認識するタスクでは，どの単語や文も同一の確率で出現されることを前提としたネットワーク文法が用いられる．しかし，大語彙連続音声を対象とした認識タスクでは，直前の単語の系列からの出現確率を定義した N グラムモデルが用いられる．通常，2 グラムや 3 グラムモデルが用いられる．N が大きい方が認識性能がよくなることが予想されるが，計算量も爆発する．N をある程度大きくしても認識性能向上が小さい一方，計算量は増えるため，3 グラム程度が適当であるとされている．

以上のように，HMM による音響モデルと言語モデルを用いて統計的に認識結果 w の推定を行うのが音声認識である．現状の音声認識技術を用いることで，接話マイクを用い，正しい日本語文法を用いた場合，90%以上の性能で音声認識が可能であるとされている．しかし，雑音のある環境，話し言葉などの崩れた日本語文法を用いた場合，音響モデルと言語モデルが対応できず，認識性能は 50%にも満たないことがある．本研究では，雑音環境下における音声認識を対象とするため，以下，雑音の頑健性向上に対するアプローチを紹介する．

5.2.2　前処理

本節では，雑音除去を目的とした前処理について示す．接話マイクを用いた音声認識では雑音の混入が少ないため，雑音除去のための前処理を行わないこともあるが，ロボットによる音声認識では雑音の混入が大きいため，雑音除去は必須であると考えられる．以下，提案されている手法を紹介する．

SS (Spectral Subtraction)

SS は最も広く用いられている雑音除去方法である．Boll[98] らによって提案された手法であり，式 (5.1) のようにスペクトル領域において推定雑音を減算する．$|S_i(\omega)|$ は推定音声信号のスペクトル，$|X_i(\omega)|$ は入力信号のスペクトル，$|\overline{N_i(\omega)}|$ は平均雑音スペ

クトルである．

$$|S_i(\omega)| = |X_i(\omega)| - |\overline{N_i(\omega)}| \tag{5.1}$$

Berouti らは，SS をスペクトルのパワー領域で減算することを提案している [99]．Boll の方法では，入力信号よりも平均雑音の方が大きいと，スペクトルが負となってしまう場合が生じる．Berouti らの手法ではこの問題を解決し，入力信号から平均雑音を引いた結果が負とならないようにフロアリングを導入している．

$$|S_i(\omega)|^2 = \begin{cases} |X_i(\omega)|^2 - \alpha|\overline{N_i(\omega)}|^2 & if \quad |X_i(\omega)|^2 - \alpha|\overline{N_i(\omega)}|^2 \geq \beta|\overline{N_i(\omega)}|^2, \\ \sqrt{\beta}|\overline{N_i(\omega)}|^2 & otherwise. \end{cases} \tag{5.2}$$

また，SS の処理効果を高めるため，SNR に依存した値を減算することも提案され，改善が報告されている [100, 101, 102]．

$$|S_i(\omega)| = \begin{cases} |X_i(\omega)| - \phi_i(\omega) & if \quad |X_i(\omega)| > \phi_i(\omega) + \beta|\overline{N_i(\omega)}|, \\ \beta|X_i(\omega)| & otherwise. \end{cases} \tag{5.3}$$

ここで，$\phi_i(\omega)$ は次のように定義する．

$$\phi_i(\omega) = \alpha_i(\omega) - sig(\rho_i(\omega))[\alpha_i(\omega) - |\overline{N_i(\omega)}|] \tag{5.4}$$

$\rho_i(\omega)$ は推定された SNR の値を示す．

SS は通常，無音区間や，音声入力前のマイクに混入する音を用いて推定雑音とする．定常雑音はこれらの区間の音を平均することで推定が可能であり，SS による除去が可能であるが，非定常雑音は雑音推定が困難なため，SS による除去が難しい．そこで，二本のマイクを用いた 2ch-SS などが提案されている [103, 104]．2ch-SS では，音声を収録するためのマイクの他に雑音を収録するためのマイクを用いて雑音を推定する．このようにすることで，非定常雑音に対しても適用が可能である．しかし，二つのマイクに混入した音は，音響伝達特性が異なるため，これらを考慮した雑音推定が必要であり，LMS アルゴリズムや RMS アルゴリズムなどを用いることで音響伝達特性の違いを吸収している．

SS をロボットの動作音へ適用した例もある [105]．ロボットの動作音は非定常であるが，音声認識時に発生するその他の雑音と比べると推定が容易である．すなわち，環境雑音は音源位置や音源数など何ら雑音に関する情報がないのに対し，ロボットの動作音は動作情報を用いることで推定が可能である．この方法では，ロボットの発する動作雑音を関節位置を入力としたニューラルネットワークにより推定することで，雑音除去を行う．小型ロボットを用いた実験において，その有効性が示されている．

また，SS を行うことで SNR は向上するが，引き残し成分があることで，認識性能が低下することがある．これは，音声特徴量の抽出段階において，引き残し成分の影

響により学習されている音声特徴量との差が拡がってしまうからであると考えられる．そこで，山出らは SS 処理後の信号に雑音を重畳し，引き残し成分を平坦化することで認識性能が向上することを報告している [106].

SS は実装が簡単であり，効果も大きいことからよく用いられている．特に，無音区間を用いることで簡単に実現できることから，式 (5.2) がよく用いられている．その際のパラメータは $(\alpha = 1, \beta = 0.1)$ などが用いられることが多い．

フィルタ

SS も一種のフィルタにより雑音を除去する方法であるが，信号処理分野で使用されているフィルタを用いることで雑音を除去する方法も提案されている [107]．短時間内で目的信号と雑音がともに定常であると仮定し，誤差を最小化するウィーナーフィルタ (Wiener Filter) を用いたアプローチや [108, 109]，ウィーナーフィルタを拡張し，非定常信号を扱うこともできるカルマンフィルタ (Kalman Filter) を用いた手法もある [110]．また，Ephraim らの提案する方法 [111] では，スペクトルパワーに基づいて適応的に音声存在確率を抽出し，この音声存在確率に基づいて入力信号に重畳された雑音を軽減する．この方法はスペクトルの連続性が考慮されているため，歪みを軽減することが期待できる．

マイクロホンアレーを用いた手法

マイクロホンアレーを用いた音源分離はロボットの音声認識の場面で数多く採用されている．これは，ロボット自身のマイクによる音声認識に対する要求が強いこと，ロボット自体のコストから考えると複数マイクのコストは低く，複数マイクを備えるスペースもロボットにはあることなどの理由によるものと考えられる．

マイクロホンアレーを用いた音声認識では，ビームフォーミング (BF:Beam Forming)[113]，独立成分分析 (ICA:Independent Component Analysis)[115]，あるいは，幾何学的音源分離 (GSS:Geometric Source Separation)[116] による手法が提案されている．BF は一般的な音源分離手法であるが，音源分離による雑音の消し残しが多く生じる．雑音の消し残しの少ない適応 BF も提案されているが [114]，計算量が膨大であるという欠点がある．ICA は音源の独立性を仮定するだけで分離を行うことができる有効な手法であるが，実環境においてはしばしばこの仮定が成立しないことがある．また，各周波数での分離信号が同じ音源に対応するように分離信号を並べ変えなければならないという permutation 問題も生じる．BF と ICA の中間的な手法として，GSS が挙げられる．GSS では音源位置とマイク位置及び音源の相関に基づいて音源分離を行うが，実環境では位置の正確な抽出が難しく，分離性能に影響を与える．

5.2.3　音声特徴量

本節では，音声特徴量について示す．音声特徴量に求められる性質として，音素の特徴を表すこと，雑音に頑健であることの二つが挙げられる．以下，提案されている音声特徴量について示す．

MFCC (Mel Frequency Cepstrum Coefficient)

音声認識では，音響特徴量として MFCC が一般的に用いられている．これは，FFT によりスペクトルに変換した後，さらに DCT を施してケプストラムに変換し，スペクトル包絡を求めるものである．スペクトル包絡の構造が，音素を決定するのに重要な役割を果たすことから，ケプストラム領域の特徴量が用いられることが多い．雑音のない環境での音声認識では，ケプストラム領域の音声特徴量を用いることで高い認識性能が達成できることがこれまでの音声認識の研究の中で実証されている．

雑音に頑健な特徴量

MFCC は雑音が含まれない信号に対しては高い認識性能を達成することができる．しかし，雑音環境下では認識性能が大きく低下することがあり，その他の音声特徴量が提案されている．

聴覚の特性を勘案した特徴量として，PLP (Perceptual Linear Predictive analysis)[117] が提案されている．PLP では，人間の聴力範囲で音声分析をして全極モデルを得るための分析を行う．スペクトルのパワーに臨界帯域分析を行い，等ラウドネス強調をかけた後にケプストラム領域への変換を行う．PLP の他に RASTA (RelAtive SpecTrAl) があるが，これは RASTA フィルタをかけることで，人間の聴覚に近い特徴量を抽出する [118]．PLP と RASTA を組み合わせた，RASTA-PLP も提案されている．これらの音声特徴量は，MFCC と比べて雑音に頑健であると報告されている．

MFCC のようなケプストラム領域の音声特徴量は，音素の特徴をよく表すが，加法性雑音による影響を強く受けるという欠点を持つ．認識対象と異なる音源から発せられた雑音はマイクでの収録時に音声と加算されるため，加法性雑音と呼ばれる．この雑音をケプストラム領域に変換すると，特定の周波数帯域にのみ重畳した雑音であっても，全ての周波数帯域に拡がってしまう．特に，音声特徴量抽出後に信頼度に応じた音声認識を行う MFT では，雑音に埋もれた周波数帯域を抽出できたとしても，それが全ての音声特徴量に拡がってしまうと無意味となる．この問題を解決するため，ケプストラムを抽出する際に帯域分割することが提案されている [123]．この方法は，DCT をする際にスペクトル周波数帯域ごとに分割し，そのグループごとに DCT を行うことで，他のグループに雑音の影響が拡がるのを抑える．

また，ケプストラム領域の特徴量を用いることをせず，スペクトル領域の特徴量を用いることも提案されている．スペクトル領域の特徴量は MFCC と比べて加法性雑音の問題を扱いやすいという面において有効である．William らは複素ガウス分布を仮定した HMM を用いた音声認識を対象とし，スペクトル領域の特徴量を提案している [124]．また，MFCC と同様の正規化をスペクトル領域で行うことにより MFCC と同程度の認識性能を達成する対数スペクトル特徴量が提案されている [126].

動的特徴量

動的特徴量は特徴量の前後のフレームとの差をとることで特徴量とするものである．動的特長量はデルタ特徴量とも呼ばれる．半母音ではスペクトルの動きそのものに音素の音響的な特徴が現れることから，動的特徴を音声特徴量の中に含めることで認識性能の向上が見込まれる．また，定常的な雑音が重畳している場合，動的特徴量は雑音の影響を受けないため，雑音に頑健な特徴量としても期待されている．一般的な音声認識では，ケプストラム領域の特徴量に加え，デルタケプストラムとデルタパワーを用いて音声認識を行う．また，デルタケプストラムのデルタ特徴量であるデルタデルタケプストラムが用いられることもしばしばある．

5.2.4　音声特徴量の正規化

音声特徴量を抽出する際に正規化が行われることがある．これは，音声信号に重畳した雑音の影響を軽減し，認識性能を向上させる目的で行われる．

CMN (Cepstrum Mean Normalization) は，ケプストラムの時間方向の平均を差し引くことで，音響伝達特性などの影響を除去する正規化手法である [119]．入力信号に含まれる雑音には，加法性雑音の他，乗算性雑音がある．乗算性雑音とは，音源からマイクまでの音響伝達特性や，マイクの伝達特性などであり，音声信号に畳み込み成分として重畳される．この雑音は，スペクトル領域に変換することで，音声信号に乗算される雑音として表現でき，対数をとることで加算として表現可能である．スペクトルの対数をとり，DCT を施したケプストラム領域で時間方向の平均を差し引くことは，定常的な乗算性雑音を除去することに相当する．これは有効な手法として，MFCC の計算の際に利用されることが多い．同様のアプローチとして，ケプストラムの分散を差し引くことにより正規化を行う CVN (Cepstrum Variance Normalization) や，CMN と CVN を組み合わせた MVN (Mean and Variance Normalization) も提案されている [120].

その他に，ヒストグラムを用い，累積頻度分布が一致するように非線形な変換を施すことにより正規化を行う方法なども提案されている [121, 122]．また，認識時の音声特徴

ベクトルを学習時の音声特徴ベクトルに写像する正規化として，SNR に依存して写像したケプストラムを推定する SDCN (SNR Dependent Cepstrum Normalization) やベクトル量子化に基づくコードブックを用いてケプストラムを推定する CDCN (Codeword Dependent Cepstrum Normalization) なども提案されている．

5.2.5　有効な音声特徴量の利用

前処理や正規化により雑音を軽減する処理が提案されているが，これらの処理を施したとしても，雑音を全て取り除くことは不可能である．そこで，音声認識デコーダ内で有効な音声特徴量を効果的に利用するアプローチが検討されている．

マルチストリーム音声認識

マルチストリーム音声認識は，複数の特徴量を用いた音声認識手法である．一つの音声の情報から複数の特徴量 (例えば MFCC，RASTA など) を用いて音声認識を行うことで，相互に雑音に埋もれた情報を補完することが期待できる．また，音声のみならず，カメラで唇の画像をとり，音声，画像の情報によりマルチストリームとすることも行われている [125].

マルチストリーム音声認識の実現には，各音声特徴量をいかに統合するかという問題がある．一つの方法として，複数の音声特徴量を一つのベクトルの中に結合することが挙げられる．これはデコーダ内での尤度計算において，HMM の内部状態は全ての特徴量について同じように遷移することに相当する．その他に，音声特徴量ごとに HMM を生成し，音素ごと，単語ごとに尤度を統合する方法などが提案されている．しかし，この方法では何らかの情報を用いて音素境界や単語境界を抽出しなければならない．

マルチバンド音声認識

マルチバンド音声認識は，複数の周波数帯域情報を利用した音声認識である．いずれかの周波数帯域の情報が雑音に埋もれてしまったとしても，それ以外の情報を利用することで認識性能の向上が期待される．主にマルチバンド音声認識では，周波数帯域ごとに HMM を用意し，それを統合する試みがなされている [128, 129]．また，統合時に信頼度に応じて重み付けをすることで，雑音に頑健な音声認識とすることが可能である．しかし，マルチストリーム音声認識と同様に，何らかの情報を用いて音素境界や単語境界を抽出しなければならない．

MFT (Missing Feature Theory)

MFT は音声信号のうち雑音や歪みのない部分の情報のみを用いて音声認識を行うアプローチである [130]．信頼性の低い部分をマスクすることにより音声認識には用いない．MFT にはマスクするかしないかの二者択一とする binary mask ベースの MFT と，マスクを連続値とする continuous mask ベースの MFT がある．continuous mask ベースの MFT と，HMM を 1 つとするマルチバンド音声認識は，同じ手法ということになるため，本論文ではこれらを MFT として表記する．MFT はマスクが正確に推定できれば他の手法と比べて有効であり，自動生成の提案についての報告もなされているが [131, 116]，何らかの雑音に関する情報等がなければマスク生成は困難であるという問題を持つ．また，相関を用いて雑音に埋もれた部分を回復する手法も提案されている [133].

MDT (Missing Data Theory)

MFT では，1 つのフレームの中で，信頼度の低い音声特徴量を除外するのに対し，MDT では，信頼度の低い音声特徴量のフレームを除外する方法である．つまり，MFT は雑音の大きな周波数領域の音声特徴量をマスクするのに対し，MDT は雑音の大きな時刻における音声特徴量をマスクする．MFT と同様に，信頼度の低いフレームをマスクして音声認識する方法 [134, 135, 136] と，信頼度の低いフレームを周辺のフレームから予測する方法がある．

5.2.6　雑音に頑健な音響モデル

これまで紹介した手法は，雑音を除去することにより認識性能を向上させる試みである．しかしこれとは逆に，音響モデルを雑音環境に適応する試みもなされている．本節では，これらの手法について説明する．

マルチコンディション学習による音響モデル

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は，雑音に頑健な音響モデルを構築する一つの方法である．音響モデルの学習は通常，雑音の含まれないクリーン音声を用いて行う．しかし，混入する雑音があらかじめ分かっている場合には，その雑音を音声に重畳し，そのデータを用いて音響モデルを学習することで，認識性能を向上させることが可能である．例えば，Lippmann らの報告では，クリーン音声の他に，早口で話した音声，大きな声で話した音声，疑問形でのピッチの異なる音声を含めた音声データを学習させることで，これらの環境での認識性能の向上を確認している

[137]．同様に雑音を重畳したデータを含めて学習することで，雑音環境下において認識性能を向上させることが可能である．特にロボットの音声認識では，静止時にマイクに混入するモータ音，ファン音は一定であるため，これらの雑音を含めた音声データを用いて音響モデルを学習させることで，認識性能の向上が期待できる．この方法は，ロボットの音声認識ではよく用いられている手法である．

既に構築された音響モデルの適応

マルチコンディション学習による音響モデルの構築は，音響モデルを作成する段階から雑音を含めた音声データを用いるが，既に作成された音響モデルを，現在の認識環境に適応させる試みもなされている．その中でも MLLR(Maximum Likelihood Linear Regression)[138] は有効な手法として捉えられている．MLLR は，HMM の正規分布の平均値や分散値を，適応データに対して最尤となるように線形変換する手法である．いま，n 次元正規分布をもつ連続型 HMM のある状態における，各ガウス分布の平均値を μ_i，共分散行列を Σ で表す（ただし $1 < i < n$）．平均を適応化する場合，

$$\xi = [1, \mu_1, \mu_2, \cdots, \mu_n]^\top \tag{5.5}$$

に対し，次式中の変換行列 W を推定することになる．

$$\hat{\mu} = W\xi \tag{5.6}$$

分散を適応化する場合には，Σ をコレスキー分解して得られる行列 L に対して，次式における変換行列 H を最尤推定することになる．

$$\hat{\Sigma} = LHL^\top \tag{5.7}$$

また，MLLR 法を行うには，あらかじめ全ての HMM のガウス分布を，音響的に近いもの同士でクラスタリングしておく必要がある．そしてこの各クラスタごとに変換行列を求め，クラスタ内のガウス分布全てに同じ変換行列を適用する．このため，適応データが少ない場合でも有効に機能するという利点があり，適応化の際にはよく用いられる手法である．MLLR 法で適応化する際には，適応データに対する音素系列情報が必要である．この音素系列が既知のときの適応を「教師あり適応 (supervised adaptation)」といい，正しい音素系列が未知の場合の適応を「教師なし適応 (unsupervised adaptation)」という．実際のアプリケーションでは，場面に応じてどちらを用いるかが異なる．例えば，あらかじめ使用者が限られている場面などにおいては，使用前にアプリケーション側からの指示に従った発話を行い，そのデータを用いて教師あり適応を行うことが可能である．しかし，不特定多数の者が使用する場合には，アプリケーションとの対話の中で認識結果を利用して教師なし適応を行うことが考えられる．

音響モデルの適応法として，MLLR の他に MAP 推定 (Maximum A Posteriori Probability Estimation) が提案されている [139]．これは，学習データが多い場合などに有効であるが，学習データが少ないと適応ができない．そこで，MAP 推定を行う有効な方法として，MLLR と組み合わせる方法が提案されている [140].

5.2.7　ロボットの動作音への適応に有効な手法の考察

雑音に頑健な音声認識手法について関連研究を紹介したが，この中でも実用的に数多く採用されているものは，前処理に SS，音声特徴量に MFCC，正規化に CMN，音響モデルに MLLR を用いる方法である．また，ロボットの音声認識では，雑音が混入するため，マルチコンディション学習による音響モデルやマイクロホンアレーを用いた手法も数多く採用されている．そこで，これらを考慮し，ロボットの動作音への適応に有効な手法について考察する．

マルチコンディション学習による音響モデルや MLLR を用いた音声認識は最も有効な手法の一つである．これらの手法は，音響モデルを雑音へ適応するため，効果的に雑音を含む音声が学習された場合には強力である．しかし，雑音が大きい環境では，音声の特徴が雑音に埋もれてしまうため，効果的な学習が期待できない．また，定常的な雑音については効果的な学習が期待できるが，非定常な雑音に対しては難しい．このため，この手法には限界があると考えられる．

従来の音声認識では，音響モデルを雑音に適応するための研究が多く行われてきた．これは，入力信号から雑音を取り除くというアプローチをとると，音声の歪みが大きくなり，結果的に音響モデルの適応を行った方が性能が出やすいからであると考えられる．しかし，ロボットにおける音声認識では，従来の音声認識で想定していた雑音よりも大きな雑音環境 (SNR 0dB 以下である場合もある) での認識が必要となる．このような環境では，音響モデルを雑音に適応しても，音声の特徴を示す成分は雑音に埋もれてしまうため，有効な情報はほとんど残っておらず，音声認識を行うことは困難である．したがって，雑音を除去する仕組みが必要となる．

ロボットにおける音声認識では，その前処理として，BF, ICA, GSS などのマイクロホンアレーを用いた音源分離を行うことが多い．ロボットの音声認識で問題となる雑音には，動作音の他，環境雑音などがある．環境雑音は非定常であり，実環境では音源位置や音源数に関する仮定を置けないため，雑音の推定にはマイクロホンアレーを用いた手法が有効である．しかし，本研究で対象とする動作音はロボット自身が発するものであり，ロボットは自己の動作情報を取得可能なため，動作音の推定が可能である．よって，マイクロホンアレーのような手法を用いなくても単一のマイクで効率的に動作音への適応ができると考えられる．

動作音を対象とし，一個のマイクで雑音への適応を行うアプローチとして，伊藤ら

による SS を用いた手法があるが [105]，反響音のある環境や，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比べ，有効性があるのかどうかは言及されていない．一般に，SS は定常雑音に対しては有効であるが，非定常雑音に対しては歪みが大きくなり，性能が劣化するため，実環境で有効とは必ずしもいえない．

非定常雑音に対しても有効な手法として，MFT を用いた手法がある．MFT を用いた方法では，信頼性の推定を正確に行うことができれば，認識性能は他の雑音適応手法と比較して大きく向上する．信頼性の推定を正確に行うためには雑音の推定が必要であるが，音声と雑音の混合音である入力信号から雑音推定を行うこと自体が音声認識と同レベルの難しさを有するという問題がある．従来の音声認識では，この信頼性推定が非常に困難であるため，MFT が有効な手法として用いられることが少なかった．しかし，本研究で対象とするロボットの動作音はその雑音推定が容易であるため，MFT が有効に利用できると考えられる．

MFT を用いる際には，音声特徴量に MFCC を用いることはふさわしくない．ロボットの動作音は加法性雑音として重畳するが，スペクトル領域である周波数帯域にのみ重畳した雑音もケプストラム領域では全ての周波数帯域に拡がってしまうからである．そこで，CMN などの MFCC と同じ正規化処理を施したスペクトル特徴量を用いることが適切であると考えられる [126]．

ロボットの動作音への適応にあたり，SS により雑音を除去することで，SNR の低い雑音環境下においてもマルチコンディション学習による音響モデルや MLLR を用いた手法が有効に機能すると考えられる．これらの手法では，定常成分の雑音への適応は可能であるが，非定常成分の雑音への適応は不十分である．そこで，MFT を組み合わせることで非定常成分への適応も可能になると考えられる．その際，MFCC を用いることはできないので，スペクトル領域の特徴量の利用が必要となる．

5.3　ロボットの動作音に頑健な音声認識手法の提案

本節では，ロボットの動作音に頑健な音声認識手法の提案を行う．提案手法ではまず入力信号から雑音除去処理を行う．動作音の混入した環境では SNR が低いため，雑音除去処理が必要である．次に，雑音除去処理での雑音の引き残し成分を平坦化するため，白色雑音の重畳を行う．SNR が高い環境では雑音除去処理による音声信号の歪みは小さいと考えられるが，SNR が低い環境ではその歪みは大きく，雑音除去処理を行うことでかえって認識性能が劣化する場合も考えうる．雑音の除去により，モータ音などの定常的な雑音の多くは取り除くことができるが，動作による非定常成分の雑音への適応は不十分であると考えられる．これに適応するため，MFT を利用した音声認識を行う．MFT のマスク生成には推定動作雑音を用い，雑音の多く重畳した箇所は信頼性が低く，音声認識への関与を低くするようにする．図 5.3 に提案する雑音適応化手法のブロック図を示す．以下，それぞれの処理について示す．

