

仮想エージェントの注視対象遷移モデルに向けての仮想空間における会話参加者の視線分析

Gaze Analysis of Conversational Participants in a Virtual Environment for Developing a Gaze Target Transition Model for Virtual Agents

石原 理央 黄 宏軒*
Rio Ishihara Hung-HsuanHuang

福知山公立大学情報学部
Faculty of Informatics, The University of Fukuchiyama

Abstract: 本研究では、参照物が共有される対話環境において、人間の視線行動を注視対象レベルで捉え、その遷移をモデル化することを目的とする。従来の視線研究は視線方向や注視点の解析に主眼が置かれており、実際にどの対象が注視され、その切り替わりがどのように生じるかについては、十分に検討されていない。本研究では、VR 空間で取得した視線データと環境情報を統合し、注視対象を推定することで、対象間の注視遷移を記述・分析する。さらに、その結果をもとに対話エージェント設計への応用可能性を検討する。

1 はじめに

近年、メタバースや VR 空間の発展に伴い、仮想空間上での社会活動やビジネス利用が急速に拡大している。すでに自治体の情報発信・住民対話空間（例：静岡県 Metaverse SHIZUOKA）、企業の仮想展示・商品体験空間（日産のメタバース展示）、ソーシャル VR プラットフォームを通じた交流（VRChat の利用事例）、さらには高齢者向け福祉活動（VR を用いたセラピー体験）など、社会活動やビジネス利用の具体的な事例が存在する。会議、接客、教育、イベント運営など、従来は対面で行われていた活動が、時間や場所の制約を受けずに実施可能となりつつある。一方で、これらの活動を常に人間が直接担うことには、時間的・人的コストの観点から限界がある。

このような背景から、仮想空間において人間の役割を部分的に代替し、利用者と自然に関わることでできる対話エージェントの必要性が高まっている。とくに、メタバース空間における案内、説明、議論支援などの場面では、「その場に存在し、会話に参加している存在」として振る舞うエージェントが求められる。

しかし、現在の多くの対話エージェントは、発話内容が適切であっても、非言語行動が乏しい、あるいは不自然であるため、利用者に対して「話を聞いてもらえ

ている」「相互に理解し合っている」という感覚を十分に与えられていない。対話においてこのような感覚は、信頼感や会話の円滑さに直結する重要な要素である。

対話相手が話を聞いていると感じられる要因の一つとして、視線行動が挙げられる。Kendon (1967) は、視線が感情表出や会話調整といった機能を担うことを指摘しており、相手や話題対象に適切に視線が向けられることで、聞き手の関与や注意が非言語的に伝達されることが知られている。逆に、視線が会話内容と無関係な対象に向けられている場合、利用者に違和感や不信感を与える可能性がある。そのため、対話エージェントに自然な視線行動を付与することは、メタバース空間における社会的エージェント実現のための重要な課題である。

一方で、自然な視線生成を実現するためには、まず人間が対話中に「どの対象を注視しているのか」、そして「その注視が時間的にどのように遷移しているのか」を理解する必要がある。従来の視線生成研究では、視線方向や頭部動作を直接制御・学習する手法が主流であり、人間の視線行動を「注視対象」という意味的な単位で捉え、その遷移構造を体系的に分析する試みは十分になされてこなかった。しかし、対話理解やエージェント設計の観点からは、単なる方向情報ではなく、「誰／何を見ているか」という抽象度での記述が重要である。

そこで本研究では、メタバースや VR 空間における

*連絡先：福知山公立大学情報学部情報学科
〒620-0886 京都府福知山市字堀 3370
E-mail:hhhuang@acm.org

対話エージェントの自然な視線生成を最終目標とし、その第一段階として、参照物が共有される VR 空間における二者対話を対象に、人間の視線行動を分析する。具体的には、対話中に人間が「どの対象を注視しているか」および「その注視が時間的にどのように遷移するか」を明らかにすることを目的とする。そのために、本研究では VR 空間での対話実験を実施し、HMD に搭載されたアイトラッキング機能を用いて視線情報を取得する。取得した視線情報と環境情報を統合することで、対話中に注視されている対象を推定し、注視対象の時間的遷移構造を分析する手法を提案する。

2 関連研究

人と人以外が会話することに関する研究として、内田ら [1] は、人、ロボット、アンドロイドに対する第一印象が、参加者の自己開示行動にどのような影響を与えるかを検証している。その結果、人はロボットやアンドロイドに対しても一定程度の自己開示を行うことが示唆された。また、人を対話相手とした場合にはポジティブな事柄に関する自己開示が多い一方で、ロボットやアンドロイドに対しては、話題がニュートラルあるいはネガティブになるほど自己開示の割合が高くなる傾向が確認された。これらの結果から、対話エージェントに対しても人は一定の社会的態度を持って接することが分かり、対話相手としての妥当性が示されている。

