

意図を読まれるエージェントは共同注意を誘発する Agents with Perceived Intentions Elicit Joint Attention

土橋 一斗^{1*} 中島 亮一² 大澤 正彦¹
Kazuto Dobashi¹, Ryoichi Nakashima², Masahiko Osawa¹

¹ 日本大学
¹ Nihon University
² 京都大学
² Kyoto University

Abstract: 本研究では、ミニマルデザインのエージェントと共同注意が生じるかを検証した。結果、目のないエージェントでは生じにくいですが、そのエージェントが「見る」機能を持つことを知らせる事前インタラクションを行うと、生じた。つまり、共同注意は他者の見た目のみならず、他者に対する認識によって影響を受ける。ここで事前インタラクションを行った場合にはユーザのエージェントに対する擬人感が高まっており、ユーザは「見る」という行為の意図性を感じ取っていた可能性が高く、共同注意はエージェントに対する擬人感や意図の程度を定量化する指標となりえる。

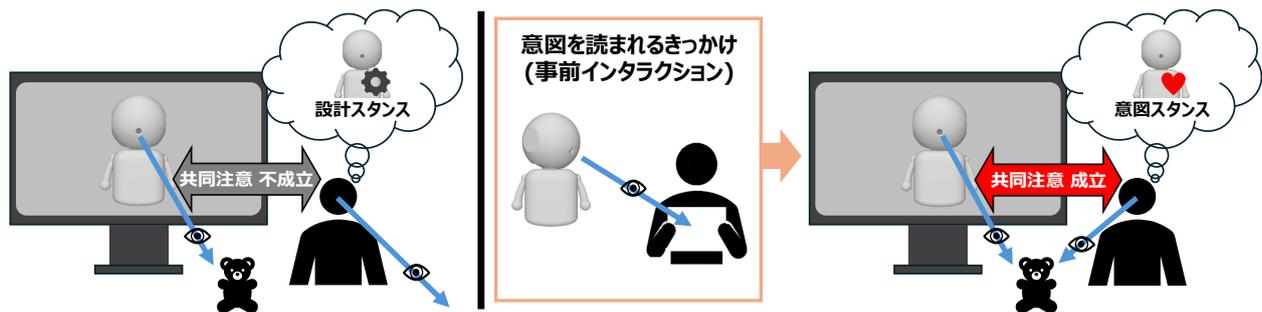


図 1: 本研究で得られた知見。目のないミニマルデザインのエージェントとの共同注意は成立しないが、事前インタラクションをした場合にのみ成立した。視線に対する意図性を感じるかが重要と考えられ、ユーザがエージェントを意図スタンスで捉えているか、設計スタンスで捉えているかを測る手がかりとなるかもしれない。

1 はじめに

近年、大規模言語モデルやコミュニケーションロボットの普及に伴い、エージェントのインタラクションは日常的なものとなりつつある。ユーザが対話型のエージェントに対して持つスタンスは、デネットが提唱する3つのスタンス [1] のうちの2つに大別できると考えられる。1つは設計スタンスである。エージェントが設計スタンスで認識される場合、ユーザは挨拶や感謝の礼儀等を省略するなど、心理的なコストをかけずにインタラクションを行うことができる。また、単刀直入で簡易的な指示出しができるとともに、リテイクが可能になることもメリットとして挙げられるだろう。も

う1つは意図スタンスである。エージェントが意図スタンスで認識される場合、ユーザはエージェントのエラーに対して許容性が向上し、適応行動や援助行動をとりやすくなる。Human-Agent Interaction(HAI)の文脈においては、エージェントを意図スタンスで認識させることは重要である。

エージェントが意図を読まれた状態でインタラクションするには2つの技術基盤が必要である。1つは、意図を読まれるエージェントを実現する技術である。人工物であるエージェントは、意図スタンスで捉えられるために何らかの工夫が必要とされてきた。例えば、ユーザにとって行動予測をしやすい振る舞いをする [2] や、あえて弱々しいデザインにすること [3] など、行動から見た目まで多岐にわたる提案がされてきた。もう1つは、ユーザがどの程度エージェントを意図スタンスで

*連絡先：日本大学文理学部
〒156-8550 東京都世田谷区桜上水 3-25-40
E-mail: chka24009@g.nihon-u.ac.jp

認識しているかを定量的に計測・評価する技術である。近年の対話型生成 AI をはじめ多くのエージェントが「意図スタンスで認識されていること」を前提に応答する仕組みになっているが、全てのユーザが等しく意図スタンスで捉えている状況は考えにくい。本来であれば意図スタンスで捉えられているかを推測し、その程度に応じて適切な応答をすべきである。例えば、Godspeed Questionnaire[4] などの事後的な主観評価や、特定のタスク文脈に適した評価 [5] が存在するが、事後的な評価ではその時々の変化や無意識的な反応を捉えることが困難であり、具体的なタスクや文脈に依存した指標では汎用性に乏しい。多くのユーザと意図スタンスで捉えられた状態でインタラクションが可能なエージェントを実現するためには、リアルタイムかつ汎用的な方法でユーザのスタンスを推定しつつ、より意図スタンスで捉えられやすくなる行動ができるインタラクションの設計が必要である。

