

複数操作可能な半自律エージェントシステムの試作

Prototyping Semi-Autonomous Communication Agents for Multiple Operation

峯岸朋弥¹ 大澤博隆¹

Tomoya Minegishi¹ and Hirotaka Osawa¹

¹筑波大学

¹University of Tsukuba

Abstract: 本研究では、一人の操作者が複数のコミュニケーションエージェントを操作し聞き手側から見て複数エージェントが情報を提示しているように見せるために、操作者が1つのエージェントを通して発話している間、操作していない他のエージェントは、対話に付随する相槌などの身体動作を自動生成する手法を提案する。現在までにグローブ型VRデバイスを用いて合計2台のコミュニケーションエージェントを操作可能なシステムを開発した。本論文では、開発したエージェントの紹介と今後の研究計画を述べる。

1 序論

近年、アクティブラーニングと呼ばれる学習手法の導入が教育現場で行われている。学習者はアクティブラーニングにより、能動的に学習することで、認知的、倫理的、社会的能力、教養、知識、経験を含めた汎用的能力の獲得が期待できる。文部科学省は、グループディスカッション、ディベート等も有効なアクティブラーニングの方法である[1]としている。参加者同士のディスカッションを行うためには複数の参加者が必要であるが、これを複数のコミュニケーションロボットを用いて擬似的に行っても同様の効果を得られ、学習者の数が少ない場合でもアクティブラーニングによる能力獲得ができる可能性がある。

アクティブラーニングのグループディスカッションやディベート等にロボットを活用するため、コミュニケーションロボットを複数用いることに着目した。コミュニケーションロボット（以下、エージェント）を複数用いるシステムでは、人がエージェントに注目し、それにより理解が深まり[2]、情報を受け取りやすくなり[3]、エージェントが人の話に正しく応答し、深く議論できていると感じる[4]ことが示されている。複数エージェントを教育現場に応用した場合、人の学習意欲向上、タスクパフォーマンス向上[5]することが分かっている。ディスカッションに導入した例では、多数派と少数派の人数が不均衡な際に発生するバイアスを、複数エージェントを参加させることで低減できることがわかっている[6]。しかしながら、複数エージェントを一人が同時に操

作することは難しく、自律動作エージェントの研究例を複数のエージェントの研究にそのまま当てはめることはできない。現在の技術では、設計者が予め定めたシナリオを元に複数台を操作した成功例しかない[7]。例えば一人の操作者が同時に複数操作するならば、発話や身体動作を同時に複数オペレーションする必要がある。一方で、複数の操作者が複数エージェントを操作するならば、発話状態や身体動作状態など、お互いの状態を把握しながら自身のエージェントを操作しなくてはならないため、稼働数が増えるほど難しくなる。人が複数台を操作し、それが聞き手から見たときに複数エージェントが存在するように見せるためには、操作者への認知負荷が高い問題を解決する必要がある。

本研究の最終目的は、一人の操作者が複数エージェントを操作し聞き手側から見て複数エージェントが

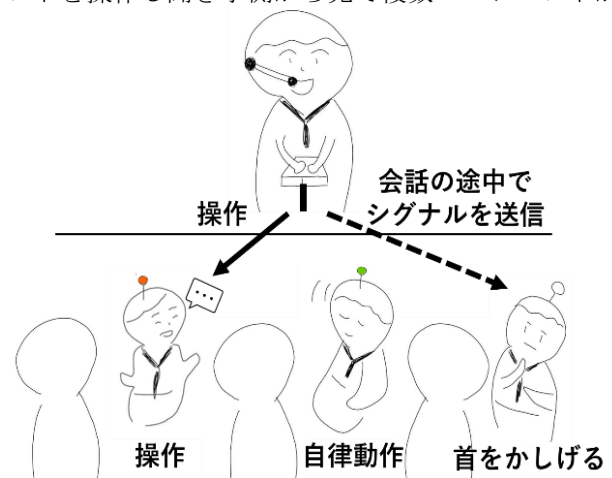


図 1 複数エージェント動作の様子

情報を提示しているように見せるために、操作者が1つのエージェントを通して発話している間、操作していない他のエージェントは、対話に付随する相槌などの身体動作を自動生成する手法を提案することである。これによって、提示情報に素早く注目させ、聞き手の理解を深めることができる。この目的のため、一人の操作者が同時に2つ以上のエージェントを操作できるインタフェース開発を達成する(図1)。また、操作者が1つのエージェントを通して発話している間、操作していない他のエージェントは、対話に付随する相槌などの身体動作(以下、受動的振る舞いと呼ぶ)を自動生成システムの開発を達成する。話者エージェント以外にシグナルを送り、操作者が任意のタイミングで首を傾げる等の動作を行えるシステムの開発を達成する。

