

人型エージェントによる不気味な笑顔表出に対する小脳の内部モデルを用いた異常検知メカニズムの記述の試み

Qualitative Description for Error Detection to Uncanny Smile of Humanlike Agent with Internal Model Formed by Cerebellum

田和辻 可昌^{1*} 近藤 佑亮² 松居 辰則³
TAWATSUJI Yoshimasa¹ KONDO Yusuke² MATSUI Tatsunori³

¹ 早稲田大学 大学院人間科学研究科

¹ Graduate School of Human Sciences, Waseda University

² 早稲田大学 人間科学部

² School of Human Sciences, Waseda University

² 早稲田大学 人間科学学術院

² Faculty of Human Sciences, Waseda University

Abstract: In observing the smile expressed by a humanlike agent, it sometimes elicits negative emotional valence. Our hypotheses were as follows: (i) human expected the facial expression of others, based on the internal model of smile expression formed by cerebellum. (ii) the movement of smile expression of humanlike agent were slightly different from that of human, and (iii) which causes the observer to detect the error when observing the smile expression of humanlike agent as it was different from expected movement. In this paper, the brain-functional model was proposed providing for explanation for the error detection.

1 はじめに

1.1 背景

近年ではアンドロイドなど、より物理的にヒトの外見に類似した人型エージェントが登場し、複雑な表情表現が可能となってきた。このことからヒトとヒトの間で行われるような複雑な非言語コミュニケーションがヒト-エージェント間で実現することが期待される。このとき、人型エージェントの表出した表情が意図したとおりにヒトに伝わるかという点は重要な課題である。

具体的には、人型エージェントが表出した表情に対して意図せず「不気味である」「怖い」といった評価を受ける可能性が示唆される。例えば、これまでもアンドロイドが笑顔を表出する際に、「どことなく不気味である」といった感覚を抱くことは経験的に共有される。このような現象は「不気味の谷」[1]として近年では広く知られている。ところが、その否定的評価に関する具体的な生起要因や評価に関する情報処理プロセスに関して、未だに明らかになっていない点が多い。

1.2 先行研究

これまで、心理学や認知科学、脳科学など様々な分野において、「不気味の谷」と考えられる否定的評価の生起要因や評価に関する情報処理プロセスを明らかにする試みがなされてきた。これまでの研究に共通している点として、否定的評価の生起に関しては、人型エージェントに対する期待と実際の不一致という枠組みで説明が可能であるという点が挙げられる。心理学的実験に基づく結果として、人型エージェントの眼を形態的に異常なほど大きくすると、観察者に否定的な印象を与えることが示唆されている [2]。「この異常なほど」という点について、「ヒトの目と比較して異常なほど」という暗黙的な眼の大きさに対する期待が人型エージェントに抱かれていることが考えられる。脳科学の分野において、Saygin らは、人間、アンドロイド、機械的なロボットのそれぞれが手を振る動作を被験者に観察させ、fMRI によって被験者の脳血流量の変化を調べた [3]。ここで得られた結果に対して、Saygin らはアンドロイドの動作と不気味の谷について以下のように考察する。すなわち、人間のような外見を持つアンドロイドは典型的な機械的動作を行わないであろうという被験者の予想に反し、実際の動作が機械的であるという

*連絡先：早稲田大学 大学院人間科学研究科
〒 359-1192 埼玉県所沢市三ヶ島 2-579-15
E-mail: wats-kkoverfay@akane.waseda.jp

齟齬が不気味の谷に関連している。つまり、多くの研究から、人型エージェントに対する否定的評価には人間的な特徴（外見・動作）を期待することと実際に知覚された特徴の間に齟齬があることで生じることが大きな要因であることが示唆されている。

1.3 目的

既往研究が示唆する「期待との不一致」を考えると、人型エージェントの笑顔動作を観察した際に生じる否定的評価も枠組みによって説明できることが期待される。そこで本研究では、人間が経験的に獲得する笑顔動作系列と、実際の人型エージェントが表出する笑顔動作系列との間に齟齬があると仮定した。この上でこの齟齬を検出する神経基盤モデルを提案することを本研究の目的とする。また、このモデルとこれまで我々が構築してきたモデルとの統合を図り、人型エージェントが表出する表情動作に対して、どのようにして否定的に評価されているか、に関する説明を試みる。

