

意図的な共同注視行動を獲得する人工システムの構築とその検証方法

金野 武司[†] 橋本 敬[†]

Construction and Verification Method of Acquisition System of Intentional Joint Visual Attention

TAKESHI KONNO[†] and TAKASHI HASHIMOTO[†]

1. はじめに

乳幼児は言葉を話し始めるより前に様々なコミュニケーション行動を獲得する。中でも社会的なコミュニケーション行動の萌芽として注目される行動が共同注視 (Joint Visual Attention) である。共同注視は、単純に親の視線の向く方向に自らも視線を向ける行動と定義されている¹⁾。しかし、この行動の持つ社会性に注目するトマセロは、共同注視には乳幼児が他者の見ようとするものを理解し、さらには他者と注意を共有するようになる過程があることを指摘している²⁾。

こういった内面の変化によって共同注視という行動が質的に変化する過程を、乳幼児の行動観察において客観的な証拠と共に提示することは難しい。脳活動計測のような手法においても、測定することの技術的な難しさと同時に、測定されたデータから乳幼児の見ようとしている状態を知ることには根本的な難しさがある³⁾。

こういった難しさの解決を試みるアプローチの1つに認知発達ロボティクスがある⁴⁾。これはロボットに乳幼児の発達過程を再現することで、ロボットの行動とその内部メカニズムを同時に観測することを可能にする手法である。長井ら⁵⁾は、このアプローチに基づいた共同注視の獲得実験を行なっている。この研究では、ロボットはヒトとの視覚的インタラクションを通じてヒトの視線方向を見る行動を学習する。ところが、その共同注視はあくまでも反射的な行動に留まるものであり、意図性のような内面の成長がロボットの中で

起こるようにはなっていない。

これに対して我々は、反射的な行動と意図的な行動を分けるメカニズムとして、反射的な行動を手段にするような注視目標 (目的) を内部に形成する計算モデルを構築している⁶⁾。この試みは、内部観測を可能にする手法であるからこそ、意図性を獲得する過程に潜むメカニズムを解明できるという考えに基づくものである。また、意図性の形成をメカニズムに内包する人工物を構築することは、ヒト-ロボット間の心的なコミュニケーションの実現につながるものであることも重要である。

我々が構築するモデルは、他者の内面を理解するようになるための前段階として、まず自らが意図的な内部状態を持つようになる過程を扱う (以降、これを意図的主体性の獲得と呼ぶ)。これは、トマセロが指摘する共同注視の発達過程を支持するものである⁷⁾。トマセロは、まず自らが意図的な主体になる過程が先行し、それを基に他者も自分と同じ意図的な主体であることを理解するようになることで、他者の意図を理解・共有するようになることを指摘している。

本研究では、ロボットに我々の計算モデルを実装して、反射的な共同注視の獲得モデルと比較するための実験環境を構築する。実験の目的は、ロボットに意図的主体性を形成する計算モデルを実装した場合と、反射的な共同注視を実装した場合で、ヒトが前者のロボットに対して有意に意図性を感じるかどうかを調査することにある。もし前者のロボットに有意に意図性を感じるのであれば、計算モデルに含まれるメカニズムの違いにその原因を求めることができるからである。

本稿では、ヒトとロボットのインタラクションによってロボットが内部に形成する意図性を検証するための

[†] 北陸先端科学技術大学院大学 知識科学研究科
School of Knowledge Science,
Japan Advanced Institute of Science and Technology

実験環境を提示することを目的とする。本稿の構成は次のとおりである。2節では、共同注視の発達過程を概観し、反射的な行動の計算モデルと意図的主体性を形成する計算モデルの違いを説明する。3節では本研究で用いるロボットの仕様と、そこに実装する基礎的な認識能力（顔やオブジェクトの認識能力）を説明すると共に、ヒトとのインタラクシオンを行なうための実験環境を説明する。最後に、4節でその実験環境においてロボットがどのように振る舞うのかを説明する。

2. 計算モデル

共同注視は、バターワースら¹⁾の実験によって8～18ヶ月程度の期間に発達する行動であることが知られている。ハードら⁸⁾は、この共同注視が5～6カ月程度の乳児で観察されることを報告しているが、この段階の共同注視は親の見ていた方向を意図的に見るのではなく、視覚刺激が持つオプティカルフローのような特徴に反射的に反応する運動だと考えられる。なぜならハードらの実験ではヒトの顔をモニタに表示し、その顔を消した後で周辺刺激を提示するという特殊な状況下において共同注視が観察されているからである[☆]。一方コークムらは、より自然な環境で観察される共同注視の獲得は8ヶ月程度から始まることを指摘すると共に、その初期段階では親の視線の先にあるオブジェクトを見るような体験を通じた条件付け的な学習が関与することを示唆している⁹⁾。

