



遠隔操作ロボットを利用した現実空間拡張コンテンツ 「R-Metaverse」 - 第 4 報 —音声システムがソーシャルプレゼンスに与える影響の調査—

”R-Metaverse” – Augmented Reality Content Using Teleoperated Robots – Report 4
—Research for the Effect of Audio System on Social Presence—

錦織勇飛¹⁾, ブイ フイ ホアン¹⁾, 細田侑也²⁾, 李周浩²⁾

Yuhi NISHIKORI, Huy Hoang BUI, Yuya HOSODA, and Joo-Ho LEE

1) 立命館大学大学院 情報理工学研究科 (〒 567-8570 大阪府茨木市岩倉町 2 - 150, nishikori.aislab@gmail.com)

2) 立命館大学 情報理工学部 (〒 567-8570 大阪府茨木市岩倉町 2 - 150, leejooho@is.ritsume.ac.jp)

概要: 本論文では、遠隔操作ロボットを用いて現実空間と仮想空間の両方にメタバース技術を展開するシステム Ritsumei Real-World Robotic Metaverse (R-Metaverse) において、ロボットの音声システムと遠隔の参加者のソーシャルプレゼンスの関係性について検証する。実験では、遠隔の参加者の音声を機体から出力するようにスピーカーを搭載した遠隔操作ロボットを使用する。実験の結果、ヘッドホンで音声を聴く場合と比較して、スピーカーを用いた音声システムがソーシャルプレゼンスの向上に貢献するという仮説は有意水準を満たさなかった。

キーワード: 遠隔操作ロボット, メタバース, 複合現実, 音声システム

1. 緒言

近年、ユーザを模したアバタが仮想空間上で交流するメタバース技術は、娯楽分野だけでなくオンライン会議や商品の展示紹介などのビジネス分野でも利用されている。異なる空間に存在するユーザが仮想空間上で効果的に交流するためには、相手がそこにいると感じる度合いをあらわすソーシャルプレゼンスを高めることが重要である。ソーシャルプレゼンスは、表情や仕草などの非言語情報を伝えられるメディアほど高く、良好なコミュニケーションに繋がる [1]。しかしながら、仮想空間上でコミュニケーションは本質的には非人格的であるため、良好な関係を構築することは容易でない [2]。そのため、ソーシャルプレゼンスを意識して非言語情報を効果的に参加者に提供できるメタバース技術を開発することが必要である。

これまでに著者らは、遠隔の参加者のアバタをあらわす遠隔操作ロボットを用いて現実空間と仮想空間の両方で現地の参加者との交流を促すメタバース技術 R-Metaverse (図 1) [3, 4, 5] を開発している。R-Metaverse では、ヘッドマウントディスプレイを介して遠隔の参加者に、携帯端末を介して現地の参加者に、現実空間と CG モデルを融合した複合現実 (MR) を提供する。遠隔の参加者は、ロボットの前面に付いているカメラで現地の様子を確認しながら、遠隔操作ロボットを操作して現地の参加者と交流する。また現地の参加者は、現実空間のロボットにアバタを投影した携帯端末を介して遠隔の参加者と交流する。

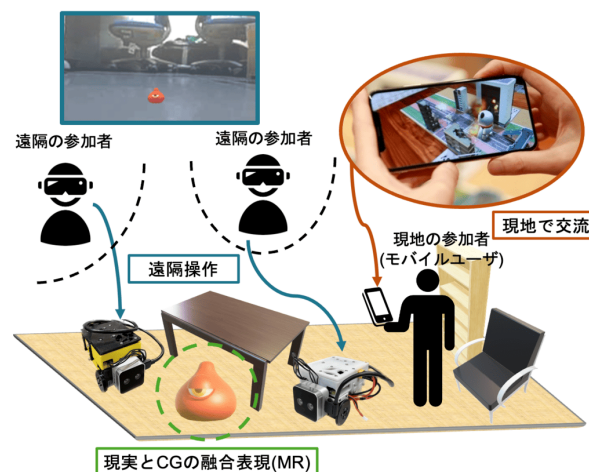


図 1: R-Metaverse の概要図

本論文では、R-Metaverse における遠隔の参加者のソーシャルプレゼンスを高める手法として、その音声をロボット本体から再生する音声システムを検証する。従来の音声システムでは、現地の参加者はヘッドホンを介して音声を聴くため、遠隔の参加者のプレゼンスを誤って知覚する恐れがある。そこで、遠隔操作ロボットを遠隔の参加者のアバタと知覚しやすくなるように、そのロボットにスピーカーを取り付けて音声を再生する。実験では、ヘッドホンとスピーカーを比較して、R-Metaverse におけるソーシャルプレゼンスへの影響について主観評価を実施する。