1.4 本稿の構成

本稿の構成は次の3つの節からなっている。第2節では、われわれがこれまで構築したモデルについて紹介する。ここでは、人型エージェントに対して異常を検知した際にヒトの情動状態がどのように遷移するのか、に関する定性的モデルを記述する。第3節では、表情動作に関する情報処理を取り入れるために、新しく小脳のモデルを追加することを提案する。この点については、以下の二点について詳細に述べる。まず、これまでのモデルを表情認知に関する情報処理に応用することができる点について述べる。次に、特に小脳のモデルは、他者の動きの理解において必要であり、どのようにして異常が検出されるのかという点についても説明可能である点について述べる。最後に、これらを踏まえてこれまで提案してきたモデル（従来モデル）と本稿で提案するモデルとの統合を行う。

2 従来モデルの概要と説明限界

これまで我々は、ヒトが人型エージェントの顔を見た際に得られる異常の度合いに基づいて、ヒトのエージェントに対する視線と情動状態（ポジティブ/ネガティブ）の変化過程を表現するモデルを脳科学の知見を基に構築してきた [4]。本節では、まずこれまでのモデルでどこまで記述されているのかについて述べる。その上で、これまでのモデルにおける説明限界について述べる。

2.1 従来モデルの概要

これまで我々は、人間が人型エージェントを観察した際に生じる否定的評価に対して、感情の二重経路 [5] による情動情報処理の結果、齟齬が生起されると考えてきた。図1にこれまで我々が提案してきたモデル（従来モデル）の概念図を示す。このモデルでは、人間が人型エージェントを観察する際に、二つの情動情報処理がなされると考えた。まず一つの情報処理系は、迅速で雑多な情報処理によって、人型エージェントを「人間」と判断する系である。もう一つの情報処理系は、これにやや遅れた詳細な情報処理で、人型エージェントを「非人間」と判断する系である。この二つの系によって導かれる判断結果の齟齬が否定的な情動形成に関わるというモデルである。このモデルは従来研究の「期待との不一致」の枠組みを説明できるモデルであると考えられる。

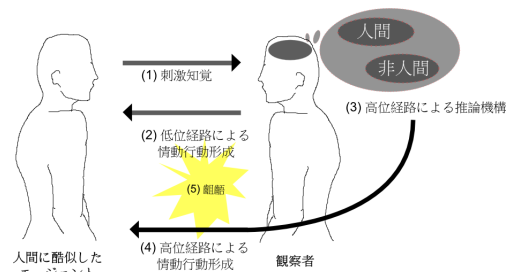


図1: これまでに提案してきた否定的情動反応形成モデル

この上で、我々は人型エージェント観察時のヒトの情動状態が時間的にどのように遷移していくかについて脳科学の知見を基にモデル化を試みた。具体的には以下の仮定に基づいてモデル化を行っている。人型エージェントには人間に比べて形態的な異常があり、知覚された情報は扁桃体の機能によって、その異常に基づく情動評価が行われる。前者は、言い換えると、形態的な異常が少なければ少ないほど「人間」、形態的な異常が多ければ多いほど「非人間」であることを表す。また、異常に基づく情動評価は、異常の大きさに対して反比例的に決定される。つまり、エージェントに対する情動評価を $V(t)$ 、異常の大きさを $abn.$ (ただし、 $0 \leq abn. \leq 1$) で与えると、

$$V(t) = -abn. + (1 - abn.) \quad (1)$$

で決定される。

これまでのモデルでは人型エージェントの目鼻口にそれぞれ形態的な異常が予め与えられている。それらを逐次観察することによって異常を検出し、式1にしたがって情動評価がなされる。この情動評価は逐次の