長井らはこういった知見を基に、視界に映るものを反射的に注視する視覚定位行動（視界に映るものを注視する行動）によって親の次にオブジェクトを見る体験を得て、これを条件付け的に学習することで共同注視を獲得するモデルを構築している¹⁰⁾。しかし、そのモデルのメカニズムは感覚と運動を直接結び付ける形になっており、結果として反射的な共同注視を学習するモデルになっている。このモデルのアーキテクチャを図式化したのが図1（上）である。

これに対して我々が構築する計算モデルでは、反射的な行動として既に機能している運動モジュール（視覚定位）に対して、これを手段とするような目的形成モジュールを前置接続する（図1下）。このアーキテクチャにおいて感覚-運動システムは、視覚定位の機能を手段として用いる内部状態の形成によって共同注視を獲得することになる。

視覚定位を適切に行使することのできる内部状態を

☆ ヒトの顔が正面に提示されていると顔から注意が移行しないことが確認されている。

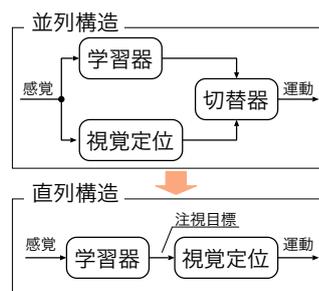


図1 反射的な行動を獲得するモデル（上）と意図的主体性を形成するモデル（下）

Fig.1 Upper architecture acquires a reflexive behavior, and lower architecture forms the intentional agency.

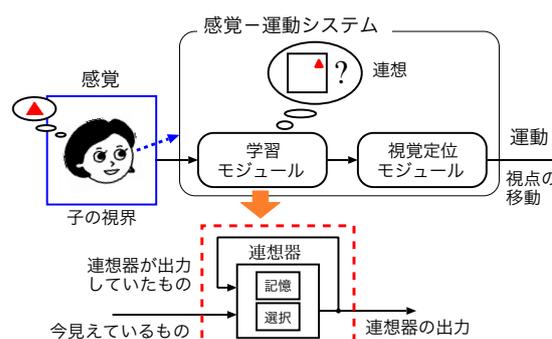


図2 親とのインタラクシオンを通じて内部に注視目標を形成する計算モデルのシステム図

Fig.2 System block diagram of forming visual targets.

形成するためには大きく分けて2つのユニットが必要である（図2の連想器内に示す記憶と選択のユニット）。1つは親とのインタラクシオンを通じて得る体験を記憶するユニットである。このユニットが親の視線方向に見えたオブジェクトを想起することによって、親の次にオブジェクトを見るという行動が実現される。もう1つは、今見えているものや連想したものを特定の優先順位で選択するユニットである⁶⁾。このユニットの詳細は紙面の都合上割愛するが、こうして構築するシステムは、注視目標として機能する内部状態を持つようになる。これによってシステムは、親の視線方向に複数のオブジェクトがある場合にも、どれを見るのかを主体的に決めることができるようになる。また、期待するオブジェクトが視線方向にないことを知ることができるようになる。

計算モデルをロボットに実装する際には、動作を比較するために3つのモデルを用意する。

- (1) 視覚定位のみのモデル（VOモデル）：
視界の最も端に映るもの（ヒトの顔もしくはオブジェクト）を注視する。
- (2) 反射的な共同注視モデル（RJVAモデル）：
ヒトの視線方向に向かって視界をある一定の距

離移動させた後、視点（視界の中央）に最も近いオブジェクトを注視する。

- (3) 意図的主体性を形成する共同注視モデル (IJVAモデル)：ヒトの視線からオブジェクトを連想し、その連想に基づいて視界を移動させる。連想したオブジェクトが見えたらそれを注視する。

これらの計算モデルをロボットに実装したとき、共同注視のモデルだけではロボットはヒトとのインタラクションを持続することが難しい。なぜなら、ロボットがオブジェクトを見たときに視界からヒトが見えなくなると、ヒトを見に戻り行動を起こせなくなるからである。これに対してヒト乳幼児は、持続的なインタラクションを支える認知的な能力を持っていることが知られている。特に重要な行動は、乳幼児がヒトの顔を選好的に注視する行動と参照視行動である。前者は早い段階から乳幼児が物よりもヒトの顔を好んで見る行動傾向として知られており、後者は乳幼児が何らかの物体を見たときに親の顔を参照する行動として知られている²⁾。また、発達初期の参照視行動は不安や喜びといった情動状態を起因とした行動だと考えられている¹¹⁾。それは参照視が、見知らぬ物を見たときには不安になって親を見るのであり、また興味のある物を見たときには喜びを原因として親を見るのではないかと考えられるからである。

