

CG エージェントを用いた音声対話システムにおける 空間共有感のための自己投影法

Self-Projection Methods for Sense of Shared Space in a CG Agent based Spoken Dialogue System

東 省吾^{1*} 上乃 聖¹ 李 晃伸¹
Shogo Higashi¹ Sei Ueno¹ Akinobu Lee¹

¹ 名古屋工業大学

¹ Nagoya Institute of Technology

Abstract: CG エージェントを用いた音声対話システムは、ロボットやアンドロイドを用いるシステムと比べて、対話に影響を与える要素の一つである空間共有感に乏しい。本研究では、画面上にユーザ自身の存在をリアルタイムに投影し、疑似的にユーザをシステム側の世界に引き込むことで空間共有感を高める手法を提案する。自己投影法としてユーザの動きを反映する CG アバターを用いる手法とユーザのカメラ映像を直接合成する手法の 2 種を比較するとともに、複数の表示方法を比較した結果を報告する。

1 はじめに

CG エージェントと相対して会話する音声対話システムにおいて、人と話すようなリアルな会話体験を提供することは重要である。CG エージェントを用いた音声対話システムは、人の形を模した対象物と会話させることで、人間に近い感覚でより自然に話せる環境を提供する狙いがある。しかし、現実世界に実体を持つロボットやアンドロイドと比較すると、存在感が乏しく人間と話すようなリアリティを得られにくい [1]。

本研究では、そのようなリアリティの無さはユーザと CG エージェントが同一空間上に存在しないことにより起こる空間共有感の乏しさが要因であると考え、注目する。空間共有感とは、対話であれば話し相手といったような、対象と空間を共にしているという感覚のことである。空間共有感を課題としている分野の一つに人対人によるオンライン対話がある。人と人が対面で会話する場合と比較して、オンライン対話では空間共有感が低下するため会話のスムーズさに影響を与えてしまう [2]。画面の中の存在である CG エージェントにおいても、オンライン対話と同様にユーザは物理的な空間を共有できず、空間共有感を得られにくい。

そこで本研究ではユーザの存在をシステムの画面内に合成表示する手法として、自己投影法を提案する。端末の前にいるユーザを Web カメラ等で撮影し、リアルタイムで画面上にクロマキー合成する自己映像システム

によって、ユーザをディスプレイ内の仮想空間に引き込み、エージェントとの空間共有感の向上を目指す。ロボットやアンドロイドの研究を参考にすると、空間共有感が向上することで、エージェントを身近に感じる、信頼性が向上する、親密度が上がる等の効果が見込まれる [3]。CG エージェントを用いた音声対話システムにおいても、空間共有感が高まることでシステムの使用感やユーザの印象に影響を与える可能性がある。本稿では特に、この自己映像の投影手法について複数の手法を検討・比較するとともに、3.1 章で述べる自己投影アバターとの比較評価を行う。

2 関連研究

人対人によるオンライン対話において空間共有感の乏しさを扱ったものがある。互いのインタラクションの様子を共有するための研究として、オンライン対話で自己アバターの対面合成を試みた研究がある [4]。仮想的な空間共有のために、対話者のアバターを対話相手の映った映像に合成する手法を取っている。対話者はアバターの振る舞いと対話相手のノンバーバル情報を把握するとともに、互いのインタラクションを把握しながら対話を行うことができる。

ユーザ自身の映像を用いた研究として、超鏡の研究がある [5]。超鏡は 1 地点の映像に対し多地点の対話者をクロマキー合成することで対話者全員が同じ映像を見ながら会話することを可能にしており、これにより

*連絡先：名古屋工業大学
愛知県名古屋市長和区御器所町
E-mail:s.higashi.444@stn.nitech.ac.jp

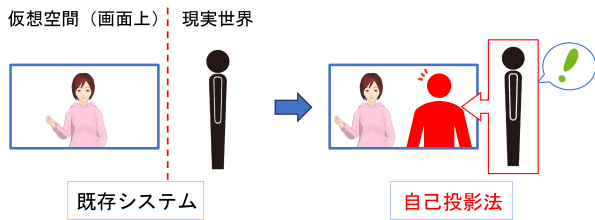


図 1: 自己投影法の概念図

対話相手と同室にいる感覚が得られる。また、超鏡における画面上の自己像と相手像の立ち位置の重要性についての研究がある [6]。自己・相手像ともに正面を向くよりも 2 つの像が斜めに向き合う方が効果的であり、自己像が相手像を向いていることが話しやすさにおいて重要であるとしている。