情動反応と考え、それらの和算 $\sum_{t \in T} V(t)$ を人型エージェントに対する認知的評価として求めた。

2.2 説明限界とモデルの発展の検討

従来モデルでは形態的な異常の度合いを予め定めており、人間がどのようにして異常を検出しているのかについては考慮されていない。特に、表情動作における異常検知を考えた場合は、この異常がどのように計算され、情動評価に影響を与えるのかについて検討する必要がある。そこで、次節では人型エージェントが行う表情動作に対する異常検知過程をより詳細にモデル化することを目指す。

3 表情動作予測と異常検知に関する 大脳-小脳モデル

本節では、これまでのモデルを表情認知に適用できることを初めに述べる。その上で (1) 表情動作系列の予測、(2) 予測動作と実際に知覚した動作との間に生じる齟齬を計算するモデルを提案する。このモデル化にあたり、まず、ヒトがどのように表情動作を生成しているかという点について、脳科学の知見に基づき大脳と小脳の連関という観点から述べる。次に、この表情動作生成に関するモデルを用いることで、ヒトは知覚した初期の表情動作から期待される他者の表情動作を計算できることについて述べる。

3.1 表情認知への従来モデルの適用可能性

ヒトの顔・表情認知においては二つの情報処理系が存在していることが示唆されている [6][7]。一つ目は、顔刺激に関する低空間周波数成分をもとに雑多かつ迅速に処理される系であり、これは皮質下経路を通り扁桃体へと伝達される。この系によって顔の全体知覚および表情認知がなされる。これに対して、もうひとつの系は先ほどの情報処理に比べると遅く、皮質経路を経て扁桃体へと伝達される情報処理系である。こちらでは顔刺激に関する高空間周波数成分をもとに顔の形態的な特徴などが処理される。これらの二つの系は従来モデルの感情の二重経路と対応すると考えられ、従来モデルを顔・表情認知に適用できることが示唆される。

3.2 表情動作における小脳の役割

これまで小脳は運動動作、特に手足の運動などの随意運動を制御する上で重要な役割を担っていることが指摘されてきた [8]。一方で、人間の表情生成に対しても小

脳が関わっていることが近年示唆されている。Ahmedらによると、情動調節障害において大脳-橋小脳経路が極めて重要な役割を果たしていると考えられ、大脳から与えられるムードや社会的状況に沿った適切な（表情生成などの）情動反応を行うように、小脳が情動反応表出の調整を行っているという仮説を立てている [9]。また、表情は表情筋によって構成されているが、これら表情筋は顔面神経（第 VII 脳神経）のひとつである。この表情筋を支配する顔面神経核は「橋の尾側レベルで三叉神経運動核の尾側に位置し」（p.1007）ている [10]。したがって、小脳は、大脳から送られる感情状態や状況に適切と考えられる表情筋動作系列を決定し、橋（脳幹）にそれらを実行させていると考えられる。

これを受けて感情から表情がどのように生成されるかに関してモデル化を行う。まず大脳皮質では感情状態 $E.S.$ (=Emotional State) と状況 S が決定され、 $X = (E.S., S)$ が出力される。小脳ではこの出力に対して適切と考えられる表情筋動作系列 $m(t)$ を決定する。ただし、 $m(t)$ は各表情筋 $m_i (i = 1, 2, \dots)$ の時刻 t における定性的状態を表す。これらの動作系列は橋を通して表情筋へと伝達され、表情が生成されると考える。すなわち、表情動作生成にいたっては、(1) 感情状態と状況に関する情報 X の決定（大脳皮質）、(2) 大脳からの情報 X に即した小脳による適切な表情動作系列の決定、(3) 橋による表情動作の実行の三段階からなる。

3.3 他者の表情動作系列の予測

次に、他者の表情動作の予測がどのように行われるかについて述べる。まず、前小節で見た表情動作生成の流れにおける第二段階、小脳による適切な表情動作系列の決定は、小脳における内部モデルによって決定されると考える。川人 [8] によると、ヒトの「速くて滑らかな多関節運動」を実現するにあたって、「中枢神経系の中に制御対象の内部モデルが存在し」ているとされる (p.117)。これまでの議論から表情生成にあっても小脳の関与が示唆されているから、表情動作生成においてもその制御のための内部モデルの存在が示唆される。