そこで本研究では、ユーザのスタンスをリアルタイムに測る汎用的な指標として共同注意に着目する。共同注意の成立には、なんらかの対象を意図を持って視線を向けているかどうかに関係するため、エージェントが意図を持った存在として認識されているか否かを判定する要素になり得る。もしも共同注意を用いてユーザのスタンスを測定できることが明らかになれば、様々な場面で適用可能な汎用性の高い評価指標を確立したといえる。また、共同注意を誘発するのによく用いられる情報として視線がある。視線は主体の意図を表現する典型的なモダリティでもあり、視線を使ったインタラクションを用いることは、意図スタンスで捉えられるきっかけとしても有望である。つまり、視線という単一のモダリティを活用して、ユーザのスタンスの計測から意図スタンスへの誘導までを一貫して扱うことができる可能性がある。

本研究の知見は図 1 としてまとめられ、本研究の貢献は主に 2 つある。

【認知科学的貢献】

見る機能があるということを知覚させると、目がないエージェントはビジュアルに頼らずとも、視線手がかり効果（共同注意に関連した現象）を誘発するという新たな知見を示した。

【工学的貢献】

共同注意を誘発する振る舞いをし、人間がそれほどの程度反応するかを計測することで、インタラクションの中で人間がエージェントの意図を読んでいる程度を定量化できる可能性を示した。

2 背景

2.1 エージェントに対するユーザのスタンス

近年、大規模言語モデル（Large Language Models: LLM）の急速な発展や普及により、人々はチャット形式のインターフェースを通じ、日々の業務から私的な相談に至るまで、LLM にあらゆる支援を求めるようになった。また、コミュニケーションロボットの社会実装も着実に進展しており、人々がエージェントとコミュニケーションをとる機会は、もはや特殊なものではなく日常の一部となりつつある。

エージェントとのコミュニケーションにおいて、ユーザがエージェントを認識するスタンスは、「意図スタンス」と「設計スタンス」に大きく分かれる [1]。「意図スタンス」は、エージェントの動作や発話に意図を見出し、あたかもエージェントが主体的に行動しているかのように解釈するスタンスを指す。「設計スタンス」は、エージェントの動作や発話が、あくまで事前に定められた設計やプログラムの結果であると解釈するスタンスを指す。

ユーザのエージェントに対するスタンスと実際のエージェントの振る舞いが一致しないと、ユーザとエージェントのコミュニケーションの質が大きく低下する。例えば、ユーザが設計スタンスに基づき、効率的な回答のみを求めている際に、エージェントが過剰に人間らしく情緒的な対応をすれば、ユーザは煩わしさを感じるだろう [6]。逆に、ユーザが意図スタンスに基づき、共感を求めている際に、エージェントが事実の列挙のみを行えば、そこには負の適応ギャップが生じ、ユーザの対話意欲は著しく減退する [7]。この問題を解決するためには、ユーザのスタンスを推定する必要があるが、これについての研究はそれほど多くなく、未解明なことも多い。

2.2 ユーザのスタンス推定

ユーザとエージェントのコミュニケーションを向上させるために、ユーザのスタンスを適切に推定することが重要である。例えばユーザの言語情報から推定する手法も考えられるが、スタンスは明示的な言葉として表れるとは限らない。例えば、「本当に？」という短い問いかけ 1 つをとっても、それが AI が出力した情報の正確性を疑うファクトチェック（設計スタンス）なのか、あるいは話の内容に対する純粋な驚きや感情の共有（意図スタンス）なのかを、文字列のみで判断することは難しい [8]。つまり言語情報のみからユーザのスタンスを推定するには限界がある。

したがって、ユーザのスタンスを言語情報以外の指標を用いて推定する必要がある。ユーザのスタンスを

推定するための言語情報以外の指標として何が考えられるだろうか。人間同士のコミュニケーションを参考に、意図的な行動時に引き起こされるいくつかの状態について議論する。

まず、人間同士の会話を考えてみる。通常、会話において相手の発話を途中で遮る行為は非礼とされるが、ユーザが対象を設計スタンスに基づいて接している場合、このような社会的配慮は働かない。必要な情報が得られた時点、あるいはエージェントの回答が求めているものと異なる判断した瞬間に、ユーザは躊躇なくシステムの発話を中断し、次の命令や訂正を行うだろう。これにより、ユーザが対象に対して敬意や人格を見出しているか否かが如実に表れる。

また、韻律にもユーザのスタンスは反映される。例えば、他者に対して語りかける際、声には自然な感情的抑揚やフィラーが含まれる。これに対し、音声認識器に対して語りかける際、ユーザは誤認識を防ごうとする意識から、明瞭かつ平坦で大きな発声を行うだろう [9]。