人間同士の対話における非言語行動の分析として、石井ら [2] は、複数の視線情報および頭部動作の情報を用いて会話参加態度を推定する手法を提案している。この研究では、注視パターン、注視時間、注視位置の移動距離、瞳孔径といった視線情報に加え、眼球検出に依存しないヘッドトラッキングによる頭部姿勢情報と会話参加態度との間に相関があることが示された。さらに、視線情報と頭部動作情報の両方を用いた推定モデルが最も高い性能と頑健性を示すことが報告されている。本研究においても、視線情報に加えて頭部動作に関する情報を収集し、視線行動分析の補助情報として活用する。

視線行動の基礎的理解に関しては、Rayner[3]や Land & Tatler[4]による研究が広く知られている。Rayner は、読字、シーン知覚、視覚探索といった課題における眼球運動を整理し、注視 (fixation) を視覚情報取得の基本単位として位置づけている。また、注視の持続時間や空間的安定性が、注意や認知処理と密接に関係していることを示している。

一方、Land & Tatler は、日常行動や自然行動を対象とした分析を通じて、人間の視線行動が環境や行為と強く結びついた能動的プロセスであることを論じている。これらの研究は、視線行動を単なる物理的な視線

方向としてではなく、「どの対象に注意が向けられているか」という意味的単位で捉える重要性を示している。

先行研究において、注視は一般にサッカードのような高速な視線移動とは区別され、一定時間以上にわたり視線が相対的に安定している区間として扱われている。Rayner (2009) は、読字や視覚探索課題において、注視の持続時間が概ね 100~500ms の範囲に分布し、特に 200~300ms 前後が典型的であることを報告している。また、自然行動を対象とした Land & Tatler (2009) の研究においても、視線が一定時間以上対象に向けられて安定している区間が、行為や注意と結びついた注視として分析されている。

一方、アイトラッキング機器の実装やデータ処理の観点からは、注視は 50~600ms 程度の持続時間をもつ視線停留として整理されることが多く、タスクや環境に応じて柔軟に閾値が設定されている。本研究では、これらの理論的知見および実装上の整理を踏まえ、短時間のサッカードやノイズ成分を除外するための下限として 50ms を設定し、視線位置が一定範囲内に収まり、かつ 50ms 以上継続した区間を注視として抽出する操作的定義を採用する。

マルチパーティ対話環境における注視対象の推定に関して、Massé ら [5] は Visual Focus of Attention (VFOA) 推定モデルを提案している。この研究では、視線方向および頭部姿勢といった非言語情報を用いて、対話参加者が「誰」あるいは「どの対象」に注意を向いているかを確率的に推定する手法が示されている。VFOA は単なる視線方向ではなく、視線が実際に向けられている対象を推定する概念であり、「何を見ているか」を扱う点で本研究の目的と密接に関連している。しかし、この研究は主に実世界の対話環境を対象としており、仮想空間における三次元オブジェクトとの関係や、環境情報を明示的に用いた注視対象同定は扱っていない。また、注視対象間の遷移構造をモデル化すること自体を主目的とはしていない。

対話における視線行動の役割については、Nakano や Ishii ら [6][7] による一連の研究がある。これらの研究では、対話中の視線遷移を分析することで、話者交替、会話へのエンゲージメント、注意の逸脱といった対話状態を推定できることが示されている。特に、視線がどの対象からどの対象へと移動するかという対象間の遷移情報が、対話理解において重要な手がかりとなる点が指摘されている。また、Ishii らは、対話エージェントが人間の注視行動を理解・利用するための枠組みを示し、注視行動を意味的に解釈することの重要性を論じている [7]。

さらに、人間同士の対話においては、相手の視線方向に影響を受けて自身の視線が移動する「追従注視 (gaze following)」が生じることが古くから指摘されている。Kendon[8] は、視線行動を単なる知覚的反応ではなく、

相互行為を組織するための行為として捉え、会話相手の視線方向が注意の配分や相互理解に影響を与えることを論じている。この立場では、視線の追従は固定的な時間閾値によって定義されるものではなく、対話の文脈に依存した即時的反応として理解される。