本研究ではこれらを参考にシステムを構築する。

3 空間共有感のための自己投影法

自己投影法の概念図を図 1 に示す。ユーザ自身を画面内にリアルタイムで表示することで、画面内の仮想空間に引き込み、エージェントとの空間共有感を高めることを目指す。本研究では、ユーザ自身を画面内に投影する方法として、ユーザの動きを反映した CG アバターである自己投影アバターを用いる手法とユーザのカメラ映像を直接合成する手法の 2 種を比較検証する。

どちらのシステムも MMDAgent-EX [7] をベースとし、対話相手の CG エージェントとしてジェネ [8] を用いた。タスクは雑談対話とし、対話モジュールとして OpenAI API (モデルは応答速度を優先するため gpt-3.5-turbo) を利用する。

3.1 自己投影アバター

自己投影アバターとは、システムのユーザと同期した動きを行うことのできる CG アバターのことを指し、手法としては、自己投影アバターを CG エージェントを用いた音声対話システムの画面上に合成する [9]。提案システムの全体図を図 2 に示す。過去の実験により、自己投影アバターを CG エージェントと同一空間かつエージェントと正面に相対した形で表現することで、アバターに対する自己認知と対話の臨場感を高められ、また会話がうまく成立しないという条件のもとで話しやすさが向上することが示されている。

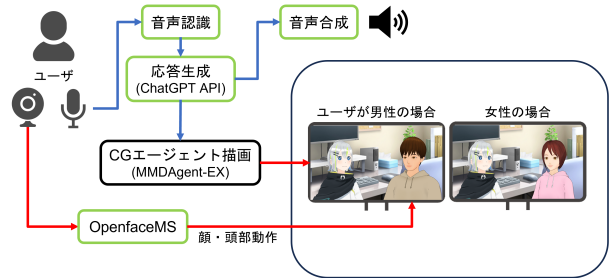


図 2: 自己投影アバターを用いたシステム

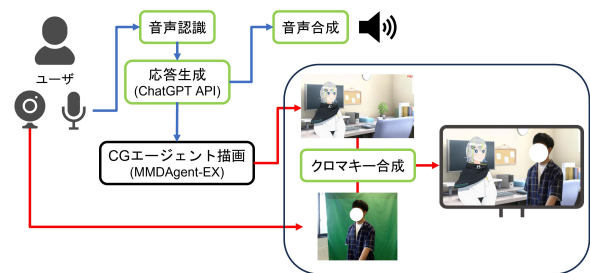


図 3: 自己映像を用いたシステム

3.2 自己映像

提案システムの全体図を図 3 に示す。自己映像の合成部分では、カメラで撮影した自己映像を対話システムの画面上に合成する。自己映像を画面上に表示する際、CG エージェントとの画質やフレームレートの差が生まれないように、カメラは解像度が 1080 p、フレームレートが 60 fps である、logicool StreamCam を使用した。システムの背景として、カメラで撮影するライブ映像と仮想背景を切り替えられるように設計する。背景にライブ映像を使用する場合、MMDAgent-EX では CG キャラクタと緑単色の背景画像を出力する。また、カメラで背景を含めてユーザを撮影する。そして、エージェントをカメラ映像にクロマキー合成する。一方背景に仮想背景を使用する場合には、MMDAgent-EX で CG キャラクタと仮想背景を出力する。また、カメラでグリーンバックの前に立ったユーザを撮影する。そして、ユーザを仮想背景にクロマキー合成する。

自己映像の提示方法として空間共有感に影響を与える要素だと予想される条件として、自己映像の配置と使用する背景の 2 つを比較する。

自己映像の配置や向きを選択として、CG エージェントと自己映像が向き合い対話しているような表現として内向きと肩越しを用意する。また、背景の選択として、CG エージェントと親和性が高い方がよいのか、または自己映像との親和性が高い方がよいのかを明ら



図 4: 内向き-ライブ映像



図 5: 内向き-仮想背景



図 6: 肩越し-仮想背景

かにする必要があるため、仮想背景とライブ映像の2種類を用意する。自己映像の配置と使用する背景の2つを比較するため、本研究では以下の「内向き-ライブ映像」「内向き-仮想背景」「肩越し-仮想背景」の3種類の方法を比較する。それぞれの表示方法を図4, 5, 6に示す。