さらに、パーソナルスペースや共同注意など、人間同士のコミュニケーションにおける心理学的な知見も、意図を持った主体を前提としている。例えば、人は無意識にパーソナルスペースを維持し、コミュニケーション相手に対し不快感を与えないよう適切な距離を取ろうとする [10]。一方、対象が単なる物体であれば、そこに社会的規範は適用されず、ユーザは操作の効率性を優先して躊躇なく物理的距離を詰める。また、人間は視線や指差しなどで対象を指すことで相手の注意を誘導するといった共同注意を行っている。共同注意は社会的な情報のやりとりであり、相手が単なる物体であれば共同注意は生じづらいと考えられる [11]。

2.3 スタンス推定指標の選定

ユーザのスタンス推定の指標において、獲得の容易さや正確さが重要となってくる。例えば、音声情報は既存のスマートスピーカやマイク等でも取得可能であるために情報の獲得は容易である。しかし、音声情報はユーザが発話している瞬間にしか取得できず、エージェントの話をしている最中のスタンスを測ることはできない。

また、対人距離は数値として正確な情報の取得が可能である。しかし、PCやスマートフォンを介した対話においては、ユーザは一定の位置から動かないことが多く、距離の変化をスタンスの指標として用いることは困難である。また、対人ではない対象（デバイス）との物理的距離を、対人距離の指標として適用することの妥当性についても不明確である。

そこで、本研究では共同注意に注目する。共同注意とは、二者が同じ対象に注意を向け、その注意の共有

を通して情報のやりとりを行う過程である [12]。発話の有無にかかわらず常時観測が可能であるため、情報の獲得容易性が極めて高い。ただし、共同注意をスタンス推定の指標として用いるためには、エージェントがバーチャルな画像であっても、身体の一部を有するという前提が必要である。次節以降では、具体的な共同注意の測定方法と、エージェントの身体性について議論する。

2.4 測定指標としての共同注意

共同注意を定量的に検証するための指標の1つとして、視線手がかり効果がある。視線手がかり効果とは、他者の視線方向に対して、観察者の注意が自動的にかつ反射的に誘導される現象である [13]。具体的には、顔画像などが特定の方向へ視線を向けた際、視線の向いていない位置にターゲットが表示される際より、視線の向いている位置にターゲットが表示される際に、ターゲットに対する検出の精度向上、検出までの反応時間が促進される効果を指す [14]。特に視線画像表示から標的刺激表示までの時間が短い場合に顕著である。

視線手がかり効果は、単に視線を物理的に知覚するだけでなく、他者が視線を向けたと思われる動作、すなわち視線を向ける意図を認知すると大きくなる。例えば、サングラスをかけた人物画像であっても視線手がかり効果は生じるが、閉眼している場合や視界が遮蔽されていると認識された場合、効果は消失または減弱することが報告されている [15]。すなわち、視線手がかり効果の大きさは、ユーザが他者に対し視覚的な知覚能力や見るといった意図を帰属させているか否かに依存し変わる。

さらに、視線手がかり効果は、対ロボットのインタラクションにおいても確認されている。Wieseらは、ディスプレイ上の顔刺激を用いた実験において、人間の顔とロボットの顔で視線手がかり効果の生起に差異があるかを調査した [11]。その結果、ロボットの顔であっても効果が生じるものの、人間の顔よりも効果は小さかった。しかし、ロボットの顔に対し「このロボットは人間が遠隔操作している」と教示すると、視線手がかり効果が増大した。逆に、人間の顔に対し「この人間の顔はマネキンである」と教示した際には、効果が減弱した。

これらの結果は、ユーザが相手が意図を持っていると認識していれば視線手がかり効果が生じ、逆に意図を持たないと認識していれば効果が生じにくいことを示す。これを逆に考えれば、視線手がかり効果が、ユーザがその相手に対し意図スタンスあるいは設計スタンスに基づいて接しているかを推定するための指標となりうることを示唆している。

しかし、エージェントの外見をどのように設計するかも考える必要がある。ロボットのデザインには、人間の姿を忠実に模したヒューマノイドや、犬や猫などの動物を模したズーモルフィック、特定の生物そのものではないが有機的な形状や動きによって生命感を感じさせるデザイン、そして要素を極限まで削ぎ落としたミニマルデザインなど、多岐にわたるアプローチが存在する。人間に近い外見を有したヒューマノイドや、可愛らしいキャラクターのような外見を採用した場合、ユーザはその対象を外見情報から擬人化してしまう可能性がある。これでは、ユーザが本来持っていたスタンスが外見の強さによって強制的に変わってしまい、スタンス推定が困難になる。つまりエージェントの外見は、ユーザのスタンスの移行が引き起こされないものでなければいけない。そこで本研究では、あえて情報量を極限まで削ぎ落としたミニマルデザインに注目する。