一方、心理学的実験に基づく研究では、他者の視線方向に対する追従反応は比較的短時間で生起することが報告されている。Frischen ら [9] は、視線による注意喚起効果に関するレビューを行い、他者の視線提示後、おおよそ 300~700ms の範囲で注意の追従が生じることを示している。これらの知見は、視線追従が自動的かつ社会的意味を持つ反応であることを示唆している。

しかし、これらの研究の多くは静的刺激や単純化された実験課題を用いたものであり、実際の対話環境や仮想空間において、具体的な注視対象（人物、物体、環境内オブジェクト）への追従行動を時系列的に分析した研究は限られている。

本研究では、仮想空間内の対話環境において、参加者同士の注視対象の時間的關係に着目し、相手の注視後、短時間内に同一対象への注視が生じた場合を追従注視として捉える枠組みを採用する。これにより、対話中における視線の相互影響を定量的に分析することを目的とする。

3 提案手法

本章では、第 4 章で実施する VR 対話実験により取得される視線および環境情報を前提として、注視対象および注視対象遷移を推定・分析するための枠組みを示す。

本研究では、対話環境における人間の視線行動を「注視対象」とその時間的遷移として捉え、注視対象およびその遷移構造を推定・記述するための分析枠組みを提案する。

本手法は、自然な視線生成を最終目標とした基礎的研究として、人間が会話中に「どの対象を注視しているか」および「その注視がどのように切り替わるか」を明らかにすることを目的とする。

提案手法は、

1. 視線および環境情報の取得（実験の実施）
2. 視線と 3D 環境情報を統合した注視対象の推定
3. 推定された注視対象列に基づく注視対象遷移の抽出と分析

上記の 3 段階から構成される。

3.1 手法の概要

本研究で提案する手法の全体的な流れを示す。本手法では、対話中に取得した視線情報と VR 環境内の 3D 環境情報を統合することで、各時刻における参加者の注視対象を推定し、その時系列的变化を注視対象遷移として記述・分析する。

まず、VR 空間内で対話を行う参加者から、HMD に搭載されたアイトラッキング機能を用いて視線方向ベクトルおよび頭部姿勢などの視線関連情報を取得する。同時に、対話相手の身体部位や対話中に参照される可能性のある 3D オブジェクトについて、位置・形状・識別情報などの環境情報を記録する。

次に、取得した視線方向と環境内オブジェクトとの空間的關係に基づき、各フレームにおける注視対象を推定する。具体的には、視線方向に基づいて Raycast 処理を行い、最初に衝突したオブジェクトを注視対象として割り当てる。これにより、単なる視線方向ではなく、「誰（または何）を見ているか」という意味的な注視対象レベルでの視線行動を推定する。

最後に、推定された注視対象を時系列データとして整理し、注視対象がどの対象からどの対象へと切り替わったかを注視対象遷移として抽出する。これらの遷移を分析することで、参照物が共有される対話場面における人間の視線行動の構造的特徴を明らかにする。

注視対象の定義 注視対象は、視線原点から視線方向ベクトルに基づいて Raycast を行い、最初に衝突した 3D オブジェクトと定義する。VR 空間内の各オブジェクトには、あらかじめ注視対象カテゴリを示すタグを付与しており、本研究で用いた主なカテゴリは以下の通りである。

- Human：対話相手の頭部および上半身
- TV：参照物として地図を添付
- Object：机や椅子などの周辺物体
- None：床や壁を含む具体的な注視を定義できなかった領域

なお、本研究で用いたデータにおいては、周辺物体への注視は机のみが検出された。また、TV に関しては画面の右上、右下、左上、左下をそれぞれ TV1、TV2、TV3、TV4 とした。

4 視線および環境情報の取得実験

実験の目的 本実験の目的は、参照物が共有される対話場面において、参加者がどの対象を注視し、その注視がどのように遷移するかを記録・分析するための視線データおよび環境データを取得することである。

参加者 参加者は大学生3ペア（計6名）であり、普段から会話を行う間柄のペアを募集した。内訳は女性ペア1組、男性ペア2組である。VR利用経験は4名あり、未経験は2名。視力条件は1名裸眼、他は矯正視力。眼鏡使用者は実験中外した状態。

実験課題および会話内容 実験では、参加者2名がペアとなり、VR空間内で互いに向かい合った状態で対話課題を行った。対話は合計3回実施され、それぞれ異なる目的を持つ会話テーマを設定した。

1回目の対話では、VR環境への慣れを目的とし、「夏休みの予定」をテーマとした軽い雑談を行った。

2回目の対話では、「海外からの留学生が来た際のおもてなしプラン」をテーマに、具体的な内容を話し合ってもらった。この対話は、次の対話課題につながる準備段階として位置づけられる。