配置として以下の2種を用いる。

- 内向き
自己像がこちらを向いた鏡のような表現であるため違和感が少なく受け入れやすいと予想される。
- 肩越し
画面上でエージェントと自己像が対面し空間を共有している様子を観察できることから、空間共有感が高まることが期待できる。

背景として以下の2種を用いる。

- 仮想背景
背景とエージェントの親和性が高く、またユーザが仮想世界にいるように見せられる。
- ライブ映像

背景と投影されたユーザの自己映像の親和性が高く、またエージェントが現実世界にいるように見せられる。

4 実験1：自己映像の配置・背景の比較実験

4.1 実験条件・タスク

「内向き-ライブ映像」「内向き-仮想背景」「肩越し-仮想背景」の3手法について、それぞれの存在感や空間共有感、話しやすさについて主観評価実験を行い評価する。

実験タスクは、被験者の話しかけから始まる1分30秒間の雑談対話タスクであり、約5ターン程度の会話が行われる。雑談テーマは最近あった出来事についてであるが、被験者から異なる話題が提供されればその話題で対話が行われる。被験者数は大学生・大学院の情報系学生13名である。実験環境は被験者のみが入室する区切られた静かな空間である。

4.2 評価方法

本実験では、自己映像の存在を認知できたかどうかを調べる項目、CGエージェントとの空間共有感や実在感、話しやすさ、およびシステムの使用感に関する項目を準備する。空間共有感に関する質問項目は、田中らによる人対人の遠隔対話におけるソーシャルテレプレゼンスの研究[10]で使用された評価項目の一部を使用する。すべての質問は7段階のリッカート尺度として集計を行う。

4.3 実験結果と考察

作成した主観評価アンケートの項目と、それぞれの平均スコアのグラフを図7に示す。なお、項目末尾に(-)のある逆転項目については、スコアを反転して棒グラフを作成しており、有意水準5%のt検定の結果、有意差の見られた項目のグラフには*を付与している。

背景の選択としてライブ映像と仮想背景の比較をした結果、t検定により、「CGキャラクターがあたかもこちらの現実世界にいるような感じがした」という項目において、ライブ映像が有意に高い評価を得ており、ライブ映像によってCGエージェントが現実世界に存在するよう見せることに成功している。また、有意な差は見られなかったものの、引き込みに関する2項目すべてと空間共有感に関する3項目で、仮想背景の方が高い評価となる傾向が見られた。

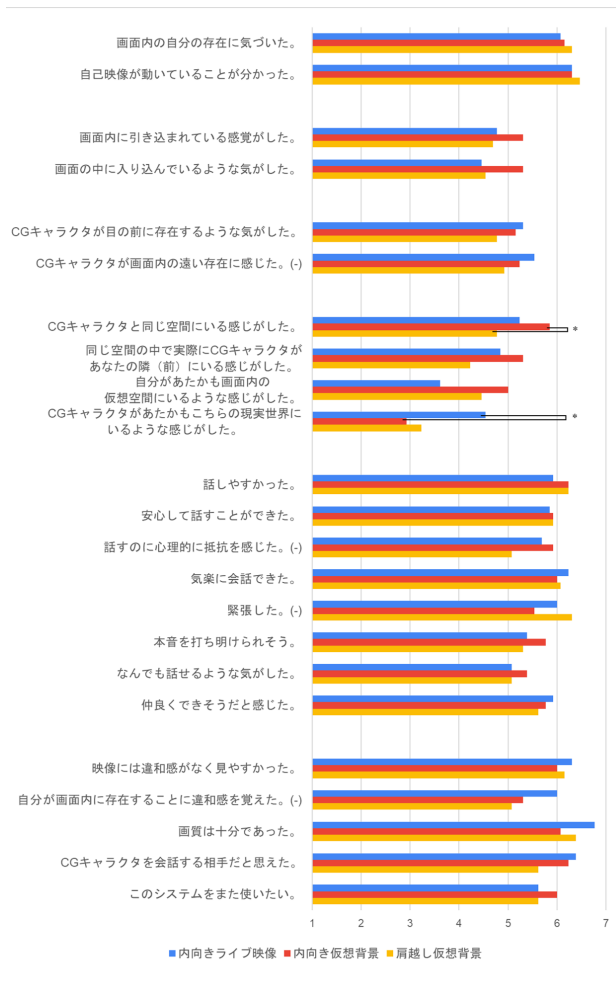


図 7: 3手法主観評価アンケート結果

次に、自己映像の配置として内向きと肩越しの違いを比較した結果、「CGキャラクターと同じ空間にいる感じがした」という項目において、内向きの配置が有意に高い評価を得たことから、内向きの空間共有感への有効性が示された。

