豊かなノンバーバルコミュニケーションのための HMD を用いた没入型音声対話システム

Immersive 3D-CG Spoken Dialogue System using HMD for Rich Non-verbal Communication

宮下 陸[1]　上乃 聖[1]　李 晃伸[1]
Riku Miyashita[1]　Sei Ueno[1]　Akinobu Lee[1]

[1] 名古屋工業大学
[1] Nagoya Institute of Technology

Abstract: 音声対話システムにおける CG 対話エージェントは，ロボットなどの実体をもつエージェントと比較すると，身体性や空間共有感のなさから，存在感に乏しいとされている．本研究では，CG エージェントの表示に Head Monted Display を用いることで，身体性と空間共有感をユーザに与えられるシステムを提案する．このシステムで対話することで，CG エージェントのノンバーバルな情報がユーザに円滑に伝わり，高い存在感を感じられることが期待される．実際に本システムを用いて 35 名の被験者が対話を行い，エージェントに対する認知について評価した．

1　はじめに

音声対話システムは，音声認識技術や自然言語処理技術による応答生成精度の向上，合成音声のリアリティの向上を背景に高い自然性を実現しつつある．音声対話システムの会話型 UI に用いられるキャラクターは，ロボットと，画面上に表示される CG 対話エージェントの 2 つに分けられる．CG 対話エージェントはデジタル表現であるためロボットに比べて表現の制約が少なく，可搬性や運用性が高い [1]．しかし，ユーザと物理空間を共有できず直接的なインタラクションが行えないため，存在感に乏しいという課題がある [2]．

ところで，一般的な人対人の遠隔コミュニケーションにおいて，離れた場所にいる会話相手があたかも目の前にいるかのような感覚である，ソーシャルテレプレゼンスの実現の研究が数多く行われている．これらの人対人を対象にしたコミュニケーションの研究のアプローチは，人対エージェントとの対話にも適用できることが多く，非常に関係が深い．ソーシャルテレプレゼンスは，ノンバーバル表現が相手に即時に伝達される事が重要だとされている [3]．そこで，ソーシャルテレプレゼンス改善のアプローチの一つに，ジェスチャーを伝えるために Head Mounted Display を用いて仮想空間上で対話を行うものがある [4]．

また，近年，HMD を利用し，仮想空間でコミュニケーションを行えるプラットフォームが開発されている．VRChat [5] は，ユーザが仮想空間で楽しく交流することを目的に作られたプラットフォームである．VRChat のユーザ数は月間アクティブユーザ数が 2000 万人を超えており，また，2017 年のリリース以来，アクティブユーザ数を伸ばし続けている．音声チャットやジェスチャーなどを使用しながら，アバターを通じて自分を表現し，アクティビティを通したコミュニケーションを楽しむことができる．このように，遠隔地同士でも，HMD を利用し，仮想空間で対話相手のアバターと空間を共有しコミュニケーションをとることができる．

そこで，本研究では，HMD を CG エージェントの表示に用い，ユーザ自体を対話エージェントの存在する仮想空間へ連れ込む対話システムを構築した．このシステムで対話することで，CG エージェントのノンバーバルな情報がユーザに円滑に伝わり，高い存在感や生命感を感じられることが期待される．

2　関連研究

人対人の遠隔コミュニケーションを対面のコミュニケーションに近づける研究として，ソーシャルテレプレゼンス (social telepresence) の研究が数多く行われている．ソーシャルテレプレゼンスは，対面でのコミュニケーションにおける相手の存在感であるソーシャルプレゼンスが遠隔コミュニケーションにおいても再現された感覚のことである．つまり，ソーシャルテレプレゼンスとはメディアを介して行われる擬似的な対面会話のリアリティのことであり，実際には離れた場所

にいる会話相手があたかも目の前にいるかのように感じる感覚のことである [6].

ソーシャルプレゼンスの概念は，古くより研究されてきた社会心理学の人対人コミュニケーション論の中に起源があるとされている．Mehrabian らは，一般的な対面での人対人コミュニケーションにおいて，より親密さを感じるためには，ノンバーバルな情報が，同時双方向的にタイムラグなく伝わることが必要である可能性を示唆した [7].