3回目の対話では、2回目で決定したおもてなしプランについて、VR空間内に設置したTVに表示された地図を参照しながら、訪問順序や動線を含めた最終決定を行ってもらった。本研究では、明示的な参照物が存在するこの3回目の対話データを主な分析対象とする。

実験手順および参加者への説明 実験開始前に、参加者に対して実験内容および手順の説明を行った。参加者は、5分間の会話を1回、7分間の会話を2回実施した。各会話の間には休憩時間を設け、実験前にはHMDの装着およびキャリブレーションを含むセットアップを行った。

実験中の行動を記録するため、参加者にはHMDを装着してもらい、視線情報に加えて、VR内の画面録画およびWebカメラによる顔映像を記録した。

4.1 実験装置・環境

本実験では、参加者が操作する2体のキャラクターを用意した。キャラクターの造形はVroidを用いて作成し、Blenderにより目や眉の動きを表情キー（BlendShape）として実装した^{1,2}。例として、キャラクターに左を見る動作をさせる場合、図1の様にEye_Left_Right, Eye_Right_Rightという二つのBlendShapeの値を変化させることで実現できる。

次に、UnityおよびSteamVRを用いて会話環境（Unity 3D）を構築した³。VR内の内装は、参加者の視界と現実空間との乖離による事故を防止するため、実際の実



図1: キャラクターと BlendShape

験環境を模倣して作成した。図2に構築した実験環境を示す。

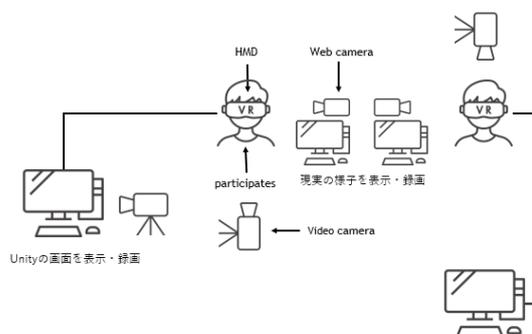


図2: 構築した実験環境

実験では2台のPCと2台のHMDを使用した。参加者2名はそれぞれ自身の動きをトラッキングしたキャラクターを操作し、互いのトラッキングデータをネットワーク経由で送受信することで、同一のVR空間内で対話を行う。この環境には、参加者キャラクター、トラッキングデータに基づいてキャラクターを制御するプログラム、データ送受信プログラム、およびデータ記録・保存プログラムが含まれる。

実験中は以下の処理が繰り返し行われる。

1. HMD およびコントローラにより参加者の動作データを取得する。
2. 取得したトラッキングデータを自身のUnity環境内で再現する。
3. 自身のトラッキングデータを相手側のUnity環境へ送信する。
4. 相手側のトラッキングデータを受信し、Unity環境内で再現する。

¹Blender Foundation: Blender Documentation, <https://docs.blender.org/manual/en/latest/>

²pixiv Inc.: Vroid Studio Guidelines, <https://vroid.com/en/studio/guidelines>

³Unity Technologies: Unity マニュアル (VR/XR 開発ガイド), <https://docs.unity3d.com/ja/current/Manual/VR0verview.html>

視線計測にはアイトラッキング機能を備えたヘッドマウントディスプレイ VIVE Pro Eye を用いた⁴。Vive Pro Eye は HTC 社製の VR ヘッドセットであり、視線計測用の赤外線センサーを内蔵している。本装置を用いることで、対話中の視線情報を高精度に取得することが可能である。

参加者 2 名分のデータを同時に取得するため、本実験では 2 台の PC および HMD を使用し、トラッキングデータの送受信には NetMQ ライブラリを用いた⁵。NetMQ は ZeroMQ の C# バインディングであり、効率的かつ低遅延なデータ通信を実現できる。

実際の実験の様子は図 3 のようになった。



図 3: 実験環境：(上) 現実環境における実験中の様子，(下) 対応する Unity 環境

4.2 取得データの内容

本実験では、VR 空間内での対話中における参加者の視線行動、身体動作、および表情情報を記録し、注視対象推定および注視対象遷移分析に用いるためのデータを取得した。

視線関連データ HMD に搭載されたアイトラッキング機能を用いて、各フレームにおける参加者の視線情

⁴HTC Corporation: VIVE Pro Eye Support & Setup Guide, <https://www.vive.com/jp/support/vive-pro-eye/>. 本機器の視線計測機能は Tobii のアイトラッキング技術に基づいている。

