

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

西村 義隆[†] 簗津真一郎[†] 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^{††}

[†] 東京大学大学院情報理工学系研究科

〒113-8656 東京都文京区本郷 7-3-1

^{††} (株) ホンダ・リサーチ・インスティテュート・ジャパン

〒351-0114 埼玉県和光市本町 8-1

E-mail: [†]{nisshi,mino,dohi}@mi.ci.i.u-tokyo.ac.jp, ^{††}ishizuka@i.u-tokyo.ac.jp,

^{†††}{nakano,funakoshi,Johane.Takeuchi,yuji.hasegawa,tsujino}@jp.honda-ri.com

あらまし 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多くのモダリティを持つため、人間と同じような仕事ができることが期待される。我々は、ヒューマノイドロボットのこの点に着目し、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のできる記述言語MPML-HR (Multimodal Presentation Markup Language for Humanoid Robots) の開発を行ってきた。MPML-HRを用いることでプレゼンテーションコンテンツの開発が容易となったが、このコンテンツはロボットからの一方向のプレゼンテーションを行うものであった。人間によるプレゼンテーションでは、視聴者からの質問も想定されるため、インタラクションができることはプレゼンテーションコンテンツにおいて重要であると考えられる。本稿では、プレゼンテーションにおいて想定される視聴者からの簡単な問いかけにも答えられるよう、インタラクション機構の実装を行った。

キーワード ヒューマノイドロボット、マルチモーダルプレゼンテーション、記述言語

Yoshitaka NISHIMURA[†], Shinichiro MINOTSU[†], Hiroshi DOHI[†], Mitsuru ISHIZUKA[†], Mikio NAKANO^{††}, Kotaro FUNAKOSHI^{††}, Johane TAKEUCHI^{††}, Yuji HASEGAWA^{††}, and Hiroshi TSUJINO^{††}

[†] Graduate School of Information Science and Technology, University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN

^{††} HONDA Research Institute Japan Co., Ltd.

8-1 Honcho, Wako-shi, Saitama 351-0114, JAPAN

E-mail: [†]{nisshi,mino,dohi}@mi.ci.i.u-tokyo.ac.jp, ^{††}ishizuka@i.u-tokyo.ac.jp,

^{†††}{nakano,funakoshi,Johane.Takeuchi,yuji.hasegawa,tsujino}@jp.honda-ri.com

Abstract Recently many kinds of robots are paid much attention. Among them, humanoid robots are expected to be able to work as humans do. Humanoid robots are suitable for a tool for presentation, because they can use many modalities and are attractive. We have been developing MPML-HR (Multimodal Presentation Markup Language for Humanoid Robots), for easily describing humanoid robot presentations. Previous versions of MPML-HR, however, do not feature interaction with audiences. In this paper, we propose the new version of MPML-HR ver. 3.0 that includes the interaction for the presentation.

Key words Humanoid robot, Multimodal presentation, Description language

1. はじめに

近年、ロボットが注目され、さまざまな種類のロボットが研究、開発されている。特にヒューマノイドロボットは多くのモダリティを持ち、人間の代わりとしての活躍が期待されている。ロボットは実空間上で動作するため、視聴者への印象が強く、興味をひきやすい。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実現するためには、ロボットの制御用プログラムを用いることが必要である。制御用プログラムは下位レベルの操作を行うため、記述が複雑であり、専門的な知識を必要とする。そこで我々は、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容易に作成することを目指し、MPML-HR [1] [2] の開発を行った。MPML-HR は下位レベルのインタフェースをあらかじめ構築しておき、ユーザに中位レベルで記述させることで、誰でも簡単にプレゼンテーションコンテンツを作成することを可能とする。

MPML-HR を用いることでロボットからの一方的なプレゼンテーションコンテンツの作成は可能となったが、インタラクションを含むコンテンツの作成はできなかった。インタラクションはプレゼンテーションの重要な要素であり、プレゼンテーションに関する質問に答えることでプレゼンテーションを分かり易くさせ、充実したコンテンツとすることができると考えられる。MPML-HR はアニメーションエージェントをプレゼンタとしたコンテンツを作成する記述言語 MPML [3] [2] をヒューマノイドロボット用に拡張したものである。MPML は、特定の箇所において音声インタラクションを受け付ける機能を有していたが、任意の時点での音声入力を受け付けることはできなかった。

