

演出を用いたロボット用音声対話システムの設計と実装*

鳴海 真里子[†]
慶應義塾大学 理工学部[‡]

今井 倫太[§]
慶應義塾大学 理工学部[‡]
科学技術振興事業団 さきがけ研究 21[¶]

{narumi, michita}@ayu.ics.keio.ac.jp

1 はじめに

近年、多数の知能ロボットが開発されており、ロボットの持つコミュニケーション能力が注目されている。本研究ではロボットが自発的にコミュニケーションの「演出」を行うことにより、人間同士の対話にみられる「共感」を人間・ロボット間の対話においても実現させる。

1.1 コミュニケーションロボット: Robovie



図 1 Robovie

本研究では人間と対話を行うロボットとして Robovie[1] を用いた。概観を図 1 に、ハードウェア構成を表 1 に示す。Robovie はコミュニケーション機能に重点を置いて設計された擬人化が容易なヒューマノイド型ロボットである。

表 1 Robovie の構成

アクチュエータ	頭部 (3 自由度)、腕 (4 自由度) 2 輪独立駆動方式の車輪・キャスター
センサ	ステレオマイクروفोन 全方位視覚センサ ステレオカメラ (両眼の位置、 2 自由度で注視方向が制御可能) 接触センサ (上半身 18 個下半身 10 個) 超音波センサ (24 個) 赤外線センサ・温度センサ

1.2 共感の実現

本研究ではロボットが人間を同じ状況に注目させるために行う行動を「演出」と定義する。ロボットがある状況 S を感覚 F として認識したとき ($Robot(S, F)$)、人間を共感させるためには、演出により人間を自分と同じ状

況 S に注目させたのち感覚 F を人間に伝える。

$$(S, F) \rightarrow direct(S), affect(F) \quad (1)$$

これにより、(2) 式のように人間が状況 S に対してロボットと同じ感覚 F を持ったとき ($Human(S, F)$)、人間はロボットと共感したとする。

$$\begin{aligned} Robot(S_R, F_R), Human(S_H, F_H) \\ direct(S_R) \rightarrow S_R = S_H \\ affect(F_R) \rightarrow S_R = S_H, F_R = F_H \end{aligned} \quad (2)$$

1.3 目的

本研究の目的は (1) 式で表される演出を用いた音声対話システムを実装し、人間にロボットと共感したと感じさせることである。

2 音声対話システム

(1) 式で表される演出を実現するため、図 2 に示された構造を持つ音声対話システムを実装した。

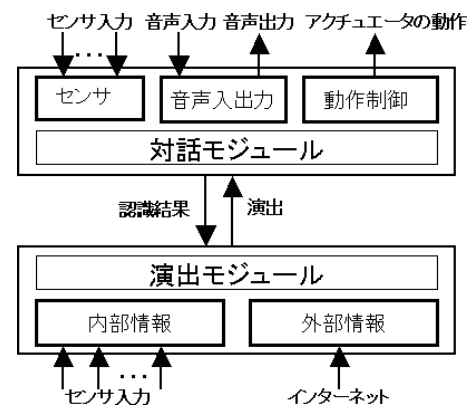


図 2 システム構成

音声対話システムは演出モジュールと対話モジュールで構成される。演出モジュールは与えられた (S, F) に基づいて選択された $direct(S) \cdot affect(F)$ の実行を管理する。対話モジュールは演出モジュールによって決定された行動を実行し、結果をフィードバックする。

2.1 演出モジュール

演出モジュールは、センサ等ロボット自体が収集する内部情報とインターネットに接続して得られる外部情報を保持する。今回入力として用いた内部情報は、接触センサ・温度センサ・超音波センサの 3 種類である。外部情報は特定の HTML ファイルを取り込み解析するこ

*Design and Implementation of Dialog System with Direction for Robot

[†]Mariko Narumi

[‡]Faculty of Science and Technology, Keio University

[§]Michita Imai

[¶]Precursory Research for Embryonic Science and Technology, Japan Science and Technology Corporation

とで情報を抽出する。本システムにおいて $direct(S) \cdot affect(F)$ は以下のように定式化できる。

$$direct(S) \rightarrow act(S, I), sense(R) \quad (3)$$

$$affect(F) \rightarrow act(F, I), sense(R) \quad (4)$$

I は現在の演出で注目させたい状況 S に応じて選択された情報、 R は人間の反応の認識結果である。 $act([S, F], I)$ は状況 S または感覚 F と情報 I に基づいてロボットの行動が決定されることを表す。 $direct(S)$ 及び $affect(F)$ は必要に応じて繰り返される。演出モジュールは状況 S または感覚 F ・情報 I ・反応の認識に関する時間制限を対話モジュールに渡す。