そののちに，Short らは，遠隔コミュニケーションにおいて，音声のみのコミュニケーションよりも，相手の映像を使ったコミュニケーションのほうがより親密さが高まることを発見した [3]．そこで，遠隔コミュニケーションの社会心理学を論じる中で，Mehrabian らの研究に基づき，ソーシャルプレゼンスを，「相互作用における相手の存在感の度合い」として定義し，これが遠隔コミュニケーションにも必要であると述べた．

そして，ビデオ会議の代わりに，HMD を用いて仮想空間上で対話することで，ソーシャルプレゼンスを改善することができることが報告されている [4]．Head Mounted Display (HMD) とは，ユーザーが自身の頭部に装着し使用するディスプレイである．デバイスには，ディスプレイ画面，オーディオ出力，そしてセンサーが内臓されている．センサーによる頭部のトラッキング機能を使用して，ユーザーの頭部動作に追従して映像や音響が遷移するため，没入感のある体験が可能である．

3　提案手法

本研究では，CG キャラクターとのノンバーバルコミュニケーションの強化のために，高精細で写実的な CG キャラクターを表示する HMD を用いた没入型インタフェースを用いることを提案する．この提案手法の実現のために，HMD を用いて CG エージェントを表示するシステムを構築した． 提案システムの全体図を図 1 に示す．システムには，高精細な人体モデルと豊かな表現力を持つ，3D-CG エージェント “Rubica” [8] を利用する．モデルは，軽量化のために，肌表現の解像度や頭髪を改変し利用する．HMD デバイスとしては，HTC VIVE Pro 2 を使用した．まず，HMD の口元にあるマイクに入力された発話音声を，Whisper [9] によってテキスト化する．テキスト化された入力発話は，Silero-vad [10] を用いて発話終了検知を行い，得られたユーザ発話テキストを応答速度が高速な GPT-3.5-turbo [11] に入力し，応答文を生成する．応答文生成時に，同時に笑顔，喜び，驚き，怒り，悲しみ，呆れの 6 種から選択し，感情ラベルの出力を行う．出力された感情ラベルは，Unreal Engine [12] で作成された

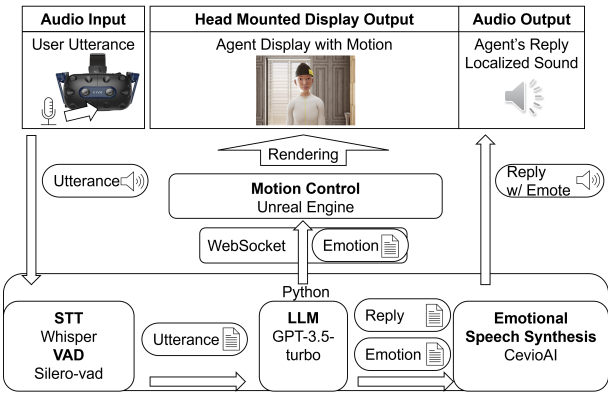


図 1: 没入型音声対話システムの構成

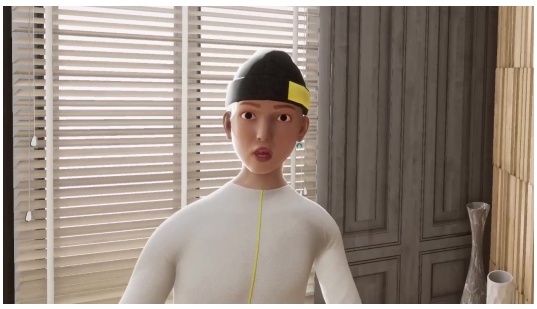


図 2: 驚きの感情表現

モーションコントロール部に送られ，感情に合わせた表情変化やジェスチャーを行う．図 2 に，驚きの感情表現を行うエージェントのようすを示す．そして，応答文は感情ラベルと共に Cevio AI [13] へ送られ，感情音声合成を行い，エージェントの発話とする．

4　評価実験

4.1　実験設定

本実験では，HMD による表示が，空間共有感や視覚的な相互作用可能性にどのような影響を与え，ノンバーバルコミュニケーションをどの程度円滑にするのかを検証するため，前述のシステムをベースとして，エージェントの表示手法と身体動作の有無を変化させた 4 つの音声対話システムを構築し，比較する．以下に，それぞれの手法の特徴を説明する．

- 平面感情無-エージェントの表示は平面ディスプレイを用いて行う．また，応答内容の感情によって，表情変化や，ジェスチャーを行わない．
- 平面感情有-エージェントの表示は平面ディスプレイを用いて行う．そして，応答内容の感情によって，表情変化や，ジェスチャーを行う．

