

一人称視点のターンテイキングダイナミクスが 人／計算機の識別に与える影響

A study on the effect of first-person turn-taking dynamics on human / computer identification

金野 武司[†], 竹田 亮大[‡]

Takeshi Konno, Ryodai Takeda

[†] 金沢工業大学 工学部 電気電子工学科, [‡] 金沢工業大学 工学部 電子情報通信工学科

Kanazawa Institute of Technology, College of Engineering,

[†] Electrical and Electronic Engineering,

[‡] Electronics, Information and Communication Engineering

konno-tks@neptune.kanazawa-it.ac.jp, b1700499@planet.kanazawa-it.ac.jp

概要

人と機械のインタラクションで実現が困難なことの1つに、ターンテイキングと呼ばれる、対話やジェスチャーにおける主従関係の自然な入れ替わりがある。先行研究では、人どうしのダイナミクスに倣って動作させた計算モデルが、相手の動きに応じて主従関係を切り替えるようになっていなかったにもかかわらず、相手が人間であるのか計算機であるのかが区別できなくなったことが報告されている。我々は、相手とインタラクションする視点が三人称から一人称が変わることで、その計算モデルに違和感を持つようになるのではないかと予想した。しかし、実験の結果、視点の違いはその違和感に気づかせることはなく、それとは別に、計算機が持つ主従関係に合わせやすくする効果があることを示唆する結果を得た。

キーワード: Turn-taking, Nonverbal communication, Human agent interaction

1. はじめに

近年、ソフトバンク社のペッパーやユニロボット社のユニボなど、対話だけでなく、身振り手振りを交えて人とインタラクションするロボットが身近なものになってきている。しかしそういったロボットは、急に動き出したり不意に話し出したりするなど、そのインタラクションに違和感を覚えることが多い。

この違和感の原因はターンテイキングダイナミクスにあると考えられる。人間どうしの場合にそのダイナミクスは、対話における話し手と聞き手の入れ替わりであったり、ジェスチャーのような視覚的なインタラクションでのリーダー／フォロワーの入れ替わりとして自然に表れる。しかし、その入れ替わりのパターンを生成する仕組みは未だよく分かっていない。

入江ら [1, 2] はこの問題に対して、特に視覚的なインタラクションでのターンテイキングダイナミクスに着目し、人どうしと同様のリーダー／フォロワーの切り替えタイミングを再現する計算モデル（入江らはこれをターンテイキングモデルと呼んだ）を構築した。その計算モデルと人間とのインタラクション実験を行った結果、実験の参加者は、その計算モデルを人間と区別することができないことが示された。

しかし、我々はその計算モデルが、人が行う自然なインタラクションを再現しているとは思わない。なぜならその計算モデルはリーダー／フォロワーの入れ替わるタイミングがあらかじめ決められていたからである¹。これを対話に置き換えれば、一方が話しているときに他方が話し出そうとしても、計算モデルはそれを一切無視することに他ならない。つまり、入江らの計算モデルは、相手の動きに応じてリーダー／フォロワーを切り替えるようになっていないのである。この意味で、我々は入江らの計算モデルを真のターンテイキングモデルと区別するため、Turn pre-Schedule Model（以降 TpS モデル）と呼ぶ。

入江らの実験において、人はなぜ計算モデルとの区別がつかなかったのだろうか？ その原因として我々は、視点の違いに着目する。入江らの実験では、モニターに映る自分と相手の球が左右に動き、三人称視点で他者とインタラクションする。この場合、球を自分が操作しているという自己主体感はあるとしても、自分の身体の一部であるという自己所有感は得られないことが知られている [3]。

ここで我々は、視点を一人称にすることで自己所有

¹人どうしのパターンから抽出された確率分布（一般化パレート分布）に従って、リーダー／フォロワーそれぞれの時間を事前に生成する。

感を球の操作者に持たせることができ、それがインタラクションパターンへの感性を上昇させることで、TpS モデルが内包する問題に参加者が気付くようになるのではないかと考えた。そこで本研究では、入江らの実験環境を三人称視点から一人称視点に変更し、TpS モデルが人間と区別されるかどうか検証することを目的とした。

