

三者対話における遠隔操作型ロボットへの半自律視線制御支援システムの提案

A model of eye gaze in social robots for three-party interaction

新谷 太健^{1,2*} 石井 カルロス寿憲²
Taiken Shintani¹ Carlos T. Ishi²
劉 超然² 石黒 浩^{1,2}
Chaoran Liu² Hiroshi Ishiguro¹

¹ 大阪大学 大学院基礎工学研究科 システム科学領域

¹ Systems Science and Applied Informatics, Osaka University

² 国際電気通信基礎技術研究所

² Advanced Telecommunications Research Institute International

Abstract: In recent years, remote control of interactive agents has been studied. Gaze control is one important factor in dialogue robots. In this research, we propose a dialogue agent system that generates appropriate gaze movements using "speech act" and controls gaze to support remote operators in order to realize a smooth dialogue in three-party conversation.

1 はじめに

遠隔地の人との会話において、携帯電話やスマートフォンを用いた対話は日常的に行われており、離れた位置に住む人とも通話することが可能となっている。近年、この通話方法について携帯電話での対話の拡張として遠隔動作型対話エージェントの研究が盛んに行われている。たとえば、Telenoid は携帯電話を用いて音声のみで通話するより対話エージェントを用いて通話を行った方が有益であることを示している [1]。また、技術の進歩によりモーショントラッキングや顔検出の技術の進歩によりバーチャルエージェントが操作者の動きや表情をリアルタイムで再現・生成を行うことが可能である。このように、携帯電話以外のエージェントを用いた通話手法が社会に現れ始めている。この対話エージェントが操作者の声のみから自動で動作生成することができる、操作者は従来の電話を使用する感覚で対話を行うことができるようになり、また、対話エージェントと会話を行うユーザはヒト同士で話す感覚で対話を行うことが可能になる。例えば、車内から知人らと会話するとき、操作者は携帯電話・マイクから通話を行い、対話を行う知人らは対話エージェントを通じて動的な対話を楽しめる。

人同士の対話を行うように、対話エージェントに求

められる機能として各関節の動き、ジェスチャーの生成、表情の生成、視線の生成など様々な要素があるがそれらのうち、視線の動きに重心を当ててみる。

視線の動きは対話を円滑に進めるうえで最も重要な要因の一つであり、視線に関する研究は様々になされている。Kendon や Argyle らは視線には会話の流れを制御する機能、対話相手にフィードバックを与える機能、感情情報を伝える機能、共存感覚を与える機能、対話相手に正もしくは負の評価を与える機能があると論じている [2] [3]。

そのため、対話エージェントにおいて視線制御は重要な機能となるため、視線を扱っていく。

次に、対話においても様々な対話状況が想定される。1対1の対面対話に関しては数多くの研究者が研究しているが、三者対話以上の複数人対話を扱っているものは少ない。2者対話における視線制御システムは対話相手一人のみを考慮すればよく、話し手・聞き手、ターンテイキングの認識が比較的容易であるが、三者対話では対話相手のどちらに視線を向けるべきであるのか、目を逸らすべきであるのか次に誰が話してよいのかターンテイキング認識などが困難なものになる。そのため、視線制御システムには2者対話とは別の戦略が必要である。

談話機能と動作の関係性は、頷きや首傾げなどの頭部動作や、ジェスチャーの出現頻度と関連していることが報告されている [4-6]。

また、これらの研究では、ミニマルデザインのテレ

*連絡先：大阪大学 大学院基礎工学研究科 システム科学領域
〒560-8531 大阪府豊中市待兼山町 1-3
E-mail: shintani.taiken@irl.sys.es.osaka-u.ac.jp

ノイドや、人に酷似したアンドロイド ERICA などを用いて動作生成の評価を行っている。

本研究では、談話機能と視線パターンの関連性について対話データを用いて分析を行った。本研究で用いる談話機能の分類は、[4,5] で使用したラベルセットに基づいている。

また、本研究では様々存在する対話エージェントのうち図 1 に示す社会的会話ロボットである CommU に実装していく。



図 1: 社会的会話ロボット CommU

2 ベースラインシステム

本章ではベースラインシステムの説明を行う。Mutluらは、体面对話、三者対話における発話者のデータセットを収録し、データセットより二人の人と視線をそらす際に対話者二人の間を見る前提のもと、どのような割合でみるべきかという割合のモデルを提案している [7]。Kinect により顔の向きを推定し、これを視線の方向として視線の分布を作成し、割合を算出している。Mutluの論文では、対話相手それぞれの顔、体、対話相手の間の空間の合計 5 箇所のうち、どれだけ見ていたかの割合を載せていたが、本研究のベースラインシステムでは、すべての対話相手の顔と体を見ていた割合の合計を半分にし、見る割合を算出している。この時の、対話相手 A、対話相手 B、対話相手間を見る割合は 35%、35%、30%である。