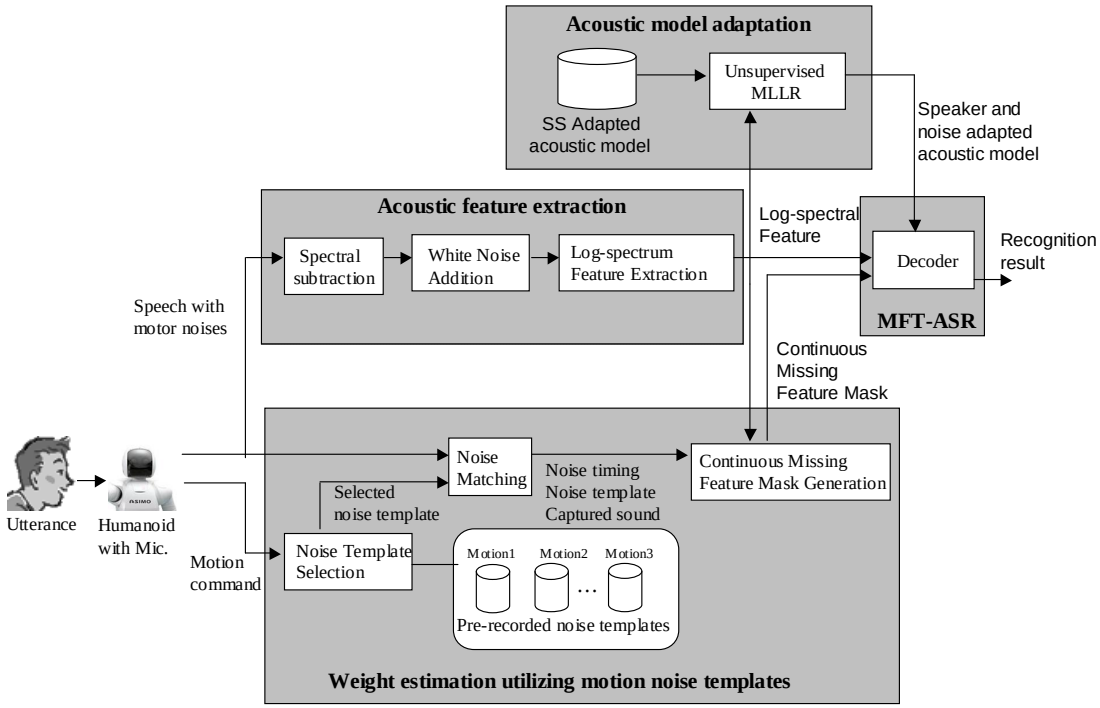


図 5.3: 提案する音声認識手法

5.3.1　雑音除去処理

入力信号の SNR は低い (0dB 以下である場合もある) ため，このような環境で音声認識に有効な音声特徴量を抽出することは難しい．そこで，入力信号の SNR を改善するため雑音除去処理を行う．雑音除去処理には式 (5.8) に示される SS を用いた．これは，Boll によって提案された SS を基調としているが，スペクトルが負となることを防止するため，フロアリング項 (式 (5.8)$\sqrt{\beta}|\bar{N}|$) を導入している．

$$|X(f)| = \max\{|X(f)| - \sqrt{\alpha}|\bar{N}|, \sqrt{\beta}|\bar{N}|\} \tag{5.8}$$

$X(f)$ は入力信号のスペクトルを示し，$\bar{N}$ は入力信号に重畳している雑音信号の平均スペクトルを示す．α, β は SS を行う際のパラメータであるが，本研究では一般的によく用いられている値 ($\alpha = 1, \beta = 0.1$) を用いた．

5.3.2　白色雑音重畳

雑音除去処理は SNR を向上させるが，同時にスペクトル歪みを生み出す．このスペクトル歪みが認識性能に悪影響を及ぼす．雑音除去手法に関わらず，背景雑音の状況によっては大きな歪みを生じることがあり，音声認識ではスペクトル歪みに対する処理が必要である．特に本研究の対象とするロボットの動作雑音では，雑音パワーが大きく，歪みが大きくなる傾向がある．そこで，本研究ではこのスペクトル歪みを軽減するため，雑音除去処理の後に薄く白色雑音を重畳させることとした．同様の方法は山出 [106] らの報告にも述べられており，定常雑音を加えることで，雑音の引き残し成分を平坦化し，認識性能を高めることが期待される．

白色雑音の重畳には，入力信号のある程度のレベルの白色雑音を加えることが歪みを抑制するのに役立つと考え以下のような式を用いた．

$$y'(t) = y(t) + \frac{2p}{T}\sum_{t=1}^{T}|y(t)| \cdot \mathrm{random(t)} \tag{5.9}$$

$y(t)$ は雑音除去処理後の信号であり，random(t) は -1 から 1 までの任意の実数値をランダムに返す関数である．本研究では $p = 0.1$ とした．すなわち，平均して入力信号の 1 割程度の大きさの白色雑音が加わることとなる．

5.3.3　音響モデルの雑音除去処理への適応

ロボットの音声認識では，学習に，定常雑音を含めた音声データを用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が有効である．ロボットは定常時で

もモータ音やファン音を発するため，この雑音を含めて学習することでクリーン音声データのみで音響モデルを学習する場合と比べ，認識性能が向上する．定常雑音が常に発せられているロボットでは，マルチコンディション学習による音響モデルは通常の音声認識で想定されているクリーン音響モデルと等しいとも考えられる．

しかし，マルチコンディション学習による音響モデルを用いると，雑音除去処理を行った際に認識性能が低下することがある．この原因として，雑音除去処理によりスペクトル構造が歪むことや，雑音除去処理によりロボットが定常的に発する雑音までもが取り除かれ，学習時の音声データと認識時の音声データに大きな差が生まれることなどが考えられる．

本研究ではこの問題を解決するため，雑音除去処理により雑音が除去された音声データを用いて音響モデルを学習することとした．これにより，雑音除去処理による認識性能の低下を防ぐことが期待できる．

5.3.4　対数スペクトル特徴量の抽出

白色雑音を重畳した後に音声特徴量を抽出する．音声特徴量には音声認識に一般的に用いられる MFCC ではなく，対数スペクトル特徴量 [126, 127] を用いた．動作音などの雑音は，スペクトル領域において加算される．しかし，従来用いられている MFCC はスペクトルをさらに DCT したケプストラム領域であるため，ある周波数帯域に加算された雑音は全ての特徴量に影響を与えてしまう．MFT を用いた音声認識では，雑音に埋もれた信頼性の低い周波数帯域を抽出することが必要であるため，ケプストラム領域の音声特徴量よりもスペクトル領域の音声特徴量の方が都合がよい．MFCC ではケプストラム領域に変換された後，C_0 項の除去，リフタリング，CMN [119] の 3 つの正規化処理が行われる．これらの正規化処理は音声認識性能を向上させる上で重要であることが知られているため，使用した対数スペクトル特徴量においても，対数スペクトル領域において同様の正規化処理を施している．

以下，正規化処理について概説する．初めに C_0 項の除去であるが，これは対数スペクトルの平均を引くことによって同等の処理を行う．フレーム f における i 次元目のスペクトルを S_{fi} とすると，C_0 正規化後の対数スペクトルは以下のように示される．

$$S^1_{fi} = S_{fi} - \frac{1}{N}\sum_{j=1}^{N} S_{tj} \tag{5.10}$$

次に，リフタリング処理ではスペクトル構造の山と谷が強調されるため，対数スペクトル特徴量の正規化では以下のリフターにかける．

$$H(z) = 1 - 0.9z^{-1} \tag{5.11}$$

$H(z)$ のインパルス応答を h_x とし，

$$S^2_{fx} = h_x * S^1_{fx} \tag{5.12}$$

として処理を行う．$*$ は畳み込み演算を表す．最後に CMN であるが，以下のように対数スペクトルの時間方向の平均を減算することで同じ処理が行われる．

$$S^3_{fi} = S^2_{fi} - \frac{1}{F}\sum_{j=1}^{F} S^2_{ji} \tag{5.13}$$

5.3.5　MFT マスクの生成

MFT マスクはフレームごと，周波数帯域ごと (音声特徴量の次元ごと) に生成される．自動的なマスクの生成は Raj らの報告 [131] や山本らの報告 [116] がある．しかし，完全に理想的なマスクを生成することは現実的には困難である．本研究では，ロボット自身の動作情報は動作前に取得できるため，これに基づいて動作音の推定を行う．動作音の推定については，あらかじめ収録したテンプレート雑音と現在入力されている動作音との時間的なマッチングにより行う．そして，入力された信号と推定された動作音に基づいてマスクの生成を行う．詳細については次に示す．

テンプレート雑音の選択

あらかじめ収録した動作音をテンプレート雑音としてデータベース化する．本研究では，34 種類の動作音を用意した．ロボットが動作を行う際には，データベースから動作種類に応じたテンプレート雑音を選択する．現在発せられている動作音はこのテンプレート雑音と同じであると仮定し，テンプレート雑音を用いた雑音推定を行う．

雑音マッチング

テンプレート雑音の選択が行われても，その雑音と現在発せられている雑音が時間的にはマッチしていない．そこで，時間的に雑音をマッチングさせる必要が生じる．マッチングは以下の方法により行われる．$T_d(f)$ をテンプレート雑音のスペクトル系列，$I_d(f)$ を入力信号のスペクトル系列とする．f はフレームとし，d は周波数軸方向のスペクトルの次元とする．D を 1 フレームの窓長 (サンプル数) とすると，$0 \leq d \leq \frac{D}{2}$ である．また，テンプレート雑音における各次元のスペクトルの最大値を M_d とする．

ここで，入力信号 $I_d(f)$ について，M_d を超えるものは音声信号が重畳しており，ミスマッチの要因となると考え，そのようなスペクトル系列の値を 0 とする．

$$I'_d(f) = \begin{cases} I_d(f) & if \quad I_d(f) \leq M_d, \\ 0 & otherwise. \end{cases} \tag{5.14}$$

マッチングは I'_d と T_d の相互相関をとることにより行った．最も相関が高いフレーム s_d は

$$s_d = \underset{\tau}{\mathrm{argmax}} \sum_{f=0}^{N-1} I'_d(f) T_d(f - \tau) \tag{5.15}$$

である．得られた s_d の $0 \leq d \leq \frac{D}{2}$ のうち，最も s_d の値の数が大きいものを s_{match} としてマッチングに用いる．

マッチング後の推定雑音 $E_d(f)$ は，

$$E_d(f) = T_d(f - s_{match}) \tag{5.16}$$

により得られる．

マスクの生成

まず，マッチングされたテンプレート雑音 $T(s_{match})$ は対数スペクトルに変換される．変換された対数スペクトルの雑音を $n(k, f)$ とする．k は次元 (周波数軸方向) を示し，f はフレーム (時間軸方向) を示す．同様に，入力された雑音を含む対数スペクトルを $y(k, f)$，雑音除去処理後，白色雑音を重畳した対数スペクトルを $p(k, f)$ とする．推定された音声信号は以下のように表される．

$$c'(k, f) = y(k, f) - n(k, f). \tag{5.17}$$

マスク $m(k, f)$ は以下のように計算される．

$$m(k, f) = \frac{|C'(k, f) - \mathrm{median}_k(C'(k, f))|}{P(k, f) - C'(k, f)} \tag{5.18}$$

$\mathrm{median}_k(a(k))$ は $a(k)$ の中央値を得る関数である．$P(k, f)$ および $C'(k, f)$ は対数スペクトル $p(k, f)$ および $c'(k, f)$ に正規化処理を施したものである．$m(k, f)$ が大きな値になることを防ぐため，閾値 t_{th} を設けた．したがって，$m(k, f)$ のとる範囲は 0 から t_{th} である．t_{th} は実験的に 5.0 とした．

さらに，MFT マスクの正規化を行う．この正規化は，MFT を用いた音声認識を行うことで挿入ペナルティなどの最適パラメータの変化を抑えるために行う．正規化後

の MFT マスクを $w(k,f)$ とし，1 フレームにおける $w(k,f)$ の合計が音声特徴量の次元数 K と同じになるように正規化を施す．

$$
\begin{aligned}
w(k,f) &= \frac{m'(k,f)}{\sum_{k=1}^{K} m'(k,f)} \\
m'(k,f) &= \begin{cases} m(k,f) & if \quad m(k,f) < t_{th}, \\ t_{th} & otherwise. \end{cases}
\end{aligned}
\tag{5.19}
$$

5.3.6　MFT に基づく尤度の計算方法

MFT は非定常な雑音に対しても効果がある．雑音除去処理や白色雑音の重畳によって SNR は改善されるが，MFT を用いることでさらに非定常な雑音成分に対しても効果があると期待できる．しかし，テンプレート雑音と実際に生じた雑音に大きな差がある場合には効果は薄い．

MFT では信頼性の高い特徴成分に対しては大きな重みを，信頼性の低い特徴成分に対しては小さな重みを用いて尤度の計算を行う．MFT を用いない従来の音声認識では，音素モデル q_l，音声特徴量 $\mathbf{s_f}$ の尤度は以下の式によって与えられる．

$$
L(\mathbf{s_f}|q_l) = \sum_{i=1}^{M} L(\mathbf{s_f}(i)|q_l). \tag{5.20}
$$

MFT を用いた尤度計算は，マスクを $\omega(k,f)$ として以下のように定義される．

$$
L(\mathbf{s_f}|q_l) = \sum_{i=1}^{M} \omega(i,f) L(\mathbf{s_f}(i)|q_l). \tag{5.21}
$$

5.4　評価実験

本節では 5.3 節で示した提案手法の有効性を示すため，従来手法と比較した実験結果について示す．

5.4.1　実験条件

Honda ASIMO を用いて評価実験を行った．ASIMO の左マイクを用いて音声の収録を行い，デコーダには大語彙連続音声認識エンジン Julius[142, 143] およびマルチバンド版 Julius[144] を用いて孤立単語認識による評価を行った．評価用データには ATR 音素バランス単語を用いた．音素バランス単語には男性 12 話者，女性 13 話者の合計 25 話者の音声データが含まれ，1 話者あたりの発話数は 216 である．各発話は “いきおい”, “いよいよ” などの単語発声である．

音響モデルの構築には男性 9 話者女性 10 話者の合計 19 話者の音声データ (学習セット A_1) を用いた．このデータは無響室において 100 cm の距離から収録を行い，音圧の変化にも柔軟に対応できるようにするため，SNR のレベルを変化させて (+5 dB, +10 dB, +15 dB) 学習を行った．

テスト用のデータは男性 3 話者女性 3 話者の合計 6 話者の音声データ (テストセット R_1) を用いた．このデータは音響モデルの学習とは異なる話者から構成されている．収録は 7 m (W) × 4 m (D) × 3 m (H) の部屋において行った．実用的な環境においても性能を発揮するか検証するため，家庭のリビングを想定した大きさの部屋で，反響音のある環境で収録を行った．話者とロボットのマイクの距離は 50 cm, 100 cm, 150 cm, 200 cm の 4 距離である．

ロボットの動作雑音については，34 種類の動作を用いて認識実験を行った．この動作音は ASIMO の電源を投入し，動作を全く行っていない定常雑音 1 種類と「バイバイ」や「お辞儀」などの上半身の動作を主とするジェスチャ雑音 25 種類および「直進」や「回転」など足を用いた動きを主とする歩行雑音 8 種類より構成される．テストセット R_1 に動作音を重畳したものをテストセット R_2 とする．

提案手法と従来の有効な手法であ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の比較を行うため，マルチコンディション学習用のデータを用意した．マルチコンディション学習は A_1 のデータに加え，ASIMO の電源を投入したときのモータ音やファン音などの定常雑音が重畳されたデータ A_2，動作雑音 (動作 $1 \leq N \leq 34$) が重畳されたデータ $A_{3(N)}$ を用いた．認識実験では，以下の 4 つの音響モデルを用意した．

AM-1 学習セット A_1 を用いたモデル (クリーンモデル)

AM-2 学習セット A_1 と A_2 を用いたモデル (マルチコンディション学習モデル 1)

AM-3 学習セット A_1 と $A_{3(N)}$ を用いたモデル （マルチコンディション学習モデル 2）

AM-4 学習セット A_1 と A_2 に雑音除去処理を施した A_4 を用いたモデル

AM-5 学習セット A_1 と A_4 に白色雑音を重畳した $A_{5(p)}$ を用いたモデル

音響モデル [AM-3] は雑音環境ごとに作成しているため，全部で 34 種類のモデルが存在する．さらに，音響モデル [AM-5] は重畳する白色雑音の大きさを変化させているため，式 (5.9) に示す $p = \{0.05, 0.1, 0.2, 0.4\}$ とした 4 種類のモデルが存在する．

評価実験において比較した方法を表 5.1 にまとめる．ベースラインとして 3 つの音響モデルを用いた認識実験を行った．**A** では音声認識において一般的に用いられるクリーンモデルを用いた．**B** および **C** では雑音に頑健な手法としてマルチコンディション学習モデルを用いた．**B** では定常雑音のみを用いて音響モデルを学習しているのに対し，**C** ではロボットの動作音，すなわち非定常雑音も用いて音響モデルの学習を行った．

実験は，以下の 4 つより構成される．

白色雑音重畳の効果検証

雑音除去処理と白色雑音重畳を用いることで有効な SS の効果が現れることの検証を行う．**D** は雑音除去処理を施し，マルチコンディション学習による音響モデル (マルチコンディション学習モデル 1) を用いて音声認識を行った．これは，マルチコンディション学習による音響モデルに単に SS を組み合わせたものである．**E** は雑音除去処理を行った後の音声データを用いて音響モデルを学習し，音声認識を行った．**F** から **I** は雑音除去処理の後，白色雑音を重畳した音声データを用いて音響モデルの学習を行い，音声認識を行った．これらは式 (5.9) に示す p の値を変化させている．

MFT の効果検証

次に，提案手法の MFT を用いることで認識性能が向上することの検証を行った．雑音除去処理の後，白色雑音を重畳した音声データを用いて音響モデルを学習する手法 **G** と，**G** で MFT による音声認識を用いる **J** から **L** の手法を比較する．白色雑音の重畳は $p = 0.1$ を用いた．p の最適値は，距離や動作によって異なるため，中間的な値として $p = 0.1$ を用いることとした．

MFT を用いた音声認識では次の 3 つの条件のもとにマスクの計算を行った．**J** で用いている条件は実環境を想定した条件であり，雑音マッチングの際に，雑音と音声が混入した入力信号とテンプレート雑音とのマッチングを行う．テンプレート雑音はあら

かじめ収録した雑音であるため，入力信号の雑音と同じ種類の動作音であるが，同一ではない．テンプレート雑音と入力信号の雑音は，マッチングした時間より 0 ms から 200 ms の間でランダムにずらして重畳させてある．**K** は **J** よりも理想的で雑音マッチングのしやすい条件である．この条件では雑音区間が完全に抽出できたことを想定し，雑音区間のみでテンプレート雑音と入力信号の雑音のマッチングを行う．この条件でも，入力信号の雑音とテンプレート雑音は同じ種類の動作音であるが，同一ではない．**L** は最も理想的な条件であり，実環境では想定できない条件である．この条件では，雑音が完全に既知としてマスク計算を行った．雑音が完全に既知である場合に，どの程度の認識性能となるか参考のため実験を行った．したがって，**J** および **K** は推定雑音は入力信号の雑音と同一ではないのに対し，**L** は入力信号の雑音が推定されている．

SS を用いた動作音への適応手法と比較した有効性の検証

本手法と同様に，ロボット動作音への適応を目的として提案された伊藤らの手法との比較を行った．伊藤らの手法は，関節位置および角度を入力としたニューラルネットを用いて雑音推定を行うが，ここでは 1 フレームまたは 10 フレーム平均の雑音が既知であるとし，そのフレームごとに SS 処理を行うこととした．**M** および **N** では 1 フレームごとの雑音が既知とし，**O** および **P** では 10 フレームごとの雑音平均が既知であるとして SS を行っている．**O** および **P** では 1 フレームごとの雑音が既知であるが，SS の際にフロアリングが働くため，SS 処理後のスペクトルはクリーンな音声信号のスペクトルと同一とはならない．**M** および **O** は音響モデルにクリーンモデル [AM-1] を用い，**N** および **P** は音響モデルにマルチコンディション学習モデル 1[AM-2] を用いた．

MLLR の効果検証

雑音に頑健な手法として一般的に用いられている MLLR と本提案手法との組合せについて実験を行った．本研究では，正解ラベルファイルを既知とする教師あり適応と，正解ラベルファイルを既知としない教師なし適応両方の効果について実験を行った．

C’ および **J**’ は **C** および **J** に教師なし MLLR を行ったものである．また，**C**” および **J**” は **C** および **J** に教師あり MLLR を行ったものである．この実験により，従来から有効な手法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と提案手法の，MLLR との組合せについての比較を行う．

表 5.1: 比較した手法一覧

Condition	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Multi-condition (Stationary)		✓					
Multi-condition (Motion)			✓				
SS (each 1 utterance)				✓	✓	✓	✓
SS (for motion noise)							
Adaptation for SS					✓	✓	✓
White Noise Addition						$p = 0.05$	$p = 0.1$
MFT (voice+noise matching)							
MFT (only noise matching)							
MFT (known noise)							
Unsupervised MLLR							
Supervised MLLR							
Acoustic Model	AM-1	AM-2	AM-3	AM-2	AM-4	AM-5	AM-5

Condition	**H**	**I**	**J**	**K**	**L**	**M**	**N**
Multi-condition (Stationary)							✓
Multi-condition (Motion)							
SS (each 1 utterance)	✓	✓	✓	✓			
SS (for motion noise)						1 frame	1 frame
Adaptation for SS	✓	✓	✓	✓	✓		
White Noise Addition	$p = 0.2$	$p = 0.4$	$p = 0.1$	$p = 0.1$	$p = 0.1$		
MFT (voice+noise matching)			✓				
MFT (only noise matching)				✓			
MFT (known noise)					✓		
Unsupervised MLLR							
Supervised MLLR							
Acoustic Model	AM-5	AM-5	AM-5	AM-5	AM-5	AM-1	AM-2

Condition	**O**	**P**	**C'**	**J'**	**C"**	**J"**
Multi-condition (Stationary)		✓	✓		✓	
Multi-condition (Motion)						
SS (each 1 utterance)				✓		✓
SS (for motion noise)	10 frame	10 frame				
Adaptation for SS				✓		✓
White Noise Addition				$p = 0.1$		$p = 0.1$
MFT (voice+noise matching)				✓		✓
MFT (only noise matching)						
MFT (known noise)						
Unsupervised MLLR			✓	✓		
Supervised MLLR					✓	✓
Acoustic Model	AM-1	AM-2	AM-3	AM-5	AM-3	AM-5

5.4.2　実験結果

白色雑音重畳の効果

図 5.4a)-d) に実験結果を示す．これらの図では，34 種類の動作音について，同じ種類の雑音ごとに結果を平均している．

D は SS を施し，マルチコンディション学習による音響モデルを用いて音声認識した結果を示す．これは，SS とマルチコンディション学習による音響モデルの単なる組み合わせであるが，認識性能が低い．マルチコンディション学習では，雑音を含んだ音声を用いて音響モデルを学習しているにも関わらず，SS により雑音を除去してしまうことで，高い認識性能が実現できないものと考えられる．

E は SS を施した後の音声データを用いて，音響モデルの学習を行っている．**E** は **D** と比べて認識性能が高い．SS をマルチコンディション学習による音響モデルと組み合わせると，高い認識性能を達成できないが，雑音除去が行われた後の音声データを用いて音響モデルの学習を行うことで，雑音除去処理の効果が表れることが確認できる．

F から **I** は SS により生じた歪みを軽減するため，白色雑音の重畳を行っている．音響モデルは，雑音除去処理後，白色雑音の重畳を行った音声データを用いて学習を行い，認識時にも同様の処理を入力信号に施す．**E** と **F** から **I** を比較すると，**F** から **I** の中に最も性能のよいものの多くが入っている．これより，白色雑音を重畳することにより認識性能が向上することが確認できる．また，**F** から **I** の中では認識性能が最もよい条件を一意に定めることはできない．認識性能を最大とする白色雑音の重畳の大きさ (式 (5.9) の p の値) は雑音環境によって異なるが，白色雑音を重畳することで認識性能を高めることができることを確認できる．

MFT の効果

図 5.5a)-d) に実験結果を示す．これらの図では，34 種類の動作音について，同じ種類の雑音ごとに結果を平均している．ここでは，ベースラインとして求めた音声認識結果に加えて，提案手法の認識結果を示している．**B** および **C** はマルチコンディション学習による音響モデルを用いた音声認識結果を示す．クリーンモデルと比べてマルチコンディション学習による音響モデルは有効性が大きいことが確認できる．**B** および **C** は環境によってどちらが有効であるか異なるが，全体的に見て性能がよい **C** を従来手法として提案手法との比較を行う．

MFT を用いた **J** から **L** の中で，最も実用的な手法は **J** である．そこで，**C** と比べた **J** の有意性の確認を行った．これには危険率 p 値 [141] を用いた．図 5.5a) に示す 50 cm の環境では 34 雑音環境中 28 雑音で 5% 水準で有意な結果となった．同様に，図 5.5b)

に示す 100 cm では 32 雑音，図 5.5c) に示す 150 cm では 31 雑音，図 5.5d) に示す 200 cm では 33 雑音で 5%水準で有意な結果となった．

白色雑音の重畳には，$p = 0.1$ を用いた．実験結果より，どの雑音環境，距離についても提案手法の方が従来手法であ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ものよりも高い性能を示している．これにより，提案手法の有効性が確認できる．

また，MFT を用いることで認識性能は向上しているため，MFT はロボットの動作音へ適応させるのに有効な手法であるといえる．

なお，参考までに動作雑音のない音声を入力とし，クリーン音響モデルを用いて認識した結果は，50 cm では 80.56%, 100 cm では 80.94%, 150 cm では 78.63%, 200 cm では 80.94%であった．

