

メタバース空間で対話相手に向かって 自律移動する AI アバター『ノア』の開発

小山田 悠人^{1,a)} 湯村 翼^{1,b)}

概要：メタバース空間では会話可能な AI アバター開発が進む。しかし、その多くは対話機能に限定され、アバターの動作を交えたインタラクションは困難という課題がある。そこで本研究では、ユーザーとの会話に加え、呼びかけに応じて自律移動を行う AI アバターを開発する。VRChat を対象プラットフォームとし、ユーザーが発する音声の方向をリアルタイムで算出し、音源へ向かって自律的に接近する機能を Python で実装した。ChatGPT 等と連携した対話も行い、AI が能動的に振る舞う、存在感のあるインタラクティブな体験の創出を目指す。

1. はじめに

ChatGPT に代表される大規模言語モデル (LLM) の急速な発展に伴い、AI との対話は身近な技術となった。この潮流はメタバース空間、特にソーシャル VR プラットフォームである VRChat においても例外ではなく、キャラクター性を付与された AI と音声で対話するツールやアバターが数多く開発・報告されている。これらの AI アバターは、ユーザーに新たなコミュニケーションの形やエンターテインメントを提供している。

しかし、既存の多くの AI アバターには共通の課題が存在する。その一つは、AI アバターの多くがその場に留まり、ユーザーからの呼びかけに応答するだけの静的な存在である点である。移動能力を持つアバターも存在するが、それらは事前に定められたルートを巡回するワールドギミックとして実装されている場合が多く、ユーザーとの対話に応じて動的に振る舞う機能は持たない。つまり、対話と移動という二つの要素が分断されているのが現状である。

そこで本研究では、この課題を解決するため、ユーザーとの音声対話能力と、対話相手に向かって自律的に移動する能力を両立した AI アバターの開発を行う。具体的には、ユーザーが発した声の方向をトリガーとして、AI アバターが能動的にユーザーへ接近し、対話を開始するシステムを構築する。これにより、AI が単なる応答プログラムではなく、空間内にいるという強い存在感をユーザーに与え、より自然でインタラクティブなコミュニケーション体験を

創出することを目的とする。最終的には、特定のワールドに依存せず、多くのユーザーが手軽に導入できる汎用的なツールとしての展開を目指す。

2. 関連研究

VRChat 空間ににおけるアバターの自律移動に関する研究や事例はいくつか存在する。クレイマンによる VRChat アバターを生物のように操作する人工知能の作成 [1] では、原始的な生物の行動原理に基づき、ボトムアップで移動アルゴリズムを開発している。また、別のメタバースプラットフォームである Roblox では、DeterminantAI[2] のようなプラグインが開発されており、NPC がキャラクターの位置など多角的な環境認識に基づいて自律的な行動計画を立てる。

対話システムとの連携に目を向けると、EmemeAI の VRChat 連携機能 [3] は、AI キャラクターとの対話と移動を可能にするサービスを提供する。さらに、Song らによる LLM-Driven NPCs[4] は、Unity と Discord を横断して一貫した対話体験を提供するため、クラウド上で対話履歴を管理するシステムを提案している。

これらの先行研究に対し、本研究は主に 3 つの点で独自性を有する。第一に、行動のトリガーである。先行研究がアルゴリズムや多角的な環境認識に基づいて行動を決定するのに対し、本研究はユーザーの発話方向というリアルタイムかつ社会的なインタラクションの入力に特化している。第二に、AI の動作形態である。EmemeAI などがワールドに設置されるギミックとして機能するのとは異なり、本研究の AI はプラットフォーム上を一個の独立したユーザーとして振る舞う。第三に、研究の主眼である。

¹ 北海道情報大学

a) s2221031@s.do-johodai.ac.jp

b) yumu@yumulab.org

LLM-Driven NPCs が対話の継続性に焦点を当てる一方、本研究では既存の LLM の対話能力を基盤とし、対話の文脈や音声入力と連動した移動とモーションによって、より人間らしく自然なコミュニケーションを実現することを目的としている。

3. 設計と実装

本システムは、VRChat 上で稼働する AI アバターと、その思考と行動を制御する外部の Python プログラムから構成される（図 1）。

両者は Open Sound Control(OSC) API を介して通信し、連携動作を実現する。また、VRChat の画面を Optical Character Recognition(OCR) で認識する。

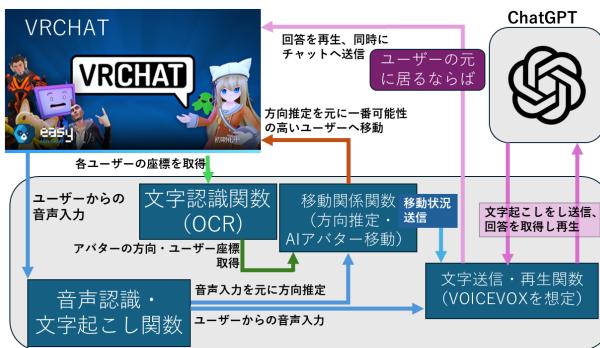


図 1 システム概要。

3.1 システム構成

本 AI アバター『ノア』は、以下のモジュールを連携させることで機能する。対象プラットフォームは VRChat を選定した。制御プログラムの実装には Python を利用した。