3 解析

実際に ATR で収録され、複数人の人の手によってラベル付けされた三者対話のデータセットより、スピーチアクトと視線に関連性があることを確かめる。表 1 に詳細を記す。対話形式は三者対話であり、全部で 6 セッションのフリートークである。1 セッション当たり 10 分から 20 分になり、同じ人が別の組み合わせで収録したこともあったため、本研究で用いたデータセットでは全部で 9 人の対話者が参加している。

表 1: 3 者対話のデータ

対話形式	三者対話
セッション数	6 セッション
対話内容	フリートーク
1 セッションの時間	10-20 分
対話者数	9 人

3.1 スピーチアクト

ここでは、談話機能であるスピーチアクトについての説明を行う。本研究では発話について、以下の 5 つに分類して考えていく。

- 相槌：「はい」や「なるほど」のように、相手の発話に対する反応としての短い発話内容を相槌と分類する。
- フィラー：「えーと」や「あー」のように、考えている最中に出る短い発話をフィラーと分類する。単語によっては、相槌の意味も兼ねることがある。
- 文の途中：相槌、フィラー以外の発話文のうち、文の途中で割り当てられる発話を分類する。
- 文末（疑問）：相槌、フィラー以外の発話分のうち、文の末尾にあたる部分。特に、疑問形になっているところを分類する。
- 文末（疑問以外）：相槌、フィラー以外の発話分のうち、文の末尾にあたる部分。特に、疑問形になっていないところを分類する。

では、これらの談話機能と視線の動きのパターンの解析結果を図 2 に示す。ここで、対象とする発話のうち p1 は最初に見た人、p2 は次に見た人を指し、env は人以外を見た時を指している。また、env/p1 のような表記の時は、対象とする発話において、人以外の位置を見る、つまり視線をそらした状態から人を見た様子を示している。

3.2 分析結果

スピーチアクト全体

図2左上に示すように、談話機能つまりスピーチアクト全体としては聞き手を見ながらもしくは視線を逸らすことが多く、その次に視線をそらした状態から聞き手の方を見る傾向にあることがわかる。

相槌

相槌の時は図2上段中央に示すように聞き手の方を見ながらもしくは視線を逸らした状態から聞き手の方を見る傾向にある。

フィラー

フィラー時は図2右上に示すように、視線を逸らす傾向にあることがわかる。

発話の途中

発話の途中は図2左下に示すように、聞き手の方を見る、視線を逸らす、逸らした状態から聞き手を見るなど様々なパターンが考えられる。

文末(疑問)

文末において疑問となるときは、図2下段中央に示すように、聞き手を見る傾向にあることがわかる。

文末(疑問以外)

文末において疑問以外となるときは図2右下に示すように疑問時と比べて聞き手を見る割合が減っていることが見て取れる。

4 提案システム

この章では提案システムについての説明を行う。先ほどの章より、スピーチアクトによって、視線パターンの分布が異なることが分かった。そのため、各スピーチアクトごとに視線パターンの分布を作成し、実際の対話中に得られたスピーチアクトから、視線パターンを生成する。

図3に実験の様子を示す。一つの部屋に CommU と 2 人の対話相手を用意し、CommU の操作者は別の部屋からこの実験者二人を含めて 3 人で対話を行う。

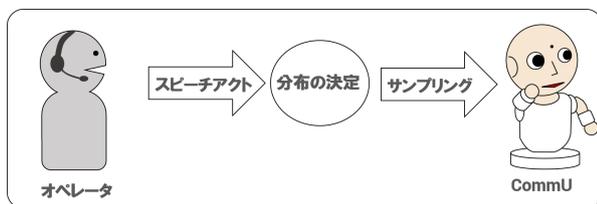


図 3: 提案システムの概略

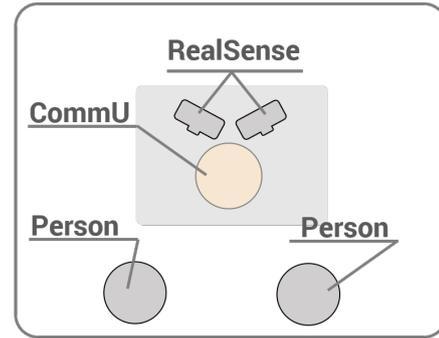


図 4: 実験の様子

5 おわりに

本稿では、3 者対話データの解析に基づいて、発話のスピーチアクトと視線の動きのパターンの関連を示し、実際に視線生成システムの提案を行った。まずは、データベースの音声を用いて、視線生成をベースラインシステムを用いたものと、提案モデルを用いて視線の動きを生成したものとを比較する予備実験を行い、その有効性を確認できた。今後はより詳細な評価実験を進める予定である。

