

ロボットとのコミュニケーションへの没入を実現するコミュニケーション技術

今井 倫太

慶應義塾大学、JST Presto

E-mail: michita@ics.keio.ac.jp

鳴海 真里子

慶應義塾大学大学院

Communication Technology for Immersion in Interaction

Michita Imai

Mariko Narumi

Keio Univ., JST Presto Keio Univ.

Abstract: This paper proposes a new communication strategy named Directed Interaction. In usual human-robot interaction, a person looks on the robot's behaviors without responding to them and seldom immerses him/herself in the interaction. Directed Theory achieves person's immersion in interaction by generating an affective utterance which stimulates a person's feeling about a physical object. The innovative idea of Directed Interaction is to employ the actual physical objects to mediate the human-robot interaction. In particular, since the affective utterance relates directly to a sort of human sensory perception, it is easy to develop a robot system to stimulates the person's feeling.

1 はじめに

多数のコミュニケーションロボットが開発され、ロボットの人間社会での活躍が期待されている [4] [7]。しかし、実用に耐え得るコミュニケーションロボットを構築するには、音声やジェスチャを用いた対話手法をさらに向上させる必要がある。本稿では、人間社会で実用に耐え得るコミュニケーションロボットの実現を目指し、人間とロボットのマルチモーダル対話について取り扱う。

人間とロボットがコミュニケーションする上で最初の障壁となるのは、ロボットとのコミュニケーションに人間自身が没入しないことである。ROBODEX[12]に代表されるロボットの博覧会にロボットを展示すると、上記の障壁が特に明白になる。博覧会に訪れる多くの人々は、ロボットからの問いかけや振る舞いに対して一言も返答せずに、ロボットの反応を観察する。一方で、ロボットからの問いかけや振る舞いに対して自然に対応する人も入場者の中にはいるのも事実である。ロボットにどのようなコミュニケーション機能を搭載すれば、より多くの人間がロボットに対して自然に受け答えをし、コミュニケーションが持続するようになるのだろうか？この問題は、音声対話や画像認識、行動制御といったロボットの各要素技術の影に隠れ、今まであまり取り扱われてこなかった。しかし、コミュニケーションが成立して初めて各要素技術が役に立つのであり、コミュニケーションロボットを構築する上

で避けて通れない重要な問題である。

コミュニケーションロボットに関する研究は、技術的な側面から社会心理学的な側面までさまざまな観点から行われてきた。例えば、人型ロボット ROBITA [8] は、話者が誰かを同定し、複数の人間とのグループ対話を可能にしている。また、人型ロボット Kismet [3] や Infanoid [6]、SIG [10] は、ロボットの視線を用いて注目対象の表出やアイコンタクトを実現している。以上の機能で実現される振る舞いは、社会性を持ったロボットの実現につながり、人間とのコミュニケーションを成立させる上で重要である。また、コンピュータ上で扱われる自律エージェントの社会性の研究も行われている。自律エージェントの研究では、社会的に意味のある決まり文句を使ったコミュニケーション [2] や、個性が現れかつ無意識の内に人間が反応してしまう要素 (例えば声の個性)[9] を積極的に用いる手法が試みられている。

上記の手法は、人間をコミュニケーションへと没入させるのに不十分である。特に、ロボットとコミュニケーションする意図を人間が持っていない場合にその不十分さは顕著である [14]。ロボットに対するコミュニケーションの意図の欠如に関する問題は、将来ロボットが日常生活の中で人間と偶然出会い、人間に話かけるといった状況での最重要課題である。この状況で注目すべき点は、ロボットがコミュニケーションを主導するところである。ロボットとコミュニケーションするように指示する第三者 (例えば説明員) がいない。一方、

ROBITA に代表される対話ロボットと人間をコミュニケーションさせるためには説明員が必要である。Kismet や Infanoid、SIG で扱われる視線による引き込みは、人間が反射的に反応する社会的シグナルであり、説明員が不在でも人間をコミュニケーションに引き込める可能性がある。しかし、社会的シグナルは、我々のロボット [5] も採用しており、それでもなおコミュニケーションに没入しない人間がいるのが事実である。自律エージェントの研究では、社会的シグナルの代わりに、決まり文句といった言語による引き込みを狙っている。しかし、動的な実環境で行動するロボットにとって、状況を適切に判断し決まり文句を生成するのは至難の技である。つまり、決まり文句の生成の欠点は、言語的文脈に強く依存している所である。また、声の個性の利用は言語や文脈に依存しないが、コミュニケーションの主導をロボットが取るにあたって直接的な問題可決とはならない。

