



話者 CG モデルを実空間アバターロボットに提示する XR 遠隔対話システム

An XR remote dialog system presenting speaker's expression on a real space avatar robot

米田悠人¹⁾, 小島優希也¹⁾, 福地庸介¹⁾, ヤエム ヴィボル²⁾, 池井 寧³⁾, 西内信之¹⁾
Yuto YONEDA, Yukiya OJIMA, Yosuke FUKUCHI, Vibol YEM, Yasushi IKEI and Nobuyuki NISHIUCHI

- 1) 東京都立大学システムデザイン研究科情報科学域 (〒191-0065 東京都日野市旭が丘 6-6, {yoneda-yuto, yukiya-ojima}@ed.tmu.ac.jp, {nnishiuc, fukuchi}@tmu.ac.jp)
- 2) 筑波大学大学院 システム情報系 (〒305-8573 茨城県つくば市天王台 1-1-1, yem@iit.tsukuba.ac.jp)
- 3) 東京大学大学院 情報理工学系研究科 (〒113-8656 東京都文京区本郷 7-3-1, ikei@vr.u-tokyo.ac.jp)

概要: XR 技術とアバターロボットを利用した遠隔対話システムを構築している. 本システムでは, MR グラス (HoloLens2) を装着した現場作業者と HMD (head mounted display) を装着した遠隔対話者が, 現地で空間オブジェクトを含む対話を円滑に行うことを目的としている. 遠隔対話者の表情は 3D アバターで再現され, 頭部・口元の動き, 視線の方向を反映した対話を行える機能を実装した. 評価実験では, 遠隔対話者と現場作業者が対話しながら共同作業を行うための本システム構成の有効性とリアルアバターの表情生成の効果を定性的に評価した.

キーワード: テレプレゼンス, コミュニケーション, 拡張現実, 遠隔対話, リアル CG アバター

1. はじめに

近年, B5G (beyond 5th generation) 技術の進展とテレワークの普及が進む中, 遠隔地とのリアルタイムコミュニケーションの重要性が高まっている. しかしながら, 現在一般的な平面ディスプレイでは空間情報が不足し, 非言語コミュニケーションが十分に行えないという課題がある. また臨場感の欠如が遠隔地の状態の共有品質を低下し, 相手とのコミュニケーションの障害となりうる状況がある.

そこで, 3次元空間情報を十分に活用できる XR システムによって円滑な遠隔地コミュニケーションを実現することが求められている. Mixed Reality (MR) 機能を高めた MR グラスの出現によりその実装範囲も拡大されてきた. 本研究では, 相手話者のリアル CG アバターをテレプレゼンスロボットのカメラ部に提示し, 現実空間で対話しているのと等価な感覚を提供するシステムの実現を目指している.

2. XR 遠隔対話システム

本研究で構築している XR 対話システムの全体構成を図 1 に示す. 左側は遠隔空間 A であり, そこにいる遠隔対話者 A は現場作業空間 B に行かずに自分のアバターロボットを B 地点に置く. アバターロボットには, 全天球カメラ 2 台 (THETA Z1, Ricoh 社) が搭載され, 360° リアルタイム立体映像 (4K, 30 fps) が遠隔対話者 A の HMD (Head Mounted Display, Vive Pro Eye, HTC 社) に提示される. こ

の映像は両眼視差を伴った映像であり, 見かけの遅延なく提示される. 現場作業空間 B では, MR グラス (HoloLens2, Microsoft 社) を装着した現場作業者 B が遠隔対話者 A のリアル CG アバターを, アバターロボットのカメラの位置に見て対話を行う[1].

遠隔対話者 A の HMD のセンサーで計測された頭部の回転運動, 眼球運動, 口領域の運動データを現場作業空間 B の MR グラスに送信し, それに基づいてリアル CG アバターにその運動を再現する.