SS を用いた動作音への適応手法と比較した有効性の検証

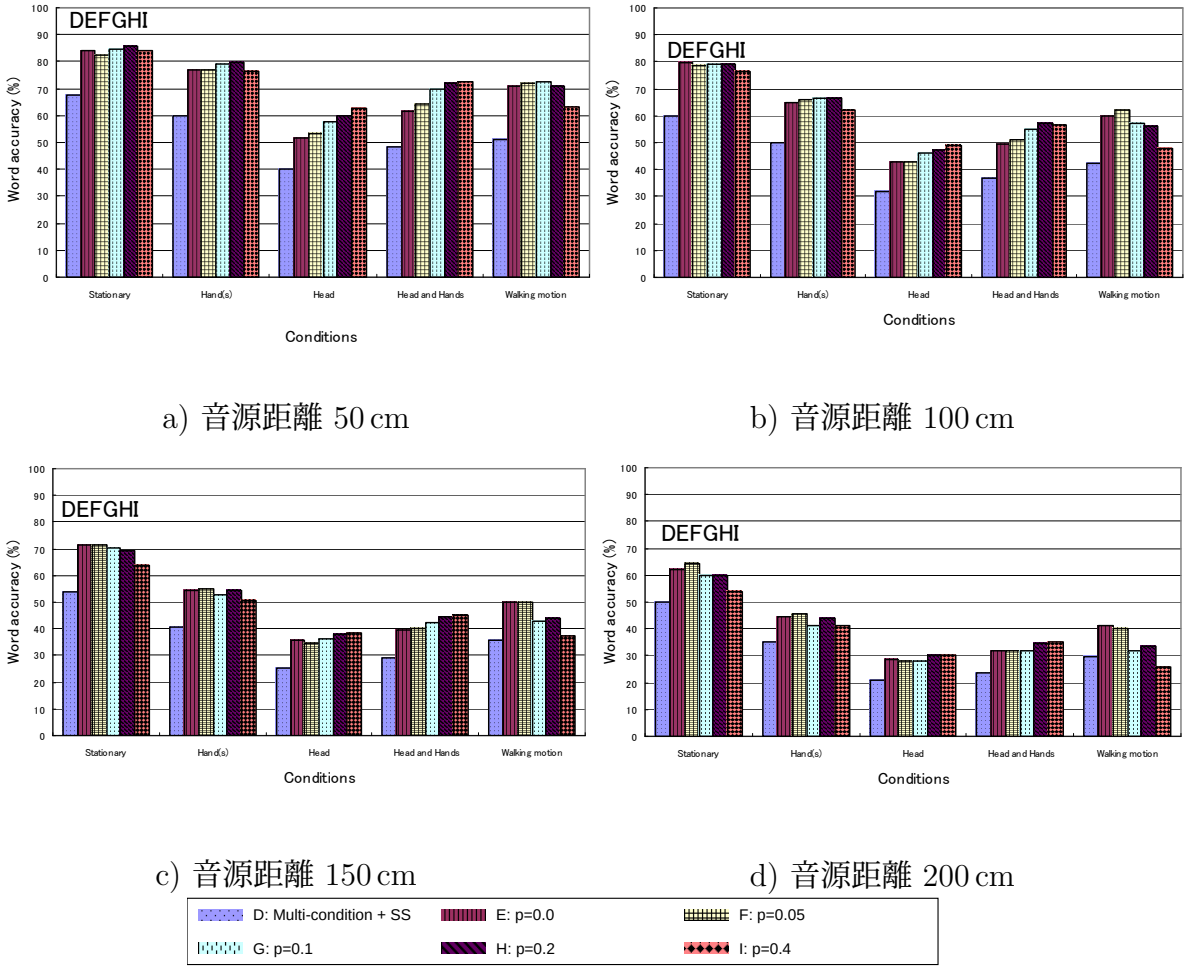
図 5.6a)-d) に実験結果を示す．これらの図では，34 種類の動作音について，同じ種類の雑音ごとに結果を平均している．SS を用いた **M** から **P** を比較すると，マルチコンディション学習による音響モデルを用いた **O** および **P** の方が高い認識性能を達成している．しかし，SS を用いない図 5.5a)-d) の **B** と比べると認識性能が低下している．また，クリーンモデルを用いて SS を施した **M** および **N** は SS を施さない図 5.5a)-d) の **A** と比べると認識性能は向上しているものの，他の手法と比較すると認識性能は低い．提案手法 **J** は，SS を用いて動作音への適応を行った **M** から **P** と比較して高い認識性能を達成していることが確認できた．

MLLR の効果

図 5.7a)-d) に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，提案手法のそれぞれに教師なし MLLR を組み合わせた際の実験結果を，図 5.8a)-d) に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，提案手法のそれぞれに教師あり MLLR を組み合わせた際の実験結果を示す．これらの図においては，34 種類の動作音のうち，動作の種類ごとに平均して結果を示している．

教師なし適応についてみると，50 cm の定常雑音では，提案手法 **J'** は従来手法 **C'** よりも認識性能が低いが，その他の環境では提案手法が有効であり，特に 200 cm の距離において提案手法が有効性が大きいことが確認できる．教師あり適応についてみると，150 cm, 200 cm の定常雑音，手の動きのあるジェスチャ，頭の動きのあるジェスチャについては提案手法 **J"** の方が認識性能が高いが，その他の環境では従来手法 **C"** の方が有効であることが分かった．また，MLLR を用いない **J** よりも MLLR を用いる **J'** および **J"** の方が認識性能が高いことが確認できた．

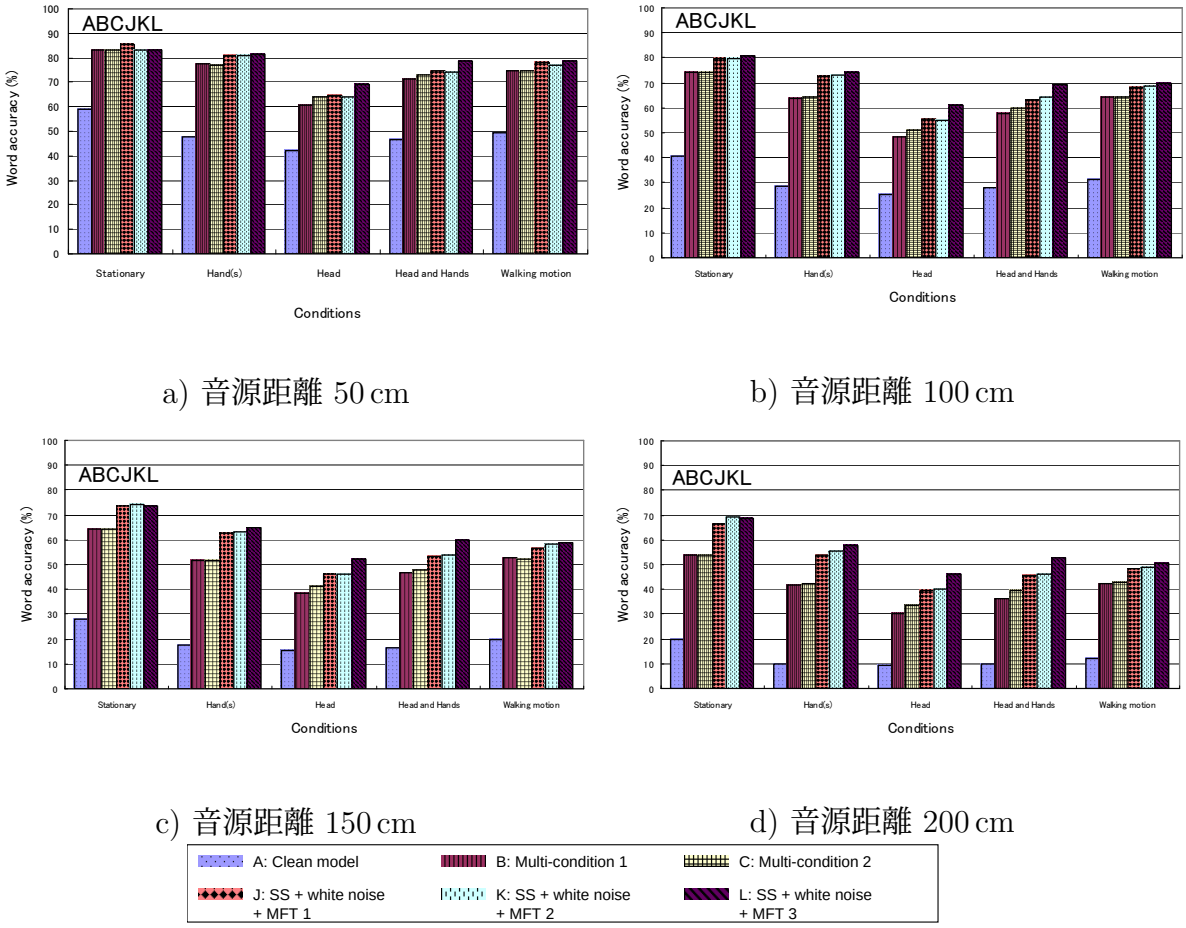
実験結果より，提案手法は有効な音響モデルの適応化手法とされる MLLR との組み合せによって認識性能は向上するものの，従来手法よりも有効なのは教師なし MLLR のみであり，教師あり MLLR では環境によって提案手法が劣る場合があることが確認できた.



a) 音源距離 50 cm　　b) 音源距離 100 cm

c) 音源距離 150 cm　　d) 音源距離 200 cm

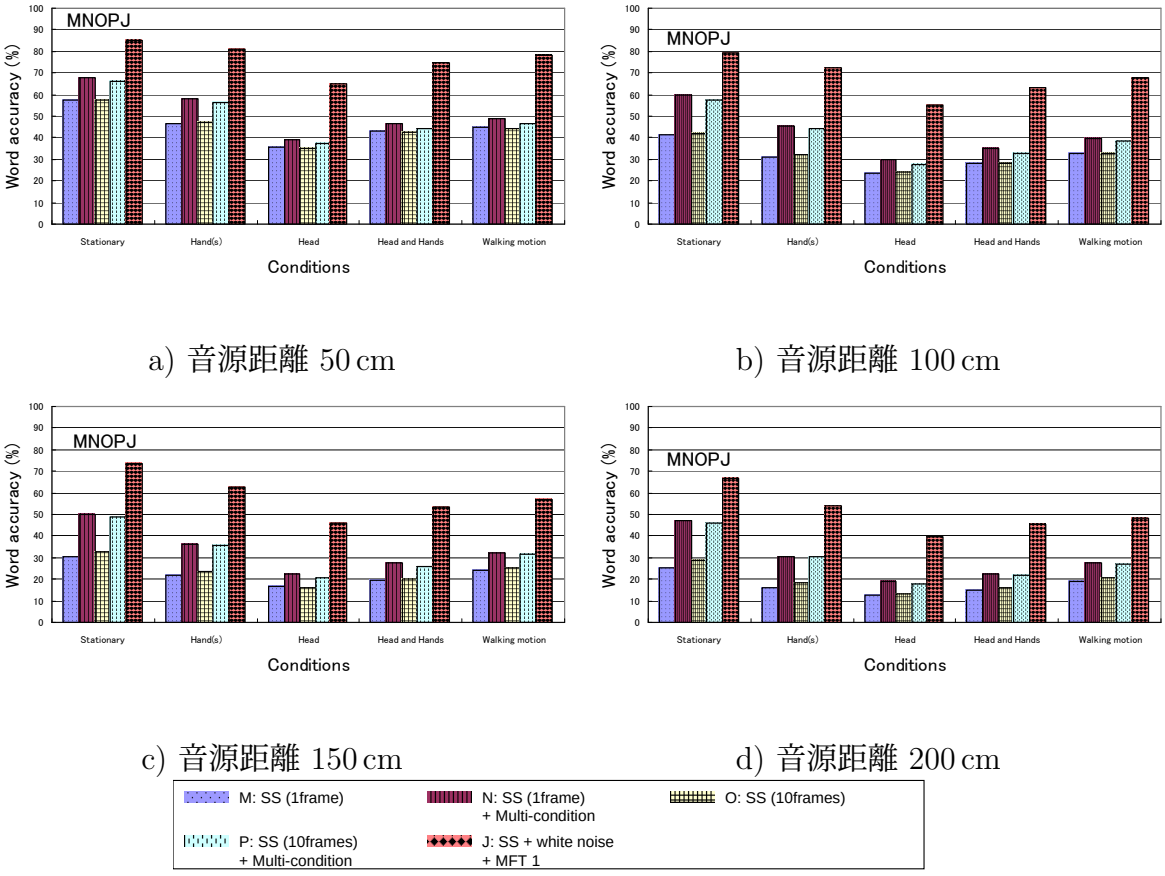
図 5.4: 白色雑音重畳の結果



a) 音源距離 50 cm　　b) 音源距離 100 cm

c) 音源距離 150 cm　　d) 音源距離 200 cm

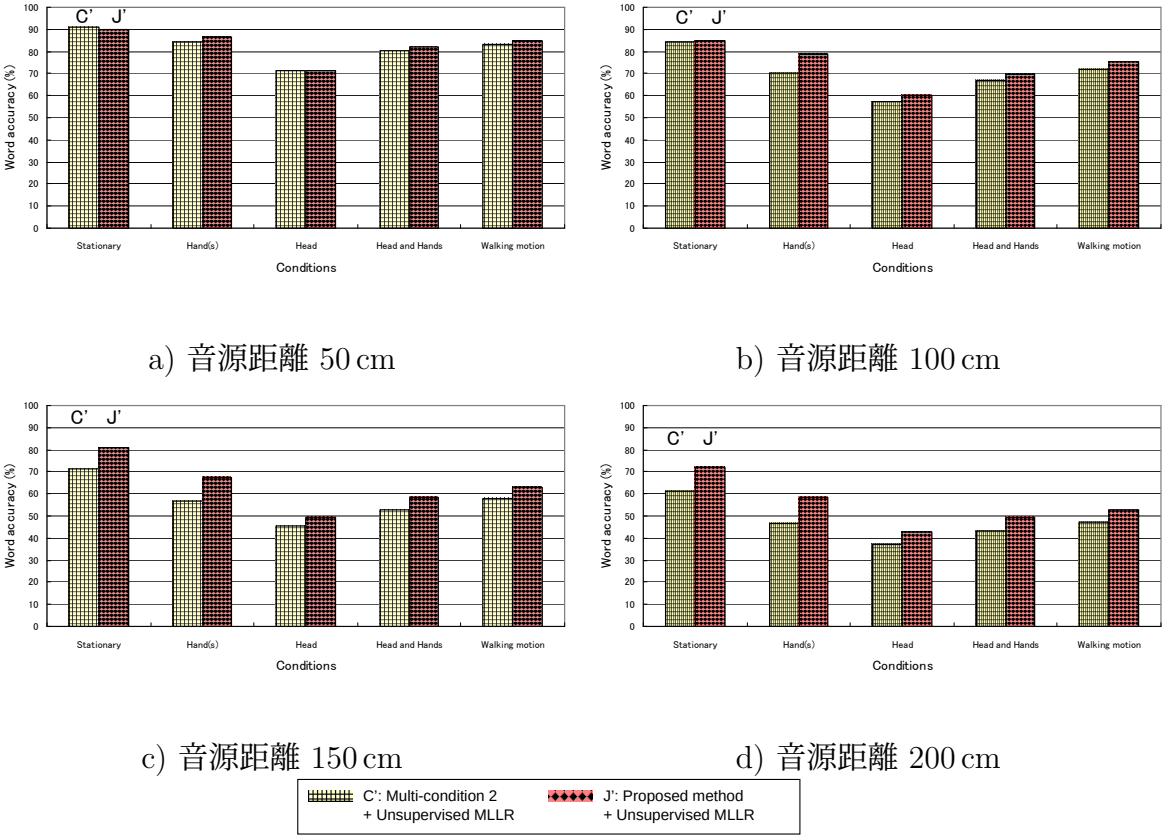
図 5.5: MFT を用いた結果



a) 音源距離 50 cm　　b) 音源距離 100 cm

c) 音源距離 150 cm　　d) 音源距離 200 cm

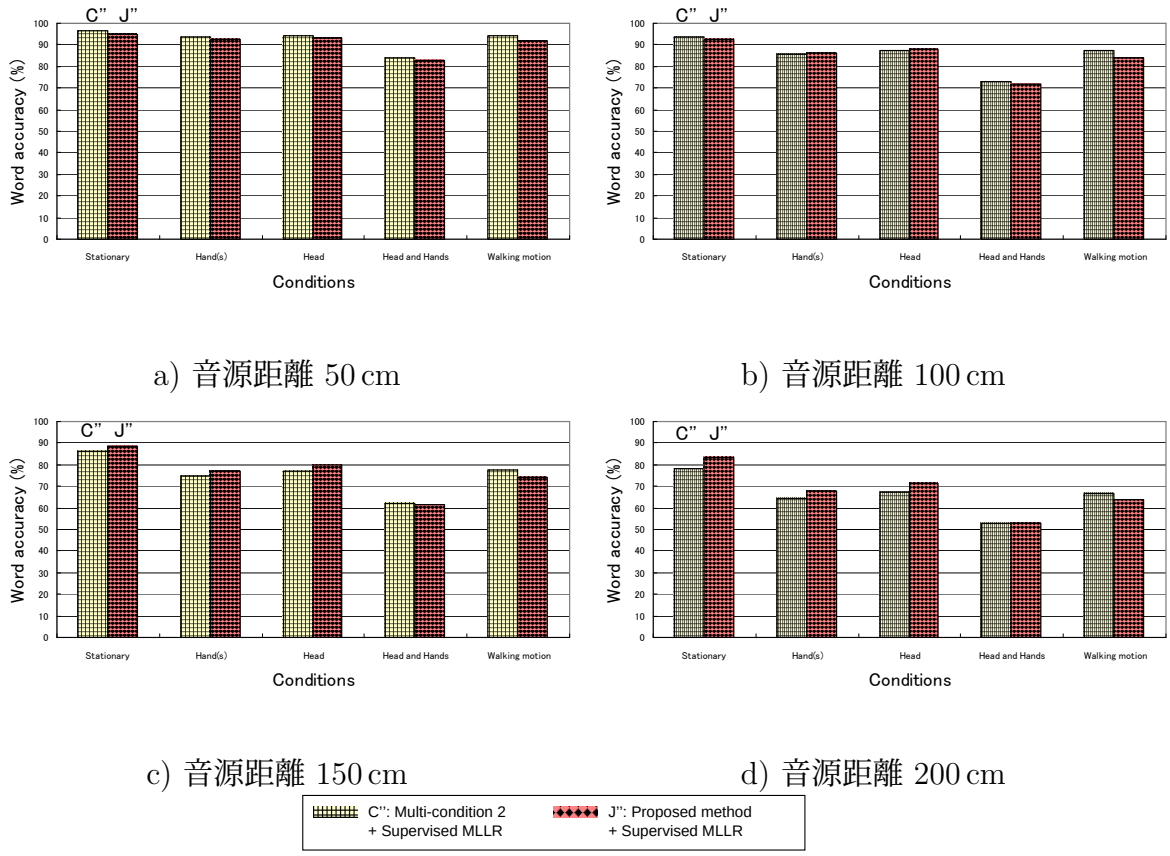
図 5.6: SS により動作音へ適応する方法との比較



a) 音源距離 50 cm　　b) 音源距離 100 cm

c) 音源距離 150 cm　　d) 音源距離 200 cm

図 5.7: 教師なし MLLR の結果



a) 音源距離 50 cm　　b) 音源距離 100 cm

c) 音源距離 150 cm　　d) 音源距離 200 cm

図 5.8: 教師あり MLLR の結果

5.4.3　考察

雑音除去処理と白色雑音重畳の効果

SS は雑音除去を行う上で有効な処理として捉えられているが，マルチコンディション学習による音響モデルを用いた音声認識では，SS を行うことで学習時の音声データと認識時の音声データの差を拡げ，認識性能の低下が起こることがある．本研究の実験では，**D** の認識性能が低く，本来有効であるはずの SS が，マルチコンディション学習による音響モデルとの組合せでは逆に認識性能を低下させることが実験結果より明らかとなった．

本研究では，マルチコンディション学習による音響モデルのような，雑音に頑健な音響モデルを用いつつ SS を有効に活用するため，SS を行った後の音声データを用いて音響モデルの学習を行った．この手法を用いた **E** では，**D** と比べて認識性能が改善されている．どの環境においても 10%から 20%程度の改善が達成できており，SS を行った後の音声データを用いて音響モデルを学習することで，雑音に頑健な音響モデルを用いつつ SS の効果が表れることが確認できる．

また，本研究では SS 処理後に白色雑音を重畳することにより SS による歪みを抑え，認識性能の向上を図った．**F** から **I** は白色雑音の重畳の大きさを変化させた結果を示している．白色雑音を重畳しない **E** と比べると，白色雑音を重畳したものの方が高い認識性能を示しているものがほとんどであり，白色雑音の重畳はスペクトル歪みを軽減し，認識性能の向上につながることが確認できる．しかし，全ての環境で最適となる白色雑音の重畳量 (式 (5.9) の p の値) を一意に見出すことはできなかった．動作の種類ごとに見ると，頭の動作を含む雑音については白色雑音の重畳を大きくした方がよいことが分かる．これは，頭の動作は短い時間のものが多く，他の動作と比べるとマイクの近くで行われるため，雑音も大きい．ところが，SS は平均雑音を用いて雑音除去を行うため，時間の短い雑音は平均すると小さくなる．よって，本来は SS に用いられる平均雑音よりも大きな雑音が重畳しているにも関わらず除去しきれない引き残し成分が大きく表れることが考えられる．頭の動作では，白色雑音の重畳を大きくすることでこの引き残し成分を平坦化し，認識性能向上につながったと考えられる．

その他の雑音では最適な白色雑音の重畳量を見出せないが，傾向として，距離が離れているものほど白色雑音の重畳を小さくした方が効果がある．一見，距離が離れているものは SNR が低いため歪みが大きく発生し，白色雑音の重畳を大きくした方が効果的とも考えられる．しかし，距離が離れた環境では入力信号と比較して雑音信号が大きく，SS のフロアリングがよく効く．このフロアリングにより歪みの発生が軽減され，白色雑音を大きく重畳しなくとも高い性能が達成できたと考えられる．白色雑音の重畳量を決定するに際し，スペクトルの大きさだけでなく，フロアリングや雑音の持続時間も考慮に入れることでより高い認識性能を達成できると考えられる．

MFT の効果

雑音除去処理および白色雑音重畳の後，MFT を用いた音声認識を行う提案手法 **J** は従来手法 **C** と比べてほぼ全ての環境で高い性能を示し，有効であることがわかる．また，MFT を用いない **G** と比べて MFT を用いる **J** はほぼ全ての環境で高い性能を示しており，MFT を用いることの効果が確認できる．

J は音声と雑音が重畳した入力信号を用いてテンプレート雑音との雑音マッチングを行い，推定雑音を求めるが，**K** は雑音信号が検出できたと仮定し，雑音信号のみを用いてテンプレート雑音とのマッチングを行う．また **L** は雑音を既知とした条件である．**K** および **L** は **J** と比べると理想的な環境であるため，認識性能も向上している．しかし，**J** も **K** および **L** と比較して同じような性能を示しており，提案手法の雑音マッチングは音声と雑音が重畳していても効果的にできることを確認できる．50 cm の環境では，**L** の方が **J** よりも性能が低くなっているものも見られる．条件 **L** は雑音が既知であるが，この条件での MFT マスクが正しい音声認識結果を得るのに最適なマスクということはできない．なぜならば，本研究で用いるマスク生成手法は，音声認識において重要と考えられるスペクトルの山と谷の重みを大きくし，さらに，雑音の小さな箇所の重みが大きくなるようマスク生成を行う．しかし，音響モデルはクリーンな音声のみで学習されたモデルではないため，このマスク生成手法が全ての入力信号に対して最もよいマスクを生成するとは限らない．したがって，雑音が既知であっても他の条件と比較して最も高い認識性能とはならない環境が現れたと考えられる．しかし，全体的に見て MFT を用いることで認識性能は向上しており，提案手法のマスク生成手法は有効な手法であると捉えることができる．

提案手法では，マルチコンディション学習による音響モデルは定常的な雑音に対して効果が高いと考え，**B** をベースとした音響モデルを用いている．すなわち，あらかじめロボットの発する定常雑音を収録しておき，この雑音と音声との重畳を行う．得られた雑音を含む音声に SS を施し，白色雑音を重畳した後にその音声データを用いて音響モデルを学習する．しかし，図 5.5c)-d) の結果を見ると，**B** よりも **C** の方が認識性能が高い場面が多く見られる．提案手法においても **C** をベースとした音響モデル，すなわち，ロボットの発する定常雑音のみならず動作音も含む音声データを用いて音響モデルの学習を行うことで，認識性能のさらなる向上が期待できると考えられる．

SS を用いた動作音への適応手法と比較した提案手法の有効性

動作音への適応手法として提案されている SS を用いた手法 (**M** から **P**) と提案手法 **J** を比較すると，提案手法の方が有効であることが確認できる．

SS を用いた手法は，フレームごとに一定の推定雑音を除去することで定常雑音を除去し，大きな効果が現れると考えられる．しかし，数フレーム，場合によっては各フ

レームごとに雑音推定を行い，除去するスペクトルを変化させると，雑音除去後のスペクトルはSNRが向上したとしても大きく歪んでいる場合があり，音声認識デコーダに入力した際に認識性能が向上しない場合が存在する．特に，大きなロボットを用いた場合，ロボット自身が発生する動作音が大きく，SSの処理時に発生する歪みも大きくなる傾向がある．また，SSはクリーン音響モデルを用いた音声認識では有効であるが，マルチコンディション学習による音響モデルを用いた音声認識では認識性能が低下する場合が存在する．このような理由から，SSを用いた動作音への適応手法は提案手法と比べて認識性能が低くなったものと考えられる．

MLLR に対する提案手法の有効性

教師なしMLLRと組み合わせた場合において，従来手法**C'**と比べて提案手法**J'**の方が高い性能を示すことが確認できる．MLLRは有効な音響モデルの適応化手法として捉えられており，多くの環境で性能が向上する．本提案手法は，MLLRとの組み合せが可能な雑音適応化手法である．

マルチコンディション学習による音響モデルとMLLRを併用した手法は実用的にも数多く用いられている．本提案手法においても，MLLRとの組み合わせた条件で効果があることで，従来手法と比べたメリットが一層大きくなると考えられる．

ロボットによるプレゼンテーションというタスクでは，不特定話者からの質問が想定される．このような場合に教師なし適応と提案手法を組み合わせることで，話者との対話の蓄積の中で音響モデルがオンライン適応され，高い認識性能の達成が可能となる．また，同様の場面はプレゼンテーションのみならず，案内ロボットにも考えられ，多くの場面において本提案手法を用いることが可能であるといえる．

しかし，教師ありMLLRとの組み合わせにおいては，従来手法**C"**の方が提案手法**J"**よりも認識性能が高い環境が数多く存在した．全体的に見ると，距離が離れており，SNRが悪い環境ほど提案手法**J"**が有効な場面が多くなるが，50 cmや100 cmの距離では従来手法**C"**の方が有効である．教師ありMLLRを用いる実用的な場面として，あらかじめ発話者が分かっている場合や，システムを利用する前にいくつかの発話をしてもらう場合が挙げられる．プレゼンテーションタスクでは，質問者が不特定である場合の方が多く想定され，教師ありMLLRを用いる場面は少ないと考えられるが，教師ありMLLRを用いる場面においては，距離や動作に応じて従来手法との切り替えを行うことでどのような環境においても有効な音声認識システムを実現することが可能である．