プレゼンテーションコンテンツに音声インタラクションを付加するための一つの方法として、各プレゼンテーションごとに特定ドメインに依存する音声対話システム (e.g. [6], [7]) を結合することが考えられる。しかし、専門的知識を持たないユーザがプレゼンテーションに対する答えを導き出せるようにシステムを調整することは容易ではない。それゆえ、VoiceXML [4] や XISL [5] などのインタラクションを対象とした記述言語の開発がされている。VoiceXML は電話による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。しかし、プレゼンテーションでは音声のみならず、スクリーンへの資料の表示や動作などによる出力も必要であるため、VoiceXML では不十分である。XISL [5] は音声、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。ヒューマノイドロボットと同様のジェスチャや移動、発話というモダリティを持つアニメーションエージェントによるインタラクションが可能であるため、XISL を用いることで目的とするインタラクションの記述は可能であるとも考えられる。しかし、現段階の XISL はアニメーションエージェントのみを対象としているため、ロボットへの拡張が必要である。また、XISL はプレゼンテーション用ではなく、インタラクション一般を目的として作

られているため、記述量が多くなりがちであるといった問題もある。音声入力による質問も想定したロボットによるプレゼンテーションを実装した例もある [8]。この研究では、ヒューマノイドロボットを用いたプレゼンテーションを行う枠組みについて実装が行われているが、記述言語のような簡単な機構でプレゼンテーションコンテンツを作成するための実装には触れられていない。

本稿では、これまで開発してきた MPML-HR を拡張し、プレゼンテーションコンテンツのインタラクションを対象とした記述言語を提案する。そして、従来の MPML-HR のロボットによるプレゼンテーションの生成とインタラクションの生成を合わせて MPML-HR ver. 3.0 とした。MPML-HR ver. 3.0 は従来の MPML-HR の「簡単な記述が可能」というメリットを受け継ぎ、充実したコンテンツを専門的な知識なしに誰でも簡単に記述できることを特徴とする。MPML-HR ver. 3.0 の割り込み機構には、対話ロボットのエキスパートモデルである RIME (Robot Intelligence based on Multiple Experts) ^(注1) [9] を用いた。それぞれのエキスパートは、対話の特定のタスクまたは物理的な動作命令に対して応答可能である。MPML-HR によるプレゼンテーションを実現するため、RIME に MPML-HR 用エキスパートを開発し、追加した。MPML-HR のスクリプトは、エキスパート知識に変換され、MPML-HR 用エキスパートを通してプレゼンテーションが実行される。インタラクションの際の音声認識に用いる言語モデルは、MPML-HR のスクリプトごとに作成する。MPML-HR エキスパートのための音声インタラクション機能は他のエキスパートと共通のものである。本システムの実装により、RIME の制御機構が MPML-HR ver. 3.0 の実現に適していることが判明した。また、RIME を用いることで、MPML-HR によるプレゼンテーションとタスク指向または非タスク指向の対話システムとを簡単に接続することが可能となった。

2. インタラクション機構実現のためのプレゼンテーション記述言語

本節では、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作成するための MPML と、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 MPML-HR について概説する。次に、インタラクション機構を含む MPML-HR ver. 3.0 について説明する。

2.1 MPML および MPML-HR

MPML はスクリー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誰でも簡単に記述することのできる記述言語である。中位レベルの言語であるため、下位レベルのエージェント制御プログラムとは独立である。したがって、ユーザは下位レベルの制御知識なくコンテンツを作ることが可能である。また、XML に準拠しており、エージェントの位置、動作、ジェスチャ、感情表現などを制御する豊富な関数が用意されている。MPML-HR は MPML をヒューマノイドロボット用に拡張し

(注1): RIME は文献 [9] では MEDBP と呼ばれていた。

たものである．拡張には例えば point タグなどがある．これは，実空間に存在するロボットがスクリーンのある座標を指示するための命令である．スクリーン上に存在するアニメーションエージェントは移動命令である move タグを用いて特定の座標を指示可能であるが，ロボットでは別の命令が必要となる．このように，エージェントの存在する空間の違いを吸収するための拡張が MPML-HR ではなされている．

2.2 インタクション機能の拡張

本研究では，プレゼンテーションにおける視聴者からの音声割り込みとして，説明の省略や，前に話した内容の再度の説明，あらかじめ想定した説明に応える機構を実現することを目的とする．これらのインタクションを実現するため，コンテンツ作成者が発話パターンや次のロボットの動作を記述できるよう，MPML-HR を拡張する必要がある．

本研究の目的とするインタクションは，プレゼンテーションの説明箇所を遷移させることで実現できる．プレゼンテーション内の特定の箇所に遷移するため，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．

視聴者からの割り込みの実現においては音声認識誤りへの注意が必要である．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，インタクション機能を導入するメリットが失われてしまう．そこで，音声認識誤りに対処するため，二つの機構を取り入れた．一つは，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．もう一つは，間違った遷移が行われた場合，視聴者からの指摘により，遷移前のプレゼンテーションに戻す機構である．