本論文では、2章で関連研究を示し、3章で現在までに開発したバーチャルエージェントの紹介を行う。4章では今後の研究計画を述べる。

2 関連研究

2.1 ロボットを活用したアクティブラーニング

アクティブラーニング教材として、ロボットを活用した方法が提案されている。ロボットに学習者が触れたり対話したりするインタラクションを行うことで学習効果を得る手法[8]や、ロボットと一緒に学習に取り組むことで学習者の意欲向上を狙う手法[9]などがある。これらアプリケーションは、教師がいない場合でも学習を一人ですすめることができること、従来の紙媒体による教科書と異なり、インタラクションを行いながら学習が行えることの利点が挙げられる。一方で、予め設計者が定めた内容の学習に限られることや、ロボットが学習者に飽きられてしまう欠点がある。

ロボットが教師役として学習者に教えることや、学習者と一緒に学ぶことは、学習者に飽きを感じさせる。Kandaらは2台のロボットを教師役として小学校に導入し、英語教育を行った[8]。最初の1週間目は盛んなインタラクションが行われていたが、2週間からはインタラクション回数が減少した。学習者はロボットの教師に飽きてしまうことが分かった。Jimenezらは、学習者と同じように学習に取り組むロボットを提案した[9]。学習を進めるうちにロボットと共に学習したほうが正答率の上昇が見られたが、学習者の学修達成率が上昇するにつれロボットを無視しだすことがわかった。Tanakaらは、学習者から

指導を受けるロボットを提案した。学習レベルが低いロボットに教えることで、学習者自身が学習できることから自発的学習を促した[10]。以上より自律ロボットをアクティブラーニングに活用することは、学習者に学修度の上昇が期待できるが、時間が経つにつれ飽きを感じさせることがわかる。

2.2 複数エージェントによるインタラクション

複数エージェントを使用することで、聞き手のタスクパフォーマンスが向上する。Shiomiらは物理的なロボットやオンスクリーンエージェントと人のインタラクションにおける影響を調べた[11]。物理的なロボットやオンスクリーンエージェントが人へ褒め言葉を贈ることによる人のタスクパフォーマンスやタスク中の感情について調査し、2台のエージェントから称賛された参加者は、1台から称賛された場合より有意にタスクパフォーマンスが向上したことを示した。しかしながら、インタラクションは称賛のみであり、人が操作する複数遠隔アバターによる通常会話における影響は明らかにされていない。

Wangらは複数の協力エージェントにおける人への影響を調べた[5]。人は複数の協力エージェントが居る場合、タスクパフォーマンスが向上し、学習意欲が向上することを示した。しかしながら、使用されたオンスクリーンエージェントと人の対話の分析が行われておらず、複数エージェントによる対話への影響は明らかにされていない。

2.3 遠隔アバターによる身体動作

聞き手は、遠隔アバターの外見よりも身体動作と会話の声質の一致に影響を受ける。Ennisらは、聞き手はエージェントの声と身体動作の不一致に対して不快を感じることを示した[12]。Ferstlらは、声と動きを一致させることで、参加者の好感度が上がることを示した[13]。外見(人アバター・ロボットアバター)、声質(録音された肉声・音声合成)、身体動作(モーションキャプチャあり・なし)でそれぞれ聞き手に対しスピーチをする実験を行い、聞き手は身体動作と声質の一致する場合快適に思うことを示した。よって実装するエージェントは、遠隔者の音声と身体動作が提示できるようにすることで、声質と身体動作を一致させることができ、聞き手が不快感なく対話が行える可能性がある。

3 提案手法

本研究では、裸眼立体視可能な 3D ホログラムディスプレイ上にコミュニケーションエージェント（以下、バーチャルエージェント）を開発した（図 2）。3D ホログラムディスプレイ上にバーチャルエージェントを実装することで、参加者が HMD などのウェアラブルデバイスを装着する必要がないことから、参加者に負担がかからない。また実装が容易であることが利点として挙げられる。