図 3: HMD 表示における実験風景

- HMD 感情無-エージェントの表示は HMD を用いて行う．また，応答内容の感情によって，表情変化や，ジェスチャーを行わない．
- HMD 感情有-エージェントの表示は HMD を用いて行う．そして，応答内容の感情によって，表情変化や，ジェスチャーを行う．

実験は 4 つのシステムに対して評価を行う．被験者は，大学生，大学院生の 35 名である．十分なデータを集めるために，一部の被験者は，1 名当たり 2 つのシステムを，片方の条件を固定した組み合わせで評価をした．その際，慣れの影響を防ぐために，2 週間の間を空けて実験を行った．また，順序はランダムにした．実験室は被験者のみが入室する静かな環境である．本実験では，HMD と平面モニターでフレームレートが一定になるようにした．

HMD を利用する場合は，現実空間における周囲の障害物が気にならないように，十分な空間を用意した．なお，エージェントに対する距離は，平面ディスプレイにおいても，HMD における仮想空間内においても，1m の距離である．これは，エージェントとの距離が，ノンバーバルコミュニケーションの円滑さに直結するため，あらかじめ距離を設定した．

実験の手順は以下の通りである．(1) 実験の事前説明およびタスク説明 (2) ディスプレイの前の立ち位置に誘導 (3) 実験器具の装着 (4) ユーザ発話をきっかけに対話開始 (5) 5 分経過後に実験者による対話終了の合図 (6) アンケートによる主観評価 (7) 感想や気になった点についてインタビュー

手順 (1) では，エージェントの発話のつじつまが合わなかったり，会話がうまくいかなかったりする場合があることを伝えた．この実験は，対話内容の評価ではなく，対話中のノンバーバルコミュニケーションに対する認知の評価が目的であるため，対話のミスはできるだけ考慮せずに評価してもらうように事前説明を行った．対話内容は，お互いに感情豊かな対話ができるように，最近嬉しかったことや悲しかったこと，驚いたことなどの感情にまつわる出来事について対話する．

本実験では，3 種の主観評価アンケートを使用した．1 つ目は，God Speed Questionnaire [14] から，SD 法による全 19 項目の質問を使用する．これは，エージェントの印象を評価するための尺度である．2 つ目は，Presence Questionnaire [15] から，全 18 項目の質問を対話システム向けに改変し使用する．これは，エージェントの仮想空間における存在感を評価するための尺度である．3 つ目は，エージェントとの話しやすさ，感情についての評価，およびシステムの総合評価に関する項目を準備した．これらは 7 段階のリッカート尺度として集計を行う．

4.2 実験結果および考察

主な評価結果を，図 4 に示す．殆どの被験者が，成立した対話を行うことができた．一部，音声認識誤りによって，ユーザの意図と異なる会話が発生したが，対話として破綻するほどではなかった．本実験の分析方法は対応のない t 検定（有意水準 5 %）とする．全ての結果を，付録に示す．

まず，身体動作のある平面表示と，身体動作のある HMD 表示の実験を比較すると，“CG キャラクタがあたかも目の前に存在するような気がした.”，“エージェントが画面内の遠い存在に感じた.，“会話相手と同じ部屋（空間）にいる感じがした.”，“同じ部屋の中で実際に相手があなたの隣（前）にいる感じがした”，“自分が，あたかも会話相手の部屋（仮想空間）にいるような感じがした.”の項目で，HMD 表示が有意に高い値となった．そして，“どれくらい自由に会話をできましたか？”，の項目では，HMD 表示が有意に低い値となった．

次に，平面表示における感情表現のある場合とない場合の結果の比較をしたが，すべての項目において，有意差がなかった．

そして，HMD 表示における感情表現のある場合とない場合の結果の比較をすると，“活気のない - 生き生きとした”の項目で，有意に感情表現のあるシステムが高い値となった．また，“エージェントの感情は豊かに感じた.”，“エージェントには感情変化があるように感じた.”，“エージェントのジェスチャーや表情は，どれくらいあなたの問いかけに関連していましたか？”の項目で，有意に感情表現のあるシステムが高い値となった．

これらの結果から，HMD によるシステムでは，エージェントの身体性をより知覚できることが分かった．また，空間共有感が高まることが分かった．つまり，エージェント表示に HMD を用いることで，システムの臨場感が上がり，エージェントの存在感が向上することが明らかになった．