- (1) 音声認識: 高速・高精度な音声認識モデル Faster Whisper を利用し、ユーザーの発話をリアルタイムでテキスト化する。
- (2) 対話生成: ChatGPT API (gpt-4o-mini モデル) を使用。事前に設定されたキャラクターのペルソナに基づき、自然な応答文を生成する。
- (3) 音声合成: オープンソースの音声合成エンジン VOICEVOX の API と連携し、生成されたテキストをキャラクターの声で発話させる。
- (4) アバター制御 (移動・動作): Python から OSC メッセージを VRChat へ送信し、アバターの移動や回転を制御する。
- (5) 情報取得 (座標・方向): AI アバターのクライアント画面に表示させた座標等の情報を OCR で読み取ることで、自己位置や他ユーザーの位置を推定する。

3.2 自律移動の実装

ユーザーへの自律的な接近は、本研究の中核機能であり、

以下の手順で実現される。

音源方向の推定

AI アバターが VRChat 内で聞いている音声を、仮想オーディオデバイスを介して Python プログラムに入力する。入力された音声の左右チャンネルの音量差を解析し、ユーザーがどの方向にいるかを推定する。但し、この方法では左右 90 度しか認識できず、前後の区別が不可能なため、単純な音量差比較だけでなく、音が聞こえた方向へ回転し続けるスキャンニング方式及び近くのユーザーの方向へ回転し、音源が正面に来るかどうかの判定を併用することで、特定精度を向上させる。

接近動作の実行

推定された音源方向に対し、アバターをその場に回転させる OSC コマンドを送信する。ターゲットとなるユーザーが正面に来たことを確認後、前進するコマンドを送信し、ユーザーとの距離を縮める。

対話の開始と終了

アバターがユーザーに接近後、本格的な対話モードに移行する。ユーザーが「さようなら」など特定のキーワードを発話すると、システムは対話終了を認識し、アバターは事前に定められた初期位置へ自動的に帰還する。

3.3 開発プロセスと技術的課題

開発初期段階では、ワールド側から HTTP リクエストのパラメータを用いて、Python プログラムに情報を送る事でユーザーの座標を取得する手法を検討した。しかし、送信頻度に制限があるという VRChat の制約により断念した。現在は、AI アバター自身が見ている VRChat クライアント画面を OCR で取得し、ワールド上にユーザーの座標を表示する事で情報を取得している（図 2）。



図 2 VRChat クライアント画面。

この過程で、音源方向推定の精度、OCR による座標認識の安定性、VRChat と制御プログラムを同時実行する際のパフォーマンスなど、複数の技術的課題が明らかになった。これらに対しては、マルチプロセス処理の導入による処理の並列化や、認識アルゴリズムの改良によって対応を進めている。

4. 結果と考察

プロトタイプの実装と動作検証を通じて、ユーザーの呼びかけに応じて AI アバターが振り向き、接近し、対話を開始するという一連の動作が実現可能であることを確認した。声を発した相手に AI が能動的に関わろうとする挙動は、静的な AI アバターと比較して格段に高い存在感とインタラクティブ性を生み出すことが示唆された。

一方で、現状のプロトタイプには以下の課題が存在する。

音源方向の推定

音声認識モデルの初回ロードや API 通信に時間を要し、応答に遅延が生じる。また、OCR の認識精度によって数字の誤検出（例えば"5"と"9"など）が見られるため、精度の向上が必須となる。

音源方向の推定

現在の実装は 1 対 1 の対話を想定しており、複数人が同時に存在する空間において、どのユーザーが発話したかを正確に識別可能かどうかは未知数である。

音源方向の推定

移動と会話は実現したが、対話内容に応じた身振り手振りや表情の変化といった、より細やかな感情表現が不足している。

音源方向の推定

現状のシステムは、仮想オーディオデバイスの設定など、導入にある程度の知識を要する。

5. おわりに

本研究では、メタバース空間における新たなインタラクションの形を提案するため、対話能力と自律移動能力を兼ね備えた AI アバターの開発に着手した。プロトタイプを通じて、ユーザーの音声に反応して自律的に移動する AI アバターの開発が進むと共に、この研究がユーザーに対してどのような影響を与えるかの問題がある。

今後の展望として、まず、対話内容と連動したモーション・表情生成機能の追加を行う。次に前述の課題解決に取り組む。具体的には、処理の高速化と安定性の向上、話者識別機能の実装、ユーザーが設定を容易に行える GUI を Flet 等のフレームワークを用いて開発し、最終的にはアバターへの導入を簡便化するモジュールとしてパッケージ化し、誰もがいつでも会える友達として AI アバターと触れ合える未来を目指し、開発を継続していく。

参考文献

- [1] クレイマン: VRChat アバターを生物のように操作する人工知能の作成、バーチャル学会発表概要集, Vol. 2024, pp. 40–41 (オンライン), DOI: 10.57460/vconf.2024.0_40 (2024).
- [2] DeterminantAI: Create Intelligent AI Characters For Your 3D Virtual World, , available from <<https://www.determinant-ai.com>> (accessed 2025-07-16).
- [3] EmemeAI: オリジナルの AI キャラクターを外部アプリに接続できる 3D チャット AI 投稿プラットフォーム『EmemeAI』3 月 27 日新機能リリース, , 入手先 <<https://www.atpress.ne.jp/news/430788>> (参照 2025-07-16).
- [4] Song, L.: LLM-Driven NPCs: Cross-Platform Dialogue System for Games and Social Platforms (2025).