図 2: 遠隔操作ロボット

表 1: スピーカの特徴

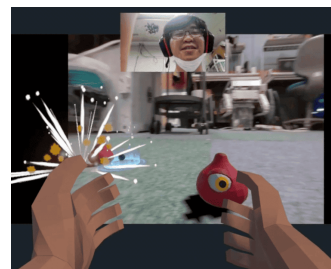
型番	LC001
メーカー	LC-doida
接続方式	AUX 3.5 mm
出力	3W
周波数応答	50-18000 Hz
電源供給	内蔵バッテリー
サイズ	幅 46 mm, 奥行 46 mm, 高さ 33 mm
重さ	50 g

2. 手法

図 2 は、R-Metaverse で使用する遠隔操作ロボットを示す。遠隔操作ロボットのサイズは幅 80 mm, 奥行 125 mm, 高さ 107 mm, 重さ 500g で、室内環境での移動に適した設計である。またロボットのコンピュータとして Raspberry Pi 4 Model B を、ロボット前面にはステレオ深度カメラ (Intel RealSense D405; Intel) を搭載している。さらに本研究では、ロボットの上面にスピーカー (LC001; LC-doida) を取り付けた。表 1 は、スピーカーの特性を示す。

図 3 は、本実験中に投影する複合現実を示す。ヘッドマウントディスプレイ (Meta Quest 3; Meta) は、現実空間における遠隔操作ロボットの位置を視点として、その様子に 3D オブジェクトを投影した複合現実をディスプレイに投影する。このとき、遠隔の参加者はコントローラでロボットを操作して、その音声を内臓マイクで収録する。収録した音声は、遠隔操作ロボットに取り付けられたスピーカーから出力される。携帯端末 (Galaxy S20 Ultra 5G; SAMSUNG) は、遠隔の参加者のアバタを遠隔操作ロボットに付与した複合現実を描画するアプリケーションを提供する。そのディスプレイには、現場の参加者の画像とその視点における仮想空間を投影する。

図 4 は、R-Metaverse のシステム構成を示す。本システムでは、遠隔操作ロボット、コンピュータ、ヘッドマウントディスプレイおよび携帯端末は ROS2 で通信する。サーバは、ROS2 側と Unity 側でそれぞれ役割が異なる。ROS2 側のサーバでは、ヘッドマウントディスプレイや携帯端末、遠隔操作ロボット、サーバ側の Unity とのデータ通信を管理する。Unity 側のサーバでは、遠隔操作ロボットや携帯端末で撮影した RGB 画像を背景にして 3D オブジェクトを配置して、現実空間と仮想空間が融合した表現である複合現実



(a) ヘッドマウントディスプレイ



(b) 携帯端末

図 3: 投影する複合現実

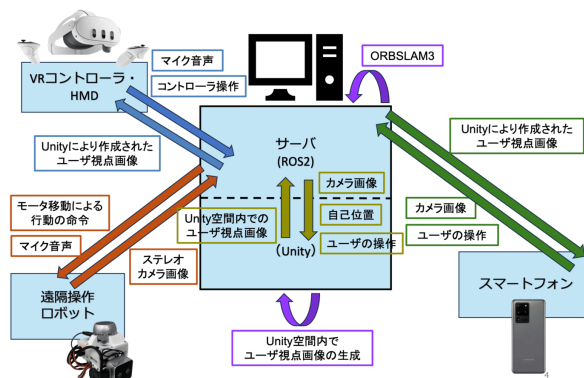


図 4: R-Metaverse のデータ処理の流れ

を作り出す。このとき、VisualSLAM[6] で推定した自己位置に合わせて視点をそれぞれ定めることで、リアルタイムで共通の仮想空間を遠隔および現地の参加者に提供できる。

3. 実験

3.1 実験条件

実験では、ロボット本体のスピーカーあるいはヘッドホンで外部参加者の音声を出力したときのソーシャルプレゼンスについて検証した。スピーカーを用いるとき、現地の参加者は耳ではなく首にヘッドホンをかけた。また、ヘッドホンを用いるときは、R-Metaverse とは別に用意した PC あるいはタブレットを介して外部の参加者と通話した。被験者である 20 代の男性 10 名は、現地あるいは遠隔の参加者として二種類の音声システムを順番に体験した。このとき、体験する音声システムの順番はランダムに設定して、それぞれ三つのタスクを被験者に課した。それらのタスクを終える度に、現地の参加者のみアンケートに回答した。