本稿では、ロボットとのコミュニケーションに人間を没入させるために演出を用いたコミュニケーション戦略を提案する。具体的には、実世界の状況や物に対する人間の五感を利用してロボットのコミュニケーションを演出する。五感を利用した演出において特に注目すべき点は、人間の五感が、言語的な意味や文脈から完全に独立している点である。

演出は、二つのステップから構成される。最初に、演出の対照となる実世界の物や状況に人間の注意を向ける行動や発話を生成する。その後、その状況や物に対して人間が感じているであろう感覚に関する発話(感覚的発話)を生成する。例えば、最初のステップとして、人間にチョコレートを食べるようにロボットが勧め、次のステップで、「美味しいでしょう?」と人間が感じる感覚を表現した感覚的発話を生成する。感覚的発話は、ロボットに対する人間の共感を引き起こしやすくするため、ロボットとのコミュニケーションへと人間を没入させる。

本稿では、演出を用いたコミュニケーションロボットを構築し、コミュニケーションの没入に演出が与える効果について被験者実験により調べる。

2 コミュニケーションへの没入

2.1 ヒューマン・ロボットインタラクションの理想形

人間とコミュニケーションロボットのインタラクションは、コミュニケーションへの没入といった観点から見ると大きく二つに分かれる(図1)。図では、ロボットが人間にお菓子をあげている場面である。図中の左の人間は、ロボットからの働きかけに対して、何を返答しているのかわからず無言になってしまっている。一方、図中の右の人間は、「ありがとう」と真剣に答えている。コミュニケーションへの没入といった意味

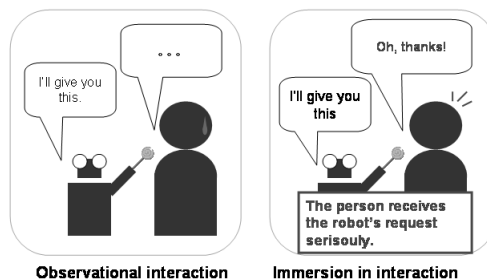


図1: 人間とロボットのコミュニケーション

では、右側のコミュニケーションが理想的である。しかし、人間は、大抵ロボットの発話を真剣に受け止めることはない。この違いが、人間同士のコミュニケーションとロボットとのコミュニケーションの大きな違いである。本稿では、右図に示すコミュニケーションの没入を実現するコミュニケーション戦略にどのような要件が必要なのか考える。

2.2 コミュニケーションへの没入とマインドリーディング

図1のコミュニケーションの違いは、人間が、ロボットとのインタラクションに没入しているかに依存する。図1の左のインタラクションでは、人間がロボットの動きを客観視(Calm Viewpoint)しており、ロボットとのインタラクションであることを意識している(図2の左の四角)。一方、インタラクションに没入しているときには、インタラクションのやりとり自体に集中しており、客観的視点は消えていると言える(図2の右の四角)。つまり人間の視点をVとすると、コミュニケーションの没入を実現するためには以下の式を実現する必要がある。

$$\text{Rule 1 } V = \text{calm_viewpoint} \rightarrow V = \text{interaction_viewpoint}$$

Rule 1 を実現するために、人間同士のコミュニケーションと人間とロボットのコミュニケーションをより詳しく比較する。図3の(a)は、人間同士のコミュニケーションを表している。人間は、他人の発話を理解する際に相手の心を読み、発話の意図や、相手が知覚している状況を察する [13] [1]。つまり、インタラクションに没入している時に人間は相手の心を読む視点を取っており、客観的な視点が消滅している。