5.5　リアルタイムでの実装

本節では，提案する動作音に頑健な音声認識手法をプレゼンテーションシステムへ組み込むことを前提とし，リアルタイムでの実装について示す．リアルタイムでの処理を行うため，提案する音声認識手法を実現する際に必要となる計算時間について考察した後，システムへの実装について示す．

リアルタイム処理に関する考察

本提案手法をプレゼンテーションシステムへ実装する際に，リアルタイムで処理できることが要求される．提案手法は，音声特徴量および MFT マスクの計算の前に，動作雑音推定のための雑音マッチングが必要である．また，通常の音声認識では必要のない MFT マスクを計算する処理も必要である．図 5.9 に提案手法で必要な処理について示す．灰色で示すブロックおよび破線で示す矢印が従来の音声認識で必要な処理のに加えて提案手法で必要な処理を示す．

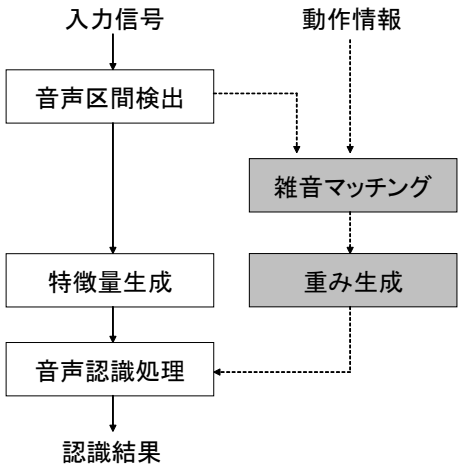


図 5.9: 提案手法の処理の流れ

実験で用いた大語彙連続音声認識エンジン Julius はリアルタイムでの音声認識を実現する．したがって，図 5.9 に示す白色で示すブロックおよび実線で示す矢印はリアルタイム処理が可能である．

提案手法で必要となるブロックの一つである重みの計算は，音声特徴量の抽出とほぼ同じ計算量で可能であり，音声特徴量の抽出は音声認識処理と比較すると微々たる計算量である．実際，音声特徴量抽出の関数のみを取り出し，Linux の time コマンドで測定したところ，音声特徴量の抽出に必要な音声ファイル 1 秒あたりの処理時間は，0.032

秒であった．これは，実際にかかった時間 (time コマンドで表示される real の時間) である．なお，計測に用いた計算機は一般的な PC であり，CPU は Pentimu4(2.4GHz), メモリは 738MB 搭載し，OS は Vine Linux 3.1 である．

提案手法で必要となるもう一つのブロックである雑音マッチングについても同様に処理時間の計算を行った．5.4 節における実験では，テンプレート雑音と実際に入力されている雑音に 200ms のズレが生じることを想定した実験を行った．そこで，ズレは 200ms 以内とし，動作コマンドを音声認識エンジンが受け取ってから雑音をバッファにためる．そして，音声認識開始時点において雑音バッファにたまった信号とテンプレート雑音とのマッチングをすることで，推定雑音を決定することとした．すなわち，動作の開始直後に音声入力があった場合，雑音マッチングに利用されるデータは 1 秒にも満たないが，動作が終わる直前に音声入力があった場合，雑音マッチングに利用されるデータはジェスチャの動作時間とほぼ同じになる．一つのジェスチャは約 5 秒程度であるため，0.5 秒から 5 秒まで 0.5 秒おきに処理時間を計測した．図 5.10 に結果を示す．横軸にマッチングに用いた雑音の長さを表し，縦軸に処理時間を表す．5 秒の雑音を用いてマッチングを行っても，処理時間は 1 秒にも満たないため，この方法で雑音マッチングを行うことで提案手法の音声認識をほぼリアルタイムで処理することが可能である．

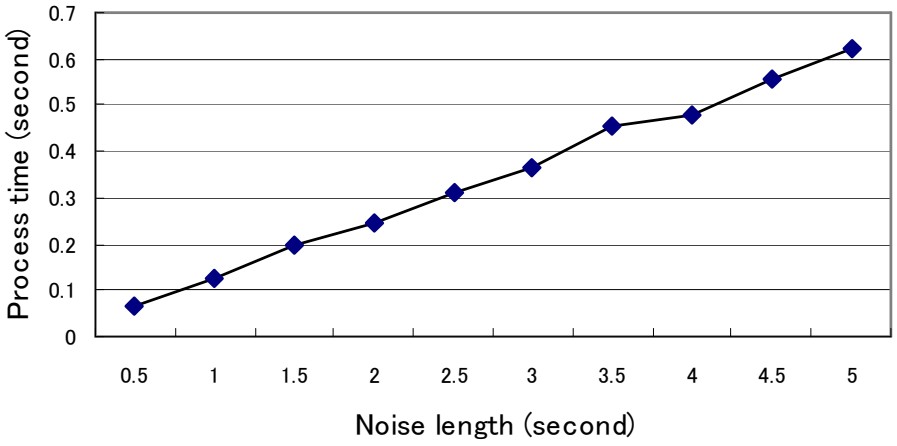


図 5.10: 雑音マッチングに要する時間

実装

第 4 章で示す MPML-HR ver. 3.0 のプレゼンテーションシステムは，音声認識エンジンに Julius を使用している．Julius をモジュールモードで立ち上げておき，音声認識結果が得られた際にはネットワーク経由でプレゼンテーションシステムへの入力が行われる．

提案する音声認識手法を MPML-HR ver. 3.0 へ組み込むため，Julius のモジュールモードに 2 種類のコマンドを追加した．1 つは PREPNOISE コマンドである．これは，テンプレート雑音の振幅を現在の雑音の振幅と合わせるためのコマンドである．音声認識システムに入力される信号の振幅は OS の設定等により変化するため，現在の雑音の振幅に合わせることでテンプレート雑音の振幅を調整する必要がある．そこで，何も動作がない静止時に PREPNOISE コマンドを発行することで，定常雑音での振幅調整を行い，テンプレート雑音の振幅を現在の雑音の振幅に合わせる．

さらに，ACTION コマンドを用意した．ロボットの動作時に ACTION\n*actionname* として Julius に送信する．これにより，Julius は入力信号を雑音バッファへためると同時に，対応するテンプレート雑音の読み込みを行う．ACTION コマンド発行後音声入力がなされると，雑音マッチングが行われ，MFT マスクの生成を経て提案手法による音声認識が行われる．

動作がない時点において音声入力が行われた際には，ロボットの静止時の雑音とマッチングが行われ，提案手法による音声認識が行われる．なお，ACTION コマンド発行後 10 秒間音声入力がなかった場合にはタイムアウトし，動作は終了したものとみなされる．タイムアウトの時間は一律 10 秒であるため，例えば，5 秒の動作に対し ACTION コマンド発行後 7 秒後に音声入力が行われた場合，テンプレート雑音には静止時の雑音が埋められることで補完される．

以上のように，Julius のモジュールモードに 2 つのコマンドを追加し，マルチバンド版 Julius を拡張することで，プレゼンテーションシステムへの組み込みが可能なように提案手法のリアルタイムでの実装を行った．

5.6　本章のまとめ

本章ではロボットの動作雑音除去を目的とした雑音適応手法の提案を行った．提案手法では，雑音除去処理と，歪みを補正するための白色雑音の重畳，MFT を用いることにより非定常雑音への適応を行う．雑音除去処理は SNR の高い環境では有効であるが，雑音が大きな環境においては歪みが大きく生じる．本研究ではこの歪みを抑えるために，白色雑音を重畳した．白色雑音を重畳させることは SNR を低下させ，一見，認識性能を下げるようにも思えるが，雑音の引き残し成分を平坦化し歪みを補正することで認識性能は大きく改善した．また，雑音除去処理は定常雑音には有効であるが，非定常雑音に対しては適応しきれない部分が存在する．この点を補完するため，MFT を用い，信頼性の低い部分が音声認識に関与する割合を低くすることにより認識性能の向上を図った．提案手法を用いることにより，従来から用いられている有効な手法であ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や SS を用いた動作音への適応手法よりも高い認識性能を達成できた．さらに，MLLR と組み合わせた実験も行い，教師なし MLLR では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有効であることを確認した．対話の中で蓄積される音声データを用いてオンライン適応することでプレゼンテーションシステムの中で提案手法と MLLR を組み合わせることが可能であり，従来手法よりも高い認識性能を得ることが可能である．また，提案する MPML-HR を用いたプレゼンテーションシステムへ組み込むため，提案する音声認識手法をリアルタイムで実装する方法についても示した．

第6章　結論

本論文では，近年注目を浴びているヒューマノイドロボットに着目し，ロボットの有するモダリティ，情報提供能力，およびネットワークへ接続可能なことから情報提供を行うタスク，特にプレゼンテーションが有効な活躍分野であると考えた．そこで第一に，スクリーンを用いたプレゼンテーションコンテンツを簡単な記述で実現すべく，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 MPML-HR の提案を行った．従来，ロボットの操作にはアセンブラなどを用いた制御プログラムが必要であり，専門的な知識を必要とする．一方で，スクリーン上のアニメーションエージェントの分野では簡単にマルチモーダルコンテンツを作成する記述言語が数多く存在する．そこで，アニメーションエージェントの記述言語をヒューマノイドロボット用に拡張することで，簡単な記述でプレゼンテーションコンテンツを作成できる枠組みの提案と実装を行った．実装にあたっては，自然なプレゼンテーションとなるよう，発話と動作を同時に実行できる機構の導入と，自律的な動作の導入を行った．これにより，簡単な記述によりヒューマノイドロボットを用いたプレゼンテーションを実現することが可能となった．また，ヒューマノイドロボットによるプレゼンテーションの印象評価を心理学的手法により行い，ロボットによるプレゼンテーションは視聴者の受ける印象がよいことを確認した．

人によるプレゼンテーションでは，質疑応答があり，不明な点に答えることで視聴者の理解が深まる．ロボットによるプレゼンテーションでも質問に答えられることは重要である．そこで第二に，プレゼンテーションシステムにインタラクション機構を導入した．プレゼンテーションで想定される質問として，再度の説明や説明の省略，分かりにくい箇所の説明などが挙げられる．これらの質問に答えるため，MPML-HR に新たな命令を導入し，説明箇所を遷移することでインタラクションを実現した．また，音声入力を含むアプリケーションでは，音声認識の誤認識の問題がある．誤認識の結果，視聴者の意図しないプレゼンテーションの遷移が行われると，視聴者に煩わしさを与え，インタラクション機能を導入する意義が薄れる．この音声認識の誤認識の問題に対処するため，信頼度を用いて音声認識結果を棄却，聞き返し，確認，受理のいずれかをすることとした．さらに，誤って遷移した場合には，ユーザからの指摘により，遷移前の状態に戻す機能を導入することで，音声認識誤りへの対応を行った．これらの機能の導入により，ロボットによるプレゼンテーションに音声インタラクション

を実現することができた．なお，インタラクションには，ロボットの対話行動制御モデルを用いてシステムの実装を行った．

インタラクション機能の導入にあたり，音声認識誤りへの対応をシステムに導入した．しかし，ロボットは動作音や環境雑音など雑音を含む環境下で音声認識をする必要があり，システムレベルのみならず，音声認識レベルでも雑音に頑健な手法を導入することが必要である．特にプレゼンテーションでは動作があり，動作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な雑音となり音声認識性能を低下させる大きな要因となる．そこで第三に，ロボットの動作音に頑健な音声認識手法の提案を行った．ロボットの動作音は，自己の動作情報を用いることにより推定が可能であるという特徴を活かし，MFT を用いた動作音への適応を提案した．ロボットの定常成分への適応には，マルチコンディション学習による音響モデルと SS を組み合わせ，非定常成分への適応には，MFT を用いる．これにより従来手法よりも高い認識性能を実現した．これらの手法を単純に組み合わせただけでは認識性能の向上は得られないが，白色雑音を重畳することで認識性能の向上を得ることができた．白色雑音の重畳は SNR を低下させ，一見，認識性能をも低下させるように思えるが，SS 処理時に発生す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．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．

本論文で提案する枠組みを用いることで，プレゼンテーションコンテンツを簡単に記述し，ヒューマノイドロボットによる実現が可能となった．しかし，ロボットをより身近なものとし，人間と協調するためには，この枠組みを活かしたさらなる応用が必要である．具体的には以下のことが考えられる．第一に，本論文で提案するプレゼンテーションはスクリーン上の資料を用いたものを対象とした．しかし，ロボットが実空間に存在するという利点を活かすためには，実際の商品の説明なども対象とし，必要な拡張を行わなくてはならない．第二に，提案するインタラクション機構はプレゼンテーションにおいてあらかじめ想定していた質問などへの対応を行うものである．しかし，より自由なインタラクションを行うためには，自由な対話を実現するエキスパートを構築し，プレゼンテーションとこのエキスパートとの切り替えを行うことが必要である．第三に，提案する音声認識手法はロボットの発する動作音へ適応することを目的とした．動作音へ適応することはプレゼンテーションにおいて重要な課題の一つであるが，周囲から発せられる環境雑音への対応や，複数の質問者から同時に質問がされた場合の対応も重要な課題である．ロボットの音声認識では，マイクロホンアレーを用いた手法が数多く提案されている．そこで，マイクロホンアレーを用いた手法と本手法を組み合わせることで，環境雑音への適応や，複数話者からの問いかけへの対応を実現することが必要である．これらの課題を解決し，今後ヒューマノイドロボットの活躍分野が広がることを期待する．

謝辞

本研究を進めるにあたり，終始暖かい御指導を受け賜りました石塚満氏 (東京大学大学院情報理工学系研究科教授) に心から感謝申し上げます.

博士論文の審査では，石塚満氏のほか，広瀬啓吉氏 (東京大学大学院情報理工学系研究科教授)，原島博氏 (東京大学大学院情報学環・学際情報学府教授)，佐藤洋一氏 (東京大学生産技術研究所助教授)，峯松信明氏 (東京大学大学院新領域創成科学研究科助教授)，苗村健氏 (東京大学大学院情報学環・学際情報学府助教授) に貴重なご意見を賜りました．ここに厚く御礼申し上げます.

土肥浩氏 (東京大学大学院新領域創成科学研究科助手) には MPML-HR を実現するきっかけを与えて頂き，ここに深く御礼申し上げます.

本研究は，(株) ホンダ・リサーチ・インスティチュート・ジャパンにおいて多くの方々の暖かいご支援に支えられて行うことができたものです．MPML-HR の実現および論文投稿時に貴重な助言を頂きました辻野広司氏 (ホンダ・リサーチ・インスティチュート・ジャパン)，研究全般に渡り御指導頂き，ミーティング時におきましても多くの助言を頂きました中野幹生氏 (ホンダ・リサーチ・インスティチュート・ジャパン)，音声認識の研究を進めるにあたり貴重な助言を頂きました中臺一博氏 (ホンダ・リサーチ・インスティチュート・ジャパン，東京工業大学大学院情報理工学研究科客員助教授)，MPML-HR の実現および心理評価実験において御指導と助言を頂きました竹内誉羽氏 (ホンダ・リサーチ・インスティチュート・ジャパン)，インタラクション機能の導入に関して多くの助言を頂きました船越孝太郎氏 (ホンダ・リサーチ・インスティチュート・ジャパン)，ASIMO の操作時に多くの助言を頂き御指導頂きました長谷川雄二氏 (ホンダ・リサーチ・インスティチュート・ジャパン)，音声認識の信号処理分野において多くの助言を頂きました中島弘史氏 (ホンダ・リサーチ・インスティチュート・ジャパン)，MPML-HR のプログラムの構築にあたり貴重な助言を頂きました鳥井豊隆氏 (ホンダ・リサーチ・インスティチュート・ジャパン)，MPML-HR の心理評価実験におきまして御指導頂きました有吉斗紀知氏 (現在，本田技術研究所) に厚く感謝申し上げます．また，心理評価実験等に協力して頂き，日頃からお世話になりましたホンダ・リサーチ・インスティチュート・ジャパンの方々に心から感謝申し上げます.

音声認識の研究におきましては，東京工業大学古井研究室にもお世話になりました．貴重な助言を頂きました古井貞熙氏 (東京工業大学大学院情報理工学研究科教授)，親

切かつ丁寧に多くの助言を頂きました岩野公司氏 (東京工業大学大学院情報理工学研究科助手) に深く感謝致します.

海外における学会での発表練習時には，奥乃博氏 (京都大学大学院情報学研究科教授) に貴重な助言を頂きました．ここに厚く感謝申し上げます.

研究に進めるにあたり，学生の方々にもご協力いただきました．心理評価実験及び MPML-HR の実装において多くの助言と協力を頂きました櫛田和貴氏 (現在，東京電力株式会社)，音声の収録，音声プログラムの操作及び Linux の操作等におきまして日頃からご協力頂きました山本俊一氏 (京都大学大学院情報学研究科博士課程)，音声の収録において何度もご協力頂きました隅谷亮太氏 (現在，東日本電信電話株式会社)，インタラクション機能の導入にあたりプログラムの作成時にご協力頂きました簑津真一郎氏 (東京大学大学院情報理工学系研究科修士課程) に深く感謝申し上げます.

日頃から石塚研究室のメンバーにもお世話になりました．同期入学でもあり，学生生活において励ましてくれた小野真裕氏 (東京大学大学院情報理工学系研究科博士課程) をはじめ石塚研究室のメンバー皆様に深く感謝申し上げます.

本研究を進めるにあたり，励まして下さった方々に改めて心から感謝致します.

参考文献

[1] 日本ロボット学会，ロボット工学ハンドブック, コロナ社，2005.

[2] 石黒浩，神田崇行，宮下敬宏，コミュニケーションロボット, オーム社，2005.

[3] 日本規格協会，JIS ハンドブック, 2006.

[4] AIBO Official Site: http://www.jp.aibo.com/

[5] 藤田雅博，景山浩二，大槻正，天貝佐登史，土井利忠，“エンターテイメントロボット AIBO の開発,” 日本ロボット学会誌，Vol.21, No.1, pp.55-56, 2003.

[6] Masahiro Fujita, “AIBO: Toward the era of digital creatures,” *The International Journal of Robotics Research*, Vol.20, No.10, pp.781-794, 2001.

[7] PARO homepage: http://paro.jp/

[8] 柴田崇徳，“人の心を癒すメンタルコミットロボット,” 日本ロボット学会誌，Vol.17, No.3, pp.943-946, 1999.

[9] NeCoRo homepage: http://www.necoro.com/home.html

[10] 田島，斉藤，柴田，“猫ロボットのインタラクションによるユーザー評価と考察,” 第 17 回日本ロボット学会学術講演会予稿集，1999.

[11] 田島年浩，“感情をもったペット型ロボット,” 日本ロボット学会誌，Vol.18, No.2, pp.188-189, 2000.

[12] 加藤一郎，“2 足歩行ロボット (WABOT-1) の開発,” バイオメカニズム 2，東京大学出版会，pp.173-174, 1973.

[13] ASIMO homepage: http://www.honda.co.jp/ASIMO/

[14] 下村秀樹，青山一美，藤田 雅博，“自律型エンタテイメントロボットと音声対話,” 人工知能学会第 36 回言語・音声理解と対話処理研究会予稿集，2002.

[15] 黒木義博，石田健蔵，山口仁一，“高度統合運動制御機能を有する小型エンターテイメントロボット,” 第 19 回日本ロボット学会学術講演会予稿集，pp.1093-1094, 2001.

[16] K. Kaneko, F. Kanehiro, S. Kajita, K. Yokoyama, K. Akachi, T. Kawasaki, S. Ota and T. Isozumi, “Design of Prototype Humanoid Robotics Platform for HRP,” in *Proceedings of International Conference on Intelligent Robotics and Systems (IROS-02)*, pp.2431-2436, 2002.

[17] 神田崇行，石黒浩，小野哲雄，今井倫太，中津良平，“研究用プラットフォームとしての日常活動型ロボット“ Robovie ”の開発,” 電子情報通信学会論文誌，Vol.J85-D-I, No.4，pp.380-389, 2002.

[18] 小野哲雄，今井倫太，石黒浩，中津良平，“身体表現を用いた人とロボットの共創対話,” 情報処理学会論文誌，Vol.42, No.6, pp.1348-1358, 2001.

[19] 神田崇行，石黒浩，小野哲雄，今井倫太，中津良平，“人-ロボットの対話におけるロボット同士の対話観察の効果,” 電子情報通信学会論文誌，Vol.J-85-D-I, No.4, pp.380-389, 2002.

[20] 森政弘，“不気味の谷,” Energy(エッソスタンダード石油広報誌), Vol.7, No.4, 1970.

[21] J. Schulte, C. Rosenbergn and S. Thrun, “Spontaneous, short-term interaction with mobile robots,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA-99)*, pp.658-663, 1999.

[22] C. Breazeal and B. Scassellati, “A context-dependent attention system for a social robot,” in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, pp.1146-1151, 1999.

[23] C. Breazeal, *Desigining Sociable Robots*, MIT Press, 2002.

[24] B. Scassellati, “Investigating models of social development using a humanoid robot,” *Biorobotics*, pp.145-167, 2000.

[25] 小林哲則，白井克彦，“ヒューマノイドロボットにおけるマルチモーダル会話インターフェース,” 電子情報通信学会技術報告，SP99-113, pp.121-126, 1999.

[26] 横山真男，青山一美，菊池英明，帆足啓一郎，白井克彦，“人間型ロボットの対話インターフェースにおける発話交替時の非言語情報の制御,” 情報処理学会論文誌，Vol.40, No.2, pp.487-496, 1999.

[27] K. Murai and S. Nakamura, “Real time face detection for multimodal speech recognition,” in *Proceedings of International Conference on Multimedia and Expo*, Vol.2, pp.373-376, 2002.

[28] K. Nakadai, T. Lourens, H. G. Okuno and H. Kitano, “Ative Audition for Humanoid,” in *Proceedings of 17th International Conference on Artificial Intelligence*, pp.832-839, 2000.

[29] A. Mehrabian, *Silent Messages*, Thomson Learning College, 1980.

[30] 松井俊浩，麻生英樹，J. Fry, 浅野太，木村陽一，原功，栗田多喜夫，速水悟，山崎信行，“オフィス移動ロボット Jijo-2 の音声対話システム,” 日本ロボット学会誌，Vol.18, No.2, pp.142-149, 2000.

[31] S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, D. Schulz, “MINERVA: a second-generation museum tour-guide robot,” in *Proceedings of International Conference on Robotics and Automation (ICRA-99)*, Vol.3, pp.1999-2005, 1999.

[32] 西村竜一, 怡土順一, 李晃伸, 松本吉央，“情報科学研究の実環境プラットフォームとしての受付案内ロボット ASKA,” 情報処理学会第 64 回全国大会予稿集，2002.

[33] 萩田紀博，“ネットワーク・ロボット研究プロジェクト,” 日本ロボット学会誌，Vol.23, No.5, pp.532-534, 2005.

[34] 成田雅彦，日浦亮太，“ネットワークを通じたロボットサービス提供のための規格:RSi,” 日本ロボット学会誌，Vol.23, No.6, pp.650-654, 2005.

[35] 土井美和子，長谷川哲夫，上野晃嗣，“サービス提供のためのネットワークロボット連携技術,” 日本ロボット学会誌，Vol.23, No.6, pp.660-664, 2005.

[36] 酒井純一，西山裕之，溝口文雄，“人型二足歩行ロボットを用いたプレゼンテーションロボットシステムの設計，” 日本ソフトウェア科学会第 21 回大会論文集，pp.1-4, 2004.

[37] B. De Carolis, V. Carofiglio, M. Bilvi and C. Pelachaud, “APML, a Mark-up Language forBelievable Behavior Generation,” in *Proceedings of International Conference on Autonomous Agents and Multi Agent Systems (AAMAS-02) Workshop Embodied Conversational Agents*, 2002.

[38] B. De Carolis, V. Carofiglio, C. Pelachaud and I. Poggi, “Interactive Information Presenta-tion by an Embodied Animated Agent,” in *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue*, pp.19-23, 2001.

[39] S. Kshirsagar, A. Guye-Vuilleme, K. Kamyab, N. Managenat-Thalmann, D. Thalmann, E. Mamdani, “Avatar Markup Language,” in *Proceedings of the Eighth Eurofraphics Workshop on Virtual Environments*, 2002.

[40] Y. Arafa, E. Mamdani, “Scripting embodied agents behaviour with CML: Character markup language,” in *Proceedings of the 7th International Conference on Intelligent User Interface (IUI-03)*, 2003.

[41] A. Beard, D. Reid, “MetaFace and VHML: A First Implementation of the Virtual Human Markup Language,” in *Proceedings of Embodied conversational agents*, 2002.

[42] VHML homepage: http://www.vhml.org/

[43] Zhisheng Huang, Anton Eliens, and Cees Visser, “STEP: A Scripting Language for Embodied Agents,” in *Proceedings of the Workshop on Lifelike Animated Agents*, 2002.

[44] STEP homepage: http://step.intelligent-multimedia.net/

[45] 林正樹，“テキスト台本からの自動番組制作〜TVML の提案,” テレビジョン学会年次大会予稿集，S4-3, pp.589-592, 1996.

[46] TVML homepage: http://www.nhk.or.jp/strl/tvml/

[47] N. Badler, R. Bindiganavale, J. Allbeck, W. Schuler, L. Zhao and M. Palmer, *Conversational Agents*, ed J. Cassell et al., MIT Press, pp.256-284, 2000.

[48] PAR homepage: http://hms.upenn.edu/software/NewPAR/par.html

[49] P. Piwek, B. Krenn, M. Schroder, M. Grice, S. Baumann and H. Pirke, “RRL - A rich representation language for the description of agent behaviour in NECA,” in *Proceedings of the Autonomous Agents Workshop on Achieving Human-like Behavior in Interactive Animated Agents*, 2000.