2.2 対話モジュール

対話モジュールは、演出モジュールから受け取った状況 S または感覚 F ・情報 I に基づいて行動（発話及びアクチュエータの動作）を決定・実行し、与えられた時間制限内で得られたセンサ入力（音声認識・接触センサ）から最も重要な意味を持つものを選択・抽象化して演出モジュールに返す。音声認識には Linux 2.2.16-rtl2.2 上で動作させた汎用大語彙連続音声認識エンジン Julian[2] を用いた。

3 動作結果例

ロボットは発話または動作を行った後人間の反応を待ち、反応が得られるか規定の待ち時間が経過した場合に次の演出に移る。対話全体を通じてロボットはアイコンタクトや簡単なジェスチャーを行う。以下に実装したシステムで行われた対話の例を示す。 R はロボット・ H は人間を表す。ただし重要な意味を持たないジェスチャーに関する記述は省略している。

[フェーズ 1]

R1: ねえねえ…おしゃべりしよ

H1: いいよ

R2: わーい…今日、寒いね

H2: (無言)

R3: (アイコンタクトしながら) でもこの部屋はあったかいでしょ

H3: うん

フェーズ 1 において、 $S = \text{「気温」}$ ・ $F = \text{「部屋の気温」}$ に関する共感を得るため以下の行動を行った。

$$direct(\text{気温}) = act(\text{気温}, \text{今日の最高気温}), \\ sense(none)$$

$$affect(\text{部屋が暖かい}) = act(\text{部屋が暖かい}, \text{室温}), \\ sense(agreed)$$

R2 はインターネットの天気予報に基づく屋外の気温に関する発話を行ったが、制限時間内に反応は得られなかった。続けて R3 で温度センサに基づく室内の気温に関する発話を行ったところ、 $sense(agreed)$ を得たことから、

Human(気温, 部屋が暖かい) と言える。

[フェーズ 2]

R4: (絵を指し示して) みてみて、絵があるよ

H4: え?

R5: 山の絵だよ

H5: あー、これね

R6: (アイコンタクトしながら) 緑がきれいでしょ

H6: そうだね

フェーズ 2 では、 $S = \text{「絵」}$ ・ $F = \text{「きれい」}$ に関する共感を得るため、以下の行動を行った。

$$direct(\text{絵}) = act(\text{絵}, \text{見ていない}), \\ sense(not_looked)$$

$$direct(\text{絵}) = act(\text{絵}, \text{見ていない}), \\ sense(looked)$$

$$affect(\text{きれい}) = act(\text{きれい}, \text{緑色の絵}), \\ sense(agreed)$$

R4 は絵を指し示して発話することで絵の存在を知らせる。文脈より「え?」という人間の反応は $sense(not_looked)$ と抽象化される。人間は対象の絵を認識できていないことがわかったので、R5 では具体的にどの絵を指しているのか限定した。人間の反応 ($sense(looked)$) から、 $Human(\text{絵}, Z)$ となる。R6 でアイコンタクトをしながら「緑色の絵がきれい」であることを発話し、 $sense(agreed)$ を得たので $Human(\text{絵}, \text{きれい})$ となり、両者の間に共感が成立したと言える。

4 おわりに

本研究では、ロボットがコミュニケーションを自発的に演出することで人間を共感させる方法を定式化し、実装を行った。動作結果例 R3・H3 や R6・H6 のように伝達されたロボットの感覚 F に対して人間の同意が得られた ($sense(agreed)$) 場合、人間はロボットと共感した ($Robot(S_R, F_R), Human(S_H, F_H), S_R = S_H, F_R = F_H$) ように感じたと考えられる。今後は入力として扱えるセンサ情報の拡張 (ステレオカメラによる顔・視線の方向認識等) や適応例の拡張を行い、心理実験を行って演出の効果についての評価を行う予定である。

参考文献

- [1] Takayuki Kanda, Hiroshi Ishiguro, Tetsuo Ono, Michita Imai and Ryohei Nakatsu: “Development and Evaluation of an Interactive Humanoid Robot “Robovie”,” IEEE International Conference on Robotics and Automation (ICRA 2002), pp.1848-1855, 2002.
- [2] 李晃伸, 鹿野清宏: “複数文法の同時認識および動的切り替えを行う認識エンジン Julius/Julian-3.3”, 日本音響学会研究発表会講演論文集, 3-9-12, 秋季, 2002.