2. 実験方法

2.1 実験室実験

実験では、二人の参加者が向かい合うように設置されたコンピューター端末の前に座った。互いの姿が見えないよう、両者の間には衝立が設置された。両者のモニターの画面上には中央に相手が操作する球オブジェクトが1つ配置された。参加者は球を手元のテンキーの2つのボタンで前後にのみ移動できた。ただし、一人称視点での操作のため、自身が操作する球を見ることはできなかった。相手が操作する球は緑色で表示され、相手の動作は直ちに反映されるようになっていた。また、球は互いにすり抜けることができないようにした。これは、入江らの実験環境と同様、視覚的動作特徴として「追う/追われる」の関係としてターンテイキングが表出しやすくなることを狙いとして設定した。入江らの実験環境からの変更点として、球が移動していることを明確にするために、球には影をつけ、下部にレールが設置された。また、球の位置を把握するために、レールの中央には赤い線、末端には水平線が引かれた。参加者は、相手のキー操作音が聞こえないように、ノイズキャンセリングヘッドホン (Sony 製 WH-1000XM3) を装着した。

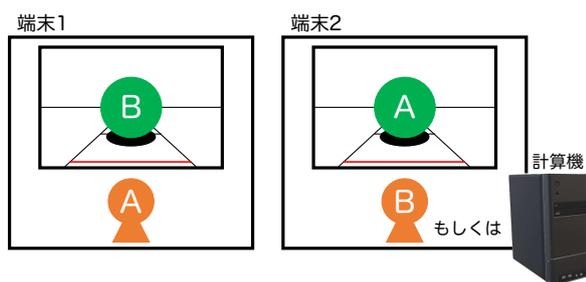


図1 実験環境。実際の画面上の球には、A,B の文字は描かれない。

2.2 実験手続き

参加者は、コンピューター端末越しに簡単なゲームに取り組むことが説明され、最初に操作に慣れるため

の練習を行なった。参加者は90秒を1セッションとして、計15セッションに取り組んだ。このうち、最初の5セッションは自由に動かすように指示され、人どうしで行われた。これを人どうし課題と呼ぶ。この課題では人どうしの自然な円の移動パターンを観察することを狙いとした。残りの10セッションでは相手が人と計算機で5回ずつランダムに入れ替わった。これを人計算機判別課題と呼ぶ。この課題で参加者は、相手が計算機になる場合には二人ともが計算機を相手とした。また参加者は、1回のセッションが終わるたびに、相手が人であったか、計算機であったかと同時にそのように判断した根拠となる動作特徴を自由記述でアンケート用紙に記述した。1セッションごとの正解についてはフィードバックしなかった。初めの5セッションが終わった後で、参加者には人であるか計算機であるかがランダムに入れ替わることが伝えられた。同時に参加者には、相手が人であるのか計算機であるかを判断すると共に、自分が人であることを相手に伝えようとするように指示された。計算機の入替わりは10セッションのうち5回でランダムに行われたが、その入れ替わりの頻度は参加者には伝えられなかった。

2.3 計算モデル

計算モデルには、入江らの研究 [1, 2] で使われた TpS モデルを使用した。TpS モデルは、予め採取した人どうしの軌道データから、先行研究と同様の方法でリーダー期間・フォロワー期間を推定し²、その頻度分布から構成した一般化パレート分布による近似確率分布を用いて確率的にリーダー時間とフォロワー時間を決定した。リーダー期間では、過去の人間が動かした軌道を再現し、フォロワー期間には、相手の動きに1.2秒の遅れを持って追従し続けるようにした³。再現する軌道は、TpS モデル作成のために実施した5ペア10名 (男性9名、女性1名、平均21.5歳、 $SD = 0.500$) の人どうし課題の結果から実験ごとに1名の軌道をランダムに選択し、該当するセッション番号および期間のデータを使用した。構成された確率分布は、三人称視点で実施された先行研究とほぼ同様の分布になることを確認した。

²時間窓 (3 秒) を設けて取り出した2つの軌道を時間的にずらしながら相関係数を算出し、そのずらした方向と相関係数の関係からリーダー・フォロワーを推定した。

³遅れ時間、およびリーダー/フォロワーを推定する際の時間窓は先行研究の設定に揃えた。

2.4 実験設備

モニターは iiyama 製 ProLiteXUB2790HS を用い、画面サイズは 597.9×336.3 [mm] (1920×1080 [pixel]) だった。実験システムは Unity で構築した。球の前後への移動範囲は 480 [pixel]⁴ とした。また、サンプリング時間は 30 [ms] とし、1 サンプリング時間あたりの球の移動距離を 6 [pixel] としたため、人が操作する球の移動は 200 [pixel/sec] の等速度運動であった。

2.5 参加者

我々は上述の実験環境と手続きの中で、TpS モデルを用いた実験を実施した。実験には、7 ペア 14 名が参加した。その内訳は男子学生 12 名、女子学生 2 名で年齢は 19 から 22 歳 (平均 21.4 歳, $SD = 0.895$) であった。参加者はいずれも金沢工業大学の学生であり、実験は全て同大学の研究室で行なわれた。