[50] RRL homepage: http://www.ofai.at/research/nlu/NECA/RRL/

[51] A. Kransted, S. Kopp and I. Wachsmuth, "MURML - A multimodal utterance representation markup language for conversational agents," in *Proceedings of International Conference on Autonomous Agents and Multi Agent Systems (AAMAS-02) Workshop on Embodied Conversational Agents*, 2002.

[52] HumanML homepage: http://www.humanmarkup.org/work/humanmlSchema.zip

[53] Helmut Prendinger and Mitsuru Ishizuka (Eds.), *Life-like Characters: Tools, Affective Functions, and Applications*, Springer, 2003.

[54] Life-like characters book online material: http://www.vhml.org/LLC/llc-book.html

[55] T. Tsutsui, S. Saeyor and M. Ishizuka, "MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions," in *Proceedings of WebNet 2000 World Conference on the WWW and Internet*, 2000.

[56] Y. Zong, H. Dohi and M. Ishizuka, "Multimodal Presentation Markup Language MPML with Emotion Expression Functions Attached," in *Proceedings of International Symposium on Multimedia Software Engineering (IEEE Computer Soc.)*, pp.359-365, 2000.

[57] N. Okazaki, S. Aya, S. Saeyor, and M. Ishizuka, "A Multimodal Presentation Markup Language MML-VR for a 3D Virtual Space," in *Proceedings of Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges (in conjunction with HF2002 and OZCHI2002)*, 2002.

[58] Santi Saeyor, Koki Uchiyama, Mitsuru Ishizuka, "Multimodal Presentation Markup Language on Mobile Phones," in *Proceedings of International Conference on Autonomous Agents and Multi Agent Systems (AAMAS-03) Workshop - Embodied Conversational Characters as Individuals*, pp.68-71, 2003.

[59] MPML homepage: http://www.miv.t.u-tokyo.ac.jp/mpml.html

[60] A. Ortony, G. L. Clore, A. Collins, *The cognitive structure of emotions*, Cambridge University Press, 1988.

[61] Justine Cassell, Joseph Sullivan, Scott Prevost and Elizabeth Churchill, *Emodied Conversational agents*, MIT Press, 2000.

[62] 藤田雅博，景山浩二，佐部浩太郎，尾山一文，“ロボットエンターテインメント用開発環境 OPEN-R について,” 日本ロボット学会誌，Vol.21, No.6, pp.596-601, 2003.

[63] 尾崎文夫，大明準治，佐藤広和，橋本英昭，松日楽信人，“オープンロボットコントローラアーキテクチャ(ORCA),” 日本ロボット学会誌，Vol.21, No.6, pp.602-608, 2003.

[64] F. Kanehiro, K. Fujiwara, S. Kajita, K. Yokoi, K. Kaneko, H. Hirukawa, Y. Nakamura and K. Yamane, “Open Architecture Humanoid Robotics Platform,” in *Proceedings of the 2002 IEEE International Conference on Robotics and Automation (ICRA-02)*, pp.24-30, 2002.

[65] 金広文男，“ヒューマノイドソフトウェアプラットフォーム OpenHRP とその応用事例,” 日本ロボット学会誌，Vol.21, No.6, pp.609-614, 2003.

[66] ORiN homepage: http://www.orin.jp/

[67] 北垣高成，末廣尚士，神徳徹雄，平井成興，谷江和雄，“RT ミドルウェア技術基盤の研究開発について-ロボット機能発現のために必要な要素技術開発,” ロボティクスシンポジア予稿集，pp.487-492，2003.

[68] RT ミドルウェア homepage: http://www.is.aist.go.jp/rt/

[69] 中田亨，佐藤知正，森武俊，溝口博，“ロボットの対人行動による親和感の演出,” 日本ロボット学会誌，Vol.15, No.7, pp.1068-1074, 1997.

[70] 尾形哲也，菅野重樹，“人間とロボットの情緒的コミュニケーションの実験的評価-アーム・ハンドによる人間との物理的インタラクション,” システム制御情報学会論文誌，Vol.13, No.12, pp.566-574, 2000.

[71] 神田崇行, 石黒浩, 石田亨, “人間ロボット間相互作用に関わる心理学的評価,” 日本ロボット学会誌, Vol.19, No.3, pp.362-371, 2001.

[72] 神田崇行, 石黒浩, 小野哲雄, 今井倫太, 中津良平, “人間と相互作用する自律型ロボット Robovie の評価,” 日本ロボット学会誌, Vol.20, No.3, pp.315-323, 2002.

[73] T. Kanda, H. Ishiguro, and T. Ishida, “Psychological analysis on human-robot interaction,” in *Proceedings of 2001 IEEE International Conference on Robotics and Automation (ICRA-01)*, pp.4166-4173, 2001.

[74] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and R. Nakatsu, "Development and evaluation of an interactive humanoid robot "Robovie"," in *Proceedings of 2002 IEEE International Conference on Robotics and Automation (ICRA-02)*, pp.1848-1855, 2002.

[75] T. Kanda, T. Miyashita, T. Osada, Y. Haikawa, and H. Ishiguro, "Analysis of humanoid appearances in human–robot interaction," in *Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-05)*, pp.62-69, 2005.

[76] J. G. Snider and C. E. Osgood, *Semantic differential techniques, a sourcebook*, 1969.

[77] Microsoft agent home page: http://www.microsoft.com/msagent

[78] C. D. Kidd and C. Breazeal, "Effect of a robot on user perceptions," in *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-04)*, pp.3559-3564, 2004.

[79] K. Shinozawa, F. Naya, J. Yamamoto, and K. Kogure, "Differences in effect of robot and screen agent recommendations of human decision-making," *International Journal on Human–Computer Studies*, Vol.62, No.2, pp.267-279, 2005.

[80] 竹林洋一, "音声自由対話システム TOSBURG2-ユーザ中心のマルチモーダルインタフェースの実現に向けて-," 電子情報通信学会論文誌, Vol.J-77-D-II, pp.1417-1428, 1994.

[81] D. Goddeau, E. Brill, J. Glass, C. Pao, M. Phillips, J. Polifroni, S. Seneff and V. Zue, "Galaxy : A Human-Language Interface to On-Line Travel Information," in *Proceedings of International Conference on Spoken Language Processing (ICSLP-94)*, Vol.2, pp.707-710, 1994.

[82] V. Zue, S. Seneff, J. R. Glass, J. Polifroni, C. Pao, T. J. Hazen, L. Hetherington, "JUPlTER: a telephone-based conversational interface for weatherinformation," *IEEE Transactions on Speech and Audio Processing*, Vol.8, pp.85-96, 2000.

[83] V. Zue, J. Glass, D. Goddeau, D. Goodineanda, L. Hirschman, M. Phillips, J. Polfroni and S. Seneff, "PEGASUS : A Spoken Dialogue Interface for On-Line Air Travel Planning," in *Proceedings of International Symposium of Spoken Dialogue*, pp.157-160, 1993.

[84] J. Glass and E. Weinstein, "SpeechBuilder: Facilitating spoken dialogue system development," in *Proceedings of European Conference on Speech Communication and Technology (Eurospeech-01)*, pp.1335-1338, 2001.

[85] S. Sutton et al., "Universal Speech Tools: The CSLU Toolkit" in *Proceedings of International Conference on Spoken Language Processing (ICSLP-98)*, pp.3221-3224, 1998.

[86] Voice XML Forum homapage: http://www.voicexml.org/

[87] K. Katsurada, Y. Nakamura, H. Yamada, T. Nitta, "XISL:A Language for Describing Multimodal Interaction Scenarios," in *Proceedings of International Conference on Multimodal Interfaces (ICMI-03)*, pp.281-284, 2003.

[88] 川本真一，下平博，新田恒雄，西本卓也，中村哲，伊藤克亘，森島繁生，四倉達夫，甲斐充彦，李晃伸，山下洋一，小林隆夫，徳田恵一，広瀬啓吉，峯松信明，山田篤，伝康晴，宇津呂武仁，嵯峨山茂樹，"カスタマイズ性を考慮した擬人化音声対話エージェントツールキットの設計," 情報処理学会論文誌, Vol.43, No.7, pp.2249-2263, 2002.

[89] AIML homepage: http://www.alicebot.org/

[90] Andre M. M. Neves and Flavia A. Barros, "XbotML: A Markup Language for Human Computer Interaction via Chatterbots," *Lecture Notes in Computer Science*, Springer, Vol.2722, pp.171-181, 2003.

[91] 杉山聡，堂坂浩二，川端豪，"音声対話によるテキスト内容の伝達方法," 情報処理学会論文誌，Vol.41, No.6, pp.1883-1894, 2000.

[92] M. Nakano, Y. Hasegawa, K. Nakadai, T. Nakamura, J. Takeuchi, T. Torii, H. Tsujino, N. Kanda, and H. G. Okuno, "A two-layer model for behavior and dialogue planning in conversational service robots," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-05)*, pp.1542-1548, 2005.

[93] 鹿野清宏，伊藤克亘，河原達也，武田一哉，山本幹雄，音声認識システム, オーム社，2001.

[94] 古井貞熙，音声情報処理, 森北出版株式会社，1998.

[95] 中村哲，“実音響環境に頑健な音声認識を目指して,” 電子情報通信学会技術報告，Vol.102, No.33, pp.31-36, 2002.

[96] 松本弘，“雑音環境下の音声認識手法,” 第 2 回情報科学技術フォーラム (FIT2003) 予稿集, 2003.

[97] C. Breazeal, *Designing Sociable Robots*, MIT press, 2002.

[98] S. F. Boll, “A spectral subtraction algorithm for suppression of acoustic noise in speech,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-79)*, pp.200-203, 1979.

[99] M. Berouti, R. Schwartz and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-79)*, pp.208-211, 1979.

[100] P. Lockwood and J. Boudy, “Experiments with a Nonlinear Spectral Subtractor (NSS), hidden Marcov models and the projection, for robust speech recognition in cars,” *Speech Communication*, Vol.11, pp.215-228, 1992.

[101] P. Lockwood and J. Boudy, “Experiments with a Non-linear Spectral Subtractor (NSS), hidden Marcov models and the projection, for robust speech recognition in cars,” in *Proceedings of European Conference on Speech Communication and Technology (Eurospeech-91)*, pp.79-82, 1991.

[102] P. Lockwood and J. Boudy, “Non-linear Spectral Subtractor (NSS) and hidden Marcov models for robust speech recognition in car noise environments,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92)*, Vol.I, pp.265-268, 1992.

[103] 鹿野清宏，中村哲，伊勢史郎，音声・音情報のディジタル信号処理, 昭晃堂，1997.

[104] 金学胤，浅野太，鈴木陽一，曽根敏男，“短時間振幅スペクトル推定を用いた 2 チャネル音声強調法における振幅スペクトル推定について,” 日本音響学会秋季研究発表会講演論文集，Vol.I, pp.499-500, 1999.

[105] A. Ito, T. Kanayama, M. Suzuki, and S. Makino, “Internal noise suppression for speech recognition by small robots,” in *Proceedings of European Conference on Speech Communication and Technology (Eurospeech-2005)*, pp.2685-2688, 2005.

[106] 山出慎吾，馬場朗，芳澤伸一，李晃伸，猿渡洋，鹿野清宏, “実環境における頑健な音声認識のための音韻モデルの教師なし話者適応,” 電子情報通信学会論文誌, Vol.J87-D-II, No.4, pp.933-941, 2004.

[107] A. Papoulis, *Probability, random variables, and stochastic process*, McGraw-Hill, 1995.

[108] B. Noe, J. Sienel, D. Jouvet, L. Mauuary, L. Boves, J. Veth and F. Wet, “Noise reduction for noise robust feture extraction for distributed speech recognition,” in *Proceedings of European Conference on Speech Communication and Technology (Eurospeech-01)*, pp.433-436, 2001.

[109] J. S. Lim and A. V. Oppenheim, “All-pole modeling of degraded speech,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol.26, pp.197-210, 1987.

[110] K. K. Paliwal and A. Basu, “A speech enhancement method based on Kalman filtering,” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP-87)*, Vol.1, pp.177-180, 1987.

[111] Y. Ephraim and D. Malah, “Speech enhancement using minimum meansquare error short-time spectral amplitude estimator,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol.32, No.6, pp.1109-1121, 1984.

[112] 武田一哉，板倉文忠，“音源分離による音声認識性能の改善,” 日本音響学会誌, Vol.53，No.11, pp.883-888, 1997.

[113] I. Hara, F. Asano, H. Asoh, J. Ogata, N. Ichimura, Y. Kawai, F. Kanehiro, H. Hirukawa, and K. Yamamoto, “Robust speech interface based on audio and video information fusion for humanoid HRP-2,” in *Proceedings of IEEE/RAS International Conference on Intelligent Robots and Systems (IROS-04)*, pp.2404-2410, 2004.

[114] S. Araki, S. Makino, R. Mukai, Y. Hinamoto, T. Nishikawa, and H. Saruwatari, “Equivalence between Frequency Domain Blind Source Separation and Frequency Domain Adaptive Beamforming,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-02)*, pp.1899-1902, 2002.

[115] H. Saruwatari, S. Kurita, K. Takeda, F. Itakura, T. Nishikawa, and K. Shikano, “Blind source separation combining independent component analysis and beam-

forming," *EURASIP Journal on Applied Signal Processing*, Vol.2003, No.11, pp.1135-1146, 2003.

[116] S. Yamamoto, K. Nakadai, J. M. Valin, J. Rouat, F. Michaud, T. Ogata, H. Komatani, and H. G. Okuno, "Making a robot recognize three simultaneous sentences in real-time," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-05)*, pp.897-892, 2005.

[117] H. Hermansky, "Perceptual linear predictive (PLP) analysis for speech," *Journal of Acoust. Soc. of America* Vol.87, pp.1738-1752, 1990.

[118] H. Helmansky and N. Morgan, "RASTA processing of speech," *IEEE Trans. Speech Audio Process*, Vol.2, pp.578-589, 1994.

[119] B. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *Journal of Acoust. Soc. of America*, Vol.55, No.6, pp.1304-1312, 1974.

[120] O. Viikki and K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition," *Speech Communication*, Vol.25, pp.133-147, 1998.

[121] J. C. Segura, M. C. Benitez, A. de la Torre, A. M. Peinado, and A. Rubio, "Non-linear transformations of the feature space for robust speech recognition," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP-02)*, Vol.I, pp.401-404, 2002.

[122] Sirko Molau, Florian Hilger, Daniel Keysers and Hermann Ney, "Enhanced Histogram Normalization in The Acoustic Feature Space," in *Proceedings of International Conference on Spoken Language Processing (ICSLP-02)*, Vol.1, pp.1421-1424, 2002.

[123] Donglai Zhm, Satoshi Nakamura, Kuldip K. Paliwal and Renhua Wang, "Maximum Likelihood Sub-band Weighting for Robust Speech Recognition," in *Proceedings of European Conference on Speech Communication and Technology (Eurospeech-03)*, pp.673-676, 2003.

[124] William J. J. Roberts and Sadaoki Furui, "Robust speech recognition via modeling spectral coefficients with HMM's with complex gaussian components," in *Proceedings of International Conference on Spoken Language Processing (ICSLP-00)*, Vol.4, pp.484-487, 2000.

[125] Satoshi Tamura, Koji iwano and Sadaoki Furui, "A stream-weight optimization method for audio-visual speech recognition using multi-stream HMMs," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP-04)*, Vol.1, pp.857-860 2004.

[126] 西村義隆, 篠崎隆宏, 岩野公司, 古井貞熙, "周波数帯域ごとの重みつき尤度を用いた雑音に頑健な音声認識," 電子情報通信学会技術研究報告, SP2003-116, pp.19-24, 2003.

[127] Y. Nishimura, T. Shinozaki, K. Iwano, and S. Furui, "Noise-robust speech recognition using multi-band spectral features," in *Proceedings of 148th Acoustical Society of America Meetings*, 1aSC7, 2004.

[128] A. Hagen and A. Morris, "Comparison of HMM experts with MLP experts in the full combination multi-band approach to robust ASR," in *Proceedings of International Conference on Spoken Language Processing (ICSLP-00)*, Vol.1, pp.345-348, 2000.

[129] H. Bourlard and S. Dupont, "A new ASR approach based on independent processing and recombination of partial frequency bands," in *Proceedings of International Conference on Spoken Language Processing (ICSLP-96)*, Vol.1, pp.426-429, 1996.

[130] J. Barker, M. Cooke, and P. Green, "Robust asr based on clean speech models: An evaluation of missing data techniques for connected digit recognition in noise," in *Proceedings of 7th European Conference on Speech Communication Technology (Eurospeech-01)*, Vol.1, pp.213-216, 2001.

[131] B. Raj and R. M. Stern, "Missing-feature approaches in speech recognition," *Signal Processing Magazine*, Vol.22, No.5, pp.101-116, 2005.

[132] A. Vizinho, P. Green, M. Cooke, L. Josifovski, "Missing Date Theory, Spectral Subtraction and Signal-to-Noise Estimation for Robust ASR: An Integrated Study," in *Proceedings of European Conference on Speech Communication Technology (Eurospeech-99)*, pp.2407-2410, 1999.

[133] B. Raj, M. Seltzer, R. Stern, "Reconstruction of Damaged Spectrographic Features For Robust Speech Recognition," in *Proceedings of International Conference on Spoken Language Processing (ICSLP-00)*, pp.341-344, 2000.

[134] M.Cooke, A.Morris, and P.Green, "Missing data techniques for robust speech recognition," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP-97)*, pp.803-806, 1997.

[135] R. P. Lippman and B. A. Carlson, "Using missing data theory to actively select features for robust speech recognition with interruptions," in *Proceedings of European Conference on Speech Communication Technology (Eurospeech-97)*, pp.37-40, 1997.

[136] L. Josifovski et al., "State based imputation of missing data for robust speech recognition and speech enhancement," in *Proceedings of European Conference on Speech Communication Technology (Eurospeech-99)*, pp.2837-2840, 1999.

[137] R. Lippmann, E. Martin and D. Paul, "Multi-style training for robust isolated-word speech recognition," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP-87)*, No.12, pp.705-708, 1987.

[138] C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density hidden markov models," *Computer Speech and Language*, Vol.9, pp.171-185, 1995.

[139] J. L. Gauvain, C. H. Lee, "Maximum a posteriori estimation for multi-variate Gaussian mixture observations of Marcov chains," *IEEE Trans. Speech and Audio Process*, Vol.2, pp.291-298, 1994.

[140] O. Siohan, C. Chesta, C. H. Lee, "Joint Maximum A Posteriori Estimation of Transformation and Hidden Marcov Model Parameters," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP-96)*, Vol.II, pp.965-968, 1996.

[141] L. Gillick and S. Cox, "Some statistical issues in the comparison of speech recognition algorithms," in *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP-89)*, pp.532-535, 1989.

[142] 河原達也，李晃伸，"連続音声認識ソフトウエア Julius," 人工知能学会誌, Vol.20, No.1, pp.41-49, 2005.

[143] Julius homepage: http://julius.sourceforge.jp/

[144] マルチバンド版 Julius homepage: http://www.furui.cs.titech.ac.jp/mband_julius/

発表文献

論文誌

1. Yoshitaka Nishimura, Kazutaka Kushida, Hiroshi Dohi and Mitsuru Ishizuka, Johane Takeuchi, Mikio Nakano and Hiroshi Tsujino, "Development of Multimodal Presentation Markup Language MPML-HR for Humanoid Robots and Psychological Evaluation," *International Journal of Humanoid Robotics*, Accepted.

2. 西村義隆，石塚満，中臺一博，中野幹生，辻野広司，"MFT を用いたロボットの動作中における音声認識," 日本ロボット学会誌，投稿中.

3. 西村義隆，簑津真一郎，中野幹生，船越孝太郎，竹内誉羽，土肥浩，石塚満，長谷川雄二，辻野広司，"インタラクション機能を有するプレゼンテーション記述言語とヒューマノイドロボットを用いた実装," 投稿準備中.

国際学会

1. Kazutaka Kushida, Yoshitaka Nishimura, Hiroshi Dohi, Mitsuru Ishizuka, Johane Takeuchi and Hiroshi Tsujino, "Humanoid Robot Presentation through Multimodal Presentation Markup Language MPML-HR," in *Proceedings of 4th International Conference on Autonomous Agents and Multi Agent Systems (AAMAS-05) Workshop 13*, pp.23-29, 2005.

2. Yoshitaka Nishimura, Kazutaka Kushida, Hiroshi Dohi, Mitsuru Ishizuka, Johane Takeuchi,Hiroshi Tsujino, "Development and psychological evaluation of Multimodal Presentation Markup Language for Humanoid Robots," in *Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids-05)*, pp.393-398, 2005.

3. Yoshitaka Nishimura, Mikio Nakano, Kazuhiro Nakadai, Hiroshi Tsujino and Mitsuru Ishizuka, "Speech Recognition for a Robot under its Motor Noise by

Selective Application of Missing Feature Theory and MLLR," in *Proceedings of ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA-06)*, pp.53-58, 2006.

4. Johane Takeuchi, Kazutaka Kushida, Yoshitaka Nishimura, Hiroshi Dohi, Mitsuru Ishizuka, Mikio Nakano and Hiroshi Tsujino, "Comparison of a Humanoid Robot and an On-screen Agent as Presenters to Audiences," in *Proceedings of International conference on Intelligent Robots and Systems (IROS-06)*, pp.3964-3969, 2006.

5. Yoshitaka Nishimura, Kazuhiro Nakadai, Mikio Nakano, Hiroshi Tsujino and Mitsuru Ishizuka, "Speech Recognition for a Humanoid with Motor Noise Utilizing Missing Feature Theory," in *Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids-06)*, pp.26-33, 2006.

6. Yoshitaka Nishimura, Shinichiro Minotsu, Hiroshi Dohi, Mitsuru Ishizuka, Mikio Nakano, Kotaro Funakoshi, Johane Takeuchi, Yuji Hasegawa, Hiroshi Tsujino, "A Markup Language for Describing Interactive Humanoid Robot Presentations," in *Proceedings of International Conference on Intelligent User Interfaces (IUI-07)*, 2007, Accepted.

研究会・全国大会等

1. 櫛田和貴，西村義隆，土肥浩，石塚満，竹内誉羽，辻野広司，"ヒューマノイドロボット用マルチモーダルプレゼンテーション記述言語 MPML-HR の開発," 第 67 回情報処理学会全国大会論文集，2005.

2. 西村義隆，櫛田和貴，土肥浩，石塚満，竹内誉羽，辻野広司，"マルチモーダルプレゼンテーション記述言語 MPML のヒューマノイドへの拡張とその心理学的評価," 電子情報通信学会技術研究報告，HCS2005-21, pp.5-10, 2005.

3. 竹内誉羽，中野幹生，辻野広司，星野厚，加藤和彦，西村義隆，櫛田和貴，土肥浩，石塚満，"ロボットの共生的対話システムとマルチモーダルな表現システムの開発と評価," 電子情報通信学会技術研究報告，NLC2005-29，pp.29-34，2005.

4. 櫛田和貴，西村義隆，土肥浩，石塚満，竹内誉羽，辻野広司，"ヒューマノイドロボットとアニメキャラクタによる共同プレゼンテーション MPML-HR Ver2.0," 電子情報通信学会総合大会予稿集，A-16-16, 2006.

5. 西村義隆，竹内誉羽，櫛田和貴，中野幹生，辻野広司，土肥浩，石塚満，“ヒューマノイドロボットとスクリーンエージェントのプレゼンテーションにおける視聴者に与える印象の比較,” Joint Agent Workshops and Symposium(JAWS-06) 予稿集, 2006. (学生奨励賞受賞)

6. 西村義隆，中臺一博，中野幹生，辻野広司，石塚満，“MFT を用いたロボットの動作音に頑健な音声認識手法の提案,” SIG-Challenge 第 24 回 AI チャレンジ研究会予稿集, pp.1-8, 2006.

7. 西村義隆，簑津真一郎，土肥浩，石塚満，中野幹生，船越孝太郎，竹内誉羽，長谷川雄二，辻野広司，“インタラクション機能を有するプレゼンテーション記述言語の開発,” 電子情報通信学会技術研究報告，HCS2006-56, pp.43-48, 2006.

8. 簑津真一郎，西村義隆，土肥浩，石塚満，中野幹生，船越孝太郎，竹内誉羽，長谷川雄二，辻野広司，“プレゼンテーション記述言語 MPML-HR における音声インタラクション機能,” 第 69 回情報処理学会全国大会論文集，投稿中.