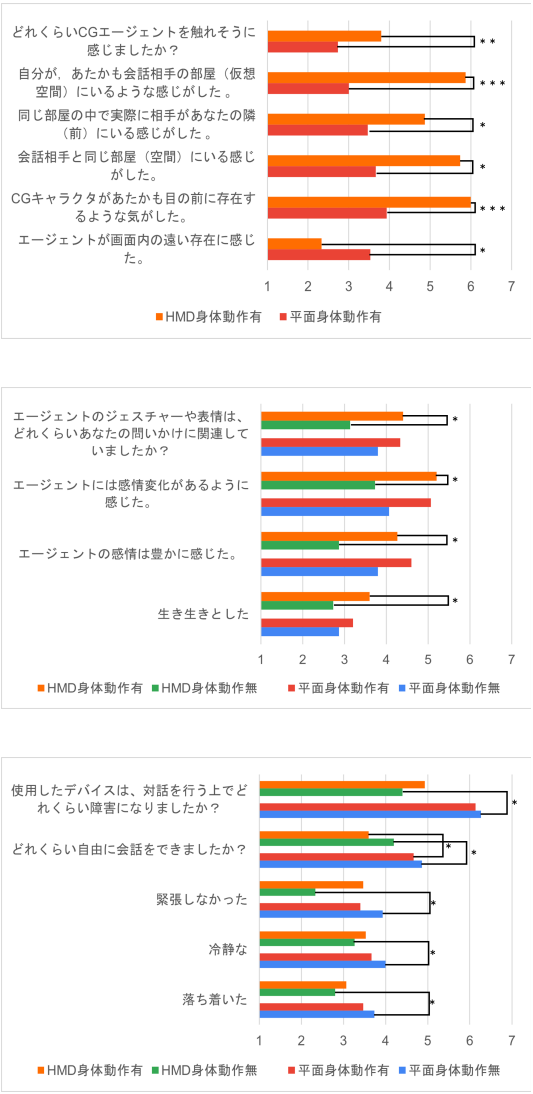


図 4: 主な評価結果

また, HMD でエージェントを表示する場合のみ, エージェントの身体動作によって，感情の有無や生命感に有意差があった．しかしこれらは，HMD を用いた身体動作の無いシステムが被験者により低く評価されたためだと推察される．HMD を用い，臨場感を高めると，エージェントのリアリティを高く感じ，それ故により自然にノンバーバルな表現を含むことをユーザは期待してしまう．そのため，身体動作のないシステムはユーザが自然に期待するふるまいを下回る，悪い意味で目立ってしまう設定になっている可能性がある．これは，過度にリアリティのある外見を持つアバターを用いると，中身が人間かコンピューターであるかに関わらず，期待を下回った際に，ソーシャルプレゼンスが低下することが報告されており [16]，同じ現象が起こった可能性がある．

また，不安や緊張を感じた被験者が，4 名いた．これは，HMD の装着の手間や重さによって，使用感が悪かったことはもちろん，視界の閉塞感を感じたという意見もあった．また，HMD を装着しノンバーバル表現のないエージェントに対峙することで，より不気味に感じられることも一因ではないかと推察する．

5 まとめ

本研究では，対人コミュニケーションにおけるソーシャルテレプレゼンスの改善手法の観点から，高精細で写実的な CG キャラクターを表示する HMD を用いた没入型音声対話システムを提案した．比較実験の結果，エージェント表示に HMD を用いることで，キャラクターとの対話の臨場感をより感じやすいことが分かった．また，HMD を用いる場合，感情表現を行わないエージェントに対して，被験者は有意に感情がないと評価した．つまり，HMD を用いることで，エージェントの感情をユーザに対して強調でき，存在感を高められる可能性が示唆された．話しやすさと臨場感や存在感の関係の検証は今後の課題である．今後の展望としては，エージェントとさらなるノンバーバルコミュニケーションを可能にすることである．

謝辞

本研究は、JST ムーンショット型研究開発事業、JPMJMS2011 の支援を受けたものです。

参考文献

[1] Jesse Fox and Andrew Gambino. Relationship development with humanoid social robots: Applying interpersonal theories to human–robot interaction. *Cyberpsychology, Behavior, and Social Networking*, 24(5):294–299, 2021.

[2] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 145–152, 2007.

[3] John Short, Ederyn Williams, and Bruce Christie. The social psychology of telecommunications. 1976.

[4] Frank Steinicke, Nale Lehmann-Willenbrock, and Annika Luisa Meinecke. A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*, pages 1–2, 2020.

[5] VRChat. `https://hello.vrchat.com/`.

[6] Marvin Minsky. Telepresence. 1980.