これらの結果から以下の考察ができる。まず、仮想背景の方が空間共有感が高い傾向にあることから、仮想背景上に自らが投影される違和感よりも、どちらかといえばCGエージェントが現実世界にいる違和感が強く感じられたと推察できる。CGエージェントに背景やシーンを合わせて投影することで知覚される違和感が緩和される可能性が示唆された。

次に、配置の比較から得られた結果から、対話における空間共有感のためには画面内でCGエージェントと自己像が向き合うだけでなく、画面の前のユーザがCGエージェントと自己像ともに等しく向き合えるような、3人で対話している状況を作り出すことが有効であると示唆された。このような配置による効果は人対人オンライン対話の自己映像の配置と話しやすさの



図 8: 自己投影アバター（内向き-仮想背景）



図 9: 投影無し通常システム

関係の研究 [11] でも示されている。また、自己映像は鏡のような見慣れた表現であるため、自分の存在を即座に認知できたことが有効に機能した可能性も考えられる。

5 実験2: 自己投影法の効果検証実験

5.1 実験システム・条件

次に、自己投影法2種類と自己投影法を使わないシステムとの比較実験を行った。対話システムに慣れていない被験者を対象とし、事前に自己投影法に関する情報は与えず、初めてシステムを使用する実環境に近い対話実験を行う。

比較するシステムは、CGキャラクターにユーザの動きを反映させた自己投影アバターを使用する図8のシステム [9] と実験1で使用した図5の内向き-仮想タイプの自己映像システム、そして自己投影法を使わない図9の通常システムの3種類である。

5.2 実験タスク

実験1と同様に、システムはユーザと雑談対話を行う。被験者は大学生・大学院生11名（男性7名、女性4名）であり、音声対話システムを使用したことがない、もしくは数回程度の使用経験がある被験者である。実験のタスク、手順、実験環境は実験1と同様である。被験者は3つのシステム（自己投影アバターシステム・自己映像システム・通常システム）を順に使用する。システムを使用する順番は被験者によってランダムである。

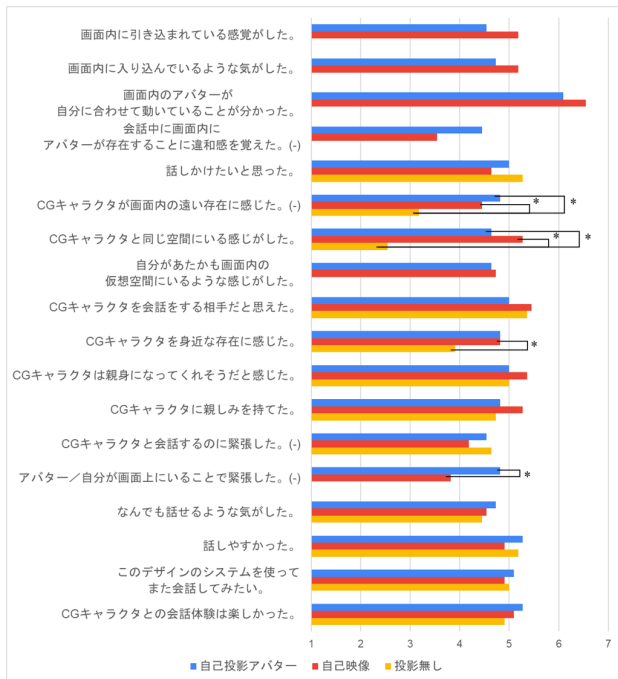


図 10: 自己投影法 2 種と通常システムの主観評価アンケート結果

5.3 評価方法

実験 2 では、評価の際に実験 1 のアンケートに身近さ、親身さ、親しみ度合いを問う項目を追加したものを使用した。また、緊張には主に自分を表す存在が画面に存在することによる緊張と会話に対する緊張があると考えられるため、2 種類の聞き方を行う。

5.4 実験結果と考察

作成した主観評価アンケートの項目と、それぞれの平均スコアのグラフを図 10 に示す。なお、項目末尾に (-) のある逆転項目については、スコアを反転して棒グラフを作成しており、有意水準 5% の t 検定の結果、有意差の見られた項目のグラフには * を付与している。

