

身振りによる対話調整機能を持つ対話ロボット

○東條 剛史 松坂 要佐 小林 哲則

早稲田大学 理工学部

〒169-8555 東京都新宿区大久保 3-4-1

tojo@tk.elec.waseda.ac.jp

1 はじめに

人間同士のインタラクションにおいて、ジェスチャやアイコンタクト、表情などといった非言語的表現は重要な意味を持っている。我々は、実空間に存在するロボットの視線表現を用いて複数話者との対話システムを構築し、非言語表現の重要性を示してきたが [1], 発話を促す効果のあるうなずきによってより円滑な対話を実現できるよう検討している。

うなずきを実現するにあたり音声認識の処理時間がロボット制御のむだ時間となり、適切なタイミングで行うことは困難になる。音声認識の中間結果を用いて端末内のマスコットの相づちやうなずきを実現するシステムも提案されているが [2], 遅れを伴う機械系を制御するにはさらに早い段階で指令値を出力する必要がある。本稿では、発話権の円滑な授受を実現するために、うなずきなどの身振りを遅れを伴う機械系の制御を考慮したタイミング推測という観点から捉え開発したロボットについて紹介する。

2 ロボット概要

研究に用いるロボット“ROBITA¹”は、2系統の CCD カメラが組み込まれ、うなずきなどの表現が可能な眼部および頭部、指示動作などの身体表現が可能な各 4 自由度の腕 2 本、および胴体と全方向移動可能な台車から構成されている。頭部には、音源方向定位用のマイク 2 系統も設置されている。発話者側に音声認識用のマイクを設置する構成をとっている。無線 LAN で接続されたワークステーションにより、ジェスチャ認識、顔個人・顔方向認識、対象物認識、音声言語理解などが可能である。

3 うなずきのタイミング推測

ロボットの制御では機械系の遅れの問題から、タイミング推測を音声認識結果が得られる以前に行う必要がある。そのため、本稿では音声信号を用いたタイミング推測を行う。人間の発話では、強調したい語を強く発話する。また、周波数は発話終了が近くなるにつれ下がっていく傾向があることが知られ

¹Real world Oriented Bimodal Talking Agent

ている。聞き手のうなずきは、強調した語の後、もしくは発話終了時に行われることが多い。したがって、音声信号の振幅およびスペクトル分布平均の下り方向微分を首ピッチ軸の制御量として利用する。

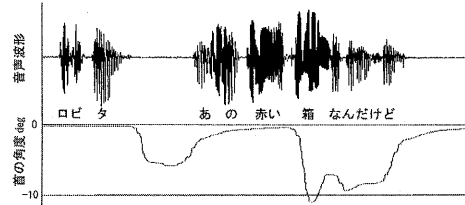
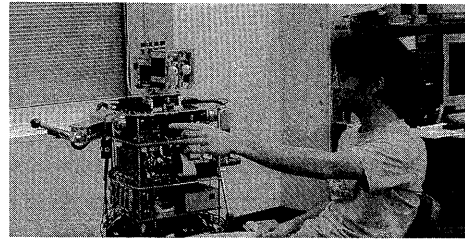


図 1: 対話時の様子

4 おわりに

非言語的表現の重要性を意識し、うなずきなどの身振りによって発話権の円滑な授受を実現するロボットを開発した。遅れのある機械系への指令値として音声信号の微分値を用いることで即応性が得られ、自然なうなずきを実現できた。今後は、文脈情報による反応の変化に対応するため、発話途中でも部分単語の認識が終わった時点で逐次出力する音声認識装置の結果も使い、内容に即した反応を示せるシステムを構築していく予定である。

参考文献

- [1] 松坂ら, “画像・音声・音響情報を統合した複数話者との会話ロボット”, 音響学会春季研究発表会, pp.95-96, (1999)
- [2] 平沢, 川端, “わかってうなずくコンピュータの試作”, 情処研報, SLP-97-19(1997)