図 1 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，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 (Fig. 1 (1))，インタクションに関する記述をする．grammar タグ (図 1 (2)) 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 (e.g.(3))，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はプレゼンテーション全体で有効である．jump タグと do タグ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

```
<?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">
    <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> .....(1)
  <grammar regram="[それは] ちがうよ"
    confirm="遷移前に戻りますか?" view="違うよ"> .....(2)
  </grammar>
  <grammar regram="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" />
  </grammar>
  <grammar regram="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" />
    <do page="昨日の天気" />
  </grammar>
  <recog page="明後日の天気"> .....(3)
    <grammar regram="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" />.....(4)
      <jump page="明日の天気" />
    </grammar>
  </recog>
</interaction>
</mpml>
```

図 1 Example MPML-HR ver. 3.0 Script

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．但し，複数ページを指定する際には，grammar 内には jump のみまたは do のみで記述する必要がある．jump タグでページが指定された場合，そのページに遷移し，遷移先ページが終了した際にはもとのプレゼンテーションには戻らない．一方，do タグでページが指定された場合，そのページに遷移し，遷移先ページが終了した際にもとのプレゼンテーションが続けられる．goback タグは，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．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 regram の内容が表示される．

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

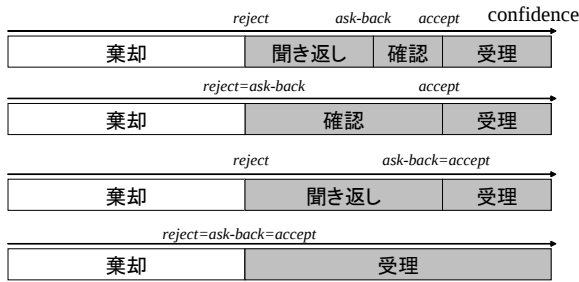


図 2 Confidence thresholds

信頼度が高かった場合には受理となる．この場合，正しく認識できている可能性が高いため，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(図 1 (4)) 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 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．

図 3 に図 1 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: 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 “last”)
 Robot: 昨日の東京は曇り空でしたね
 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？ (Ask back)
 User: 明日の天気
 Robot: はい, (Jump to “明日の天気”)
 Robot: 明日の東京は...

 Robot: 明後日の東京は雨でしょう

図 3 Transcription of an Example Presentation

3. マルチエキスパートモデルに基づくインタラクション機能

3.1 システム概要

図 4 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行動を行う．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．

対話・行動制御サブシステムは対話ロボットの会話・行動のモデルである RIME (Robot Intelligence based on Multiple Experts) [9] を基に構築した．このモデルでは，エキスパートと呼ぶ，特定のドメインの対話や特定の物理行動など遂行するためのモジュールを用いる．MPML のプレゼンテーションを実現するため，MPML 用のエキスパートを開発した．(以後，MPML-HR ver. 3.0 の代わりに MPML と表記する．) 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に変換される．

3.2 マルチエキスパートに基づく対話・行動制御サブシステム

RIME で用いるエキスパートはサブタスクごとに用意される必要がある．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，エキスパートはアクティベートされ，音声認識結果と文意を理解する．エキスパートはオブジェクト指向の枠組

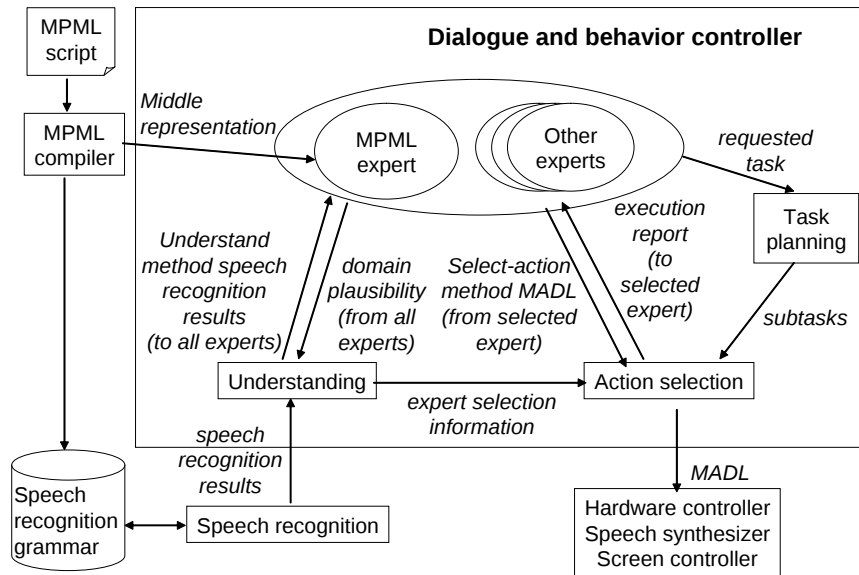


図 4 System Architecture

みのなかの一種のオブジェクトである．それぞれの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．

システムには，複数のエキスパートを利用して動作するために，3つのモジュールがある．*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．