表 2: ソーシャルプレゼンスに関するアンケート内容

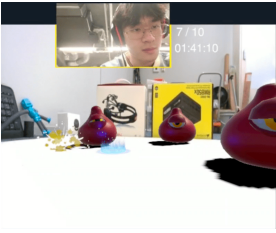
質問	内容	評価軸分類
1	相手がすぐ近くにいるように感じた	空間的存在
2	相手と一緒に行動している感覚があった	行動的連携
3	相手の声の位置が自然に感じられた	音声定位による臨場感
4	会話がスムーズに行えた	会話の流暢性
5	相手の存在を強く感じた	総合的存在感
6	楽しくコミュニケーションが取れた	感情的関与
7	相手が自分に注意を向けていると感じた	注意配分
8	相手の意図や考えを理解しやすかった	認知的理解
9	相手と心が通じ合っていると感じた	感情的理解
10	自分の行動が相手の行動に影響していると感じた	相互依存性
11	声の距離感が自然に感じられた	現実感の補強
12	相手と一緒に遊んでいるという感覚があった	協働的楽しさ
13	自由記述（感想など）	その他の意見

表 3: 比較評価に関するアンケート内容

質問	内容
1	どちらの条件のほうが、相手が近くにいると感じましたか？
2	どちらの条件のほうが、会話しやすかったですか？
3	どちらの条件のほうが、ゲームとして楽しかったですか？
4	その他、気づいたことがあれば自由にお書きください

表 2 は、ソーシャルプレゼンスに関するアンケート内容を示す。本実験では、Harms らによる Networked Minds Inventory[7] を参考に、空間的存在感や行動的連携などを評価軸として採用している。また、音声定位および声の距離感などは本研究の音声システムに関する評価のため導入している。アンケート内容の質問 1 から質問 12 までは、1 が「あてはまらない」で、7 が「あてはまる」とする 7 段階のリッカート尺度を使用した。また、質問 13 では自由記述の欄を設けた。二つの音声システムに関するアンケートを回答した後、さらに、それらシステムの好みと印象を調べるための比較評価アンケートも実施した。表 3 は、比較評価アンケート内容を示す。比較評価アンケートでは、質問 1 から質問 3 まで「ロボット本体のスピーカー」、「ヘッドホン」、「どちらでもない」の三択から被験者は選択した。また、自由記述として質問 4 を用意した。

本実験では、遠隔および現地の参加者で共同のタスクに取り組む。このとき、図 5 および図 6 は、遠隔および現地の参加者の実験の様子をそれぞれ示す。タスクでは、爆弾でスライムを 10 体倒すことを目的とした。このとき、スライムは、赤色、青色、黄色のいずれかの爆弾でしか倒せない仕様である。遠隔の参加者は、操作する右コントローラで爆弾が飛ぶ方向を定めて、左コントローラで爆弾攻撃の色を切り替えた。また、スライムの弱点となる色は現地の参加者のみがディスプレイ上で知り、その位置も俯瞰視点で確認することができる。さらに、ディスプレイをタップした位置に遠隔の参加者も確認できるマークを仮想空間上に設定できる。実験では、現地の参加者がスライムの弱点の色や位置を遠隔の参加者に伝えるように被験者に指示した。10 体のスライムを倒し終わるまでを 1 タスクとし、そのタイムを計測した。また、被験者のゲームをするモチベーションや集中力を高めるために、音声システムごとにタスクは連続で三回実施してベストタイムの更新を促した。

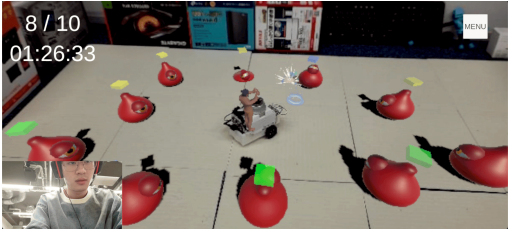


(a) アプリケーション画面



(b) 俯瞰図

図 5: 遠隔の参加者の実験の様子



(a) アプリケーション画面



(b) 俯瞰図

図 6: 現地の参加者の実験の様子

3.2 実験結果

表 4 は、ソーシャルプレゼンスに関するアンケート内容についてマン・ホイットニーの U 検定を実施した結果を示す。12 項目の主観評価すべてにおいて統計的な有意差は得られなかった ($p > 0.05$)。これらの結果より、スピーカーを用いた音声システムがソーシャルプレゼンスの向上に関して寄与しない可能性を示している。本タスクでは、遠隔操作ロボットが初期位置からあまり動かずにスライムを倒すことができていた。スピーカーの位置がほとんど動かないため、ロボット本体から音声出力されていると被験者が認識することが困難であったと考えられる。ただし、アンケート結果のいくつかの項目では、スピーカーを用いた場合の方が平均値が高い結果が得られた。質問 2 では、ロボット本体からの音声出力により、操作するユーザの存在が