付 録A　MPML-HRのシステムプログラム

本章では，図 3.5 における Distribution server のプログラム例について示す．なお，ここで示すプログラムは表 A.1 にまとめるインタフェースに従う．Distribution server は Browser, Voice, ASIMO の各サーバへここに示されるコマンドを送信することにより動作等を実行させ，終了時には ACK 信号を受信する．なお，ここに示すインタフェースは，実装に用いたインタフェースと同一ではない．

表 A.1: 各サーバの仮想インタフェース

サーバ	コマンド	動作
Browser	BROWSER.OPEN,*slide*	*slide* をスクリーンに表示
Voice	EMOTION,*emotion* SPEAK,*utterance*	*emotion* の感情で *utterance* を発話
ASIMO	PLAY,*action*	*action* のジェスチャを実行
ASIMO	MOVETO x y z	(x,y) 座標へ移動し，身体を z 回転
ASIMO	POINTTO x y	スクリーンの (x,y) を指示
ASIMO	HEAD *pan tilt*	頭を横に *pan*，縦に *tilt* 動かす

図 A.1 から図 A.4 にプログラム例を示す．このプログラムでは初めに，ソケット通信により Browser, Voice server, ASIMO control server の各サーバに接続を行う．また，中間記述ファイルの読み込みを行う．中間記述ファイルの命令系列に，同期，自律動作等の処理を加え，各サーバに命令を送信する．命令送信後は，完了を示す ACK 信号が返ってくるまで待つ．なお，自律動作の実行のため，スレッドを用いたプログラムとなっている．中間記述ファイルの命令系列がなくなると，ソケットを閉じ，終了処理を行う．

```
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <sys/socket.h>
#include <sys/time.h>
#include <netinet/in.h>
#include <netdb.h>
#include <pthread.h>

#define PORT_BR (in_port_t)70001
#define PORT_ROBOT (in_port_t)70002
#define PORT_VOICE (in_port_t)70003
#define BUF_LEN 512
#define COM_LEN 512
#define DEG_PAN 45
#define DEG_TILT 15

/*命令を格納する構造体*/
struct MPMLINT{
  char emotion[COM_LEN]="neutral";
  char speak[COM_LEN];
  char act[COM_LEN];
  char page[COM_LEN];
  int emotion_len;
  int speak_len;
  int act_len;
  int page_len;
  boolean state_br=0;
  boolean state_robot=0;
  boolean state_voice=0;
}

struct MPMLINT mpml;
char nextcommand[COM_LEN];
FILE *fp;

main()
{
  struct hostent *server_ent_br, *server_ent_robot, *server_ent_voice;
  struct sockaddr_in server_br, server_robot, server_voice;
  int soc_br, soc_robot, soc_voice;
  char hostname_br[]="browser", hostname_robot="controller", hostname_voice="voice";
  char buf[BUF_LEN];
  pthread_t thread1;

/*-----------------------------------------------*/
/*              初期化処理                       */
/*-----------------------------------------------*/
  /*ホスト名からアドレスを取得*/
  if((server_ent_br=gethostbyname(hostname_br))==NULL{
    perror("get hostname error¥n");
    exit(1);
  }
  if((server_ent_robot=gethostbyname(hostname_robot))==NULL{
    perror("get hostname error¥n");
    exit(1);
  }
  if((server_ent_voice=gethostbyname(hostname_voice))==NULL{
    perror("get hostname error¥n");
    exit(1);
  }
```

図 A.1: Distribution server プログラム例 1

```
/*相手アドレスを構造体に格納*/
  memset((char *)&server_br, 0, sizeof(server_br));
  server_br.sin_family=AF_INET;
  server_br.sin_port=htons(PORT_BR);
  memcpy((char *)&server_br.sin_addr, server_ent_br->h_addr, sever_ent_br->hlength);
  memset((char *)&server_robot, 0, sizeof(server_robot));
  server_robot.sin_family=AF_INET;
  server_robot.sin_port=htons(PORT_ROBOT);
  memcpy((char *)&server_robot.sin_addr, server_ent_robot->h_addr, sever_ent_robot->hlength);
  memset((char *)&server_voice, 0, sizeof(server_voice));
  server_voice.sin_family=AF_INET;
  server_voice.sin_port=htons(PORT_VOICE);
  memcpy((char *)&server_voice.sin_addr, server_ent_voice->h_addr, sever_ent_voice->hlength);


  /*ソケット作成*/
  if((soc_br=socket(AF_INET, SOCK_STREAM, 0))<0){
    perror("socket error¥n");
    exit(1);
  }
  if((soc_robot=socket(AF_INET, SOCK_STREAM, 0))<0){
    perror("socket error¥n");
    exit(1);
  }
  if((soc_voice=socket(AF_INET, SOCK_STREAM, 0))<0){
    perror("socket error¥n");
    exit(1);
  }

  /*サーバに接続*/
  if(connect(soc_br, (struct sockaddr *)&server_br, sizeof(server_br))==-1){
    perror("connection error¥n");
    exit(1);
  }
  if(connect(soc_robot, (struct sockaddr *)&server_robot, sizeof(server_robot))==-1){
    perror("connection error¥n");
    exit(1);
  }
  if(connect(soc_voice, (struct sockaddr *)&server_voice, sizeof(server_voice))==-1){
    perror("connection error¥n");
    exit(1);
  }

  /*中間記述ファイルのオープン*/
  if((fp=fopen("mpml_middle.txt", "r")) == NULL){
    perror("mpml middle file error¥n");
    exit(1);
  }
```

図 A.2: Distribution server プログラム例 2

```
/*-------------------------------------------------*/
/*              コンテンツの実行                   */
/*-------------------------------------------------*/
  do{
    int n;

    commandset();

    if(mpml.state_br==1){
      write(soc_br, mpml.page, mpml.page_len);
      n=read(soc_br, buf, BUF_LEN);
      mpml.state_br=0;
    }else if(mpml.state_robot==1 && mpml.state_voice==0){
      write(soc_robot, mpml.act, mpml.act_len);
      n=read(soc_robot, buf, BUF_LEN);
      mpml.state_robot=0;
    }else if(mpml.state_voice==1 && mpml.state_robot==0){
      pthread_create( &thread1, NULL, (void *)headmotion);
      write(soc_voice, mpml.emotion, mpml.emotion_len);
      write(soc_voice, mpml.speak, mpml.speak_len);
      n=read(soc_voice, buf, BUF_LEN);
      mpml.state_voice=0;
    }else{
      write(soc_voice, mpml.emotion, mpml.emotion_len);
      write(soc_voice, mpml.speak, mpml.speak_len);
      write(soc_robot, mpml.act, mpml.act_len);
      n=read(soc_robot, buf, BUF_LEN);
      mpml.state_robot=0;
      pthread_create( &thread1, NULL, (void *)headmotion);
      n=read(soc_voice, buf, BUF_LEN);
      mpml.state_voice=0;
    }
    int n;
    n=read(soc_br, buf, BUF_LEN);/*ソケット*/
    write(1, buf, n);/*標準出力*/
    n=read(0, buf, BUF_LEN);/*標準入力*/
    write(soc_br, buf, n);/*ソケット*/
  }
  while(nextcommand!="END");

/*-------------------------------------------------*/
/*                 終了処理                        */
/*-------------------------------------------------*/
  close(soc_br);
  close(soc_robot);
  close(soc_voice);
  fclose(fp);

}
```

図 A.3: Distribution server プログラム例 3

```
/*-------------------------------------------*/
/*        構造体に命令をセットする関数                 */
/*-------------------------------------------*/
void commandset(){
  if(nextcommand==""){
    if(fgets(nextcommand, COM_LEN, fp) == NULL){
      sprintf(nextcommand,"%s","END");
    }
  }
  while(1){
    if(mpml.state_br==1)break;
    else if(mpml.state_robot==1 && strcmp(nextcommand, "END")==0)break;
    else if(mpml.state_robot==1 && nextcommand[0]=='M')break;
    else if(mpml.state_voice==1 && strcmp(nextcommand, "END")==0)break;
    else if(mpml.state_voice==1 && nextcommand[0]=='S')break;

    if(nextcommand[0]=='B'){
      sprintf(mpml.page,"%s",nextcommand);
      mpml.state_br=1;
      mpml.page_len=lengthset(mpml.page);
    }else if(nextcommand[0]=='E'){
      sprintf(mpml.emotion,"%s",nextcommand);
      mpml.emotion_len=lengthset(mpml.emotion);
    }else if(nextcommand[0]=='S'){
      sprintf(mpml.speak,"%s",nextcommand);
      mpml.state_voice=1;
      mpml.speak_len=lengthset(mpml.speak);
    }else if(nextcommand[0]=='M' || nextcommand[0]=='P'){
      sprintf(mpml.act,"%s",nextcommand);
      mpml.state_robot=1;
      mpml.act_len=lengthset(mpml.act);
    }

    if(fgets(nextcommand, COM_LEN, fp) == NULL){
      sprintf(nextcommand,"%s","END");
    }
  }

}



int lengthset(char *in_string){
  int i=0
  while(in_string[i]!='¥0'){
    i++
  }
  return(i+1);
}

void headmotion(){
  int pan, tilt;

  sleep(100);
  srand((unsigned)time(NULL));
  while(mpml.state_voice==1){
    pan=(rand()%(DEG_PAN*2+1))-DEG_PAN;
    tilt=(rand()%(DEG_TILT*2+1))-DEG_TILT;
    sprintf(mpml.act,"HEAD %d,%d",pan,tilt);
    mpml.state_robot=1;
    mpml.act_len=lengthset(mpml.act);
    write(soc_robot, mpml.act, mpml.act_len);
    n=read(soc_robot, buf, BUF_LEN);
    mpml.state_robot=0;
  }
}
```

図 A.4: Distribution server プログラム例 4

付 録B　MFTを用いた音声認識手法の詳細な実験結果

本章では，5.4 節における実験結果の詳細なデータを表 B.1 から表 B.4 に示す．

また，提案手法で用いた SS の代わりに Ephraim らの提案する方法 [111] を用いて雑音除去処理を行った結果についてまとめる．Ephraim らの手法を用いた実験条件を表 B.5 に示す．ここで，大文字で示されている **A**, **B**, **C**, **M**, **N**, **O**, **P**, **C'** および **C"** は従来手法の条件であるため，表 5.1 と同様である．実験結果を表 B.6 から表 B.9 にまとめる．Ephraim らの手法では，スペクトルパワーに基づいて適応的に音声存在確率を抽出し，この音声存在確率に基づいて入力信号に重畳された雑音が軽減される．この手法では，スペクトルの連続性が考慮されているため，従来の SS よりも歪みを抑えることができる．実際に，雑音除去後の音声を聞いたところ，雑音が除去され，クリアな音声信号になっている印象を受けた．しかし，音声認識デコーダに入力した際には SS と比較して認識性能が劣る結果となった．音声認識では，入力信号の SNR が高い場合でも，歪み具合などにより正規化時にクリーンな音声信号との差が拡がってしまう場合がある．このような理由から，Ephraim らの手法では，SS と比較して認識性能が低くなったものと考えられる．

表 B.1: SS を用いた提案手法の結果 1

	50 cm					100 cm				
Condition	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Motor noise	59.19	83.34	83.34	67.60	84.26	40.44	74.23	74.23	60.11	79.56
Right hand (1)	50.93	81.25	79.63	64.74	82.10	30.71	67.67	68.60	57.87	73.15
Right hand (2)	45.45	77.63	73.84	58.10	75.46	26.00	64.97	60.80	47.84	65.97
Right hand (3)	51.32	79.63	78.55	63.89	80.25	32.56	67.75	69.60	55.02	70.45
Right hand (4)	52.70	82.64	79.94	64.51	81.02	32.95	69.14	69.83	55.40	70.53
Right hand (5)	50.70	81.02	81.17	63.35	82.02	31.10	68.52	69.99	54.63	70.99
Left hand (1)	50.31	78.32	79.79	62.04	78.32	30.48	65.05	65.36	49.54	66.05
Left hand (2)	44.60	74.69	74.38	56.02	73.46	24.31	59.88	60.27	45.53	58.88
Left hand (3)	51.55	80.25	79.32	64.74	79.17	32.72	68.21	69.29	55.40	68.14
Left hand (4)	47.45	74.62	74.85	50.93	63.20	29.32	60.42	60.81	36.04	49.92
Left hand (5)	52.01	79.40	79.94	62.04	78.47	32.72	65.43	66.05	52.47	67.06
Both hands (1)	45.07	76.70	75.08	59.11	77.86	25.31	61.19	61.50	46.53	63.43
Both hands (2)	44.29	74.54	74.31	57.57	75.70	25.47	59.18	59.88	48.15	61.96
Both hands (3)	43.60	73.61	74.46	56.10	74.23	25.62	58.10	60.65	45.76	60.34
Both hands (4)	43.37	74.08	74.38	58.88	75.08	24.62	59.88	57.87	47.30	62.35
Both hands (5)	45.91	76.93	75.47	60.19	77.70	26.01	61.58	62.50	50.16	63.58
Head (1)	45.84	63.89	67.21	40.97	54.94	23.69	50.23	51.39	30.40	43.98
Head (2)	53.25	73.23	76.16	52.86	66.98	34.57	59.80	64.12	44.99	57.25
Head (3)	50.39	74.31	74.92	50.70	67.29	31.79	62.04	61.96	40.20	55.79
Head (4)	29.94	44.91	50.16	26.39	31.87	17.83	33.88	38.27	20.06	27.55
Head (5)	30.79	46.22	52.01	29.25	37.58	17.90	34.88	40.97	24.00	30.95
Head and Hands (1)	51.24	74.38	77.24	51.62	64.59	33.57	61.81	65.51	39.12	55.17
Head and Hands (2)	51.31	66.59	70.99	41.98	55.56	31.18	55.79	59.03	31.33	45.68
Head and Hands (3)	37.50	63.81	66.67	41.59	55.10	19.53	48.85	51.85	29.17	38.81
Head and Hands (4)	46.38	77.70	77.47	61.81	77.62	27.24	63.04	63.35	49.54	64.51
Head and Hands (5)	47.38	73.30	73.46	45.30	56.02	27.47	60.27	59.57	34.19	43.83
Walking Motion (1)	50.00	74.00	73.23	49.46	67.75	29.94	62.19	64.05	41.67	59.88
Walking Motion (2)	46.91	71.30	71.22	43.21	64.12	29.40	56.95	60.65	32.80	53.24
Walking Motion (3)	49.38	74.08	74.39	50.62	70.84	30.32	64.05	64.74	42.75	59.96
Walking Motion (4)	50.78	75.85	77.47	53.32	74.16	32.33	65.51	66.82	45.37	63.81
Walking Motion (5)	52.01	76.08	75.77	52.62	74.77	34.11	67.59	66.67	43.21	64.05
Walking Motion (6)	46.99	74.85	74.70	50.39	68.75	29.63	61.81	62.35	41.90	56.95
Walking Motion (7)	49.85	75.24	76.62	54.17	70.76	32.41	66.51	66.59	46.76	61.11
Walking Motion (8)	50.85	76.62	72.92	54.09	75.00	33.41	67.75	63.97	44.76	62.04
	150 cm					200 cm				
Condition	**A**	**B**	**C**	**D**	**E**	**A**	**B**	**C**	**D**	**E**
Motor noise	27.86	64.20	64.20	54.02	71.61	19.83	54.01	54.01	49.93	61.96
Right hand (1)	19.68	56.02	56.33	48.69	63.20	12.27	45.60	46.38	43.36	53.17
Right hand (2)	15.36	50.70	49.54	38.97	55.40	9.57	41.44	38.89	31.79	44.29
Right hand (3)	21.37	56.56	56.49	47.30	61.73	12.89	46.53	46.92	41.82	51.86
Right hand (4)	20.99	56.02	57.10	45.37	60.50	12.50	46.30	49.23	41.05	48.46
Right hand (5)	18.52	55.48	57.95	47.46	59.96	11.27	46.61	46.76	42.59	49.69
Left hand (1)	18.83	54.55	53.32	41.20	54.48	10.73	42.29	44.22	35.19	45.84
Left hand (2)	13.81	47.38	47.46	35.80	46.92	6.71	36.65	37.89	30.56	38.12
Left hand (3)	21.45	55.25	54.63	46.53	58.80	12.74	44.76	47.15	40.05	49.54
Left hand (4)	16.82	50.00	50.31	27.63	38.74	10.34	39.51	39.67	21.07	27.86
Left hand (5)	19.91	54.01	52.55	41.82	55.94	11.19	43.75	44.14	36.50	46.76
Both hands (1)	15.20	51.01	49.08	37.50	53.40	7.79	39.05	38.50	31.72	43.06
Both hands (2)	15.44	49.00	47.46	39.82	51.78	8.57	38.82	40.20	31.87	42.67
Both hands (3)	16.52	45.99	48.38	36.50	49.46	8.57	35.42	39.43	30.56	38.74
Both hands (4)	15.13	46.61	47.15	37.27	51.24	7.56	36.35	36.96	31.71	41.51
Both hands (5)	15.59	49.00	50.00	40.82	53.86	8.41	39.35	41.36	34.80	44.68
Head (1)	13.74	36.89	40.13	22.61	33.95	7.02	26.78	30.63	16.59	23.84
Head (2)	21.68	51.62	51.47	37.58	48.23	14.81	42.05	42.60	32.57	42.28
Head (3)	18.45	51.01	50.62	33.26	47.76	11.81	41.13	42.29	26.93	39.74
Head (4)	11.81	24.62	30.87	14.82	21.69	7.26	19.83	25.46	12.58	17.52
Head (5)	10.57	28.47	32.87	18.91	25.85	6.56	21.22	26.39	16.36	20.61
Head and Hands (1)	20.84	52.55	53.86	33.03	45.99	14.20	41.67	45.14	27.93	37.97
Head and Hands (2)	19.91	44.14	48.46	23.54	36.81	13.66	35.88	41.05	18.91	29.86
Head and Hands (3)	9.11	38.04	39.66	20.84	28.40	4.94	27.39	30.41	15.97	21.68
Head and Hands (4)	15.21	50.23	50.54	41.98	54.09	8.49	38.97	42.59	35.42	43.98
Head and Hands (5)	16.67	47.77	46.76	25.54	33.80	9.11	37.27	38.58	21.22	24.77
Walking Motion (1)	19.29	52.09	54.32	37.11	50.16	12.12	41.13	42.13	30.87	40.43
Walking Motion (2)	18.68	45.68	48.23	27.78	43.98	11.58	36.96	38.04	21.92	34.49
Walking Motion (3)	19.29	51.47	51.00	34.11	50.31	12.19	42.21	41.75	30.25	41.28
Walking Motion (4)	19.91	55.10	54.09	38.20	53.17	12.35	45.06	44.06	32.64	44.99
Walking Motion (5)	20.91	55.09	54.02	37.66	54.56	12.42	45.99	45.60	32.33	45.30
Walking Motion (6)	17.83	51.55	51.47	33.42	47.22	10.50	40.90	40.66	28.32	38.35
Walking Motion (7)	19.91	54.40	55.02	40.28	51.01	12.97	44.37	45.99	34.11	44.06
Walking Motion (8)	20.45	55.71	51.32	36.03	51.62	12.89	43.37	42.60	28.86	41.90

表 B.2: SS を用いた提案手法の結果 2

	50 cm					100 cm				
Condition	**F**	**G**	**H**	**I**	**J**	**F**	**G**	**H**	**I**	**J**
Motor noise	82.26	84.42	85.65	84.34	85.27	60.11	79.56	78.94	76.39	79.48
Right hand (1)	80.71	82.87	82.33	79.25	83.18	57.87	73.15	73.23	66.67	75.78
Right hand (2)	73.77	75.54	77.09	73.62	80.87	47.84	65.97	62.89	57.49	72.54
Right hand (3)	79.55	80.79	80.94	77.24	81.87	55.02	70.45	70.22	63.35	76.54
Right hand (4)	81.64	83.64	83.80	81.18	83.41	55.40	70.53	72.30	68.98	75.23
Right hand (5)	79.63	81.33	81.87	78.94	83.80	54.63	70.99	71.30	65.82	75.85
Left hand (1)	78.71	81.18	83.65	80.17	81.95	49.54	66.05	70.22	65.05	73.23
Left hand (2)	73.30	75.62	77.01	72.61	78.55	45.53	58.88	61.04	59.03	68.29
Left hand (3)	80.09	82.33	82.95	80.25	83.72	55.40	68.14	71.14	68.75	74.62
Left hand (4)	63.97	68.52	71.61	70.29	78.78	36.04	49.92	55.33	53.16	68.75
Left hand (5)	78.25	82.72	82.10	79.48	82.03	52.47	67.06	69.68	64.74	74.00
Both hands (1)	76.78	78.47	79.17	76.55	80.79	46.53	63.43	66.28	61.88	71.84
Both hands (2)	76.00	77.93	77.55	73.31	78.94	48.15	61.96	62.89	57.79	69.14
Both hands (3)	74.69	76.39	77.01	74.08	78.40	45.76	60.34	62.35	58.65	67.67
Both hands (4)	75.39	77.62	78.55	75.24	78.78	47.30	62.35	63.20	58.64	69.83
Both hands (5)	77.63	78.16	79.09	75.85	81.41	50.16	63.58	65.67	60.65	71.22
Head (1)	58.49	62.50	64.74	66.05	70.76	30.40	43.98	47.61	50.54	58.57
Head (2)	69.83	73.84	76.31	76.93	75.16	44.99	57.25	62.89	63.97	65.97
Head (3)	69.29	74.54	77.32	78.55	75.93	40.20	55.79	65.13	65.36	67.21
Head (4)	31.10	36.88	39.58	47.53	51.39	20.06	27.55	28.40	32.95	41.67
Head (5)	37.50	41.36	41.75	45.30	50.23	24.00	30.95	31.79	33.03	42.83
Head and Hands (1)	68.21	75.62	78.86	79.48	77.32	39.12	55.17	64.20	65.82	67.59
Head and Hands (2)	60.50	67.52	70.45	76.47	71.22	31.33	45.68	56.18	60.03	58.49
Head and Hands (3)	58.49	64.28	66.28	64.20	70.61	29.17	38.81	50.16	46.76	56.33
Head and Hands (4)	78.40	80.17	80.86	76.16	81.56	49.54	64.51	66.90	60.58	70.76
Head and Hands (5)	56.10	62.58	64.43	65.20	73.61	34.19	43.83	49.00	48.61	62.43
Walking Motion (1)	70.99	67.67	64.58	55.79	76.39	41.67	59.88	48.92	40.28	64.43
Walking Motion (2)	65.28	63.04	59.03	51.01	73.61	32.80	53.24	43.60	36.19	60.81
Walking Motion (3)	74.38	74.85	72.30	62.89	77.62	42.75	59.96	57.10	47.46	68.36
Walking Motion (4)	74.08	73.92	71.22	60.42	79.17	45.37	63.81	55.48	45.99	69.06
Walking Motion (5)	73.46	77.78	78.01	74.62	81.10	43.21	64.05	65.59	58.95	71.69
Walking Motion (6)	71.92	71.99	69.76	61.66	76.93	41.90	56.95	54.94	45.37	67.13
Walking Motion (7)	74.70	74.85	74.08	66.75	79.17	46.76	61.11	58.72	51.39	70.22
Walking Motion (8)	72.23	76.47	76.55	71.99	79.94	44.76	62.04	63.43	57.26	72.15
	150 cm					200 cm				
Condition	**F**	**G**	**H**	**I**	**J**	**F**	**G**	**H**	**I**	**J**
Motor noise	71.22	70.60	69.29	63.74	73.69	49.93	61.96	59.80	53.94	66.75
Right hand (1)	63.97	61.81	61.42	55.25	67.75	43.36	53.17	51.70	47.07	59.73
Right hand (2)	53.32	48.23	49.08	45.83	62.96	31.79	44.29	38.89	34.26	52.01
Right hand (3)	62.04	58.80	58.57	52.94	66.75	41.82	51.86	48.69	44.99	59.11
Right hand (4)	60.58	58.64	60.50	55.79	67.67	41.05	48.46	50.62	45.68	60.03
Right hand (5)	60.34	58.34	59.42	55.02	66.51	42.59	49.69	48.69	44.37	57.80
Left hand (1)	57.95	56.10	57.18	54.71	62.89	35.19	45.84	47.76	45.76	54.86
Left hand (2)	47.92	46.45	48.30	47.61	56.72	30.56	38.12	38.43	36.81	47.23
Left hand (3)	60.88	58.72	58.87	54.86	65.05	40.05	49.54	50.00	46.15	58.95
Left hand (4)	38.35	41.82	42.90	42.83	59.73	21.07	27.86	32.33	31.33	50.85
Left hand (5)	57.41	56.49	57.80	53.78	65.97	36.50	46.76	48.46	45.53	55.71
Both hands (1)	53.71	50.31	53.78	50.46	59.80	31.72	43.06	43.91	41.44	51.70
Both hands (2)	52.47	48.15	52.32	47.53	59.65	31.87	42.67	40.82	37.81	50.54
Both hands (3)	50.62	48.54	50.16	45.99	57.18	30.56	38.74	38.89	37.12	48.38
Both hands (4)	52.09	49.31	51.39	48.61	56.41	31.71	41.51	41.05	39.28	49.16
Both hands (5)	56.33	52.01	52.63	49.39	60.57	34.80	44.68	43.21	40.74	52.24
Head (1)	32.48	32.80	35.96	37.43	45.99	16.59	23.84	23.84	25.16	37.35
Head (2)	49.31	51.62	53.71	51.62	57.64	32.57	42.28	44.75	42.91	50.39
Head (3)	49.15	51.93	53.47	53.48	56.56	26.93	39.74	43.99	44.60	51.24
Head (4)	17.21	20.45	20.76	24.31	32.56	12.58	17.52	16.13	17.98	27.16
Head (5)	25.47	24.77	24.85	25.39	36.88	16.36	20.61	21.38	21.22	30.94
Head and Hands (1)	46.37	50.16	53.09	53.32	58.65	27.93	37.97	43.67	43.52	50.00
Head and Hands (2)	35.65	42.29	44.14	47.53	49.16	18.91	29.86	36.04	39.67	42.44
Head and Hands (3)	31.72	32.64	35.57	36.04	44.37	15.97	21.68	25.23	24.31	36.04
Head and Hands (4)	56.02	51.62	53.71	50.16	62.04	35.42	43.98	43.37	41.75	53.32
Head and Hands (5)	32.10	34.57	36.73	38.04	52.70	21.22	24.77	25.85	27.63	45.06
Walking Motion (1)	48.77	37.12	38.82	30.56	53.63	30.87	40.43	26.16	17.06	45.14
Walking Motion (2)	40.59	31.56	32.80	24.54	49.23	21.92	34.49	21.38	13.74	40.36
Walking Motion (3)	52.32	43.52	45.68	37.42	56.87	30.25	41.28	34.96	25.23	48.31
Walking Motion (4)	52.63	43.98	44.45	35.73	56.87	32.64	44.99	33.65	24.39	47.30
Walking Motion (5)	53.63	51.70	51.55	48.08	62.89	32.33	45.30	44.29	37.58	53.70
Walking Motion (6)	46.99	41.21	42.44	35.11	54.87	28.32	38.35	31.17	24.31	47.53
Walking Motion (7)	53.32	45.84	47.38	40.67	58.72	34.11	44.06	36.73	29.40	50.77
Walking Motion (8)	52.09	49.08	50.54	45.45	61.50	28.86	41.90	41.13	33.18	52.47