2.5 ミニマルデザイン

ミニマルデザインとは、基本的な機能のみに絞り込み、非本質的な要素を排除する設計思想を指す [16]。最小限のパーツのみで構成されたミニマルデザインエージェントを用いることで、外見情報による擬人化を抑制し、その存在に対する解釈の余地をユーザ側に委ねることができる。すなわち、エージェントを設計スタンスで認識するか意図スタンスで認識するかは、各ユーザの特性や状況に依存して変わりうる。したがって、ミニマルデザインのエージェントとの共同注意が成立するかを測定することで、ユーザのスタンス推定の指標とすることができると考えられる。

しかし、ミニマルデザインの身体を付与したエージェントの物理的な動作が、そもそも共同注意（特に視線手がかり効果）を引き起こすかは分かっていない。特に、本研究では、エージェントによる外見情報による影響を小さくするために、目や口を持たないデザインを考える必要がある。そのような「視線」の発生源が曖昧な対象に対しても、視線手がかり効果が生じるかは検証されていない。そのため、ミニマルデザインエージェントとの共同注意が成立するかどうか、また、それが意図の帰属と関与しているかについて検討する必要がある。

そこで本研究では、「視線を示す物理的な特徴が欠如していても、対象への意図の帰属によって視線手がかり効果が生起する」という仮説を検証する。エージェントデザインにおいて、エージェントがある機能を持つことをユーザーに認識させるために、身体のパーツを付与するという考え方がある [17]。例えば目は見る機能、口は話す機能を認識させる。この考えに基づく

と、見る機能を持つと認識されたエージェントは、目がなくても視線の情報を伝達できる可能性がある。具体的には、エージェントが物体を視認できる機能を持つことを事前にインタラクションを通じて伝達する群と、機能に関する情報を与えない群を設定する。この2群間で、エージェントの動作による視線手がかり効果を比較する。もしエージェントの見る機能が重要なのであれば、事前インタラクションあり群での視線手がかり効果が事前インタラクションなし群よりも大きくなると予想される。

さらに、これらの効果の違いがスタンスによるものかを確認するために、エージェントに対する擬人化の度合いを比較する。事前インタラクションによってエージェントに意図を帰属したこと（意図スタンスで認識したこと）で視線手がかり効果が大きくなったのであれば、事前インタラクションあり群の擬人化の度合いが大きくなると予想される。以上の点を検討することで、視線手がかり効果がユーザのスタンスを反映する有効な指標となり得るかを議論する。

3 実験

3.1 使用するエージェントのデザイン

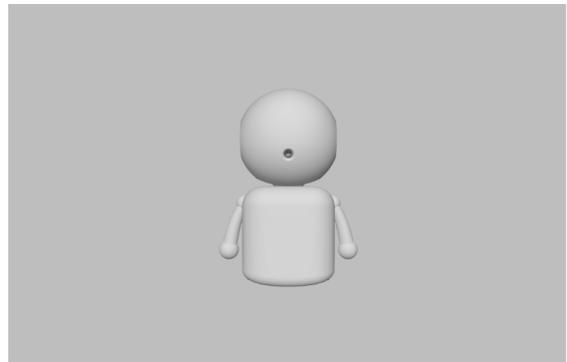


図 2: ミニマルデザインのエージェント

図 2 は本実験にて呈示されるエージェントの外見である。本エージェントの外見デザインは川崎らが開発したロボットの 3D モデルを使用した [18]。本エージェントの特徴は、顔中央に丸が一つのみ存在するミニマルデザインである。一般的な人型シルエットを持ちつつも、頭部に目や口などのパーツを有さないことで、外見情報から過度なエージェンシーを感じさせないようデザインとなっている。

3.2 事前インタラクション

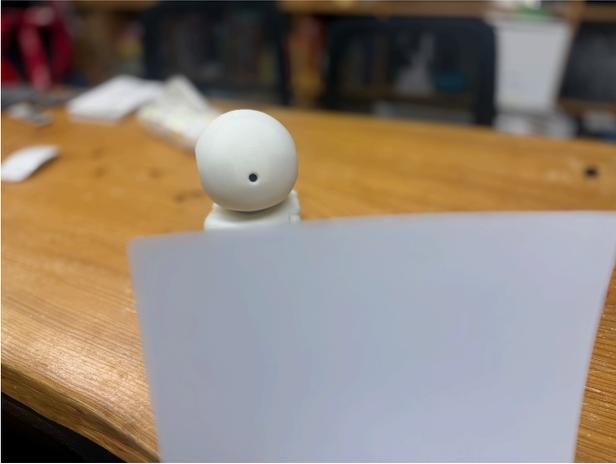


図 3: 事前インタラクション実施時の様子

本実験では、半数の実験参加者に対し、視線手がかり効果の課題の前に、このミニマルデザインエージェントとの短いインタラクションを行わせた。それにより、エージェントが対象に向けて顔を動かしそれを見ることができることを認識させた(図3)。具体的には、実験前に実物のロボットに対し「紙を見せてあげてください」と教示した。実験参加者が実際にエージェントに対し紙を提示した際に、その紙の方向にエージェントが顔を動かす動きをした。残りの半数の実験参加者には、エージェントの実物を見せず、対象を見る機能に関する情報を全く与えなかった。