3.1 ハードウェアとソフトウェア

ディスプレイは Looking Glass を使用した。これは Looking Glass Factory 社により開発された 3D ホログラムディスプレイである[14]。左右の目に異なる映像を表示することにより裸眼で立体視を可能にしている。表示領域は正面水平方向約 50 度であり、これを 45 個の視点に分けている。

表示するバーチャルロボットは Unity により開発したプログラムにより制御する。Unity はゲームエンジンであり、3D モデル制御を得意とする[15]。3D モデルが Unity に読み込まれ、Looking Glass に表示される。

3.2 3D モデルの実装

バーチャルロボットは、操作者と参加者が対話できるよう実装した。操作者は右手にグローブ型デバイスを装着し、パペット人形を操作している感覚でバーチャルロボットを操作できるよう実装した。操作者はバーチャルロボットの、頭、腕を操作できる。2 体のバーチャルロボットは、ボタンを押すことで操作の行いを切り替えることができる。

バーチャルロボットは言葉を話すことができるよう実装した。口の動きは一定だが、操作者の声に合わせて口を動かす。

3.3 操作インターフェース

本研究では、一人の操作者が同時に 2 つ以上のエージェントを操作する必要がある。現在まで複数のエージェントを、同時に操作することを目的としたインターフェースの開発は例が少ない。このために、話者エージェント以外の自動化を前提とした操作インターフェースが必要になる。エージェントを操作する方法として、キーボードやジョイスティック、タッチパネルを用いたボタン操作などが考えられる。腕や顔の向きなどを対話内容に合わせてリアルタイ



図 2 複数動作の様子（左：操作しているエージェント、右：受動的振る舞いを行うエージェント）

ム動作させるため、ジョイスティック操作では不十分であることから、パペット人形操作をベースに最適な操作方法を検討した。

現在までに、MANUS 社の VR グローブ型デバイスを使用し遠隔エージェントを開発している。エージェントの切り替えや話者エージェント以外に信号を送りたい場合はボタンが必要になるが、基本は片手で操作できるインターフェースを開発する。

3.4 受動的振る舞いの自律化

本研究は、操作者が 1 つのエージェントを通して発話している間、操作していない他のエージェントは受動的振る舞いを自立化することが目的である。現在までに、操作していないバーチャルエージェントは、相槌を行うよう実装した。

4 今後の研究計画

4.1 受動的振る舞いの自律化

本研究では聞き手から見て複数エージェントが情報を提示しているように見せるために、話者エージェント以外の頷きや相槌などの受動的振る舞いを自動化する。グループ対話のために、Clark による話者モデルの Turn-taking や、Goffman の参与構造を意識する必要がある。Turn-taking とは、話し手と聞き手を交互に交代させながら対話を行うことである[16]。参与構造は、立ち位置や話し手により発話の対象とされた場合に役割が変化するとしたものである[17]。複数エージェント操作の場面において、操作しているエージェント以外は、相槌を打ったり、操作者が送る信号のタイミングに合わせて首を傾げてみたりすることで、対話の参加者からはその場でグループ対話が行われていると見せることができると考える。また、例えば参加者と仲が悪いエージェント

が対話する場面では、エージェントが意図的に参加者を無視することも有効と考える。対話に必要な身体動作をどこまで自律動作が可能か明らかにする。自律動作しやすい部分からしづらい部分までを分類し、それぞれの課題を進めることで操作者の負担にならない半自律動作を目指す。操作者が集中していないエージェントは相槌などの身体動作を自動で行うことができ、参加者から見て自然な振る舞いを行えるようになることを予定している。

4.2 一人の操作者が複数のエージェントを操作したときの対人評価

複数エージェントの効果が操作者一人でも同様に発生するか評価する。具体例として、参加者と3台のエージェントがグループ対話を行うシナリオを用意する。操作者が1台を通して発話している間、1台は会話に対して自動的に頷き、1台は操作者が送るシグナルにより首を傾げたり首を振ったりなどの動作を行う。グループ対話後にアンケートに答えてもらい、参加者にとって自然な対話が行えたか確認する。評価項目は、議論のしやすさ、親しみやすさ、反応の自然さ、対話のしやすさ、楽しさなどについて、5段階のリッカート尺度で答えてもらう。