我々はこれまで、共同注視と参照視をセットにした交互凝視行動のモデル化を進め、意図的主体性を形成する計算モデルに情動的参照視を組み込んだモデルを構築してきた^{6),12)}。本研究ではこれをロボットに実装する。このために、ロボットには不安・中立・喜びの3つの情動状態を用意する。ロボットはオブジェクトの注視で喜び状態になり、一定時間の内にオブジェクトもしくはヒトの顔を注視できないときに不安状態になる。また、ヒトの顔を注視すると中立状態になる。参照視は、ロボットが最後にヒトの顔を見た時の自己位置（これを自己受容感覚と呼ぶ）を記憶し、不安および喜びの状態になったときにヒトの顔に付随する自己受容感覚を想起してヒトの顔を見に戻り行動として実装する。また、ヒトの顔の選好注視行動は、ロボットの視界にオブジェクトとヒトの顔が映っているときに、ヒトの顔を優先的に注視する行動として実装する。ロボットの行動アルゴリズムは、3つのモデル (VO,RJVA,IJVA) と参照視およびヒトの顔の選好注視行動を組み合わせ、図3のように構成する。このアルゴリズムにおいて、情動は喜びもしくは不安状態から始まる。これは、ヒトの顔を注視した状態からでなければ共同注視が機能しないためである。

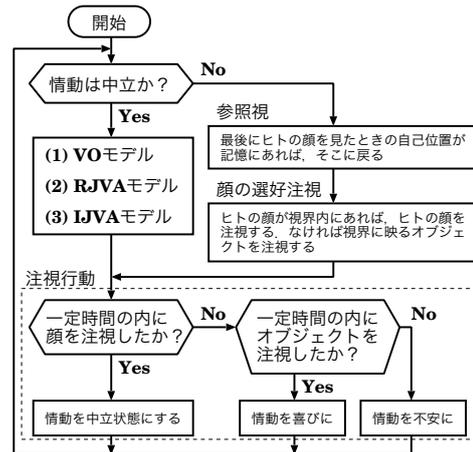


図3 ロボットの行動アルゴリズム
Fig.3 Algorithm of the robot behavior.

3. ロボットの仕様と実験環境

計算モデルを実装するロボット（視覚-運動システム）は、ステレオカメラ (PGR 社製 Bumblebee2) とパン・チルト駆動装置 (TRACLabs 社製 Biclops), およびこれらを制御する計算機 (パーソナルコンピュータ) で構成する (図4 (4. ロボット)).

このロボットに認識させるオブジェクトとヒトの顔の特徴は次のとおりである。

- オブジェクトの認識
画像から識別する色を4つまで設定し、その色の塊が、設定する閾値を超える領域をオブジェクトとして認識する。実験には4つのゴムボール (赤, 緑, 黄, 青) をインタラクション用のオブジェクトとして用意する。
- 顔の認識と視線方向の検出
顔領域の特定には OpenCV[☆]を用いる。視線方向は眼球ではなく頭部運動で代替する^{☆☆}。視界に映るヒトの顔を注視した後で、検出される顔領域の重心が移動する方向によって顔の向きを特定する。
ロボットを使った実験環境は図4のように構築する。被験者はロボットの前に座り、視覚的なコミュニケーションを行なう。このとき、ロボットの見ているもの、および見ようとしているものは計算機から取得する。また、ヒトの見ているものの計測には視線計測装置 (Tobii 社製 X120) を用い、さらにヒトの見よ

☆ <http://opencv.willowgarage.com/wiki/>

☆☆ これには2つの理由がある。1つは、8ヶ月程度の乳児では眼球運動よりも頭部運動に対して共同注視を行ないやすいからである。もう1つは単純に視線方向の検出は技術的なハードルが高いという理由である。

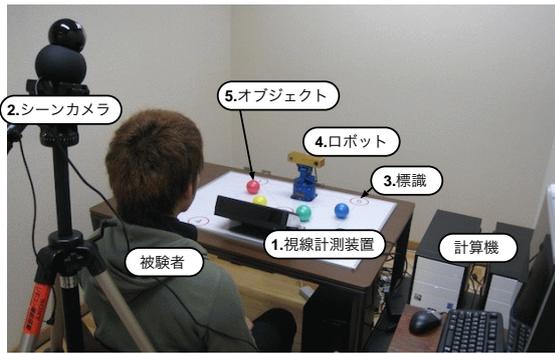


図 4 ヒトとロボットのインタラクション環境。手前に被験者が座り、その前方に 1. 視線計測装置, 2. キャリブレーション用のシーンカメラと 3. その標識, 4. ロボット, 5. オブジェクトがそれぞれ配置される。

Fig. 4 Experimental environment.