3.3 実験参加者

本実験は正常な視力または矯正視力を持ち、かつ利き手が右手である40名が参加した(18~54歳, 男性18名, 女性21名, 回答なし1名)。実験参加者はエージェントとの事前インタラクションの有無条件にランダムに振り当てられた(各20名ずつ)。

3.4 実験環境・刺激

本実験は第一著者の所属する大学施設内の比較的静かな部屋を使用し行った。モニター(23.8インチ, 1920×1080, 75Hz)の位置は観察距離約44cmとし、実験参加者間で大きく異ならないように配慮した。事前インタラクション時はエージェントが見えるように電気をつけていたが、実験課題試行中は部屋の電気を消し、モニターの画面に集中できるようにした。

3.5 実験手続き

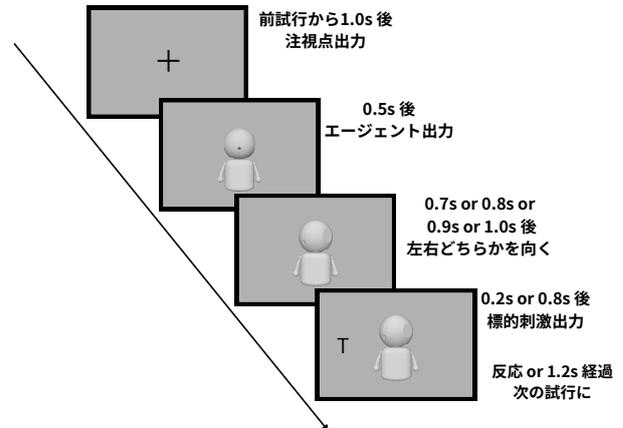


図 4: 1 試行の画面表示の順序

実験は、事前インタラクション、視線手がかり効果の課題、印象評定アンケートの順に行った。インタラクションなし群の参加者は、エージェントとの事前インタラクションを行わなかったが、それ以降はインタラクションあり群と同じ課題を行った。

心理学における典型的な視線手がかり効果の実験を、ミニマルデザインのエージェントの画像を用いて行った。実験タスクの順序を図4で表す。モニター中央に正面を向いたエージェント(423px×641px)が0.7s, 0.8s, 0.9s, 1.0s表示された後、エージェントが左右いずれかの方向を向く動作を行った。0.2s, 0.8s経過後、エージェントの左右どちらかに約13cm離れた位置に標的刺激として1文字のアルファベットT,Lのどちらかを表示した(40px×40px)。実験参加者の課題は、標的刺激がTかLかを識別することであった。エージェントが向く方向と標的刺激の表示位置を独立に操作し、それらが同じである場合を一致条件、異なる場合を不一致条件と呼ぶ。

エージェントが左右いずれかの方向を向いてから、応答または標的刺激表示から1.2sが経過するまで識別の応答を受け付けた。なお、試行間には1.0sのインターバルを設けた。応答はテンキーの2キーを人差し指, 8キーを中指で押すことで受け付けた。キーと標的刺激の対応は実験参加者間でカウンターバランスした。実験は8つのブロック(各32試行)があり、各ブロック間で自由に休憩をすることができた。本実験の前に、各条件1試行ずつの練習(全16試行)を行った。

視線手がかり効果の実験終了後、実験参加者は画面に表示されたミニマルデザインエージェントについての印象評価を行った。具体的には、Godspeed Questionnaire[4]を用いて、エージェントに対する擬人感を評価した。

3.6 分析方法

正答時の反応時間について分析する。0.2s未滿の反応時間、平均+3×標準偏差の値を越えた反応時間を示しているものは分析対象外とした。分析対象となる反応時間について、視線・ターゲットの位置（一致、不一致）、ターゲット呈示までの時間（0.2s, 0.8s）、事前インタラクションの有無の2×2×2の3要因混合計画の分散分析を行う。また、事前インタラクションの有無条件間のGodspeedの擬人感評定値を、対応なしt検定により比較する。

3.7 実験結果

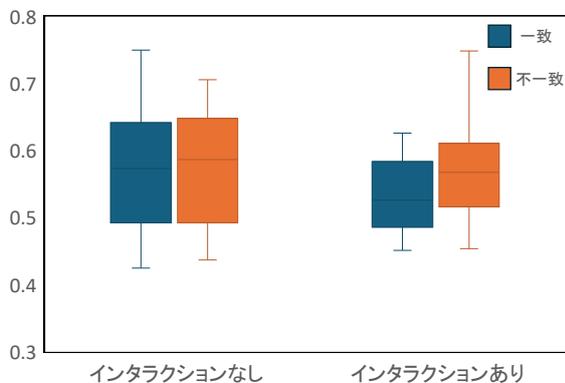


図5: 事前インタラクション有無による正答時反応時間の比較。ひげの上端から下端まで4つの区間に区切られており、それぞれの区間が全体の1/4のデータを取容している。なお、ターゲット呈示までの時間は各条件の中でまとめた。箱中の線は中央値。