⁵NetMQ Project:NetMQ Documentation, NetMQ 公式ドキュメント (C# メッセージングライブラリ), <https://netmq.readthedocs.io/>

報を取得した。取得した視線関連データは表??の通りである。

- 視線原点 (HMD 座標系)
- 視線方向ベクトル
- 視線計測の有効フラグ

これらの情報は、Unity 内のワールド座標系に変換したうえで保存し、注視対象推定処理に利用した。

注視対象データ 取得した視線方向ベクトルに基づき、VR 空間内で Raycast 処理を行い、各時刻の注視対象を推定した。Raycast により最初に衝突した 3D オブジェクトを注視対象として割り当て、そのカテゴリおよび識別情報を記録した。

頭部および姿勢データ HMD およびコントローラから取得される頭部位置と回転 (姿勢) データを記録した。視線情報の補助的情報として利用し、視線行動と身体動作の関係性分析に用いた。

表情関連データ 将来的な対話エージェントの非言語表現生成を見据え、顔の Blendshape 情報をフレーム単位で記録した。取得した表情データにはまばたき、眼球運動、眉や口の動きなどが含まれる。本研究では主に視線行動に着目し、補助的情報として扱う。

時間情報および同期 すべての取得データには、Unity の内部時間に基づくタイムスタンプを付与し、視線情報、注視対象データ、頭部姿勢データ、および表情関連データがフレーム単位で同期されるように記録した。

分析対象データ 取得した 3 回の対話データのうち、明示的な参照物 (TV 画面) が存在する 3 回目の対話データを主な分析対象とした。

5 前処理

本研究では、取得した視線および環境情報を用いて注視対象推定および注視対象遷移分析を行うにあたり、データの信頼性を確保するための前処理を行った。本節では、視線データの整形、ノイズ除去、および分析単位への変換について述べる。

5.1 視線データの整形

HMD から取得される視線データには、瞬きや計測失敗などに起因する欠損フレームが含まれる。本研究では、視線計測の有効フラグを用いて、有効でないフレームを除外した。

また、視線原点および視線方向ベクトルは、HMD のローカル座標系で取得されるため、Unity のワールド座標系へ変換を行った。これにより、視線と VR 空間内オブジェクトとの空間的關係を一貫した座標系で扱うことを可能にした。

5.2 Raycast 結果の後処理

視線方向に基づく Raycast 処理では、フレーム単位で注視対象が推定されるが、視線の微小な揺れや瞬間的な視線移動により、注視対象が短時間で頻りに切り替わる場合がある。これらは意図的な注視行動とは言えないため、以下の処理を行った。

まず、Raycast がいずれのオブジェクトにも衝突しなかったフレームについては、「未分類」として扱い、明確な注視対象が定義できない区間であるため、注視対象遷移の分析対象から除外した。

次に、同一の注視対象が連続して出現するフレームを一つの注視区間としてまとめ、注視区間の長さが一定時間未満のものはノイズとして除去した。本研究では、注視区間の最小持続時間を 50ms と設定した。この値は、先行研究における注視持続時間の一般的な下限を参考に決定した。

5.3 注視対象系列の生成

前処理後のデータから、各参加者の注視対象を時間順に並べた注視対象系列を生成した。この系列は、連続する注視区間ごとに1つの注視対象ラベルを割り当てた離散的な時系列データである。

この注視対象系列を用いることで、「どの対象からどの対象へ注視が切り替わったか」という注視対象遷移を明示的に抽出することが可能となる。

5.4 遷移データの抽出

生成した注視対象系列から、隣接する注視対象間の遷移を抽出し、注視対象遷移データとして整理した。本研究では、自己遷移（同一対象内での継続）は遷移として扱わず、異なる注視対象への切り替わりのみを遷移として定義した。

抽出した遷移データは、参加者ごと、対話セッションごとに集計し、遷移頻度および遷移確率の算出に用いた。

5.5 表情および頭部動作データの扱い

取得した表情および頭部動作データについては、視線データと同一のタイムスタンプを用いて同期を行った。本研究では主に視線行動および注視対象遷移の分析を目的とするため、表情および頭部動作データは直接的な分析対象とはせず、補助的な情報として保存した。

ただし、将来的には、注視対象の切り替えと表情変化や頭部動作との関係性を分析することで、より自然な対話エージェントの非言語行動生成への応用が可能であると考えられる。