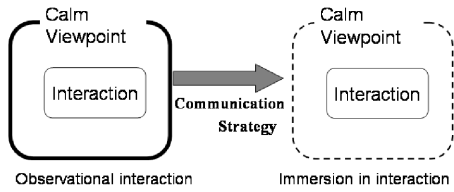


図 2: インタラクションにおける視点

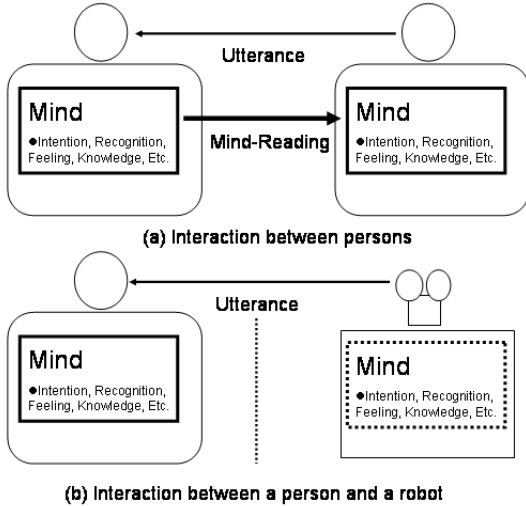


図 3: 人間同士と人間とロボットのコミュニケーションの違い

図 3 の (b) は、人間とロボットのコミュニケーションを表している。ただ単純にロボットが何かを発話しても、人間とロボット間に関係がない場合コミュニケーションへの没入が誘発されず、マインドリーディングは起こらない [11]。結果、人間は、客観的な視点を取り続ける。

図 3 で示した比較は、ロボットとのコミュニケーションへ人間を没入させるためには、ロボットの心を人間に読ませるコミュニケーション戦略が必要であることを示している。ここで、我々はロボットに心があると主張しているわけではない。人間は、ロボットとのコミュニケーションに没入している際に心を持っているかのようにロボットを扱うことを説明しているのである。

2.3 コミュニケーションロボット Robovie

本稿で用いたコミュニケーションロボット Robovie を図 4 に示す。Robovie は、片腕に 4 自由度、首に 3 自由度持ち、視線の動きとジェスチャを用いて人間とコミュニケーション可能である。また、車輪で場所を移動できる。

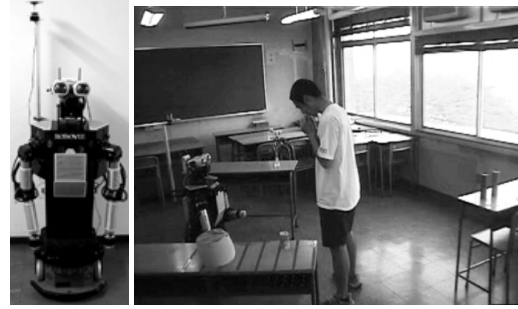


図 4: コミュニケーションロボット Robovie

3 演出によるインタラクション

本稿では、ロボットとのコミュニケーションへと人間を没入させるためのコミュニケーション戦略としてコミュニケーションの演出を提案する。具体的には、ある実世界の物や事象に対して人間を注目させ、そこで人間が感じる五感に関する発話 (感覚的発話) をロボットが生成する。演出されたコミュニケーションを用いて、ロボットの心を人間に読ませ、コミュニケーションへの没入を実現する。人間は、感覚的発話を受け取ると、現在自分が感じているものと同じ感覚をロボットが持つと思い、ロボットの心を読む視点を自然と持つことになる。ただし、ロボットは、実際に感覚を持っているわけではないことをここで強調しておく。ロボットが感覚を持っているのではなく、感覚的な発話を与えられた人間が、ロボットが感覚を持っているとあくまでも想像しているだけである。

また、人間がどんな五感を知覚しているのかを、ロボットが状況を確認し、推定するのは困難である。しかし、演出では、予め人間を特定の状況へと注目させているので、人間が知覚している対象をある程度限定することが可能となっている。

3.1 感覚的発話によるマインドリーディング

演出の戦略を式で表すと以下の通りとなる。

$$Direction(Physical_event), Utterance(H_feeling) \quad (1)$$

ここで、*Physical_event* は、実世界のイベント (周囲の物やイベント) を表す。例えば、天候、食べる行為、何かを見る行為、何かを触れる行為である。また、*Direction(Physical_event)* は、*Physical_event* を発生させるロボットの行動を表す。例えば、天気への注意喚起や、食べることを勧めること、何かへ視線を向けることを勧めること、何かへ触れることを勧めることといったものがある。