全体の正答率は98%と高かった。正答時の反応時間について(図5)、視線・ターゲット位置の主効果($F(1, 38) = 20.201, p < 0.001, \eta_p^2 = 0.347$)、および事前インタラクションの有無との交互作用が認められた($F(1, 38) = 5.740, p = 0.022, \eta_p^2 = 0.130$)。事前インタラクションあり群では、視線とターゲットの呈示位置が一致している時(一致条件)よりも、不一致の時(不一致条件)の方が反応時間が長くなった($p < 0.001$)。一方、事前インタラクションなし群では、一致条件と不一致条件における反応時間に差は見られなかった($p = 0.721$)。その他の主効果と交互作用は有意ではなかった($F_s(1, 38) < 2.051, p_s > 0.160, \eta_p^2_s < 0.052$)。

事前インタラクションの有無条件間でGodspeed評価尺度の擬人感評定値を比較した。その結果、事前インタラクションあり群のほうがなし群よりもエージェントに対する擬人感が有意に高くなった($t(38) = 2.49, p = 0.017$)。

4 考察

4.1 スタンスによる視線手がかり効果の変容

実験の結果、物理的な目や口などの顔パーツを持たないミニマルデザインのエージェントであっても、そのエージェントが対象に対して顔を動かし見ることができることを認識させた場合にのみ、視線手がかり効果が生起することが明らかになった。また、事前インタラクションを実施した群における擬人感の評定値が高かった。本実験課題では、全く同一の物理的刺激(回旋動作)を用いているため、本実験結果は物理的刺激の特性のみでは説明できない。これらのことから、本ミニマルデザインのエージェントは、事前インタラクションなしではエージェントの低い物理刺激と認識され、エージェントの回旋動作そのものだけでは、注意を誘導するには不十分だったと考えられる。一方で、ユーザは事前インタラクションによって、エージェントの回旋動作を単なる機械的な運動としてではなく、何らかの目的や意思を持つ振る舞いだと解釈したのではないかと考えられる。

すなわち、本研究の結果は、物理特性以外の、意図の帰属に基づく視線手がかり効果の変容を示唆している。視線手がかり効果は共同注意と強く関連した現象であることを考えると、ユーザが対象を意図スタンスに基づいて解釈した場合は共同注意が機能し、逆に設計スタンスに基づいて解釈した場合は共同注意が機能しない。よって、ミニマルデザインのエージェントとの共同注意は、意図スタンス・設計スタンスと関連したものだと言える。

4.2 測定システムの実装提案

ミニマルデザインのエージェントとの共同注意の生起は、ユーザの持っているスタンスそのものに大きく依存すると考えられる。このことから、共同注意の発生状況はユーザのスタンスを測るための有効な客観的指標となり得る。つまり、共同注意を観測することで、ユーザがエージェントに対して「意図スタンス」か「設計スタンス」のどちらのスタンスを取っているかを推定できる可能性がある。

では、この指標を生成AIとのチャット型の対話インタラクションに実装するには、どのような要件が求められるだろうか。本提案では、主に計測の暗黙性と操作の非要求性という2つの観点が必要不可欠であると考えられる。第1に、ユーザに計測されていると意識させない暗黙性が必要である。もし観察を自覚されれば、ユーザは本来感じていないエージェントを感じたように意図的に振る舞うかもしれない。第2に、共同注意の検知のみを目的としたユーザの主体的な操作を排除する

必要もある。本研究を含む従来の実験手法ではキー押し等の反応を求めていたが、これは測定のためにユーザが能動的に行う行動であり、自然な対話には本来不要である。そのためシステム側がその状態を自動的に検知できる手法が求められる。

これらの要件を満たすために、カメラを用いてユーザの視線方向を検知することを提案する。共同注意は、相手の視線に追従した視線移動によっても測定することができる [19]。具体的には、生成 AI とのチャット画面に本実験のようなミニマルデザインエージェントを配置する。そのエージェントは、任意のタイミングで視線をテキストボックスや参照ボタンなどへと向ける。ユーザへの回答生成時や重要情報の提示時などインタラクションにおいて意味のある動きだけではなく、インタラクションとは無関係な動きもする。そして、カメラを用いてユーザの視線をリアルタイムに取得し、エージェントの視線移動に追従してユーザの視線も移動したか（共同注意が成立したか）を判定する。この手法であれば、ユーザは追加の主體的な操作（クリックやキー押し）を行う必要がなく、単に画面を見ているだけで測定が完了する。そのためユーザに違和感を与えることなくスタンスの推定が可能となる。