6 データ分析・可視化

本分析では、収集した全参加者データの中から、例として参加者 A および B の1回分のセッションデータを用いる。これは、解析手法の説明および視線行動パターンの例示を目的としており、個別の注視比率や遷移パターンを理解しやすくするためである。

さらに、対話実験時の状況を再現し、注視対象や行動パターンを直感的に把握するための Unity による再現環境も構築した。

6.1 実験状況の再現

取得した視線および注視対象データを用いて、実験の状況を Unity 上で一部再現する可視化環境を構築した。本再現環境では、実験における参加者の頭部位置・回転および視線方向データを CSV 形式で読み込み、Unity 上においてキャラクターの動作および視線の再現を行った。具体的には、CSV ファイルから時刻ごとの視線方向ベクトルおよび頭部の位置・姿勢データを取得し、線形補間および球面線形補間 (Slerp) を用いることで、時間軸に沿ったスムーズな動きを再現した。再現された視線は Raycast による注視対象との交差判定を行い、注視位置には可視化用の青い球体を配置することにより、対象物への注視点を直感的に確認できるようにした。

さらに、ユーザーインターフェースとしてスライダーを設置し、任意の時刻における視線再生や再生速度の変更が可能となっている。これにより、特定の注視パターンや行動パターンを時間軸に沿って観察することが可能となり、注視対象遷移や個人差の理解を補助する手段として有効である。図 4 は実際の環境である。

6.2 注視データの基礎集計

全データを通して算出した結果、実験全体の記録時間に対して注視が発生していた時間は約 85% を占めて

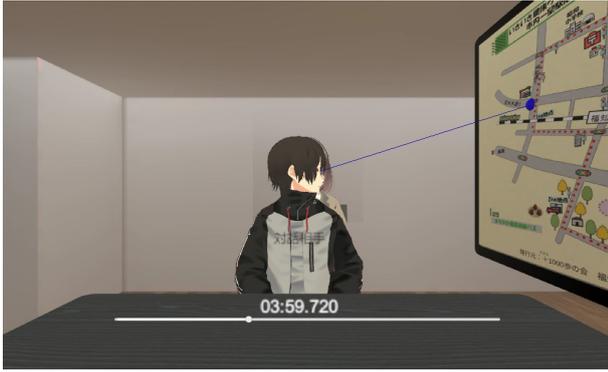


図 4: Unity 上で再現した参加者の視線および注視対象の一例

おり、参加者の視線は大部分の時間において特定の対象に向けられていたことが確認された（表 1）。

表 1: 実験全体における注視時間と非注視時間の割合

| 項目 | 時間 [s] | 割合 [%] |
|-------|----------|--------|
| 注視時間 | 12316.74 | 84.85 |
| 非注視時間 | 2199.77 | 15.15 |
| 合計 | 14516.51 | 100.00 |

一方で、残り約 15%は注視に該当しない時間であり、視線が対象間を移動している区間や、明確な注視対象が存在しない状態も一定量含まれていることが分かる。このことから、本実験環境は視線行動を十分に引き出しつつも、探索的な視線移動も含む現実的な対話状況を再現できていると考えられる。

6.3 注視時間割合の全体傾向

カテゴリ別に注視時間の割合を算出した結果、Human カテゴリへの注視が全体の約 44%を占め、最も高い割合を示した（表 2）。

表 2: カテゴリ別に見た注視時間の割合（全データ）

| カテゴリ | 注視時間 [s] | 割合 [%] |
|-------|----------|--------|
| Human | 5397.97 | 43.8 |
| None | 4795.83 | 38.9 |
| TV | 1732.78 | 14.1 |
| Table | 390.16 | 3.2 |

次いで None カテゴリが約 39%と高く、参加者の視線が必ずしも常に明示的な対象に向けられているわけではなく、空間的な探索や対象間遷移が頻繁に生じていることが示唆される。一方、Table への注視は全体の約 3%と低く、環境内に存在するものの、視線行動に

おいては背景的な役割を果たしている対象であると考えられる。

また、TV カテゴリは特定の条件下でのみ提示されたにもかかわらず、全体の注視時間の約 14%を占めており、視線を強く引き付ける視覚刺激として機能していることが分かる。この結果は、対話環境における外的刺激が視線行動に大きな影響を与える可能性を示している。

6.4 テーマ別注視行動の比較

さらに、実験条件ごとの違いを検討するため、各テーマにおける注視行動の比較を行った。テーマ別によりカテゴリごとの注視時間割合を算出した結果を表 3 に示す。

表 3: テーマ別に見たカテゴリごとの注視時間割合 [%]

| Theme | Human | TV | Table | None |
|--------|-------|------|-------|------|
| Theme1 | 52.1 | - | 2.6 | 45.3 |
| Theme2 | 51.1 | - | 4.4 | 44.5 |
| Theme3 | 29.5 | 40.4 | 2.1 | 28.0 |