2.1 対話者リアル CG アバター

遠隔対話者 A のアバターの頭部 3D モデルは, 光学キャプチャで構築した. 計測には iPhone 13 (LiDAR) と Scaniverse (Niantic, Inc.) を用いた. このモデルから図 2 の

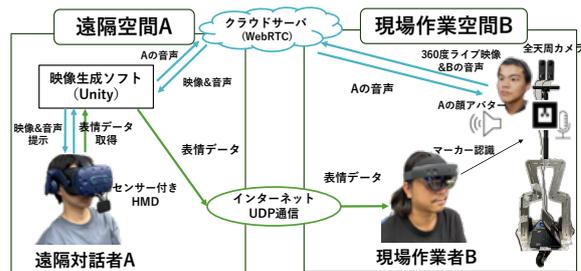


図 1: XR 遠隔対話システムの構成

ブレンドシェイプ用のモデルを作成した[2, 3].



図 2 : 作成した 3D モデル (右は、眼球と口元の変位を適用した場合)

2.2 遠隔対話者 A の表情情報

遠隔対話者 A の表情を CG アバターに再現するため、眼球、頭部および口元の運動 (変位) を HMD に内蔵された眼球運動センサと HMD に追加したリップトラッカー (VIVE Facial tracker : HTC 社) を利用した。取得したデータは眼球 4 自由度、頭部 6 自由度、口 11 自由度である。

2.3 通信処理

本研究では、2 つの通信方法を用いた。現場作業空間 B の 2CH の 360° リアルタイム映像と音声はクラウドサーバを経由して WebRTC で遠隔空間 A に送信される。音声は双方向である。遠隔対話者 A の表情情報は UDP で現場作業空間 B に送信した。

2.4 現場作業空間 B での CG アバターの表示法

遠隔空間 A の遠隔対話者 A のリアル CG アバターのモデルを現場作業空間 B の HoloLens2 に内蔵し、アバターロボットのカメラ位置に提示した。カメラ位置は、認識ライブラリ (Vuforia Augmented Reality SDK, PTC inc.) を用いて下部の AR マーカーで検出し、アバターロボットの運動に追従した。

3. 評価実験

3.1 実験目的

遠隔地にいる人が現場空間で共同作業を行うための本システム構成の有効性と話者の CG モデル (頭部) を提示する効果を定性的に評価した。屋内に設定した現場空間の実オブジェクトと共有 VR オブジェクトを対話で指示して、その写真をホワイトボードから探すモデル作業を対象とした。現場や工場などでのメンテナンス作業を共同で行う状況を想定した実験である。

実験参加者は健康な大学 (院) 生、12 名 (男子 8 名、女子 4 名、21-24 歳) である。

3.2 実験参加者と提示条件

実験参加者を遠隔対話者 A と現場作業空間 B の 2 グループに分け、A、B の 2 人を 1 組として 6 組が実施した。現場作業空間 B が装着する MR グラス (HoloLens2) に表示する遠隔対話者 A のアバターは図 3 に示す 2D 表示と 3D 表示の 2 種類とした。2D 表示は従来のテレプレゼンスロボットの操縦者表示を想定した条件である。



図 3 : 遠隔対話者 A のアバターの提示 (左 : 2D, 右 : 3D)

3.3 実験に用いたタスク

現場作業空間 B に、紙コップ、ペン、書籍、ハサミ、ホチキス、テープの実空間オブジェクト (図 4) を用意するとともに、それらの白黒写真を 3 種類のサイズ (最大 25 cm から最小 9 cm くらいまで) で印刷し、ホワイトボード (図 5, 縦横 180 cm) に貼付した。ボードの表側には 17 種類の実オブジェクトの写真を角度をランダムに変えて配置し、裏側には 3 種類の紙コップのバーチャルオブジェクトと 4 種類の円筒バーチャルオブジェクトのモノクロ写真をランダムな角度で配置した。遠隔対話者 A、現場作業空間 B は同じバーチャルオブジェクトを見ることができたため、図 6 に示すコップと円柱のオブジェクトをアバターロボットの前方向約 30 cm の位置に配置した。

タスクは、現場作業空間 B 役の実験参加者が、実オブジェクトのいずれかを手に持ち、アバターロボットのカメラ (顔) の前に出し、記憶してもらい、それを視野から外した後に、その写真をホワイトボードの写真群の中から探