4.3 スタンス推定に基づくインタラクションの最適化とその展開

前節で提案した視線情報の計測を生成 AI のインタフェースに実装することで、ユーザのスタンスが判別可能となる。つまり、エージェントの視線に追従しやすい場合には意図スタンス、追従しない場合には設計スタンスだと判定する。この判別結果に基づき、ユーザのスタンスに応じたコミュニケーション戦略を選択することで、インタラクションの最適化につながる。

ユーザがエージェントを設計スタンスで認識している場合、エージェントは過剰な挨拶や共感的な前置きなどを抑制し、機能性を最優先する。検索結果の提示やタスク処理を迅速かつ端的に行うことで、ユーザは不要なコミュニケーションコストを支払うことなく、効率的にツールを利用できる。

一方で、ユーザがエージェントを意図スタンスで認識している場合は、エージェントは共感的な応答や文脈の深掘り、あるいはユーザを肯定するといった社会的な振る舞いを強化する。ユーザが社会的な振る舞いを求めている際に、情緒的サポートを提供することで、ユーザの期待を満たし、対話を維持することが可能となる。

さらに、インタラクションを通じて引き起こされるスタンスの時間的変化も追跡できるかもしれない。人間関係と同様、ユーザのエージェントに対する認識は一定

ではない。最初は設計スタンスで接していたユーザが、対話を重ねる中で親しみを覚え意図スタンスへと変化する場合もあれば、その逆もまた然りである。ユーザの視線情報をリアルタイムで、かつ継続的に観測することで、いつ、どのタイミングでユーザが心を開いたか（あるいは心を閉ざしたか）というスタンスの変容を検知できる。これにより、ユーザとの信頼関係の深まりに合わせて徐々に対話深度を変えていくような、長期的かつ動的なインタラクション設計ができるだろう。

この考えを発展させれば、このインタラクションデザインは、ユーザのエージェントに対するスタンスを誘導することもできるかもしれない。エージェントの視線移動は、任意のタイミングで生じるが、システムでの情報提示時など、インタラクションの文脈において自然なタイミングで起こすこともできる。そのため、見てほしい情報に視線を移動させるという振る舞いを高頻度で行うことで、ユーザが視線にエージェントの意図を見出し、その視線に追従しやすくなるように誘導できるかもしれない。これによって、意図スタンスでシステムを認識してほしい状況があった場合に、設計スタンスで認識しているユーザのスタンス移行を達成できるかもしれない。

逆に、強すぎる意図スタンスでの認識は、エージェントとの危険な関係につながる可能性もある。近年、AI に対する過度な擬人化や没入は、政治的誘導や詐欺への脆弱性、プライバシーの無防備な開示、あるいは批判的思考の停止によるエコーチェンバー現象といった深刻なリスクが指摘されている [20, 21]。また、面倒な人間関係を避けて AI にのみ安らぎを求める現実逃避も懸念される [22]。もちろん、AI が心の逃げ場として機能すること自体は精神衛生上の有用性があるが、その依存度が健全な境界線を超えそうになった場合は介入を行うべきである。本研究で提案する指標は、過度な依存状態を検知するために有効である。そして、ミニマルデザインエージェントが見た目による意図スタンスへの誘導をしづらいという想定を考えると、あえて設計スタンスとして捉えられるような振る舞いを強調して提示することで（例えば、視線移動をしなくなる）、適度な距離感へと引き戻すこともできるかもしれない。これは、AI と人間が共生する社会における倫理的な安全装置として機能することが期待できる。

本研究が提案するインタラクションシステムは、生成 AI だけに限定されるものではない。本手法は物理的身体を持つソーシャルロボットや画面上のアバターやキャラクターなど、あらゆるエージェントにおいて、共通して適用可能である。つまり、本提案システムは、人間とエージェントのインタラクション全般の質を向上することに寄与できる。

5 おわりに

本研究では、可能な限り見た目の要素をなくしたミニマルデザインのエージェントが、視線による共同注意を誘発するかを検討した。その結果、目に相当するパーツがない場合でも、見る機能を持つとユーザに認識させることで、視線のような情報による共同注意が成立することが明らかになった。また、その際、エージェントに対する擬人化の度合いが高くなることも明らかになった。これらのことから、ユーザがミニマルデザインエージェントを擬人化している、つまり意図スタンスで認識していると、共同注意が成立しやすいと考えられる。

これに基づくと、共同注意を測定することでユーザのエージェントに対する認識を推定できる可能性がある。つまり、共同注意が生じている場合には意図スタンスだと判定し、共同注意が生じていない場合には設計スタンスだと判定できる。この判定結果に基づき、ユーザの認識に合わせたエージェントの振る舞いを変えることができれば、より効果的なユーザとエージェントのインタラクションを実現できるだろう。