Theme_1 および Theme_2 では、Human カテゴリへの注視時間が全体の約 50%以上を占めており、参加者の視線が主として対話相手に向けられている傾向が確認された。これらの条件では、対話相手が視線行動における主要な注視対象となっていることが示唆される。

一方で Theme_3 では、Human カテゴリへの注視割合が約 30%まで低下し、代わって TV カテゴリへの注視が増加する傾向が見られた。これは、対話環境内に視覚的に顕著な外的刺激が存在することで、参加者の視線が対人対象以外へと分散した可能性を示している。

これらの結果から、対話相手は常に最優先で注視される対象であるとは限らず、環境内の視覚的要因によって注視の優位性が変化する可能性があると考えられる。

6.5 追従注視の定義と抽出方法

本研究では、対話中において一方の参加者の注視行動が、他方の参加者の注視行動に影響を与える現象、すなわち「追従注視 (gaze following)」の有無を検討する。追従注視は、対人相互作用における共同注意の成立過程を反映する重要な指標であり、視線行動の時間的関係性を考慮することで定量的に評価可能である。

本研究における追従注視は、「相手がある対象を注視し始めた後、一定時間内に自分が同一対象へ注視を遷移させた場合」と定義する。この定義は、視線が他者の注視対象に引き込まれるという現象を、時間遅延を伴う注視遷移として捉えるものである。

まず、視線ログおよび環境情報ログを時刻同期し、同一対象への連続注視が 50ms 以上継続した区間を注視イベントとして抽出した。各注視イベントについては、注視開始時刻、終了時刻、注視対象、および注視継続時間を記録し、注視行動を離散的なイベント列として表現した。

次に、ペア内の 2 名の参加者について、同一テーマ内における注視イベントを比較した。ある参加者 A が特定の対象を注視し始めた時刻を基準とし、参加者 B が同一対象を注視し始めるまでの時間差を算出した。この時間差が正であり、かつ 0.7 秒以内である場合、参加者 B は参加者 A の注視に追従したとみなし、これを追従注視イベントとして抽出した。

以上の手続きにより、追従注視イベントについて、テーマ別の発生回数および注視遷移の平均遅延時間を算出した。これにより、各実験条件において、参加者が相手の注視行動にどの程度影響を受けているかを定量的に評価することが可能となる。

追従注視の統計 抽出された追従注視イベントに基づき、各テーマ・注視対象ごとの追従回数と平均潜時を算出した。解析結果の概要を Table 4 に示す。

表 4: テーマ別・注視対象別追従注視イベントの概要

| Theme | 注視対象 | 回数 | 平均遅延 [s] |
|---------|-------------|-----|----------|
| Theme_1 | Head | 72 | 0.48 |
| Theme_1 | Upper chest | 470 | 0.51 |
| Theme_1 | Table | 13 | 0.51 |
| Theme_1 | None | 436 | 0.51 |
| Theme_2 | Head | 107 | 0.50 |
| Theme_2 | Upper chest | 798 | 0.50 |
| Theme_2 | Table | 39 | 0.47 |
| Theme_2 | None | 749 | 0.50 |
| Theme_3 | Upper chest | 401 | 0.50 |
| Theme_3 | TV1 | 13 | 0.54 |
| Theme_3 | TV2 | 22 | 0.53 |
| Theme_3 | TV3 | 92 | 0.49 |
| Theme_3 | TV4 | 42 | 0.48 |
| Theme_3 | Table | 9 | 0.52 |
| Theme_3 | None | 211 | 0.50 |

結果として、平均潜時はおおむね 0.48~0.54 秒であり、心理学的先行研究の 0.3~0.7 秒と整合的であった。また、Human (Head および Upper chest) への追従が最も多く、次いで TV 系、Table や None への追従が観測された。このことは、対話参加者が相手の注視対象に影響を受ける傾向が強く、社会的注視行動の反映として妥当であることを示唆している。

7 考察

本研究では、VR 空間における二者対話を対象として、参加者の視線行動を「注視対象」レベルで抽出し、

その時間的遷移を分析した。取得データおよび遷移分析の結果から、以下の知見が得られた。

第一に、対話中の視線行動は単なるランダムな方向移動ではなく、会話相手や参照物に応じた意味的な構造を持つことが示唆された。特に、TV に表示された参照地図への注視は、会話の内容に応じて増減し、注視対象遷移のパターンにも反映されることが確認された。これは、視線が対話理解や情報処理に密接に関連していることを示しており、意味的注視単位での分析の有用性を支持する結果である。