表 B.3: SS を用いた提案手法の結果 3

	50 cm					100 cm				
Condition	**K**	**L**	**M**	**N**	**O**	**K**	**L**	**M**	**N**	**O**
Motor noise	83.34	83.41	57.56	67.90	57.72	41.13	60.03	41.74	79.48	80.79
Right hand (1)	81.64	82.72	51.01	66.83	51.86	35.11	54.25	36.58	76.85	77.86
Right hand (2)	79.79	81.72	41.44	55.25	41.13	27.39	42.75	27.47	71.61	74.23
Right hand (3)	82.56	81.64	47.61	58.80	47.23	31.72	46.53	32.56	76.39	76.47
Right hand (4)	83.10	82.80	51.31	66.21	53.01	35.19	53.63	35.57	77.01	76.93
Right hand (5)	82.18	82.41	50.93	62.65	50.77	34.18	52.16	34.80	76.00	76.70
Left hand (1)	82.10	81.95	48.77	60.04	50.00	34.26	45.99	35.34	75.47	76.62
Left hand (2)	78.46	79.48	46.61	57.80	47.22	29.63	45.91	31.18	69.37	71.99
Left hand (3)	84.01	82.41	51.70	62.66	51.70	33.49	49.85	35.81	76.01	76.93
Left hand (4)	77.55	79.48	40.59	47.76	40.98	26.16	34.19	26.93	68.60	71.22
Left hand (5)	81.48	82.41	50.62	61.35	52.01	35.34	49.54	37.42	74.93	75.54
Both hands (1)	79.78	80.33	45.14	54.86	46.45	30.64	43.21	32.18	72.53	73.77
Both hands (2)	78.94	80.33	42.98	54.94	43.91	27.86	42.05	29.79	69.75	71.61
Both hands (3)	79.86	80.40	41.82	50.08	41.36	27.40	37.74	28.24	69.29	70.76
Both hands (4)	79.17	79.79	42.44	50.39	42.82	27.01	39.59	27.39	69.44	71.30
Both hands (5)	80.17	80.33	45.53	57.33	46.76	29.48	44.83	30.56	71.99	73.46
Head (1)	70.30	74.15	41.75	41.59	41.13	25.54	30.40	27.16	58.10	63.27
Head (2)	74.54	78.01	47.76	53.09	47.69	33.80	41.90	34.65	65.90	72.77
Head (3)	74.00	78.48	39.74	44.60	36.81	26.70	33.57	24.92	66.13	73.07
Head (4)	51.08	56.87	22.61	26.16	22.15	15.13	18.68	14.82	40.59	47.76
Head (5)	50.24	57.41	27.17	30.87	28.09	18.06	23.92	18.14	44.45	48.23
Head and Hands (1)	76.16	80.63	46.07	51.16	45.45	31.10	39.51	31.95	68.90	73.84
Head and Hands (2)	70.06	77.47	44.06	38.73	42.21	29.79	29.56	29.71	58.57	68.91
Head and Hands (3)	68.75	75.24	35.57	39.05	35.42	19.45	26.47	20.45	57.18	61.89
Head and Hands (4)	80.71	80.87	48.62	58.72	48.77	33.65	47.92	33.18	72.92	73.31
Head and Hands (5)	74.15	77.86	41.82	46.15	40.82	27.39	32.41	25.62	63.28	69.14
Walking Motion (1)	75.47	76.78	42.67	50.23	42.90	31.87	41.67	32.33	65.74	64.90
Walking Motion (2)	73.23	75.08	42.36	44.52	41.44	30.25	33.87	31.64	61.88	64.74
Walking Motion (3)	77.32	79.32	44.14	46.38	43.52	32.87	39.97	32.02	68.83	69.14
Walking Motion (4)	78.86	78.24	45.60	53.71	45.14	34.19	44.60	34.34	70.99	70.29
Walking Motion (5)	78.48	80.94	45.68	48.23	45.61	33.34	39.12	33.11	72.30	75.47
Walking Motion (6)	76.70	78.17	42.98	45.76	43.29	31.02	36.96	31.02	68.13	67.75
Walking Motion (7)	77.55	79.56	46.92	51.86	46.38	32.72	42.75	32.95	71.22	72.30
Walking Motion (8)	79.40	80.17	46.84	49.00	46.22	34.26	40.21	34.57	71.92	73.77
	150 cm					200 cm				
Condition	**K**	**L**	**M**	**N**	**O**	**K**	**L**	**M**	**N**	**O**
Motor noise	73.92	73.85	30.56	50.23	32.72	25.39	46.91	28.71	69.14	68.44
Right hand (1)	68.60	70.14	26.39	45.14	28.32	20.99	39.51	23.00	62.66	61.89
Right hand (2)	62.19	64.12	19.29	34.18	20.14	12.96	26.39	15.13	54.40	57.49
Right hand (3)	67.52	66.51	24.00	38.51	23.92	18.37	33.49	20.53	60.73	60.80
Right hand (4)	68.37	68.60	25.46	42.67	27.16	19.14	37.04	21.76	61.11	62.35
Right hand (5)	67.44	68.52	24.38	41.67	27.09	19.06	35.04	21.61	58.72	61.65
Left hand (1)	64.43	68.06	25.23	37.89	26.62	18.44	31.49	20.76	55.86	59.65
Left hand (2)	57.48	59.42	19.83	34.34	21.53	13.43	30.63	16.05	49.15	50.16
Left hand (3)	67.52	67.83	23.85	41.98	25.62	19.76	35.26	22.30	59.57	61.88
Left hand (4)	60.19	61.88	17.13	25.24	17.98	12.73	20.14	13.20	52.39	55.79
Left hand (5)	64.82	68.13	25.39	39.74	27.55	19.37	33.03	23.07	55.87	59.57
Both hands (1)	61.42	64.66	21.15	35.34	22.92	14.66	28.40	17.44	53.48	55.48
Both hands (2)	60.50	62.50	19.45	33.95	21.45	14.43	27.70	15.67	51.63	54.94
Both hands (3)	58.34	60.73	18.06	28.48	18.37	13.04	22.53	13.97	48.92	52.94
Both hands (4)	58.03	60.57	17.52	31.79	19.06	12.89	25.00	14.66	50.39	52.47
Both hands (5)	61.81	63.58	21.92	36.42	23.15	14.89	29.87	17.59	53.40	55.56
Head (1)	47.61	52.24	17.59	21.92	17.83	12.58	16.75	12.97	37.96	43.98
Head (2)	57.79	64.20	22.76	33.42	23.69	18.75	28.48	19.76	52.93	57.72
Head (3)	57.57	63.43	18.75	25.31	16.29	13.97	21.61	12.81	49.93	57.64
Head (4)	31.10	40.28	11.96	13.28	11.65	8.72	11.35	9.11	27.09	34.42
Head (5)	36.66	41.20	12.42	18.91	11.73	10.26	15.51	10.49	31.72	36.27
Head and Hands (1)	58.64	64.82	23.77	31.87	24.00	18.91	26.70	19.60	51.55	58.95
Head and Hands (2)	49.46	59.19	22.69	23.07	22.30	17.13	19.06	18.44	43.29	53.17
Head and Hands (3)	44.91	51.08	11.88	18.90	12.89	7.80	15.59	9.03	35.26	43.13
Head and Hands (4)	63.74	63.66	22.07	39.58	23.92	17.21	32.26	19.06	54.63	56.64
Head and Hands (5)	52.78	59.41	16.44	23.46	18.44	13.43	18.99	13.58	46.07	53.01
Walking Motion (1)	53.40	53.55	25.08	35.42	25.24	19.21	29.94	20.84	44.21	45.91
Walking Motion (2)	51.01	52.39	21.68	28.70	23.54	17.59	22.84	18.91	39.66	44.53
Walking Motion (3)	58.87	57.95	24.31	29.94	27.01	18.37	26.78	20.14	49.69	50.70
Walking Motion (4)	58.80	57.87	24.77	36.65	25.85	17.59	30.17	20.68	49.54	48.77
Walking Motion (5)	64.36	64.89	25.08	32.72	26.55	19.68	28.78	21.61	55.25	57.57
Walking Motion (6)	55.87	58.26	22.46	29.02	24.31	17.90	24.92	19.14	46.69	48.54
Walking Motion (7)	60.88	60.96	24.62	34.11	25.39	19.52	28.32	21.69	50.69	52.32
Walking Motion (8)	63.12	65.05	24.46	31.56	26.16	19.60	27.01	20.99	53.78	55.48

表 B.4: SS を用いた提案手法の結果 4

	50 cm					100 cm				
Condition	**P**	**C'**	**J'**	**C"**	**J"**	**P**	**C'**	**J'**	**C"**	**J"**
Motor noise	65.90	90.78	90.00	96.74	94.65	57.49	84.03	85.04	93.95	92.64
Right hand (1)	64.66	87.45	89.23	93.18	92.48	52.39	77.37	83.65	83.95	88.53
Right hand (2)	54.17	81.17	86.05	96.05	94.42	40.51	66.67	78.76	89.85	90.08
Right hand (3)	55.63	85.51	88.14	94.65	93.72	45.53	75.35	81.55	88.07	90.31
Right hand (4)	64.66	88.06	88.61	94.27	91.86	52.94	77.29	82.95	88.68	85.51
Right hand (5)	60.11	87.75	88.68	93.10	92.17	48.92	76.51	82.64	82.10	84.65
Left hand (1)	57.18	86.75	87.44	93.34	93.18	43.98	71.86	80.23	86.44	87.83
Left hand (2)	57.18	82.25	84.19	95.50	93.03	45.60	65.58	73.96	89.46	90.16
Left hand (3)	62.35	85.97	88.14	94.96	94.03	48.07	75.66	79.93	88.37	89.85
Left hand (4)	46.99	81.17	83.33	96.51	94.96	32.87	67.05	71.79	90.23	91.01
Left hand (5)	59.88	87.13	87.76	93.57	90.54	47.61	72.87	81.01	84.88	82.40
Both hands (1)	55.87	82.79	86.20	92.95	92.56	43.06	67.75	79.07	83.72	86.36
Both hands (2)	54.25	80.39	85.66	92.48	93.10	41.13	65.59	75.27	82.87	84.88
Both hands (3)	48.69	80.47	84.96	91.94	92.72	36.65	65.20	75.12	83.57	86.52
Both hands (4)	50.23	80.31	85.35	92.64	89.85	38.35	64.19	76.59	85.43	78.61
Both hands (5)	57.10	82.87	86.21	89.69	86.13	44.60	66.83	78.76	78.45	74.27
Head (1)	40.20	73.33	75.28	93.88	92.64	29.94	57.06	61.09	84.19	88.30
Head (2)	51.55	84.89	82.56	94.81	94.34	39.66	72.33	74.04	90.08	90.93
Head (3)	38.43	85.04	84.03	95.66	93.49	28.09	69.61	74.27	86.43	88.76
Head (4)	26.39	55.43	56.59	93.72	91.79	18.60	43.26	44.96	87.52	85.58
Head (5)	30.64	57.68	56.90	94.19	92.64	22.92	44.88	46.98	87.37	85.97
Head and Hands (1)	47.22	85.04	83.80	72.41	72.72	35.42	74.81	74.35	61.01	57.68
Head and Hands (2)	36.73	79.00	78.84	73.49	71.16	26.78	67.83	65.74	62.72	57.60
Head and Hands (3)	36.81	72.64	77.21	87.91	87.44	24.54	55.20	62.40	75.82	75.20
Head and Hands (4)	57.80	84.65	87.68	93.72	93.26	46.92	69.77	77.13	84.65	87.76
Head and Hands (5)	42.83	81.01	81.71	91.47	90.47	30.09	65.89	69.15	82.17	82.10
Walking Motion (1)	49.92	82.02	84.27	93.95	89.77	41.28	71.79	72.10	86.67	81.09
Walking Motion (2)	43.06	78.76	80.47	91.79	88.61	33.34	66.28	66.36	85.51	77.29
Walking Motion (3)	44.60	83.33	85.66	94.19	92.56	37.19	71.71	75.12	86.51	84.73
Walking Motion (4)	50.70	84.50	86.20	94.34	92.10	43.06	73.72	75.74	87.06	83.26
Walking Motion (5)	45.30	84.73	86.44	95.04	93.34	37.97	74.19	80.00	89.69	88.45
Walking Motion (6)	43.06	83.18	85.20	93.49	91.79	34.80	69.23	75.51	85.50	82.95
Walking Motion (7)	49.46	85.20	85.04	95.50	92.79	41.59	75.12	77.83	89.15	85.35
Walking Motion (8)	45.91	83.03	87.13	94.58	92.79	37.73	71.32	78.69	88.84	87.45
	50 cm					100 cm				
Condition	**P**	**C'**	**J'**	**C"**	**J"**	**P**	**C'**	**J'**	**C"**	**J"**
Motor noise	48.69	71.48	80.78	86.12	88.68	45.99	61.09	71.86	78.07	83.49
Right hand (1)	44.37	62.25	72.64	70.70	77.75	37.58	52.02	66.05	60.78	68.07
Right hand (2)	33.49	54.11	67.83	79.07	84.42	27.86	42.95	55.12	67.91	77.83
Right hand (3)	38.04	61.47	72.02	77.37	81.86	33.95	51.48	63.03	66.67	70.93
Right hand (4)	41.59	62.48	73.88	81.32	75.74	36.73	54.27	65.66	70.16	66.75
Right hand (5)	40.36	63.49	73.26	70.00	73.10	36.27	51.40	64.88	59.00	63.41
Left hand (1)	37.27	58.45	69.30	74.58	79.23	30.63	48.22	60.08	63.18	69.46
Left hand (2)	33.26	51.71	61.09	77.75	82.79	29.48	40.70	50.70	69.54	74.73
Left hand (3)	39.51	60.62	71.48	78.76	83.64	35.88	53.49	61.86	68.99	76.44
Left hand (4)	24.38	54.19	60.93	80.93	86.28	20.68	42.64	52.79	69.61	76.59
Left hand (5)	38.43	57.21	71.17	74.73	71.09	33.49	47.99	59.46	64.73	62.71
Both hands (1)	35.81	54.81	65.74	73.80	76.98	30.09	43.33	55.20	62.87	67.60
Both hands (2)	33.88	51.71	65.27	70.85	74.97	27.32	42.64	55.97	61.48	65.04
Both hands (3)	28.55	52.02	61.94	69.23	76.05	23.38	41.94	52.41	59.93	64.50
Both hands (4)	30.79	52.33	61.16	74.88	67.44	24.69	41.01	52.48	64.42	60.00
Both hands (5)	36.42	53.02	64.73	66.36	62.41	30.48	45.12	56.90	55.35	50.93
Head (1)	21.68	43.95	48.37	72.17	78.53	16.67	32.56	38.68	61.78	67.68
Head (2)	30.87	59.38	63.33	79.92	84.58	27.70	50.55	56.59	68.84	75.35
Head (3)	21.68	55.97	63.41	73.80	81.32	18.36	46.90	56.75	65.27	72.71
Head (4)	12.97	32.79	35.51	78.30	77.52	11.89	27.06	27.99	69.31	69.46
Head (5)	16.75	35.59	37.83	79.85	76.90	14.74	29.07	32.95	70.16	70.47
Head and Hands (1)	29.63	61.47	63.26	52.33	48.91	25.00	51.40	54.35	45.27	40.93
Head and Hands (2)	21.99	54.89	56.05	51.32	48.92	18.29	45.97	48.07	41.71	42.09
Head and Hands (3)	18.52	42.02	47.44	63.65	61.24	14.43	31.24	38.69	52.33	52.48
Head and Hands (4)	38.51	55.66	67.29	72.10	78.61	32.64	47.21	57.99	64.42	68.68
Head and Hands (5)	21.76	51.01	56.98	71.16	69.92	17.90	41.16	49.07	61.09	61.32
Walking Motion (1)	35.96	59.69	58.92	77.52	71.40	30.09	47.29	49.07	65.04	58.92
Walking Motion (2)	27.78	52.25	52.87	74.81	65.04	23.00	40.62	43.80	66.51	55.74
Walking Motion (3)	30.71	56.83	63.26	76.98	74.73	26.01	47.06	53.34	66.05	65.28
Walking Motion (4)	35.57	59.61	62.95	77.06	73.88	30.17	47.83	53.96	66.59	63.33
Walking Motion (5)	31.95	60.47	69.15	79.92	79.23	28.55	51.01	58.45	68.69	70.23
Walking Motion (6)	27.93	56.21	62.48	74.66	73.33	24.31	45.20	52.33	62.95	62.72
Walking Motion (7)	33.18	59.77	64.58	79.85	76.98	28.94	50.16	56.21	68.84	65.35
Walking Motion (8)	31.02	55.97	67.29	78.06	78.14	26.16	46.82	56.28	67.06	69.31

表 B.5: ポストフィルタを用いた実験条件

Condition	**A**	**B**	**C**	**d**	**e**	**f**	**g**
Multi-condition (Stationary)		✓					
Multi-condition (Motion)			✓				
SS (each 1 utterance)							
Post Filter				✓	✓	✓	✓
SS (for motion noise)							
Adaptation for SS					✓	✓	✓
White Noise Addition						$p = 0.2$	$p = 0.4$
MFT (voice+noise matching)							
MFT (only noise matching)							
MFT (known noise)							
Unsupervised MLLR							
Supervised MLLR							
Acoustic Model	AM-1	AM-2	AM-3	AM-2	AM-4	AM-5	AM-5

Condition	**h**	**i**	**j**	**k**	**l**	**M**	**N**
Multi-condition (Stationary)							✓
Multi-condition (Motion)							
SS (each 1 utterance)							
Post Filter	✓	✓	✓	✓			
SS (for motion noise)						1 frame	1 frame
Adaptation for SS	✓	✓	✓	✓	✓		
White Noise Addition	$p = 0.6$	$p = 0.8$	$p = 0.6$	$p = 0.6$	$p = 0.6$		
MFT (voice+noise matching)			✓				
MFT (only noise matching)				✓			
MFT (known noise)					✓		
Unsupervised MLLR							
Supervised MLLR							
Acoustic Model	AM-5	AM-5	AM-5	AM-5	AM-5	AM-1	AM-2

Condition	**O**	**P**	**C'**	**j'**	**C"**	**j"**
Multi-condition (Stationary)		✓	✓		✓	
Multi-condition (Motion)						
SS (each 1 utterance)						
Post Filter				✓		✓
SS (for motion noise)	10 frame	10 frame				
Adaptation for SS				✓		✓
White Noise Addition				$p = 0.1$		$p = 0.1$
MFT (voice+noise matching)				✓		✓
MFT (only noise matching)						
MFT (known noise)						
Unsupervised MLLR			✓	✓		
Supervised MLLR					✓	✓
Acoustic Model	AM-1	AM-2	AM-3	AM-5	AM-3	AM-5

表 B.6: ポストフィルタを用いた提案手法の結果 1

	50 cm					100 cm				
Condition	**A**	**B**	**C**	**d**	**e**	**A**	**B**	**C**	**d**	**e**
Motor noise	59.19	83.34	83.34	36.34	71.84	40.44	74.23	74.23	30.56	60.34
Right hand (1)	50.93	81.25	79.63	28.24	66.98	30.71	67.67	68.60	20.83	54.86
Right hand (2)	45.45	77.63	73.84	25.62	63.35	26.00	64.97	60.80	18.83	48.08
Right hand (3)	51.32	79.63	78.55	15.90	62.35	32.56	67.75	69.60	12.27	47.76
Right hand (4)	52.70	82.64	79.94	26.39	66.74	32.95	69.14	69.83	19.37	52.78
Right hand (5)	50.70	81.02	81.17	25.08	67.82	31.10	68.52	69.99	18.60	53.40
Left hand (1)	50.31	78.32	79.79	18.98	64.66	30.48	65.05	65.36	14.12	46.84
Left hand (2)	44.60	74.69	74.38	25.31	58.95	24.31	59.88	60.27	18.14	44.06
Left hand (3)	51.55	80.25	79.32	15.44	60.65	32.72	68.21	69.29	11.19	47.15
Left hand (4)	47.45	74.62	74.85	18.60	55.33	29.32	60.42	60.81	13.74	41.52
Left hand (5)	52.01	79.40	79.94	23.38	65.74	32.72	65.43	66.05	16.90	49.23
Both hands (1)	45.07	76.70	75.08	16.29	60.73	25.31	61.19	61.50	11.19	43.98
Both hands (2)	44.29	74.54	74.31	23.15	55.87	25.47	59.18	59.88	15.59	41.90
Both hands (3)	43.60	73.61	74.46	19.83	55.56	25.62	58.10	60.65	14.90	41.90
Both hands (4)	43.37	74.08	74.38	22.69	56.56	24.62	59.88	57.87	15.12	41.90
Both hands (5)	45.91	76.93	75.47	19.37	57.72	26.01	61.58	62.50	12.27	44.29
Head (1)	45.84	63.89	67.21	26.00	55.56	23.69	50.23	51.39	18.21	42.29
Head (2)	53.25	73.23	76.16	21.53	52.40	34.57	59.80	64.12	16.82	42.36
Head (3)	50.39	74.31	74.92	17.60	53.16	31.79	62.04	61.96	12.12	41.28
Head (4)	29.94	44.91	50.16	8.72	24.77	17.83	33.88	38.27	5.40	18.52
Head (5)	30.79	46.22	52.01	7.87	25.93	17.90	34.88	40.97	4.94	17.44
Head and Hands (1)	51.24	74.38	77.24	23.23	56.41	33.57	61.81	65.51	16.51	44.06
Head and Hands (2)	51.31	66.59	70.99	12.50	45.99	31.18	55.79	59.03	9.11	36.65
Head and Hands (3)	37.50	63.81	66.67	16.13	49.54	19.53	48.85	51.85	10.80	36.04
Head and Hands (4)	46.38	77.70	77.47	22.38	59.65	27.24	63.04	63.35	16.05	45.06
Head and Hands (5)	47.38	73.30	73.46	16.05	54.48	27.47	60.27	59.57	10.73	39.90
Walking Motion (1)	50.00	74.00	73.23	33.72	62.65	29.94	62.19	64.05	29.56	51.47
Walking Motion (2)	46.91	71.30	71.22	28.09	59.11	29.40	56.95	60.65	23.38	46.53
Walking Motion (3)	49.38	74.08	74.39	33.11	62.97	30.32	64.05	64.74	27.16	49.85
Walking Motion (4)	50.78	75.85	77.47	42.44	66.52	32.33	65.51	66.82	35.11	52.86
Walking Motion (5)	52.01	76.08	75.77	37.20	66.21	34.11	67.59	66.67	30.48	54.63
Walking Motion (6)	46.99	74.85	74.70	31.41	60.27	29.63	61.81	62.35	24.85	48.23
Walking Motion (7)	49.85	75.24	76.62	37.12	65.59	32.41	66.51	66.59	28.94	51.93
Walking Motion (8)	50.85	76.62	72.92	34.80	66.13	33.41	67.75	63.97	28.01	55.02
	150 cm					200 cm				
Condition	**A**	**B**	**C**	**d**	**e**	**A**	**B**	**C**	**d**	**e**
Motor noise	27.86	64.20	64.20	22.69	49.15	19.83	54.01	54.01	20.22	43.44
Right hand (1)	19.68	56.02	56.33	17.60	44.14	12.27	45.60	46.38	14.58	37.58
Right hand (2)	15.36	50.70	49.54	14.97	36.11	9.57	41.44	38.89	11.65	28.70
Right hand (3)	21.37	56.56	56.49	9.50	39.20	12.89	46.53	46.92	7.56	32.64
Right hand (4)	20.99	56.02	57.10	13.66	41.98	12.50	46.30	49.23	12.19	34.88
Right hand (5)	18.52	55.48	57.95	13.58	42.28	11.27	46.61	46.76	11.89	35.50
Left hand (1)	18.83	54.55	53.32	10.58	39.43	10.73	42.29	44.22	8.65	31.64
Left hand (2)	13.81	47.38	47.46	14.74	35.42	6.71	36.65	37.89	12.81	26.63
Left hand (3)	21.45	55.25	54.63	7.49	36.88	12.74	44.76	47.15	6.56	31.41
Left hand (4)	16.82	50.00	50.31	9.26	31.10	10.34	39.51	39.67	7.87	23.23
Left hand (5)	19.91	54.01	52.55	13.51	40.28	11.19	43.75	44.14	10.96	32.64
Both hands (1)	15.20	51.01	49.08	8.57	33.87	7.79	39.05	38.50	6.49	26.86
Both hands (2)	15.44	49.00	47.46	11.89	32.02	8.57	38.82	40.20	9.26	24.38
Both hands (3)	16.52	45.99	48.38	10.42	32.87	8.57	35.42	39.43	8.26	25.70
Both hands (4)	15.13	46.61	47.15	11.03	32.33	7.56	36.35	36.96	8.03	24.31
Both hands (5)	15.59	49.00	50.00	10.34	34.73	8.41	39.35	41.36	8.80	27.40
Head (1)	13.74	36.89	40.13	11.58	30.95	7.02	26.78	30.63	8.03	24.39
Head (2)	21.68	51.62	51.47	11.27	34.72	14.81	42.05	42.60	9.11	29.87
Head (3)	18.45	51.01	50.62	8.64	31.64	11.81	41.13	42.29	7.18	26.70
Head (4)	11.81	24.62	30.87	3.55	13.81	7.26	19.83	25.46	2.47	11.81
Head (5)	10.57	28.47	32.87	2.78	14.12	6.56	21.22	26.39	2.63	12.19
Head and Hands (1)	20.84	52.55	53.86	12.20	36.81	14.20	41.67	45.14	10.57	29.78
Head and Hands (2)	19.91	44.14	48.46	5.09	27.86	13.66	35.88	41.05	4.48	23.77
Head and Hands (3)	9.11	38.04	39.66	8.03	26.39	4.94	27.39	30.41	6.71	18.37
Head and Hands (4)	15.21	50.23	50.54	12.04	35.42	8.49	38.97	42.59	9.57	27.78
Head and Hands (5)	16.67	47.77	46.76	8.41	30.87	9.11	37.27	38.58	6.02	23.15
Walking Motion (1)	19.29	52.09	54.32	23.00	42.44	12.12	41.13	42.13	20.45	34.73
Walking Motion (2)	18.68	45.68	48.23	18.06	39.89	11.58	36.96	38.04	15.59	33.26
Walking Motion (3)	19.29	51.47	51.00	23.30	41.13	12.19	42.21	41.75	20.14	33.80
Walking Motion (4)	19.91	55.10	54.09	28.25	44.83	12.35	45.06	44.06	24.93	36.65
Walking Motion (5)	20.91	55.09	54.02	25.31	44.22	12.42	45.99	45.60	21.99	37.65
Walking Motion (6)	17.83	51.55	51.47	19.37	38.04	10.50	40.90	40.66	16.82	32.80
Walking Motion (7)	19.91	54.40	55.02	25.31	44.06	12.97	44.37	45.99	22.15	36.88
Walking Motion (8)	20.45	55.71	51.32	24.85	46.14	12.89	43.37	42.60	19.76	38.28