[7] Morton Wiener and Albert Mehrabian. *Language within language: Immediacy, a channel in verbal communication*. Ardent Media, 1968.

[8] 李晃伸 and 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. In **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 第** *96* **回** *(2022.12)*, page 27. 一般社団法人 人工知能学会, 2022.

[9] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR, 2023.

[10] silero-vad. `https://github.com/snakers4/silero-vad`.

[11] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[12] Unreal Engine. `https://www.unrealengine.com`.

[13] Cevio. `https://cevio.jp/`.

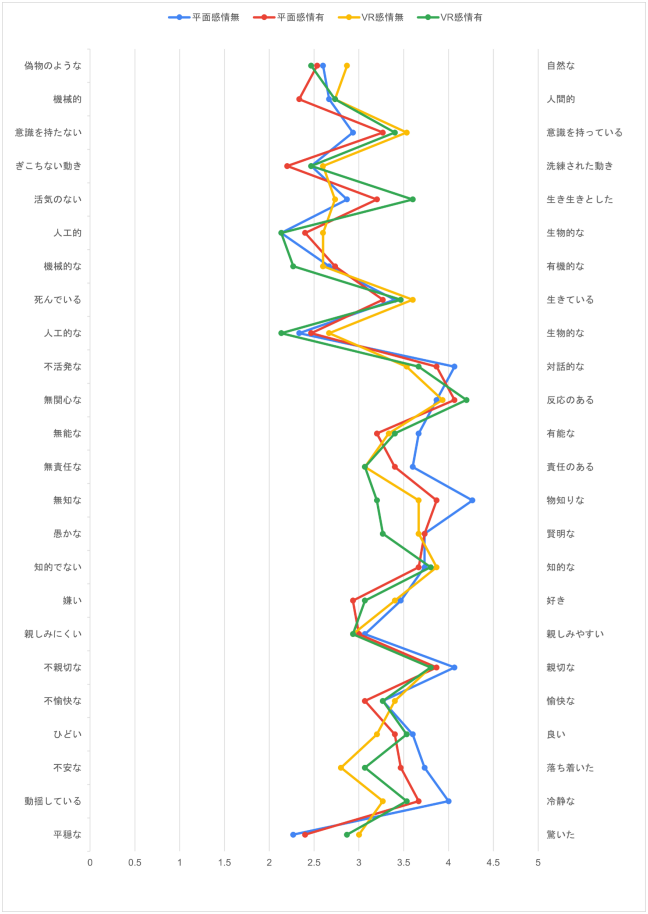
[14] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81, 2009.

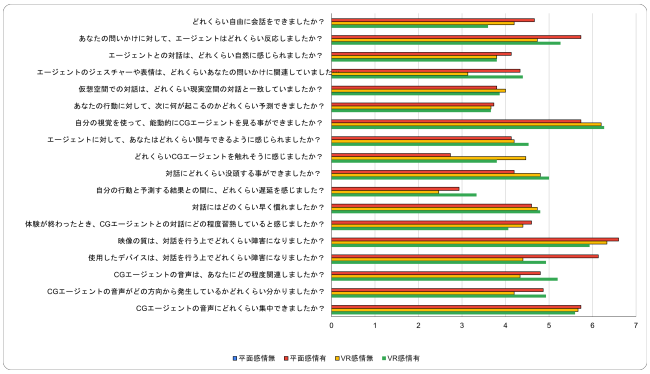
[15] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[16] Kristine L Nowak and Frank Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5):481–494, 2003.

A　付録

図 5 にすべての主観評価の質問と結果を示す,





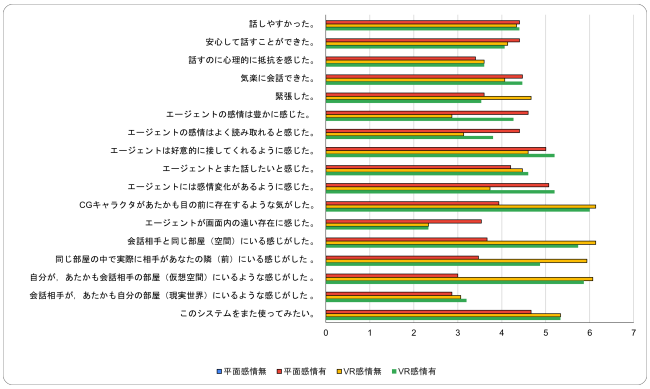


図 5: すべての主観評価結果