紙コップ、本、ファイル、ペン、消しゴム、テープ、ホチキス、はさみ

図 4 : 実験に用いた実空間オブジェクト



図 5 : ホワイトボード表側 (左)、裏側 (右)



図 6 : 対話者間で共有したバーチャルオブジェクト (左 : HMD の視野, 右 : HoloLens2 の視野)

してもらった作業である。

3.4 実験手順

現場作業者 B が提示したオブジェクトの写真がホワイトボードに何個存在するかを遠隔対話者に探索・回答させるタスクを以下の手順で実施した。最初に実オブジェクト、次にバーチャルオブジェクトを対象とした。

1. 現場作業者 B が、任意の実オブジェクトを一つ手に取り、カメラの前に提示する。
2. 遠隔対話者 A は、HMD で 360 度リアルタイム立体視映像で実オブジェクトを視認する。
3. 遠隔対話者 A は視認した実オブジェクトの写真がホワイトボードにいくつあるかをロボットの視点を移動しながら探し、口頭で位置を説明して回答する。現場作業者 B は遠隔対話者 A のアバターが探している様子を HoloLens2 で視聴観察する。

このタスクをホワイトボードの両面で行う。バーチャルオブジェクトの場合は、現場作業者 B は共有バーチャルオブジェクトの中から一つを指定し、その写真がホワイトボードにいくつ存在するか探索した。

アバターロボットの視点移動は、VIVE コントローラによる遠隔操作で行うが、本実験時に位置精度が不十分だったため、対話者の口頭指示で実験者が移動した。

3.5 評価項目

上記の作業について、遠隔対話者 A と現場作業者 B に次の質問紙で回答を求めた。

遠隔対話者 A の質問紙を以下に示す。これは現実感 (Q1)、実空間オブジェクトの明確さ (Q2)、バーチャルオブジェクトの明確さ (Q3)、選択肢印刷物の明確さ (Q4)、空自己位置把握 (Q5) の観点からなっている。回答には Visual Analog Scale (VAS) を用いた。質問文と VAS のアンカー (左端, 右端) を次に示す。

- Q1. 現場参加者 B と対話を行ったときの現実感・臨場感はどうでしたか？ [現場作業者 B の様子が全くわからなかった - 現場作業者 B が目の前にいるのと同様の感覚だった。]
- Q2. 対話相手が示した実空間オブジェクトについて明確にわかりましたか？ [何を指示しているのか全くわからなかった - 現場で実際にオブジェクトを見たのと同じだった。]
- Q3. 対話相手が示したバーチャルオブジェクトについて明確にわかりましたか？ [何を指示しているのか全くわからなかった - 明確にどのオブジェクトか理解できた]
- Q4. 視聴映像から、相手空間の回答選択肢は、明確にわかりましたか？ (視点移動を含む) [全く見えず回答できなかった - 現場で見ているのと同程度理解できた]
- Q5. 場所を変えて対話したとき、自分の位置はよくわかりましたか？ [方向と位置が全くわからなかった - どこで作業しているのか位置と方向は完全に分かっていた]

現場作業者 B の質問紙を以下に示す。これは存在感 (Q1)、アバターの視線運動の明確さ (Q2)、リアルアバターの頭部運動の明確さ (Q3)、バーチャルオブジェクト定位 (Q4) の観点からなっている。質問文と VAS のアンカー (左端, 右端) を次に示す。アバターの 2D 表示, 3D

表示の両者について回答させた。

- Q1. 遠隔対話者のリアルアバターを見たときの存在感はどうでしたか？ [人間とは思えなかった - 実際の人間が目の前に存在しているように感じられた]
- Q2. リアルアバターの視線運動が対話対象を示す明確度はどうでしたか？ [不明確で全く理解できなかった - はっきりと疑いなく理解できた]
- Q3. リアルアバターの頭部運動が対話対象を示す明確度はどうでしたか？ [不明確で全く理解できなかった - はっきりと疑いなく理解できた]
- Q4. バーチャルオブジェクトは明確に定位し視認できましたか？ [何があるのか分からなかった - はっきりとその場に存在しているように感じた]