自己投影法である自己投影アバターと自己映像と、投影無しの通常システムを比較した結果、t 検定により、「CG キャラクタと同じ空間にいる感じがした」「CG キャラクタが画面内の遠い存在に感じた」という項目で、自己投影法の両方が通常システムよりも有意に高い評価を得た。また、「CG キャラクタを身近に感じた」という項目で自己映像が有意に高い評価を得た。このことから、自己投影法によって空間共有感を高められるという有効性が示された。また、心理的な距離を縮められる可能性も示唆された。

次に提案手法の 2 種間で比較した結果、「アバター／自分が画面上にいることで緊張した」という項目で、自

己映像は自己投影アバターよりも有意に高い評価となったことから、自己投影アバターは自分の顔が画面に表示されることにより生まれる緊張感を軽減できることが示された。

なお、本研究では自己投影法による空間共有感の改善を確認するにとどまったため、空間共有感と対話のクオリティの関連については今後の検討が必要である。

6 むすび

本研究では、音声対話システムにおける CG エージェントとの空間共有感を改善すること目的とした、会話中のユーザを CG エージェントと共に画面内に投影する自己投影法の 2 種類を比較評価した。具体的にはユーザの動きを反映した CG キャラクタである自己投影アバターを合成する手法と、ユーザを撮影したカメラ映像である自己映像を合成する手法を用いた。被験者 11 人に対する雑談タスクでの評価の結果、自己投影法である自己投影アバター、自己映像のどちらを用いる場合でも、空間共有感と CG エージェントの存在感が高まり、CG エージェントを身近な存在に感じさせる効果もみられた。

自己映像の配置や背景の比較により、投影された自己映像と CG エージェントが斜め内側に向き合う配置、また CG エージェントと同質の仮想背景を用いることで、空間共有感を高めることが分かった。

今後の展望として、空間共有感をさらに向上させるために非言語的なインタラクションを実装することが考えられる。ユーザの動きにエージェントが反応を示すといった、ユーザの発話に対して表情やモーションを一方向的に表出するだけにとどまらないインタラクションを画面上で表現することで、さらなる空間共有感の向上や話しやすさ、対話の臨場感の向上が期待できる。

謝辞

本研究は、JST ムーンショット型研究開発事業、JP-MJMS2011 の支援を受けたものです。

参考文献

- [1] Powers, Aaron and Kiesler, Sara and Fussell, Susan and Torrey, Cristen. Comparing a computer agent with a humanoid robot. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 145–152, 2007.

- [2] 田中一晶. 空間共有ロボティクス. 計測と制御, Vol. 61, No. 3, pp. 209–213, 2022.
- [3] Kidd, C.D. and Breazeal, C. Effect of a robot on user perceptions. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No.04CH37566)*, Vol. 4, pp. 3559–3564, 2004.
- [4] 石井裕, 渡辺富夫. C211 自己アバタを対面合成した身体的ビデオコミュニケーションシステム. 可視化情報学会誌, Vol. 23, No. Supplement1, pp. 357–360, 2003.
- [5] 森川治ほか. 超鏡: 魅力あるビデオ対話方式をめざして. 情報処理学会論文誌, Vol. 41, No. 3, pp. 815–822, 2000.
- [6] 森川治, 山下樹里, 福井幸男, 佐藤滋. 超鏡対話における対話者の配置と話しやすさの関係. インタラクション, Vol. 2001, pp. 179–186, 2001.
- [7] Akinobu Lee. MMDAgent-EX. <https://github.com/mmdagent-ex/MMDAgent-EX>, December 2023.
- [8] 李晃伸, 石黒浩. 自律・遠隔融合対話システムのための高生命感・高存在感 CG エージェントの開発. 人工知能学会研究会資料 言語・音声理解と対話処理研究会, Vol. 96, p. 27, 2022.
- [9] 東省吾, 李晃伸. 自己投影アバターによる引き込みを用いた 2D-CG 音声対話システム. 2T-P2, 2022.
- [10] 田中一晶, 西村庄平, 耿星, 中西英之ほか. 空間の移動感と物体の共有感による鏡型ビデオ会議のソーシャルテレプレゼンスの強化. 情報処理学会論文誌, Vol. 60, No. 2, pp. 419–428, 2019.
- [11] 森川治, 山下樹里, 福井幸男, 佐藤滋. 超鏡対話における対話者の配置と話しやすさの関係. インタラクション, pp. 179–186, 2001.