また、*H_feeling* は、人間が感じている感覚を表す。*Utterance(H_feeling)* は、*H_feeling* を表現する感覚的発話を生成することを示す。

戦略 (1) は、最初 *Direction(Physical_event)* によって実世界のイベントに人間を注目させる。例え

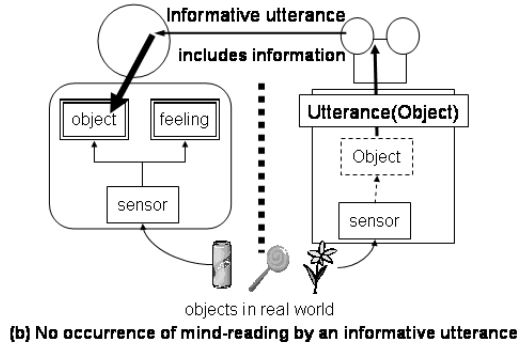
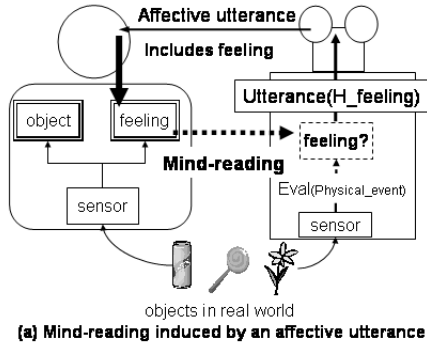


図 5: 演出によるコミュニケーション

ば、人間にお菓子をあげながら、「食べて！」と勧める行動である。つまり、 $Direction(Physical_event)$ は、実世界のイベントに対して人間の五感を向ける。次に、 $Utterance(H_feeling)$ によって、人間が $Direction(Physical_event)$ によって感じると予想される感覚 $H_feeling$ に関する感覚的発話を生成する。例えば、「美味しいでしょ？」といった物がある。

図 5(a) は、ロボットの心を人間が読む過程を示している。図では、ロボットが $Direction(Physical_event)$ の実行した後に $Physical_event$ に対する感覚を人間が持った場面である。また、ロボットが、感覚的発話 $Utterance(H_feeling)$ を生成している。例えば、お菓子の場合、「美味しいでしょ？」が $Utterance(H_feeling)$ に相当する。次に、人間は、与えられた感覚的発話 (Affective Utterance) を、お菓子自体に対する感覚 (feeling) と結びつける。さらに、同じ感覚をロボットが持っていると思像する (図 5(a) の点線矢印)。この想像が、人間の視点を客観的なものからインタラクションの視点へと変える。つまり、**Rule 1** の実現につながる。

3.2 情報発話によるインタラクション

図 5(b) は、ロボットとのコミュニケーションに人間が没入していない場面を示す。図では、以下のコミュ

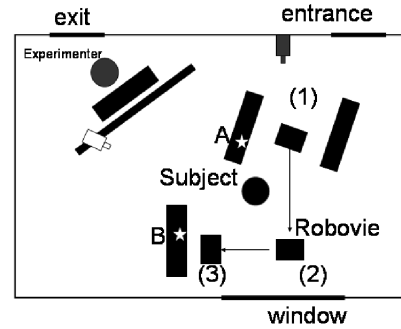


図 6: 実験環境

ニケーション戦略にしたがって発話が生成されている。

$Direction(Physical_event), Utterance(Object)$ (2)

戦略 (1) と (2) の違いは、感覚的発話を用いるかどうかである。戦略 (2) では、実世界の物やイベントを説明する情報の発話 ($Utterance(Object)$) を代わりに生成している。例えば、お菓子に関する情報の発話には「キットカットだよ。」といったお菓子の名前を教える物がある。図 5(b) では、人間がロボットの心を読む現象が起こらない。なぜなら、情報の発話 (Informative Utterance) は、人間の感覚よりも物体自体の認識 (Object) と結びつけられるからである。

4 実験

本章では、人間のロボットのコミュニケーションに対する演出の影響を検証するために行った被験者実験について述べる。つまり、感覚的発話によって人間はコミュニケーションへと没入するかどうか調べる。