3.6 実験結果及び考察

3.6.1 遠隔対話者 A の質問紙への回答

遠隔対話者 A (HMD 側) に対する質問の回答を図 7 に示す。誤差棒は標準誤差である。遠隔対話者 A は、現場作業者 B の現実感 (Q1: 不明～目の前) について、中間値より高い評価値であり、現場空間の B の様子はかなり分かっている。また指示されたオブジェクト (Q2, Q3: 不明～明確) は現場作業者 B の明確さより高い評価である。これは実オブジェクトはカメラのレンズに実物を接近して見せて観察させていること、バーチャルオブジェクトは CG のため明確さが十分高いためと考えられる。特に、実オブジェクトが CG とほぼ同じ明確さで視認できていることは、本 XR 遠隔対話システムのリアルタイム映像が遠隔地の様子を観察するのに実用的な有効性を持つことを示唆している。全天周映像を使っているが、作業目標はホワイトボード上だけだったために、自己位置把握 (Q5: 不明～明確) は付帯情報だが認識は高かったと思われる。

ホワイトボード上の回答選択肢の印刷物の明確さ (Q4 不明～明確) は他より低かった。これは印刷物が白黒でコントラストが低いこと、カメラと HMD の解像度が肉眼より低いこと、現場の照明が不十分であったことなどで視認性が下がったと考えられる。

3.6.2 現場作業者 B の質問紙への回答 (2D/3D 表示)

現場作業者 B (HoloLens2 側) に対する質問の回答を図 8 に示す。誤差棒は標準誤差である。図 8 より、全ての質問項目で 3D リアルアバターが 2D 表示よりも明確である

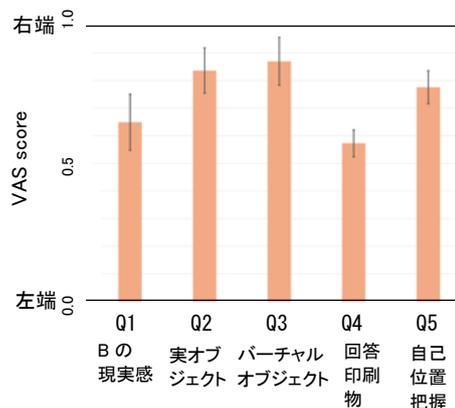


図 7: 遠隔対話者 A の回答, n=6, SE.

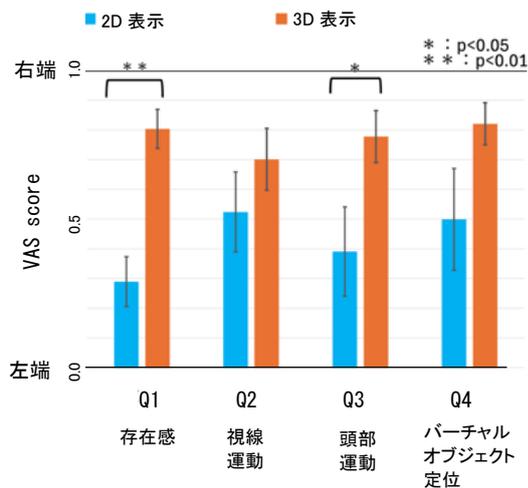


図 8: 現場作業員 B の回答, n=6, SE.

との結果である。Q1, Q2, Q3, Q4 の各質問項目について 2 水準 (2D/3D) 対応ありの t 検定 (MS Excel, Mac ver.16.86) を行った。その結果、遠隔対話者 A のアバターの存在感 (Q1) については、有意水準 1% ($p = 0.00036$) で 3D が高く、アバターの立体表示の効果が表れている。遠隔対話者 A のアバターの頭部回転によって、A が話題の対象を認識しているかどうかの明確さ (Q3) は、有意水準 5% ($p = 0.0274$) で 3D 表示の有効性が示されている。さらに、話題とするバーチャルオブジェクトの定位 (Q4) は、有意差傾向 ($p=0.077$) であり、3D 表示の場合がより明確であると考えられる。これら Q1, Q3, Q4 のスコアはほぼ 0.8 であり、明確さの水準はかなり高かったと考えられる。遠隔から現場空間に参加している人のリアリティが 2D 表示より大幅に高められていることは本手法の有効性を示唆している。