表 4: ソーシャルプレゼンスに関するアンケート内容
に対する U 検定結果

質問	ロボット本体のスピーカー	ヘッドホン	p 値
1	M=4.70, SD=1.89	M=4.90, SD=1.79	0.816
2	M=6.60, SD=0.52	M=6.00, SD=0.94	0.113
3	M=5.60, SD=1.35	M=4.60, SD=1.58	0.146
4	M=5.60, SD=1.84	M=5.90, SD=1.10	0.872
5	M=6.00, SD=0.94	M=5.80, SD=1.63	0.404
6	M=6.40, SD=0.70	M=6.30, SD=0.48	0.576
7	M=5.40, SD=1.71	M=5.70, SD=1.25	0.782
8	M=5.80, SD=0.92	M=6.10, SD=0.74	0.487
9	M=6.40, SD=0.52	M=6.00, SD=0.67	0.165
10	M=6.70, SD=0.48	M=6.60, SD=0.52	0.648
11	M=5.50, SD=1.72	M=4.30, SD=1.95	0.133
12	M=6.70, SD=0.48	M=6.10, SD=0.74	0.054

表 5: 被験者 10 名に対する比較評価アンケートの選択
結果

質問	スピーカーが良い	ヘッドホンが良い	どちらでもない
1	7	2	1
2	2	7	1
3	6	2	2

空間的に感じられ、共同行動している感覚が強まった可能性があることを示唆している。本実験で扱う協力型のゲームタスクにおいては、ユーザ間での自然な相互作用の感覚が得られやすくなることが望ましいため、空間内で音声適切に定位することが重要である。また質問 12 では、有意水準 0.05 に近い値を示した。共遊感覚は、空間的な共有感や臨場感に影響されるため、ロボットから音声の直接聞こえることで一緒に遊んでいる感覚がより強く得られたと考えられる。本実験では、被験者の数が限られていたため、今後は被験者を増やしてこれらの項目について深く検討することに取り組む。

表 5 は、比較評価アンケートの結果を示す。まず、スピーカーを使用した場合、現地の参加者は空間的近さの知覚やゲームの楽しさについて高く評価した。ユーザ間の空間的なつながりを強化してや協力型ゲームにおける楽しさを高めることは、音声を通じた相互作用に効果があったと考えられる。一方で、ヘッドホンを使用した場合、音声の明瞭性が優れていた。周囲の音を遮断して音声をクリアに聞き取れるため、協力型タスクにおいて重要な情報伝達が正確に行われたと考えられる。現地の参加者は遠隔操作ロボットから一定以上離れていたため、スピーカーの配置や音量を調整して明瞭性と空間的存在感を両立することが今

後の課題である。

4. 結言

本論文では、R-Metaverse において、遠隔の参加者の音声の出力形式がソーシャルプレゼンスに与える影響を検証した。実験では、ロボット本体に搭載したスピーカーとヘッドホンの二種類の音声システムを用いて、協力型のゲームタスクを実施した。実験の結果、音声システムによるソーシャルプレゼンスの違いは確認されなかった。ただし、一部のアンケート内容より、スピーカーで音声を出力することで、ゲームタスクの楽しさが向上することが示唆された。今後は、スピーカーの配置や音量設定など見直して、より多くの被験者で主観評価実験を実施する。

参考文献

[1] J. Short, E. Williams, and B. Christie. The Social Psychology of Telecommunications. Wiley, (1976).

[2] J. B. Walther and M. R. Parks. Cues Filtered Out, Cues Filtered In: Computer-Mediated Communication and Relationships. Handbook of Interpersonal Communication, 3rd Edition, Sage, Thousand Oaks, pp. 529–563, (2002).

[3] 児玉遼太郎, チャンディントゥアン, 李周浩. テレグジスタンスロボットを用いた R-Metaverse の提案. 日本ロボット学会学術講演会, (2022).

[4] 錦織勇飛, 藤井康之, チャンディントゥアン, 李周浩. 遠隔操作ロボットを利用した現実空間拡張コンテンツ「R-Metaverse」- 第 2 報 - R-Metaverse におけるモバイルユーザとの交流-. ロボティクス・メカトロニクス講演会, (2024).

[5] 錦織勇飛, 藤井康之, チャンディントゥアン, 李周浩. 遠隔操作ロボットを利用した現実空間拡張コンテンツ「R-Metaverse」- 第 3 報 - 3DCG アバタの違いによるソーシャルプレゼンスの変化-. システムインテグレーション部門講演会, (2024).

[6] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel and J. D. Tardós, ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimodal SLAM, IEEE Trans. Robot., vol. 37, no. 6, pp. 1874–1890, (2021).

[7] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence, (2004).