

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

簗津真一郎[†] 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[‡]
[†] 東京大学大学院情報理工学系研究科
[‡] (株) ホンダ・リサーチ・インスティテュート・ジャパン

1 はじめに

我々はこれまで、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容易に作成することを目指し、MPML-HR(Multimodal Presentation Markup Language for Humanoid Robots)の開発を行ってきた[1]。

MPML-HRを用いることでロボットによるプレゼンテーションが実現可能である。しかし、ロボットからの一方向の説明を行うのみで、インタラクションを含むコンテンツを作ることができなかった。インタラクションはプレゼンテーションの内容を理解する上で重要であると考えられる。

そこで本稿では、MPML-HRを拡張し、音声インタラクション機能を導入する。新しいMPML-HRをMPML-HR ver. 3.0とした。

2 プレゼンテーション記述言語 MPML-HR

MPML-HRは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である。中位レベルの言語であるため、ユーザは下位レベルの制御知識なく簡単に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。

具体的な命令として、ジェスチャを行うplay、移動を行うmove、発話を行うspeak、スクリーンの指示を行うpointなどがある。また、emotionを用いることで感情を含む音声合成を行うことが可能であり、スラ

イド一枚に対応するプレゼンテーションはpageで囲まれる範囲に記述し、これを組み合わせることでコンテンツを作成する。

3 インタラクション機能の拡張

本稿では、プレゼンテーション中のインタラクション機能の導入にあたり、説明の省略、前に話した内容の再度の説明およびあらかじめ想定した説明に応えることを目的とする。これらのインタラクションは、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。そこで、スライド1枚に相当するプレゼンテーションを記述するpageタグの先頭に説明箇所を遷移させることでインタラクションを実現する。

図1にサンプルスクリプトを示す。インタラクションに関する記述はinteraction内で行う(図1(1))。音声認識文法の指定にはgrammarを用い(図1(2))、recogでページを指定することで有効な範囲を記述する(図1(5))。範囲指定されなかった文法はコンテンツ全体で有効となる。説明箇所の遷移はjump(図1(7))またはdo(図1(4))の記述により行う。jumpで指定した場合には遷移先のコンテンツ終了後もとの説明には戻らないが、doで指定した場合にはもとの説明に戻る点で異なる。

4 音声認識誤りの扱い

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処が必要である。そこで、2つの機構を取り入れた。1つは、音声認識結果に対する信頼度を用いることで、音声入力受付時に棄却、聞き返し、確認、受理のいずれかを行う。聞き返しは再度の発話を要求し、確認は認識結果が正しいかどうかの問い合わせを行う。これらの閾値はシステム内であらかじめ定義されているが、confidence-thresholdsを用いてユーザが指定することも可能である(図1(6))。さらに、gobackを用いることで、視聴者からの誤り遷移の指摘により説明箇所を遷移前の状態に戻す機構を導入した(図1(3))。

Voice Interaction Function of MPML-HR

Shinichiro MINOTSU[†], Yoshitaka NISHIMURA[†], Hiroshi Dohi[†], Mitsuru ISHIZUKA[†], Mikio NAKANO[‡], Kotaro FUNAKOSHI[‡], Johane TAKEUCHI[‡], Yuji HASEGAWA[‡] and Hiroshi TSUJINO[‡]

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

Bunkyo-ku, Tokyo 113-8656, Japan

[‡] Honda Research Institute Japan Co., Ltd.

Wako-shi, Saitama 351-0188, Japan

```

<?xml version="1.0" encoding="UTF-8"?>
<mpml>
<body>
<page ref="/slide.ppt#1" id="明日の天気">
<synch>
<speech>明日の東京は晴れでしょう</speech>
<play act="victory" />
</synch>
</page>
<page ref="/slide.ppt#2" id="明後日の天気">
<speech>明後日の東京は雨でしょう</speech>
</page>
<end />
<page ref="/slide.ppt#3" id="昨日の天気">
<speech>昨日の東京は曇り空でした</speech>
</page>
</body>
<interaction> .....(1)
<grammar recgram="[それは] ちがうよ" confirm="遷移前に戻りますか?">.....(2)
<goback />.....(3)
</grammar>
<grammar recgram="きのう [のてんき]" confirm="昨日の天気ですか?">
<do page="昨日の天気" />.....(4)
</grammar>
<grammar recgram="あさって [のてんき]" confirm="明後日の天気ですか?">
<do page="明後日の天気" />
</grammar>
<recog page="明後日の天気"> .....(5)
<grammar recgram="あした [のてんき]" confirm="明日の天気ですか?">
<confidence-thresholds reject="0.6" ask-back="0.8" accept="0.9" />.....(6)
<jump page="明日の天気" /> .....(7)
</grammar>
</recog>
</interaction>
</mpml>

```

図 1: Sample script of MPML-HR ver. 3.0

5 対話例

図 2 に図 1 を実行した際の対話例を示す。プレゼンテーションは記述された順に実行されるので、明日の天気から説明が始められる。最初にユーザから、昨日の天気の問い合わせがあるが、音声認識誤りによりロボットは明後日の天気の説明を始める。ここで、遷移が違うことをユーザが指摘するとロボットはそれを認識し、もとの説明(明日の天気)に戻る。ユーザが再度昨日の天気について問い合わせると、今度はロボットが正しく認識し、昨日の天気の説明をする。ここで、昨日の天気の説明はスクリプト内で do によって指定されているため、昨日の天気の説明が終了した後、もとの説明に戻る。

次のブロックでは、問い返しについての例が示されている。ユーザが明日の天気について問い合わせるが、信頼度が低いため、ロボットからの問い返しが行われる。ユーザが再度発話すると、次は高い信頼度とともに音声認識がなされ、意図した遷移が行われる。

6 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の実装

MPML は対話ロボットの会話・行動モデルである RIME (Robot Intelligence based on Multiple Experts)[2] を用いて実現した。このモデルでは、エ

```

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: はい, (Jump to "明日の天気")
Robot: 明日の東京は...

```

図 2: Transcription of an Example Presentation

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。そして本稿では、MPML のプレゼンテーションを実現するため、MPML エキスパートを開発し、ASIMO 上で動作するようにした。

7 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。

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を用いているが、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-HR の簡単な記述が可能という長所を活かしつつ、より充実し、自由度の高いインタラクション機構について検討を考えている。

参考文献

- [1] Y. Nishimura, et al., "Development and psychological evaluation of multimodal presentation markup language for humanoid robots," in Proc. of IEEE RAS Humanoids, 2005, pp. 393.398.
- [2] M. Nakano, et al., "A two-layer model for behavior and dialogue planning in conversational service robots," in Proc. IEEE/RSJ IROS, 2005, pp. 1542.1548.