マルチモーダル情報による相槌の認識とロボット対話への応用

田崎 豪 † 山口 健 ‡ 戸田 充彦 ‡ 駒谷 和範 ‡ 尾形 哲也 ‡ 奥乃 博 ‡
† 京都大学工学部情報学科 ‡ 京都大学大学院情報学研究科知能情報学専攻

1 はじめに

ロボットが社会に普及するにつれて、ロボットが人間と円滑な対話をすることが求められる。人間同士の対話では、聞き手は話し手に、話し手は聞き手に協調することで円滑な対話を実現している。しかし、現在のロボット対話においては、ロボットが人間に合わせてもらう一方的なものが多く、相互に協調したインタラクションを実現しようとする研究は少ない。人間と協調したインタラクションを試みたロボットの例として渡辺のインタロボット [1] があげられる。これは、人間の行為が起こった時間などの物理量をそのまま用い、発話パラメータやロボットの挙動を変化させることで人間との協調を行った。

本研究では、人間同士の対話において、協調する手段である相槌を主に取り上げる。人間の話し手は、聞き手の相槌によって聞き手の心地よい話のリズムや理解状態を知り、協調を行う。そこで、これまでの研究で用いられてきた、相槌が起こったタイミングという物理量を用いた協調に加え、相槌から推定する理解状態という、意識的な情報を用いて協調したインタラクションを目指す。理解状態を推定した協調的なインタラクションを行うことにより、以下の効果が期待できる。

- 聞き手の理解状態に応じた発話内容を選択する。
- 複数あるロボットのモダリティを聞き手に合った話のリズムに乗せて同期的に用いることが可能になる。
- 呼吸感が無く相槌が打ちづらい音声合成による発話の「間」を聞き手に合わせて強調し、相槌を促進する。

本報告では、協調したインタラクションを行うための前段階として、ヒューマノイドロボット SIG(図 1) に聞き手の理解状態の推定機能を導入することを目的としている。

図 1: ヒューマノイドロボット SIG

2 理解状態推定手段として用いる人間の行為

本研究では聞き手の理解状態を推定する手段として、相槌と頭部ジェスチャを取り上げる。

相槌の定義は、メイナード [2] の述べた「聞き手の発話のうち、話し手の発話権を奪取する意図のない短い発話のこと」とする。さらにメイナードは相槌には 6 つの機能があると述べている。しかし、これらの機能はメイナード自身が、「曖昧であり、どの機能になるかは程度の差でしかない」と述べているように、明確なものでない。したがって、本研究ではメイナードの述べた機能のうち 4 つをまとめ、以下の 3 つを相槌の機能として考える。

- 続けてという機能
- 感情を表す機能
- 情報の追加、訂正、要求などをする機能

また、縦方向の頭部移動であるうなずきと、横方向の移動である傾げ、首振りを頭部ジェスチャとする。その機能として、縦移動は対話において肯定的姿勢を、横移動は否定的姿勢を示す。

3 聞き手の理解状態の推定

3.1 聞き手の理解状態

相槌と頭部ジェスチャが持つ機能から、2 軸の理解状態の平面を作成する。その軸は、理解しているかいないかを表す軸と、感情を表す軸である。感情の軸では、対話に対して積極的、能動的といった、好意的な感情を正にとり、消極的、受動的といった、非好意的なものを負にとる。さらにどちらでもないものを 0 とした 3 種類に分類する。

本研究ではこの平面上の位置で理解の状態を決定する。また、それぞれの軸の名前を understand 軸 (正と負の 2 種類)、active 軸 (正、0、負の 3 種類) とする。

例えば

($active$ 軸, $understand$ 軸) = (負, 正)

であれば話を理解してはいるが、あまり気が乗らない状態を表す。

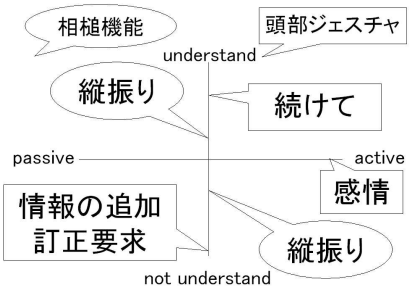

図 2: 理解状態と相槌と頭部ジェスチャの対応

相槌機能とは、図 2 のように対応する。つまり、understand 軸が正、負となる相槌は、それぞれ続けてという機能、情報の追加、訂正要求などをする機能に対応する。また、active 軸は感情を強く出す機能をもつ相槌と対応する。頭部ジェスチャとは、understand 軸の正部分と首の縦振りを、understand 軸の負部分と首の横振りを対応させる。

Backchannels recognition based on multimodal information and its application to robot dialogue. *Tsuyoshi Tasaki, Takeshi Yamaguchi, Mitsuhiko Toda, Kazunori Komatani, Tetsuya Ogata, Hiroshi G. Okuno (Kyoto University)*

3.2 理解状態の推定法

相槌から推定された理解状態と頭部ジェスチャから推定された理解状態を取捨選択し、聞き手の理解状態を決定する (図 3)。取捨選択は相槌認識の結果と頭部ジェスチャ結果が矛盾した場合にのみ行う。