3. 結果

人は一人称視点におけるインタラクション環境において、TpS モデルを判別することができたのだろうか。これを確認するために、人計算機判別課題の 10 セッションで、相手が人であるか計算機であるかを答えたアンケートの正答率を一人称視点と三人称視点で比較した (図 2. 三人称視点での結果は入江ら [1] の実験データを使用した)。

一人称視点において、相手が人だった場合と計算機だった場合の正答率の平均はそれぞれ 0.486 , 0.571 であり、ほぼチャンスレベルであった。また、三人称視点での正答率の結果と統計的に有意な差は見られなかった。この結果は、一人称視点においても、TpS モデルに対して相手が人か計算機かを識別できなかったことを示す。

では、識別できなかった TpS モデルには、軌道のダイナミクスにも違いがなかったのだろうか。これを確認するために、先行研究と同様に 2 つの動作軌道の相関係数及びリーダー比均等率による分析を行い、三人称視点における特性と比較した。ここで、相関係数は軌道の同調度を調べるためのものである。また、リーダー比均等率は先述したリーダー/フォロワー期間の推定から、互いがリーダーを行っていた時間の均等度合いを計算したものである。この比率が 1 に近づけば、二人はリーダーを同程度の時間担っていたと推定される。結果を図 3 及び図 4 に示す。

⁴仮想空間内の奥行き方向への移動量であり、二次元平面に射影されたレールの移動量とは異なる。

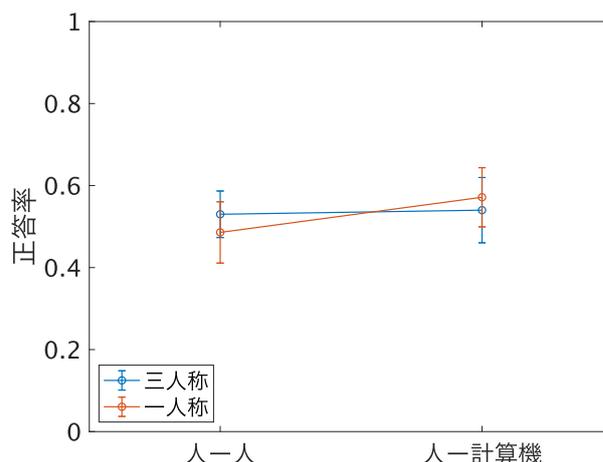


図 2 人計算機判別課題での正答率

それぞれのグラフを確認すると、三人称視点の場合には、相手が人の場合と計算機の場合で、相関係数およびリーダー比均等率に違いがあったが、その違いが一人称視点ではなくなってしまっていることがわかる。この結果は、一人称視点になることで、ダイナミクスの特徴としても人と区別できなくなっていたことがわかる。

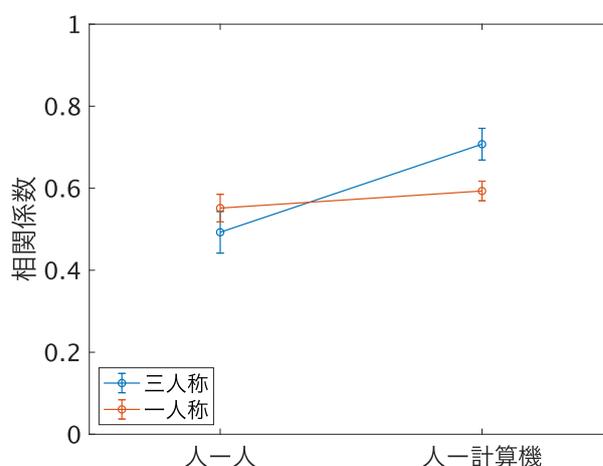


図 3 相関係数による 2 つの軌道の同調度

4. 議論

三人称と一人称の視点の違いに対して、相手が人であるか計算機であるかの正答率についてはどちらもチャンスレベルに留まった。これは、参加者は一人称視点であれば TpS モデルが持つ不備 (リーダー・フォロワーの切り替えが相手の動きに応じて行われない不自然さ) に気づくのではないかとした我々の想定を支持しないことを示す。