うとしているものの計測には音声報告を用いる[☆]。

4. 実験におけるロボットの振る舞い

視界にヒトの顔と複数のオブジェクトが映っている状態からロボットが動き始めると、ロボットはヒトの顔の選好注視行動によってヒトの顔を注視する。ヒトの顔を注視すると情動は中立状態になり、3つのモデルのいずれかが動作することになる。このとき、VOモデルは単純に視界の端に映っているものを注視し、RJVA および IJVA モデルはヒトの視線方向にあるオブジェクトを注視することになる。その後、ロボットの情動状態はオブジェクトを注視してもしなくても中立状態ではなくなるため、参照視によってヒトの顔を見に戻ることになる。つまり、3つのモデルのいずれにおいても、ロボットの視点はヒトの顔とオブジェクトの間を交互に行き来し、モデルごとの行動の違いはヒトの顔を見た後にのみ現れることになる。

ヒトとのインタラクションでロボットの意図性を判断することが難しいのは、ヒトがロボットを意図的な主体と捉える傾向があるからである。これに対して本稿で提示する実験では、VOモデルでの応答を、ヒトが補ってしまう意図性のベースラインとし、そこからRJVAモデルとIJVAモデルの計測データを比較することができる。これによって、ロボットの内部に形成される目標状態が意図的なコミュニケーションにどのように寄与するのかを明らかにすることができるのではないかと考えられる。

謝辞 本研究は財団法人日産科学振興財団および文部科学省科学研究費 (No.21700285) の助成を受けま

[☆] 赤いオブジェクトを見ようとしている場合には「赤」と発声してもらい、逆にロボットの見ているものを推定しようとしているときには「赤かな」などと発声してもらうよう被験者に依頼する。

した。ここに記し謝意を表します。

参 考 文 献

- 1) Butterworth, G. and Jarrett, N.: What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy, *British Journal of Developmental Psychology*, Vol. 9, pp.55-72 (1991).
- 2) Tomasello, M.: Joint attention as social cognition, *Joint Attention: Its Origins and Role in Development* (Moore, C. and Dunham, P., eds.), Lawrence Erlbaum, pp.103-130 (1995).
- 3) Paterson, S., Heim, S., Friedman, J., Choudhury, N. and Benasich, A.: Development of structure and function in the infant brain: Implications for cognition, language and social behaviour, *Neuroscience & Biobehavioral Reviews*, Vol.30, No.8, pp.1087-1105 (2006).
- 4) Asada, M., MacDorman, K., Ishiguro, H. and Kuniyoshi, Y.: Cognitive developmental robotics as a new paradigm for the design of humanoid robots, *Robotics and Autonomous Systems*, Vol.37, pp.185-193 (2001).
- 5) 長井志江, 細田 耕, 森田章生, 浅田 稔: 視覚注視と自己評価型学習の機能に基づくブートストラップ学習を通じた共同注意の創発, *人工知能学会論文誌*, Vol.19, No.1, pp.10-19 (2004).
- 6) 金野武司, 橋本 敬: 乳幼児の視線: 交互凝視行動の計算論的研究, *認知科学*, Vol.15, No.2, pp. 233-250 (2008).
- 7) Tomasello, M.: *The cultural origins of human cognition*, Harvard University Press, Cambridge (2000).
- 8) Hood, B., Willen, J. and Driver, J.: Adult's Eyes Trigger Shifts of Visual Attention in Human Infants, *Psychological Science*, Vol.9, No.2, pp.131-134 (1998).
- 9) Corkum, V. and Moore, C.: Development of joint visual attention in infants, *Joint Attention: Its Origins and Role in Development* (Moore, C. and Dunham, P., eds.), Lawrence Erlbaum, pp.61-84 (1995).
- 10) Nagai, Y., Hosoda, K., Morita, A. and Asada, M.: A constructive model for the development of joint attention, *Connection Science*, Vol.15, No.4, pp.211-229 (2003).
- 11) 小沢哲史: 社会的情報収集行動の起源と発達, 読む目・読まれる目 (遠藤利彦, 編), 東京大学出版会, pp.139-156 (2005).
- 12) 金野武司, 橋本 敬: 社会的参照の獲得における参照視行動の役割—構成的アプローチによる検討—, *日本認知科学会第 24 回大会発表論文集*, pp.80-81 (2007).