表 B.7: ポストフィルタを用いた提案手法の結果 2

	50 cm					100 cm				
Condition	**f**	**g**	**h**	**i**	**j**	**f**	**g**	**h**	**i**	**j**
Motor noise	80.48	82.26	81.87	80.40	81.87	69.75	70.99	73.07	67.67	76.47
Right hand (1)	76.62	77.16	78.01	76.62	77.78	63.19	63.58	63.12	61.81	67.83
Right hand (2)	71.53	71.06	70.99	66.28	70.84	55.02	55.48	53.55	48.38	59.03
Right hand (3)	71.76	74.62	75.54	73.61	76.70	59.57	60.34	60.34	55.33	66.44
Right hand (4)	77.94	79.24	78.86	77.70	79.48	62.12	64.05	63.66	60.96	68.44
Right hand (5)	77.32	77.17	77.32	75.77	78.94	63.35	63.27	63.27	59.11	68.14
Left hand (1)	76.70	76.70	77.01	76.00	79.55	58.88	60.73	61.65	57.80	66.75
Left hand (2)	69.29	70.76	69.29	68.37	71.92	51.78	52.63	54.09	49.85	57.57
Left hand (3)	76.01	77.16	77.16	78.01	78.55	61.65	62.97	63.66	61.57	69.75
Left hand (4)	64.59	66.90	67.83	63.50	64.74	49.46	50.39	50.93	45.99	52.55
Left hand (5)	76.31	77.32	78.09	75.08	79.48	61.73	61.66	61.81	60.11	67.37
Both hands (1)	73.00	74.00	74.77	72.15	77.47	55.40	55.33	57.03	52.62	64.28
Both hands (2)	70.61	70.29	68.67	68.91	71.84	51.86	51.62	53.01	49.15	58.41
Both hands (3)	67.98	68.67	67.36	65.36	70.37	51.32	50.77	50.39	48.77	57.95
Both hands (4)	68.52	69.06	68.91	68.29	73.15	51.32	52.78	53.09	53.09	58.80
Both hands (5)	70.45	71.76	71.53	71.45	73.54	55.25	57.25	56.64	51.55	61.04
Head (1)	60.42	64.28	63.89	60.42	67.28	45.60	46.69	47.30	44.22	53.40
Head (2)	66.05	67.90	72.23	69.75	70.99	53.40	54.94	58.18	53.17	59.26
Head (3)	67.06	72.15	75.46	69.06	73.77	51.85	54.71	58.42	48.61	60.50
Head (4)	33.49	35.42	41.28	36.89	43.06	23.23	24.46	26.54	24.46	31.18
Head (5)	36.11	37.89	42.90	38.12	41.74	26.08	27.32	30.25	25.93	33.49
Head and Hands (1)	69.06	73.92	77.01	72.46	74.62	52.71	57.64	59.88	54.17	60.81
Head and Hands (2)	61.11	67.98	74.15	67.21	70.22	46.76	54.48	54.94	46.61	57.64
Head and Hands (3)	56.87	59.34	58.49	54.40	62.58	40.66	42.98	42.36	35.73	47.61
Head and Hands (4)	71.15	73.69	74.85	71.61	77.01	57.18	55.87	56.48	53.78	64.82
Head and Hands (5)	59.88	58.95	59.42	52.93	57.95	44.99	43.14	42.67	38.35	46.68
Walking Motion (1)	64.20	59.49	53.78	56.41	63.97	49.85	42.52	39.35	44.29	48.23
Walking Motion (2)	60.50	55.17	51.31	53.01	60.58	44.83	40.05	34.57	40.59	47.07
Walking Motion (3)	68.83	66.28	61.65	60.80	67.75	53.17	49.69	46.22	49.54	55.64
Walking Motion (4)	72.53	69.22	65.59	65.05	70.68	58.03	53.70	47.92	49.93	57.33
Walking Motion (5)	75.54	73.85	73.38	72.15	75.39	61.73	60.50	58.41	56.64	65.21
Walking Motion (6)	69.68	65.51	64.90	62.81	68.91	53.55	49.31	47.53	48.38	56.41
Walking Motion (7)	73.46	68.13	67.75	65.90	71.99	57.33	53.71	50.62	52.16	58.64
Walking Motion (8)	74.07	73.31	72.77	71.53	73.15	61.27	59.03	56.95	55.48	63.89
	150 cm					200 cm				
Condition	**f**	**g**	**h**	**i**	**j**	**f**	**g**	**h**	**i**	**j**
Motor noise	58.80	60.19	60.65	57.02	66.05	52.86	52.32	54.17	49.85	59.72
Right hand (1)	52.16	50.77	51.47	49.77	57.49	43.67	43.36	43.06	45.06	53.24
Right hand (2)	44.06	42.05	42.67	38.43	46.38	34.34	32.80	31.71	32.03	38.82
Right hand (3)	49.08	49.08	48.15	44.83	56.33	41.05	40.05	39.43	40.51	47.61
Right hand (4)	49.15	51.24	49.85	48.77	56.33	42.28	42.36	44.45	42.29	50.31
Right hand (5)	49.92	49.16	48.15	47.38	55.48	42.52	42.13	42.13	40.82	48.84
Left hand (1)	48.92	49.62	47.77	47.23	55.40	39.43	39.97	40.20	41.59	48.54
Left hand (2)	40.28	41.44	42.75	39.36	46.14	32.57	31.64	32.33	34.03	40.13
Left hand (3)	48.62	50.16	49.93	49.92	58.80	42.21	42.52	40.97	43.52	51.70
Left hand (4)	37.97	38.74	39.58	33.72	41.29	27.47	29.01	29.17	27.78	33.64
Left hand (5)	49.08	49.39	49.69	49.54	55.87	41.59	41.82	42.05	43.68	48.38
Both hands (1)	43.06	44.91	43.91	42.13	51.39	35.34	35.19	34.72	35.81	43.37
Both hands (2)	41.05	41.44	40.05	41.21	47.99	31.72	32.41	32.02	33.95	39.12
Both hands (3)	40.90	39.89	39.28	37.66	46.15	30.40	30.02	30.87	32.33	38.20
Both hands (4)	40.13	41.75	41.82	40.36	46.69	30.63	31.64	31.95	32.26	38.66
Both hands (5)	43.67	45.22	44.37	42.60	50.23	35.80	36.88	35.57	36.04	42.44
Head (1)	32.95	33.95	34.03	33.26	39.12	25.24	25.00	24.46	25.39	30.79
Head (2)	41.52	41.90	45.22	40.97	47.84	36.04	35.73	37.97	35.65	42.52
Head (3)	40.97	43.91	46.76	35.19	49.38	32.87	35.27	37.43	29.32	43.98
Head (4)	16.59	17.21	18.52	18.91	22.84	13.82	14.43	15.20	15.13	20.06
Head (5)	19.83	21.07	22.22	19.21	25.85	16.98	17.29	18.06	16.44	22.92
Head and Hands (1)	44.68	48.54	50.62	41.82	52.24	37.27	39.51	41.36	36.04	45.45
Head and Hands (2)	34.88	40.51	43.98	34.26	47.84	28.78	34.18	37.19	31.02	42.36
Head and Hands (3)	27.01	30.79	30.25	26.55	35.81	20.06	20.37	20.45	20.37	27.32
Head and Hands (4)	44.83	45.45	44.68	41.98	51.39	34.88	36.04	35.34	34.88	43.83
Head and Hands (5)	32.33	31.56	31.79	29.17	36.50	25.62	22.77	22.84	23.54	29.17
Walking Motion (1)	38.51	32.41	27.62	36.11	39.97	29.86	22.15	19.14	27.32	29.94
Walking Motion (2)	34.26	27.78	23.31	32.87	34.26	24.69	17.44	14.20	23.15	27.08
Walking Motion (3)	41.98	38.89	35.19	41.44	44.06	32.18	29.25	23.85	33.65	34.96
Walking Motion (4)	44.68	41.28	37.35	42.05	47.23	36.11	30.25	27.01	34.72	37.97
Walking Motion (5)	51.16	48.46	46.84	46.76	54.94	41.82	39.58	37.20	41.21	45.91
Walking Motion (6)	41.83	39.43	36.65	38.66	44.60	33.95	28.94	26.01	29.86	36.50
Walking Motion (7)	46.76	41.67	40.20	43.37	47.69	36.27	32.56	28.40	35.58	38.43
Walking Motion (8)	48.84	46.53	43.83	45.45	51.16	39.58	36.19	34.41	37.73	43.99

表 B.8: ポストフィルタを用いた提案手法の結果 3

	50 cm					100 cm				
Condition	**k**	**l**	**M**	**N**	**O**	**k**	**l**	**M**	**N**	**O**
Motor noise	82.49	82.33	57.56	67.90	57.72	77.09	76.70	41.13	60.03	41.74
Right hand (1)	79.25	79.63	51.01	66.83	51.86	70.22	70.22	35.11	54.25	36.58
Right hand (2)	70.83	77.24	41.44	55.25	41.13	59.57	63.66	27.39	42.75	27.47
Right hand (3)	77.39	77.47	47.61	58.80	47.23	67.91	69.14	31.72	46.53	32.56
Right hand (4)	79.63	80.87	51.31	66.21	53.01	68.75	70.45	35.19	53.63	35.57
Right hand (5)	78.32	78.78	50.93	62.65	50.77	68.06	70.22	34.18	52.16	34.80
Left hand (1)	78.40	79.01	48.77	60.04	50.00	68.06	68.52	34.26	45.99	35.34
Left hand (2)	70.76	75.93	46.61	57.80	47.22	58.03	61.42	29.63	45.91	31.18
Left hand (3)	78.94	79.71	51.70	62.66	51.70	69.14	69.76	33.49	49.85	35.81
Left hand (4)	64.97	75.16	40.59	47.76	40.98	51.93	64.12	26.16	34.19	26.93
Left hand (5)	79.17	79.48	50.62	61.35	52.01	67.67	69.76	35.34	49.54	37.42
Both hands (1)	77.24	78.86	45.14	54.86	46.45	64.74	64.51	30.64	43.21	32.18
Both hands (2)	72.69	75.39	42.98	54.94	43.91	59.88	61.42	27.86	42.05	29.79
Both hands (3)	72.69	73.30	41.82	50.08	41.36	57.80	59.03	27.40	37.74	28.24
Both hands (4)	72.84	75.70	42.44	50.39	42.82	57.95	60.57	27.01	39.59	27.39
Both hands (5)	74.93	76.00	45.53	57.33	46.76	60.96	62.42	29.48	44.83	30.56
Head (1)	66.20	72.92	41.75	41.59	41.13	52.16	59.34	25.54	30.40	27.16
Head (2)	70.84	78.70	47.76	53.09	47.69	59.57	69.14	33.80	41.90	34.65
Head (3)	72.76	78.86	39.74	44.60	36.81	60.27	70.38	26.70	33.57	24.92
Head (4)	42.82	54.48	22.61	26.16	22.15	31.72	41.75	15.13	18.68	14.82
Head (5)	42.21	56.48	27.17	30.87	28.09	33.41	44.45	18.06	23.92	18.14
Head and Hands (1)	74.69	80.17	46.07	51.16	45.45	60.65	70.14	31.10	39.51	31.95
Head and Hands (2)	71.07	77.01	44.06	38.73	42.21	55.71	67.36	29.79	29.56	29.71
Head and Hands (3)	62.66	68.60	35.57	39.05	35.42	47.69	54.86	19.45	26.47	20.45
Head and Hands (4)	76.93	77.55	48.62	58.72	48.77	63.12	64.20	33.65	47.92	33.18
Head and Hands (5)	57.49	72.92	41.82	46.15	40.82	45.68	61.58	27.39	32.41	25.62
Walking Motion (1)	65.36	65.44	42.67	50.23	42.90	50.00	50.08	31.87	41.67	32.33
Walking Motion (2)	60.65	63.04	42.36	44.52	41.44	47.23	49.69	30.25	33.87	31.64
Walking Motion (3)	70.30	67.28	44.14	46.38	43.52	56.33	55.02	32.87	39.97	32.02
Walking Motion (4)	71.38	73.38	45.60	53.71	45.14	56.48	58.57	34.19	44.60	34.34
Walking Motion (5)	74.38	77.16	45.68	48.23	45.61	64.43	65.59	33.34	39.12	33.11
Walking Motion (6)	69.37	70.30	42.98	45.76	43.29	57.10	58.18	31.02	36.96	31.02
Walking Motion (7)	71.38	72.53	46.92	51.86	46.38	58.65	60.26	32.72	42.75	32.95
Walking Motion (8)	73.00	75.39	46.84	49.00	46.22	63.58	66.36	34.26	40.21	34.57
	50 cm					100 cm				
Condition	**k**	**l**	**M**	**N**	**O**	**k**	**l**	**M**	**N**	**O**
Motor noise	65.36	67.67	30.56	50.23	32.72	60.50	60.88	25.39	46.91	28.71
Right hand (1)	58.03	57.87	26.39	45.14	28.32	51.86	51.08	20.99	39.51	23.00
Right hand (2)	46.84	50.70	19.29	34.18	20.14	38.66	42.98	12.96	26.39	15.13
Right hand (3)	57.41	58.18	24.00	38.51	23.92	48.23	49.31	18.37	33.49	20.53
Right hand (4)	56.94	58.65	25.46	42.67	27.16	49.77	52.93	19.14	37.04	21.76
Right hand (5)	55.79	56.72	24.38	41.67	27.09	47.69	49.00	19.06	35.04	21.61
Left hand (1)	55.10	56.79	25.23	37.89	26.62	47.77	49.08	18.44	31.49	20.76
Left hand (2)	46.14	50.54	19.83	34.34	21.53	38.97	42.90	13.43	30.63	16.05
Left hand (3)	57.33	60.34	23.85	41.98	25.62	50.00	51.39	19.76	35.26	22.30
Left hand (4)	41.05	53.09	17.13	25.24	17.98	33.41	45.22	12.73	20.14	13.20
Left hand (5)	56.02	57.87	25.39	39.74	27.55	47.38	48.92	19.37	33.03	23.07
Both hands (1)	51.08	52.62	21.15	35.34	22.92	44.68	45.22	14.66	28.40	17.44
Both hands (2)	48.07	48.69	19.45	33.95	21.45	39.35	41.21	14.43	27.70	15.67
Both hands (3)	45.99	47.77	18.06	28.48	18.37	37.27	39.74	13.04	22.53	13.97
Both hands (4)	47.46	49.77	17.52	31.79	19.06	39.43	40.97	12.89	25.00	14.66
Both hands (5)	49.08	51.93	21.92	36.42	23.15	43.06	45.30	14.89	29.87	17.59
Head (1)	38.97	46.38	17.59	21.92	17.83	30.87	39.43	12.58	16.75	12.97
Head (2)	48.31	57.33	22.76	33.42	23.69	42.44	51.70	18.75	28.48	19.76
Head (3)	48.84	58.26	18.75	25.31	16.29	41.90	50.62	13.97	21.61	12.81
Head (4)	23.77	31.56	11.96	13.28	11.65	19.91	28.40	8.72	11.35	9.11
Head (5)	27.16	35.19	12.42	18.91	11.73	22.30	29.40	10.26	15.51	10.49
Head and Hands (1)	52.39	60.50	23.77	31.87	24.00	45.99	54.25	18.91	26.70	19.60
Head and Hands (2)	46.84	57.33	22.69	23.07	22.30	41.67	52.24	17.13	19.06	18.44
Head and Hands (3)	36.19	42.44	11.88	18.90	12.89	25.47	33.49	7.80	15.59	9.03
Head and Hands (4)	52.01	53.47	22.07	39.58	23.92	44.83	45.22	17.21	32.26	19.06
Head and Hands (5)	36.81	50.00	16.44	23.46	18.44	30.02	43.29	13.43	18.99	13.58
Walking Motion (1)	39.82	39.27	25.08	35.42	25.24	29.48	29.09	19.21	29.94	20.84
Walking Motion (2)	34.26	35.73	21.68	28.70	23.54	25.47	26.31	17.59	22.84	18.91
Walking Motion (3)	44.29	44.22	24.31	29.94	27.01	35.80	34.88	18.37	26.78	20.14
Walking Motion (4)	46.53	46.99	24.77	36.65	25.85	37.50	38.82	17.59	30.17	20.68
Walking Motion (5)	54.09	55.25	25.08	32.72	26.55	47.15	46.53	19.68	28.78	21.61
Walking Motion (6)	46.07	45.99	22.46	29.02	24.31	36.42	36.58	17.90	24.92	19.14
Walking Motion (7)	47.76	48.54	24.62	34.11	25.39	38.89	39.35	19.52	28.32	21.69
Walking Motion (8)	51.32	53.01	24.46	31.56	26.16	43.91	45.30	19.60	27.01	20.99

表 B.9: ポストフィルタを用いた提案手法の結果 4

	50 cm					100 cm				
Condition	**P**	**C'**	**j'**	**C"**	**j"**	**P**	**C'**	**j'**	**C"**	**j"**
Motor noise	65.90	90.78	87.60	96.74	94.81	57.49	84.03	82.49	93.95	91.24
Right hand (1)	64.66	87.45	83.26	93.18	91.79	52.39	77.37	74.65	83.95	86.98
Right hand (2)	54.17	81.17	77.91	96.05	88.14	40.51	66.67	64.65	89.85	77.91
Right hand (3)	55.63	85.51	82.64	94.65	91.79	45.53	75.35	71.94	88.07	84.27
Right hand (4)	64.66	88.06	84.50	94.27	92.56	52.94	77.29	75.97	88.68	87.37
Right hand (5)	60.11	87.75	82.41	93.10	91.71	48.92	76.51	74.27	82.10	85.82
Left hand (1)	57.18	86.75	83.49	93.34	92.64	43.98	71.86	71.94	86.44	84.11
Left hand (2)	57.18	82.25	65.20	95.50	88.14	45.60	65.58	62.72	89.46	77.29
Left hand (3)	62.35	85.97	84.66	94.96	91.94	48.07	75.66	74.58	88.37	86.28
Left hand (4)	46.99	81.17	71.79	96.51	84.96	32.87	67.05	58.45	90.23	73.03
Left hand (5)	59.88	87.13	84.58	93.57	92.17	47.61	72.87	72.64	84.88	85.27
Both hands (1)	55.87	82.79	83.10	92.95	91.01	43.06	67.75	68.45	83.72	81.55
Both hands (2)	54.25	80.39	78.61	92.48	88.45	41.13	65.59	63.96	82.87	77.91
Both hands (3)	48.69	80.47	77.44	91.94	88.22	36.65	65.20	62.72	83.57	77.76
Both hands (4)	50.23	80.31	79.00	92.64	90.31	38.35	64.19	64.27	85.43	79.46
Both hands (5)	57.10	82.87	80.39	89.69	90.70	44.60	66.83	66.98	78.45	80.70
Head (1)	40.20	73.33	72.02	93.88	83.34	29.94	57.06	58.14	84.19	71.86
Head (2)	51.55	84.89	78.84	94.81	88.53	39.66	72.33	66.90	90.08	81.86
Head (3)	38.43	85.04	81.01	95.66	89.23	28.09	69.61	65.43	86.43	79.92
Head (4)	26.39	55.43	50.86	93.72	66.98	18.60	43.26	34.73	87.52	53.18
Head (5)	30.64	57.68	47.91	94.19	67.29	22.92	44.88	35.66	87.37	54.73
Head and Hands (1)	47.22	85.04	80.62	72.41	90.47	35.42	74.81	67.29	61.01	82.95
Head and Hands (2)	36.73	79.00	77.06	73.49	89.23	26.78	67.83	63.72	62.72	79.15
Head and Hands (3)	36.81	72.64	67.21	87.91	81.79	24.54	55.20	50.55	75.82	65.27
Head and Hands (4)	57.80	84.65	82.41	93.72	91.17	46.92	69.77	68.99	84.65	80.55
Head and Hands (5)	42.83	81.01	66.90	91.47	84.73	30.09	65.89	52.95	82.17	70.62
Walking Motion (1)	49.92	82.02	73.88	93.95	84.19	41.28	71.79	56.51	86.67	70.16
Walking Motion (2)	43.06	78.76	71.94	91.79	81.48	33.34	66.28	55.04	85.51	67.60
Walking Motion (3)	44.60	83.33	77.52	94.19	87.06	37.19	71.71	63.41	86.51	75.89
Walking Motion (4)	50.70	84.50	79.23	94.34	87.68	43.06	73.72	65.58	87.06	77.99
Walking Motion (5)	45.30	84.73	82.64	95.04	92.02	37.97	74.19	72.25	89.69	83.80
Walking Motion (6)	43.06	83.18	77.83	93.49	87.60	34.80	69.23	63.49	85.50	75.59
Walking Motion (7)	49.46	85.20	81.63	95.50	88.76	41.59	75.12	67.67	89.15	79.38
Walking Motion (8)	45.91	83.03	82.48	94.58	89.46	37.73	71.32	69.61	88.84	81.24
	50 cm					100 cm				
Condition	**P**	**C'**	**j'**	**C"**	**j"**	**P**	**C'**	**j'**	**C"**	**j"**
Motor noise	48.69	71.48	69.92	86.12	83.88	45.99	61.09	63.57	78.07	77.76
Right hand (1)	44.37	62.25	61.40	70.70	75.04	37.58	52.02	56.28	60.78	67.29
Right hand (2)	33.49	54.11	51.17	79.07	64.96	27.86	42.95	43.57	67.91	55.82
Right hand (3)	38.04	61.47	59.38	77.37	74.66	33.95	51.48	50.16	66.67	64.57
Right hand (4)	41.59	62.48	60.31	81.32	75.20	36.73	54.27	53.72	70.16	67.52
Right hand (5)	40.36	63.49	59.31	70.00	75.58	36.27	51.40	52.64	59.00	67.13
Left hand (1)	37.27	58.45	60.16	74.58	74.42	30.63	48.22	51.86	63.18	65.82
Left hand (2)	33.26	51.71	50.08	77.75	64.73	29.48	40.70	43.18	69.54	56.44
Left hand (3)	39.51	60.62	62.56	78.76	76.67	35.88	53.49	55.35	68.99	70.70
Left hand (4)	24.38	54.19	44.65	80.93	61.63	20.68	42.64	36.51	69.61	50.39
Left hand (5)	38.43	57.21	60.08	74.73	72.72	33.49	47.99	51.55	64.73	64.34
Both hands (1)	35.81	54.81	55.20	73.80	70.00	30.09	43.33	47.44	62.87	60.70
Both hands (2)	33.88	51.71	51.71	70.85	65.74	27.32	42.64	42.48	61.48	57.44
Both hands (3)	28.55	52.02	50.54	69.23	64.42	23.38	41.94	41.94	59.93	53.03
Both hands (4)	30.79	52.33	50.47	74.88	66.36	24.69	41.01	42.64	64.42	55.27
Both hands (5)	36.42	53.02	52.87	66.36	66.98	30.48	45.12	45.58	55.35	58.68
Head (1)	21.68	43.95	43.02	72.17	59.46	16.67	32.56	34.89	61.78	49.15
Head (2)	30.87	59.38	53.11	79.92	72.18	27.70	50.55	46.20	68.84	63.72
Head (3)	21.68	55.97	55.12	73.80	69.85	18.36	46.90	47.60	65.27	63.49
Head (4)	12.97	32.79	25.04	78.30	42.25	11.89	27.06	19.92	69.31	34.03
Head (5)	16.75	35.59	25.74	79.85	40.93	14.74	29.07	22.95	70.16	37.13
Head and Hands (1)	29.63	61.47	55.12	52.33	71.71	25.00	51.40	50.16	45.27	65.51
Head and Hands (2)	21.99	54.89	53.65	51.32	69.54	18.29	45.97	47.44	41.71	62.72
Head and Hands (3)	18.52	42.02	36.36	63.65	53.57	14.43	31.24	27.29	52.33	41.94
Head and Hands (4)	38.51	55.66	55.43	72.10	69.85	32.64	47.21	48.38	64.42	60.62
Head and Hands (5)	21.76	51.01	40.39	71.16	59.69	17.90	41.16	32.48	61.09	50.62
Walking Motion (1)	35.96	59.69	45.19	77.52	58.14	30.09	47.29	35.59	65.04	46.98
Walking Motion (2)	27.78	52.25	40.23	74.81	52.79	23.00	40.62	30.78	66.51	42.17
Walking Motion (3)	30.71	56.83	50.00	76.98	62.71	26.01	47.06	41.24	66.05	52.17
Walking Motion (4)	35.57	59.61	53.03	77.06	63.64	30.17	47.83	43.02	66.59	53.88
Walking Motion (5)	31.95	60.47	59.15	79.92	72.95	28.55	51.01	50.16	68.69	63.26
Walking Motion (6)	27.93	56.21	49.62	74.66	62.72	24.31	45.20	41.55	62.95	52.48
Walking Motion (7)	33.18	59.77	52.10	79.85	64.73	28.94	50.16	43.64	68.84	55.74
Walking Motion (8)	31.02	55.97	55.43	78.06	68.14	26.16	46.82	47.75	67.06	58.84