第二に、追従注視の観点では、参加者同士の視線が互いの注視対象に影響を受ける様子が確認された。相手が特定の参照物を注視した直後に、自身も同対象へ注視を切り替えるパターンが一定の頻度で観察され、対話中の非言語的な相互調整の存在が示された。これにより、自然な対話エージェント設計においても、相手の注視行動をリアルタイムで反映する視線生成が有効であることが示唆される。

第三に、注視対象遷移の個人差も観察された。特に、VR 利用経験や注意傾向の違いにより、同一課題でも注視対象遷移の頻度やパターンに差が見られた。これは、対話エージェントが汎用的に自然な視線行動を生成する際に、利用者の個性や文脈に応じた調整の必要性を示している。

これらの結果は、従来の視線研究が注視点や視線方向に着目していたのに対し、本研究の「注視対象遷移」アプローチが、対話理解やエージェント設計における実用的知見を提供できることを示している。また、VR 環境における視線データの取得・解析が、非言語行動の理解とモデリングに有効であることも確認された。

8 おわりに

本研究では、参照物が共有される VR 空間における二者対話を対象に、人間の視線行動を注視対象レベルで抽出し、注視対象遷移を分析する手法を提案した。取得データの解析により、対話中の視線行動が意味的構造を持ち、相互に影響し合うこと、さらに個人差が存在することが明らかとなった。

これらの知見は、対話エージェントに自然な視線行動を付与するための基礎的情報として活用可能である。特に、相手の注視行動に追従する視線制御や、参照物への適切な注視タイミングの生成など、社会的・認知的文脈を考慮した視線生成モデルの設計に有用である。

今後の課題としては、より多人数対話環境への拡張、異なるタスクや参照物の種類に応じた注視遷移パターンの一般化、さらには表情や身体動作と視線行動の統合分析が挙げられる。これらを通じて、仮想空間にお

ける対話エージェントの自然性と信頼性を高めるためのモデル構築を進めることが期待される。

[11] 賀好 昭仁 著, 作って学べる Unity 本格入門 (Unity 6 対応版), 技術評論社, 2024, ISBN: 978-4-297-14514-9.

参考文献

- [1] 内田 貴久, 高橋 英之, 伴 碧, 島谷 二郎, 吉川 雄一郎, 石黒 浩: ロボットによる傾聴を通じた自己開示の促進, 日本認知科学学会大会, 2017 年 9 月.
- [2] 石井 亮, 大西 亮太, 中野 由紀子, 西田豊明: 視線と頭部動作に基づくユーザの会話参加態度の推定, 情報処理学会論文, Vol. 52, No. 12, pp3626-3636, 2011 年.
- [3] Rayner, K.: Eye movements and attention in reading, scene perception, and visual search, *Quarterly Journal of Experimental Psychology*, Vol. 62, No. 8, pp. 1457–1506, 2009.
- [4] Land, M., & Tatler, B.: *Looking and Acting: Vision and Eye Movements in Natural Behaviour*, Oxford University Press, 2009.
- [5] Massé Benoit, Ba Siléye, Horaud Radu: Tracking Gaze and Visual Focus of Attention of Multiple People, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11), 2711 - 2724, 2018.
- [6] Yukiko I. Nakano, Ryo Ishii: Estimating user's engagement from eye-gaze behaviors in Human-agent conversations, *Proceedings of the 15th international conference on Intelligent user interfaces* 139-148, 2010-02-07.
- [7] Yukiko I. Nakano, Ryo Ishii, Toyoaki Nishida: Gaze Awareness in Conversational Agents, *ACM Transactions on Interactive Intelligent Systems* 3 (2), 1-25, 2013-07.
- [8] A. Kendon: Some functions of gaze-direction in social interaction, *Acta Psychologica*, Vol. 26, pp. 22–63, 1967.
- [9] A. Frischen, A. P. Bayliss, and S. P. Tipper: Gaze cueing of attention: Visual attention, social cognition, and individual differences, *Psychological Bulletin*, Vol. 133, No. 4, pp. 694–724, 2007.
- [10] Harrison Ferrone 著; 吉川 邦夫 訳: *Unity 3D ゲーム開発ではじめる C#プログラミング, インプレス*, 2021, ISBN: 978-4-295-01245-0.