5 まとめ

本研究は、一人の操作者が複数エージェントを操作し聞き手側から見て複数エージェントが情報を提示しているように見せることが目的である。よって操作者が1つのエージェントを通して発話している間、操作していない他のエージェントは、対話に付随する相槌などの身体動作を自動生成する手法を提案するために、現在開発中のエージェントの紹介を行った。今後は開発したエージェントと作成したシナリオを用いて、受動的振る舞いの自律化と対人評価を行う予定である。

謝辞

本研究は JSPS 科研費 JP20H01585, JST 次世代研究者挑戦的研究プログラム JPMJSP2124 の助成を受けたものです。

参考文献

[1] 文部科学省, “用語集,” <https://www.mext.go.jp/component>

/b_menu/shingi/toushin/_icsFiles/afieldfile/2012/10/04/1325048_3.pdf, Retrieved December 1, 2021.

- [2] F. Tanaka, T. Takahashi, S. Matsuzoe, N. Tazawa, and M. Morita, “Telepresence robot helps children in communicating with teachers who speak a different language,” *ACM/IEEE Int. Conf. Human-Robot Interact.*, pp. 399–406, 2014.
- [3] Y. Yoshikawa and H. Ishiguro, “複数体のロボットによる音声認識なし対話の可能性,” *IEICE Commun. Soc. Mag.*, vol. 10, no. 3, pp. 179–183, 2016.
- [4] T. Iio, Y. Yoshikawa, and H. Ishiguro, “Pre-scheduled turn-taking between robots to make conversation coherent,” *HAI 2016 - Proc. 4th Int. Conf. Hum. Agent Interact.*, pp. 19–25, 2016.
- [5] C. Wang, X. Wan, N. Yoshida, and T. Yonezawa, “Stimulation of Motivation for Learning by Multiple Agents in Group Training,” *HAI 2020 - Proc. 8th Int. Conf. Human-Agent Interact.*, pp. 25–31, 2020.
- [6] K. Seta, T. Narumi, M. Yokoyama, T. Tanikawa, S. Yoshida, and M. Hirose, “Divided presence: Improving group decision-making via pseudo-population increase,” *HAI 2018 - Proc. 6th Int. Conf. Human-Agent Interact.*, pp. 260–268, 2018.
- [7] K. Hayashi et al., “Humanoid robots as a passive-social medium: A field experiment at a train station,” *HRI 2007 - Proc. 2007 ACM/IEEE Conf. Human-Robot Interact. - Robot as Team Memb.*, pp. 137–144, 2007.
- [8] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, “Interactive robots as social partners and peer tutors for children: A field trial,” *Human-Computer Interaction*. vol. 19, no. 1-2, pp. 61–84, 2004
- [9] F. Jimenez, M. Kanoh, T. Yoshikawa, and T. Furuhashi, “Feasibility of learning by observing in collaborative learning with robots which changes the learning method,” *Trans. Japanese Soc. Artif. Intell.*, vol. 32, no. 2, pp. 1–12, 2017.
- [10] F. Tanaka and S. Matsuzoe, “Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning,” *J. Human-Robot Interact.*, vol. 1, no. 1, pp. 78–95, 2012.
- [11] M. Shiomi, S. Okumura, M. Kimoto, T. Iio, and K. Shimohara, “Two is better than one: Social rewards from two agents enhance offline improvements in motor skills more than single agent,” *PLoS One*, vol. 15, no. 11 November, pp. 1–14, 2020.
- [12] C. Ennis, R. McDonnell, and C. O’Sullivan, “Seeing is believing: Body motion dominates in multisensory conversations,” *ACM SIGGRAPH 2010 Pap. SIGGRAPH 2010*, vol. 1, no. 212, pp. 1–9, 2010.

- [13] Y. Ferstl, S. Thomas, C. Guiard, C. Ennis, and R. McDonnell, "Human or Robot?: Investigating voice, appearance and gesture motion realism of conversational social agents," *Proc. 21st ACM Int. Conf. Intell. Virtual Agents, IVA 2021*, pp. 76–83, 2021.
- [14] Looking Glass Factory · The World's Leading Holographic Display. Retrieved December 17, 2020 from <https://lookingglassfactory.com>
- [15] Unity · Real-Time Development Platform. Retrieved December 17, 2020 from <https://unity.com>
- [16] H. H. Clark and C. R. Marshall, "Definite knowledge and mutual knowledge," *Elem. discourse Underst.*, pp. 10–63, 1981.
- [17] E. Goffman, "Replies and responses," *Lang. Soc.*, vol. 5, no. 3, pp. 257–313, 1976.