それぞれのエキスパートには内部状態にアクセスするためのメソッドが必要である．ここではMPMLエキスパートに関連するもののみ説明する．*initialize* メソッドはエキスパートが呼び出されたときに生成され，内部状態を初期化する．*understand* メソッドは音声認識結果を受け取った際に *understanding* モジュールから呼び出され，音声認識結果に基づいて情報を更新する．*select-action* メソッドは，*action selection* モジュールから継続的に呼び出され，発話割り込み待ちの状態でなければコンテンツの情報に基づいて一つ動作を出力する．

3.3 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 block が生成される．図5に図1のMPML スクリプトをコンパイルした中間記述の例を示す．

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．

3.4 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2.2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ぬ．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，

```

<block id="1" next="2">
  <interruption-handle regram="gram_0_1" goto="goback." confirm="遷移前に戻りますか？"></interruption-handle>
  <interruption-handle regram="gram_0_2" goto="3_r." confirm="昨日の天気ですか？"></interruption-handle>
  <interruption-handle regram="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 5 An Example of Intermediate Representation

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，発話者の発話は Yes または No の意味合いを想定するため，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．

4. ま と め

本稿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．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開発を行った．現段階の実装はプレゼンタにロボットを用いたものであるが，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．

5. 今後の課題

本稿で示した実装は，音声認識エンジンが算出する信頼度のみに用いているが，この信頼度の精度を向上させ，信頼性の低い発話はきちんと棄却できるようにしたいと考えている．また，MPML の簡単な記述が可能というメリットを活かしつつ，より充実したインタラクション機構について検討を考えている．

文 献

- [1] Y. Nishimura, K. Kushida, H. Dohi, M. Ishizuka, J. Takeuchi, and H. Tsujino, "Development and psychological evaluation of multimodal presentation markup language for humanoid robots," in *Proc. of IEEE RAS Humanoids*, 2005, pp. 393–398.
- [2] J. Takeuchi, K. Kushida, Y. Nishimura, H. Dohi, M. Ishizuka, M. Nakano, and H. Tsujino, "Comparison of a Humanoid Robot and an On-Screen Agent as Presenters to Audiences," in *Proc. of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 3964–3969.
- [3] H. Prendinger, S. Descomps and M. Ishizuka, "MPML: A Markup Language for Controlling the Behavior of Life-like Characters," *Journal of Visual Languages and Computing*, 2004, Vol.15, No.2, pp. 183–203.

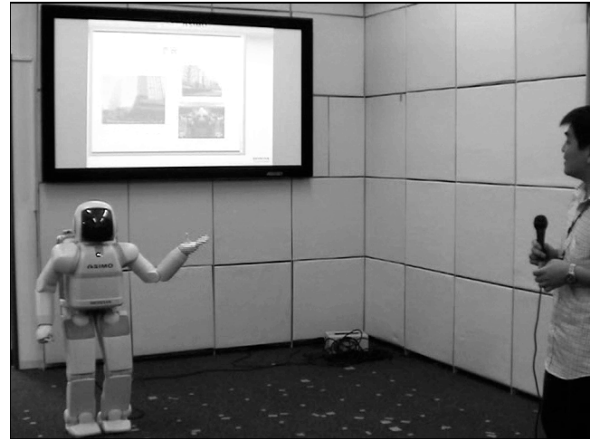


図 6 ASIMO interaction

- [4] Voice XML Forum homepage: <http://www.voicexml.org/>
- [5] K. Katsurada, Y. Nakamura, H. Yamada, T. Nitta, "XISL: A Language for Describing Multimodal Interaction Scenarios," *Proc. of ICMI*, 2003, pp. 281–284.
- [6] J. Glass and E. Weinstein, "SPEECHBUILDER: Facilitating spoken dialogue system development," in *Proc. Eurospeech*, 2001, pp. 1335–1338.
- [7] S. Sutton et al., "Universal Speech Tools: The CSLU Toolkit" in *Proc. ICSLP*, 1998, pp. 3221–3224.
- [8] 酒井, 西山, 溝口, "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, 2004, pp. 1–4.
- [9] 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 *Proc. IEEE/RSJ IROS*, 2005, pp. 1542–1548.
- [10] T. Kawahara, A. Lee, K. Takeda, K. Itou, and K. Shikano, "Recent progress of open-source LVCSR engine Julius and Japanese model repository," in *Proc. ICSLP*, 2004, pp. 3069–3072.