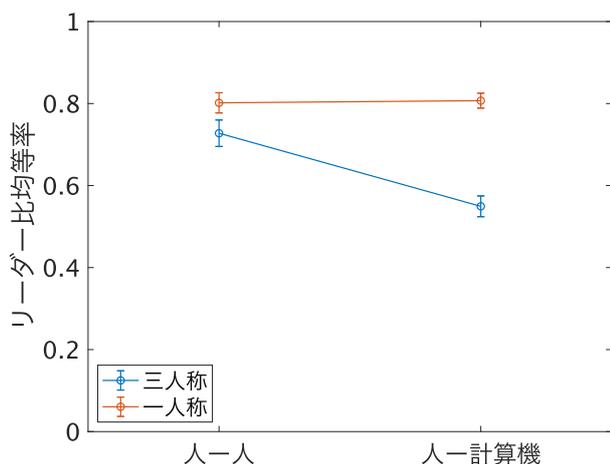


図4 リーダー比均等率

また興味深いのは、相関係数とリーダー比均等率から見る一人称視点のインタラクションのダイナミクスは、三人称視点のときとは異なり、人を相手にした場合のダイナミクスと同じ数値を示し、区別がなくなったことである。TpSモデルが、あらかじめ用意された確率分布に従ってリーダー・フォロワーの時間を生成することを考えれば、一人称視点でも三人称視点でも、計算機の動きの特徴は変わることがない。この事実から推定されるのは、一人称視点である方が、人は相手に合わせるが多かったのではないだろうかということである。TpSモデルは、例えば人がフォロワーであるところからリーダーになろうとしても、その動きに応じてフォロワーに変わることがない。このとき一人称視点での参加者は、すぐに相手(TpSモデル)に合わせてフォロワーに戻ったのではないだろうか。リーダーからフォロワーになる場合も同様で、参加者は結果的に計算機の切り替えに合わせる傾向が強くなったのではないかと考えられる。

このように考えると、我々が想定した、視点の切り替えによって得られる自己所有感の向上が、TpSモデルに潜む違和感に気づかせるという考えは成立しそうにない。なぜなら、本研究の結果は、インタラクション場面における自己所有感の向上は、相手が持つ役割への同調圧力の上昇となって現れることを示唆していると考えられるからである。これが違和感として現れるとすれば、それは人か計算機かの違いとしてではなく、相手に対するやりにくさや頑固さといった印象としてなのかもしれない。

適度なリーダーっぽさとしての自律性(相手への従わなさ)と、フォロワーっぽさとしての相手の動作への追従が示されるとき、人間はその動作ダイナミクスだけからは相手が人間であるのか、それとも機械であ

るのかを判別できないのかもしれない。しかし、人間のターンテイキングの自然さは、逐次的なインタラクションの積み上げから立ち現れる上位のルール(本研究での課題ではリーダー/フォロワーという役割)を暗黙的に共有しながら、それを時間と共に明示的なルールなしに変遷させていける点にある[4]。この現象は、明確な目的が設定されない雑談や、子どもたちの遊びの変遷に顕著に現れる。上位に形成されるルールを実験課題として明示的に与えることは簡単であるが、それは実験者が計算機にルールを埋め込むことと何ら変わりがなく、人工知能が抱える問題の解決には寄与しない。本研究では、上位に形成される自律的なルールとその変遷に対して、違和感を明確に感じるような実験課題を構築することが今後の課題である。

5. 結論

我々は、視覚的な二者間インタラクションにおいて、三人称と一人称の視点の違いが、インタラクションに潜む違和感への感度を上げることを予想した。特に我々が注目したのは、ターンテイキングにおいてリーダーとフォロワーを「相手に応じて」切り替える仕組みの不在がもたらす違和感に気づくかどうかであった。

しかし、我々が実施した実験において、参加者はその違和感に気づくことはなかった。また、三人称視点と比べて、一人称視点でのダイナミクスは、同調度やリーダー/フォロワーの時間頻度において、より人間を相手にした場合のダイナミクスと区別がつかないものになった。この結果から我々は、インタラクションにおける一人称視点は、相手の動きに対する違和感への気づきを与えるのではなく、相手の役割に合わせる傾向を強める働きがあるのではないかと考える。

文献

- [1] 入江諒, 金野武司 『円図形を介した視覚的同調動作における主従関係の相関関数による分析』, 日本認知科学会第35回大会発表論文集, pp.354-360, 2018.
- [2] R. Irie and T. Konno. Effects of turn-taking dynamics without contingency: A visual interaction experiment. In Proceedings of the 7th International Conference on Human-Agent Interaction (HAI2019), pp.239-240, 2019.
- [3] 瀧見 彰太, 坂本 大介, 小野 哲雄 『身体像の投射を用いた「自己所有感」と「自己主体感」のゆらぎ』, 2017年度日本認知科学会第34回大会, OS03-2, pp.139-147, 2017.
- [4] 飯塚博幸, 安藤英由樹, 前田太郎. 身体的相互作用におけるコミュニケーションとターンテイキングの創発. 電子情報通信学会論文誌 A, 95(1), pp.165-174, 2012.