4.1 実験環境

図 6 は、実験環境を示す。図中の (1) から (3) は、インタラクション中の Robovie の場所を示す。A(お茶) と B(お土産) は、演出 $Direction(Physical_event)$ の際に Robovie が用いる小道具を表す。また、Robovie は手にチョコレートを持っている。実験中被験者は、実際にお茶を飲んだりやチョコレートを食べたりできる。

4.2 被験者

男女の大学生 21 名の被験者を、実験群 11 名・対照群 10 名にランダムに配分する。

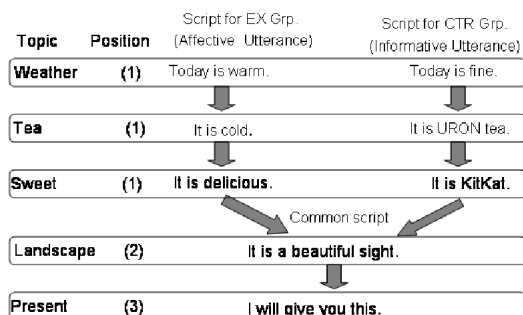


図 7: 実験で用いたトピックと発話

4.3 実験条件

実験には二つの条件が用意されている。一つは、被験者が感覚的発話を与えられる条件である。この条件は、実験群に与えられる。もう一つは、対照条件であり被験者が情動的発話を与えられる。この条件は、対照群に与えられる。図 7 は、演出のトピックとロボットが生成する発話を示している。天気とお茶、お菓子のトピックは、実験条件に従って実験群と対照群に異なる発話を与えられる。その後の二つのトピック(景色とお土産)は、実験群と対照群共に共通の発話となる。

4.4 実験手順

実験は、以下の 4 ステップで行う。

1. 実験者は被験者をロボットの前に連れていき、「ロボットの相手をするように」と教示して退出する。
2. ロボットが動作を開始し、実験条件によって表に示した異なる対話シナリオに基づいて被験者とインタラクションを行う。このインタラクションは、図 6 の (1) の位置で、周辺の机の上に置かれた小道具を用いて行われる。
3. 各実験条件に共通な対話シナリオでインタラクションを行う。その後図 7 に示す各条件に共通なインタラクションを、図 6 の (2) と (3) の位置に移動し、小道具 B を用いて行う。対話シナリオの最後で、ロボットが被験者に出口から退出するよう指示する。
4. 退出した被験者は、別室でロボットとのインタラクションについての質問票に答える。

評価は、被験者の行動観察と質問票への回答を用いて行う。実験では、感覚的発話の効果が現れるのに時間がかかることを考慮し、図 7 の 3~5 目目のトピックでの被験者の行動に注目する。質問票では、ロボット

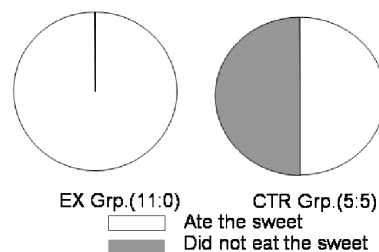


図 8: お菓子に関する被験者の反応

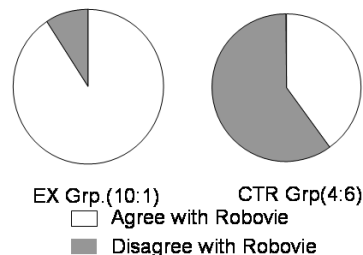


図 9: 景色に対する被験者の反応

の発話の理解度・自分自身のロボットに対する反応とその理由について、被験者に回答を求めた。

4.5 実験結果

3 目目のトピックにおける結果: 図 8 は、お菓子を食べた被験者の数を実験群と対照群に分け示した。また、分散分析によって実験結果には有意差があることが分かった ($p < 0.05$)。結果は、実験群の被験者が、Robovie の意図を読みお菓子を食べたことを示している。図 4 は、感覚的発話の影響によって被験者がお菓子を食べている場面である。

4 目目のトピックにおける結果: 図 9 は、Robovie の景色に関する発話に同意した人数を示している。この結果に対しても有意差が得られた ($p < 0.05$)。結果、実験群の被験者は、Robovie の意図を読み、景色に関する発話に同意したと言える。