さらに、共同注意が生じやすいような振る舞いや介入をエージェントが行うことで、ユーザを意図スタンスでの認識へと誘導できるかもしれない。また、介入を工夫することで、過度にエージェントを擬人化してしまうユーザを適切なスタンスへ誘導することもできるかもしれない。つまり、ユーザのエージェントに対する期待を管理して、過度な意図スタンスでの認識や、設計スタンスへの固着（エージェントを単なる道具として扱いつづけること）を抑えることが可能になるだろう。これは適切なスタンスの維持につながる。よって、飽きや失望によってインタラクションが断絶されることなく、継続的かつ豊かな関係性を構築できるようなエージェントが実現できるだろう。本研究の知見はそのための第一歩であるが、これを発展させることで、ユーザから適度に意図を読まれるエージェントの実現に繋がることが期待される。

参考文献

- [1] Daniel C. Dennett, “The Intentional Stance”, MIT Press, 1989.
- [2] Anca D. Dragan, Kenton C.T. Lee, and Sidhartha S. Srinivasa, “Legibility and predictability of robot motion”, Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI '13), pp. 301–308, 2013.
- [3] 岡田 美智男, 『弱いロボット』, 医学書院, 2012.
- [4] Christoph Bartneck, “Godspeed Questionnaire Series: Translations and Usage”, International Handbook of Behavioral Health Assessment, pp. 1-35, 2023.
- [5] 寺田 和憲, 社本 高史, 梅海 櫻, 伊藤 昭, “意図的人工物”, 第 21 回人工知能学会全国大会論文集, 2D5-7, pp. 1–4, 2007.
- [6] Jennifer Goetz, Sara Kiesler, Aaron Powers, “Matching robot appearance and behavior to tasks to improve human-robot cooperation”, Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), pp. 55-60, 2003.
- [7] Takanori Komatsu, Ryo Kurosawa, Seiji Yamada, “How does the difference between users’ expectations and a robot’s actual processing ability influence users’ impression of the robot?”, International Journal of Social Robotics, Vol.4 No.2, pp. 193-202, 2012.
- [8] Catherine Lai, “Prosodic cues for backchannels and short questions: really?”, Proceedings of the 4th International Conference on Speech Prosody, pp. 413-416, 2008.
- [9] Michelle Cohn, Bruno Ferenc Segedin, and Georgia Zellou, “Acoustic-phonetic properties of Siri and human-directed speech”, Journal of Phonetics, Vol. 90, Article No. 101123, 2022.
- [10] Edward T. Hall, *The Hidden Dimension*, Doubleday, 1966.
- [11] Eva Wiese, Agnieszka Wykowska, Jan Zwickel, Hermann J. Müller, “I see what you mean: How attentional selection is shaped by ascribing intentions to others”, PLOS ONE, Vol.7 No.9, pp. e45391, 2012.
- [12] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, Henrike Moll, “Understanding and sharing intentions: The origins of cultural cognition”, Behavioral and Brain Sciences, Vol.28 No.5, pp. 675-691, 2005.
- [13] Chris K. Friesen, Alan Kingstone, “The eyes have it! Reflexive orienting is triggered by nonpredictive gaze”, Psychonomic Bulletin & Review, Vol.5 No.3, pp. 490-495, 1998.

- [14] Mario Dalmaso, Luigi Castelli, Giovanni Galvano, “Social modulators of gaze-mediated orienting of attention: A review”, *Psychonomic Bulletin & Review*, Vol.27 No.5, pp. 833-855, 2020.
- [15] Pines Nuku, Harold Bekkering, “Joint attention: Inferring what others perceive (and don’t perceive)”, *Consciousness and Cognition*, Vol.17 No.1, pp. 339-349, 2008.
- [16] Nobuyoshi Matsumoto, Hiroyuki Fujii, Michio Okada, “Minimal design for human-agent communication”, *Artificial Life and Robotics*, Vol.10 No.1, pp. 49-54, 2006.
- [17] Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović, “Human-Robot Interaction: An Introduction”, Cambridge University Press, 2024.
- [18] 川崎 邦将, 大澤 正彦, 今井 倫太, 長田 茂美, “認知的制約付き擬人化キャラクターに着目したコミュニケーションロボットの設計と開発”, 人工知能学会 第 6 回汎用人工知能研究会 (SIG-AGI), SIG-AGI-006-02, 2017.
- [19] Eileen M. Mansfield, Teresa Farroni, and Mark H. Johnson, “Does gaze perception facilitate overt orienting?”, *Visual Cognition*, Vol. 10, No. 1, pp. 7–14, 2003.
- [20] Niloofar Mireshghallah, Maria Antoniak, Yash More, Yejin Choi, and Golnoosh Farnadi, “Trust No Bot: Discovering Personal Disclosures in Human-LLM Conversations in the Wild”, arXiv:2407.11438, 2024.
- [21] Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, and Mor Naaman, “Co-Writing with Opinionated Language Models Affects Users’ Views”, *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23)*, Article No. 111, pp. 1–15, 2023.
- [22] Tianling Xie and Iryna Pentina, “Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika”, *Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS ’22)*, pp. 2046–2055, 2022.