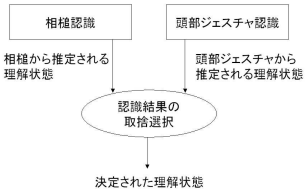

図 3: 理解状態の推定

矛盾した場合、協調したインタラクションを行うために改善すべき理解状態を、あらかじめ指定しておき、その理解状態を示した結果を選択する。

相槌については以下の手順で理解状態の推定を行う。使用した対話データは男性話者 2 名による 3 分程度の実対話 2 つと重点領域研究音声対話コーパス [3] 中の 4 対話の計 6 対話である。

3.2.1 相槌のラベリング法

相槌に与えるラベルは、上記の understand 軸 2 種、active 軸 3 種の組合わせである 6 種類の理解状態とする。

相槌は、相槌を打つ人が、他人に自分の意志を伝えるものであり、かつ人によってさまざまな解釈をする可能性がある。よって本研究では二人にラベリングしてもらい、結果が一致したもののみを正解データとし、学習用データとして用いる。ラベラーが迷った場合には、不明というラベルを付与してもらった。(a 軸は active 軸、u 軸は understand 軸を表す)

理解状態 (a 軸, u 軸)	ラベラー A	ラベラー B	一致データ
(正, 正) のデータ数	61	42	32
(0, 正) のデータ数	273	303	245
(負, 正) のデータ数	4	4	4
(正, 負) のデータ数	2	4	0
(0, 負) のデータ数	14	17	11
(負, 負) のデータ数	0	0	0
不明	22	6	-

表 1: 対話データに対するラベル付けの結果

表 1 から、理解状態が (正, 負) の判定は、人間の場合でも判断がわかれている。また、理解状態が (負, 負) の場合の相槌は、人間同士の協調的な対話ではあまりみられないことがわかる。したがって相槌の認識対象とはしないことにする。また、理解状態 (負, 正) は、人間同士の対話ではあまり見られず、ロボットに対しては、無視をする (行為無し) という形で示されると考えられる。よって実際の対話では、ロボットが、聞き手の行為を予測しているタイミングで無視されたという状況を認識することで、この状態を推定する。具体的には、相槌が呼気段落終了時に打たれやすいという知見を用い、ロボットが間をとることで仮想的に作った呼気段落終了時に連続して閾値回数以上相槌が打たれなければ、無視されたとする。

3.2.2 相槌の認識法とその認識結果

相槌 1 回当たりの発話時間、発話中の初期立上り以上の発話時間、基本周波数概形、有声区間個数、第一有声区間傾斜の 5 つを特徴量とする (図 4)。

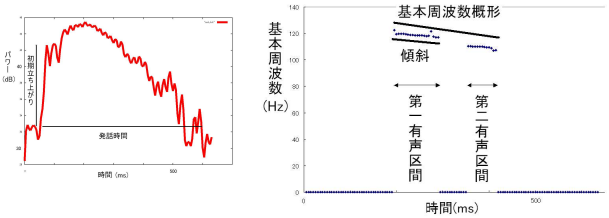

図 4: 相槌認識のための特徴抽出の例

ここで、初期立ち上がりとは発話開始から初めてパワーが落ちるまでをいい、有声区間とは基本周波数がとれた区間をいう。また、基本周波数概形は全ての有声区間をみあわせた全体の傾斜である。

時間による二つの特徴からテンプレートマッチングをし、理解状態をある程度絞り込む。どのテンプレートからの距離も閾値以下のものは noise とした。ここで絞られた理解状態と、周波数による三つの特徴から人手で決定木により理解状態を決定する。今回認識対象とした相槌により理解状態は、(a 軸, u 軸)=(正, 正)(0, 正)(負, 正)(0, 負) の 4 つである。学習データのクローズド実験による認識結果を表 2 に示す。

正解＼認識結果	(正, 正)	(0, 正)	(負, 正)	(0, 負)	noise	認識率 (%)
(正, 正)	23	7	1	0	1	71.9
(0, 正)	13	201	4	17	10	82.0
(負, 正)	0	0	4	0	0	100
(0, 負)	0	0	0	8	3	72.7

表 2: 相槌の認識結果

頭部ジェスチャについては、頭の縦方向の移動と頭の横方向の移動を、オプティカルフローと肌色領域の位置による閾値処理で認識する。認識率は縦振り、横振りともにほぼ 80%であった。

4 おわりに

本報告では、ロボットが対話において人間と協調していくために必要となる、人間の行為と理解状態の推定法について述べた。今後は、人間の行為の時刻のみを用いた物理量からの協調と、理解状態推定による行為の意味からの協調を加えた場合で比較実験を行い、理解状態推定の有効性を示す予定である。本研究の一部は、科研費、21 世紀 COE の支援を受けた。

参考文献

[1] Tomio Watanabe, Masashi Okubo, and Hiroki Ogawa. A speech driven embodied interaction robots system for human communication support. Proc. of IEEE, SMC2000, pp.852-857, 2000.

[2] 泉子.K. メイナード. 会話分析. くろしお出版, 1993.

[3] 平成 6 年度文部省科学研究費補助金重点領域研究. 音声言語概念の統合的処理による対話の理解と生成に関する研究. 音声対話コーパス, Vol.1-4.