なお、本研究には多くの課題がある。例えば、実際にスピーチアクトをリアルタイムの発話から判別する研究は少ないため、この識別システムの作成が必要である。

謝辞

本研究は、JST/ERATO (JPMJER1401) および JST/MIRAI (JPMJMI18C6) の助成を受けたものである。データ分析にご協力いただいた村瀬妙子氏および中西京子氏に感謝する。

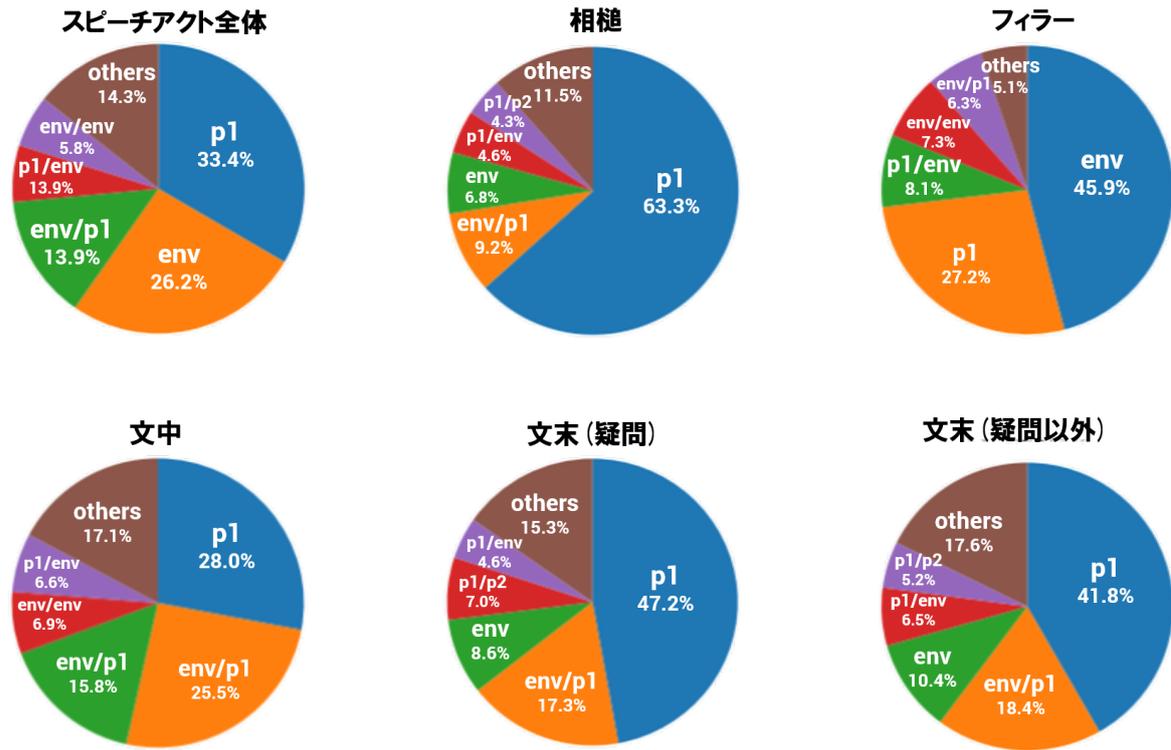


図 2: 視線パターン

参考文献

- [1] Kohei Ogawa, Shuichi Nishio, Kensuke Koda, Koichi Taura, Takashi Minato, Carlos Toshinori Ishii, and Hiroshi Ishiguro. Telenoid: Telepresence android for communication. In *ACM SIGGRAPH 2011 Emerging Technologies*, SIGGRAPH '11, New York, NY, USA, 2011. Association for Computing Machinery.
- [2] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, Vol. 26, pp. 22–63, 1967.
- [3] Michael Argyle and Mark Cook. Gaze and mutual gaze. 1976.
- [4] Carlos T Ishi, Daichi Machiyashiki, Ryusuke Mikata, and Hiroshi Ishiguro. A speech-driven hand gesture generation method and evaluation in android robots. *IEEE Robotics and Automation Letters*, Vol. 3, No. 4, pp. 3757–3764, 2018.
- [5] Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Commun.*, Vol. 57, pp. 233–243, February 2014.
- [6] Chaoran Liu, Carlos T. Ishi, Hiroshi Ishiguro, and Norihiro Hagita. Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pp. 285–292, New York, NY, USA, 2012. Association for Computing Machinery.
- [7] Bilge Mutlu, Takayuki Kanda, Jodi Forlizzi, Jessica Hodgins, and Hiroshi Ishiguro. Conversational gaze mechanisms for humanlike robots. *ACM Trans. Interact. Intell. Syst.*, Vol. 1, No. 2, pp. 12:1–12:33, January 2012.