5 目目のトピックにおける結果: 実験群と対照群の被験者双方ともロボットの発話に従って机の上のお土産を手にした。図 10 は、実際に出口までお土産を持って退出した被験者の人数と、机にお土産を置いて帰った人数を示している。この結果に対しては、実験群がお土産を持って帰りやすい傾向があることが分かった ($p < 0.1$)。また、お土産を置いて帰った被験者に理由をアンケートで質問したところ、「これは実験だから。」、「次の実験でお土産を使うと思ったから。」と実験を意識した解答が得られた。

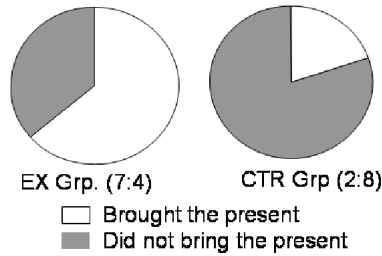


図 10: お土産に対する被験者の反応

以上の結果から、実験群の被験者は対照群の被験者よりも「ロボットの相手をする」という役割を確実に演じており、インタラクションへ没入していたと言える。

5 まとめ

本稿では、ロボットとのコミュニケーションへと人間を没入させるために、コミュニケーションの演出を提案した。具体的には、ロボットが、実世界の状況や物に人間を注目させ、その際に人間が持つ五感を表現した感覚的発話を生成する。感覚的発話の影響によって、人間は、ロボットに対して共感し、自然とロボットとのコミュニケーションに没入することとなる。また、本稿では、実際に演出がコミュニケーションの没入を実現することを被験者実験によって検証した。

参考文献

- [1] S. Baron-Cohen. *Mindblindness: An essay on Autism and Theory of mind*. Mit Press, 1987.
- [2] T. Bickmore and J. Cassell. Relational agents: A model and implementation of building user trust. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 396–403, 2001.
- [3] C. Breazeal. *Designing Sociable Robots (Intelligent Robots and Autonomous Agents)*. MIT Press, 2002.
- [4] T. Kanda, H. Ishiguro, M. Imai, and T. Ono. Body movement analysis of human-robot interaction. In *International Joint Conference on Artificial Intelligence (IJCAI 2003)*, pages 177–182, 2003.
- [5] T. Kanda, H. Ishiguro, M. Imai, T. Ono, and K. Mase. A constructive approach for developing interactive humanoid robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002)*, pages 1265–1270, 2002.
- [6] H. Kozima and A. Ito. Towards mindreading by an attention-sharing robot. In *Proc. of third int. Symp. on Artificial Life, and Robotics (AROBO III' 98)*, pages 478–481, 1998.
- [7] Y. Kuroki, T. Fukushima, K. Nagasaka, T. Moridaira, T. Doi, and J. Yamaguchi. A small biped entertainment robot exploring human-robot interactive applications. In *12th IEEE Workshop Robot and Human Interactive Communication (RO-MAN 2003), CD-ROM 5A5*, 2003.
- [8] Y. Matsusaka, S. Fujie, and T. Kobayashi. Modeling of conversational strategy for the robot participating in the group conversation. In *Proc. ISCA-EUROSPEECH2001*, pages 2173–2176, 2001.
- [9] C. Nass and K. M. Lee. Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction and consistency-attraction. *Journal of Experimental Psychology: Applied*, 7:171–181, 2001.
- [10] H. G. Okuno, K. Nakadai, and H. Kitano. Non-verbal eliza-like human behaviors in human-robot interaction through real-time auditory and visual multiple-talker tracking. In *Proceedings of the Third International Workshop on Cognitive Robotics (CogRob-2002)*, pages 59–65, 2002.
- [11] T. Ono and M. Imai. Reading a robot's mind: A model of utterance understanding based on the theory of mind mechanism. In *Proceedings of AAAI-2000*, pages 142–148, 2000.
- [12] ROBODEX. <http://www.robodex.org/>.
- [13] D. Sperber and D. Wilson. *Relevance: Communication and Cognition*. Oxford: Basil Blackwell, 1986.
- [14] 今井, 小野, 中津, 安西. 協調伝達モデル: 関係性に基づくヒューマンロボットインタフェース. *信学論 A*, J85-A(3):370–379, 2002.