アバターの視線運動が話題オブジェクトを示す明確さ (Q2) は、やや評価値が低く、2D と 3D の有意差は認められなかった。理由としては以下が考えられる。現場作業員 B は、遠隔対話者 A のアバターと正面から向き合っておらず、斜め前から観察し、視線の詳細映像を得ていなかったと考えられる。これは遠隔対話者がホワイトボードの貼付写真の詳細を視認するために、ボードと正対しておりかつ HMD の視野がそれほど広くないため、眼球の運動は小さかったためである。遠隔対話者 A の行動は視線で判断するより頭部の運動を見る傾向であったと考えられる。

2D 表示は、平面ディスプレイがアバターロボットに搭載される場合を想定しており、側方からはディスプレイの表示は見えにくい配置となる。このため、2D 表示の視認性については分散が大きくなったと考えられる。

4. 終わりに

本研究では、遠隔の現場空間に送ったアバターロボットで、現場の人との共同作業を円滑に行うために、アバターロボットを操作する人の CG アバターを XR デバイスでロボットのカメラ位置に表示し、共有した空間オブジェクトに関する対話が可能な環境を構築した。

臨場感の高い対話とするため、HMD を装着した遠隔対話者の 3 次元の CG リアルアバターの視線、頭部運動、発話の口元表情を MR グラスを装着した現場作業員が観察できる設計とした。

HMD を装着した遠隔対話者と MR グラスを装着した現場作業員の両者において、3D 表示を用いた対話と環境把握は概略として良好であり、本システムは遠隔対話に有効であったことが示唆された。遠隔対話者には、立体視映像を提供することで空間オブジェクトの明確さは十分であったと言える。しかし、屋内環境における現場作業員と探索対象のモノクロ印刷物は、カメラの解像度と照明不足により、明確さがやや低くなった。環境の操作が可能な現場では、条件の改善により明確さは高めることが可能と考えられる。

現場作業員の MR グラスでは、屋内条件下では映像は見やすく、3D のアバターとバーチャルオブジェクトの提示は十分に明確であったが、屋外条件の場合は、XR デバイスの性能に依存することになるであろう。アバターの現実感眼は眼球運動、頭部運動、口元の運動が貢献したと考えられるが、表情として自然な再現となるには顔のモデルの制御をより詳細に行う必要がある。

今後の課題として、遠隔対話者の表情の情報を計測し、自然な表情を生成する手法を構成する必要がある。それにより作業の効率だけでなく、コミュニケーションとして品質の高いインタラクションを行うことが可能となると考えられる。さらに、遠隔地から参加する対話者として、複数の人物が参加する場合、各個人のリアルアバターを提示することで、個人のアイデンティティを暗黙の背景とした対話がより円滑となると考えられる。

さらに、アバターロボットの実装では、遠隔対話者が代理身体としてのロボットを自由に操縦できるようにすること、カメラの解像度と照明に関する改善方法を検討することが必要である。また XR システムの基礎的機能としては、遠隔対話の品質を確保する通信の仕様について引き続き検討を行う予定である。

謝辞 本研究は科研費 JP19K20325, 21H04883, 18H04118, 21K19785, 東京都立大学ローカル 5G プロジェクトおよび電気通信普及財団の助成を受けたものである。ここに謝意を表す。

参考文献

- [1] 菊地勇輔, Vibol Yem, 長井超慧, 池井 寧: MR 対話支援型テレプレゼンスシステムに関する研究, 第 25 回日本バーチャルリアリティ学会大会論文集, 3B1-5, 2020.
- [2] 加藤綾斗, 菊地勇輔, ヤム ヴィボル, 池井 寧: VR アバターの表情生成手法に関する研究, 第 26 回日本バーチャルリアリティ学会大会論文集, 3C1-4, 2021.
- [3] Ryoto Kato, Yusuke Kikuchi, Vibol Yem, and Yasushi Ikei. 2022. Reality Avatar for Customer Conversation in the Metaverse. In Human Interface and the Management of Information: Applications in Complex Technological Environments: Thematic Area, HIMI 2022, Held as Part of the 24th HCI International Conference, HCII 2022.