この表情動作系列の初期値 $m(0)$ が観察されると、この観測値と内部モデルを用いることで、その表情動作の予測ができると考えられる。ただし、ここで用いる内部モデルは他者が有する内部モデルではなく、あくまで観察者が持つ内部モデルを用いると考えた。

まず、知覚された初期表情動作 $m(0)$ に対して迅速な情報処理がなされ、感情状態 $E.S.*$ が推定される。さらに、ここで状況 S は他者と共有されているから、他者における大脳皮質からの入力 $X* = (E.S.*, S)$ が推定される。ここから $X*$ および $m(0)$ に沿った表情動作系列 $m*(t)$ が推定される。

3.4 予測動作と知覚動作間の誤差計算

次にこの予測動作と知覚動作間の誤差計算について述べる。初期動作に対して内部モデルを通して予測される動作系列 $m^*(t)$ と逐次知覚される動作系列 $m(t)$ の間で乖離度を求めることで生成された表情動作系列の異常度 $abn.$ が計算されると考えられる。つまり、

$$abn.(t) = M^+ (\|m^*(t) - m(t)\| - \theta). \quad (2)$$

ただし、ここで $\|x\|$ は x のノルムを表す。また、 θ は乖離誤差を異常と見なすかどうかを判別する閾値であり、 $y = M^+(x)$ は x と y の間に定性的な単調増加関係があることを表す。

3.5 従来モデルとの統合

これまで提案したモデルでは、目や鼻、口に設けられた各顔特徴部位の異常度に対し、扁桃体が情動評価を行っていると考えてきた。これに対し本研究では、この異常度が、小脳で学習された表情動作生成の内部モデルに基づく予測表情動作と、逐次知覚される表情動作との間の乖離度によって計算されると考えた。この乖離度は式2によって異常度に変換され扁桃体に伝達され、情動状態が更新されると考えられる。

4 まとめと今後の課題

本稿ではヒトが人型エージェントの笑顔に対して生じる否定的評価（「どこことなく不気味である」）といった評価がどのような情報処理に基づいて生成されるかに関するモデルを脳科学的知見に基づいて構築した。具体的には、表情の動作の側面に着目し、表情動作生成に関する内部モデルを用いた予測表情動作と、知覚される表情動作の間で逐次誤差が計算される点を従来モデルから発展させた。この内部モデルにおいては小脳がその形成に関わっていることを既往の神経科学的知見から検討した。

今後はこの表情動作系列を生成する内部モデルがどのように獲得されるのか、またその表現形式について検討する必要がある。前者については、人間の乳児が生まれてまもなく他者の表情を模倣する [11] という知見から、乳幼児における模倣によって表情動作系列の学習が行われることが考えられる。また、表現形式に関しては、表情筋の定性的状態、すなわち表情筋が伸ばされているか、縮んでいるかを定性値として表現し、小脳はその制御項を学習すると考えられる。

参考文献

- [1] 森 政弘: 不気味の谷, エナジー誌, Vol. 7, No. 4, pp. 33-35 (1970) <http://www.getrobo.com/> (2013年2月16日参照)
- [2] Seyama, J., Negayama, R.S.: The uncanny valley: Effect of realism on the impression of artificial human faces, *Presence:Teleoperators and Visual Environments*, Vol. 16, No. 4, pp. 337-351 (2007)
- [3] Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, J., Frith, C.: The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions, *Cognitive and Affective Neuroscience*, Vol. 7, No. 4, pp. 413-422 (2012)
- [4] 田和辻 可昌, 村松 慶一, 松居 辰則: 脳の機能的結合に関する定性表現を用いた人型エージェントに対する情動状態記述の試み, 人工知能学会論文誌, Vol. 30, No. 5, pp. 626-638 (2015)
- [5] ジョセフ・ルドゥー (LeDoux, J.): 『エモショナル・ブレイン: 情動の脳科学』松本 元, 川村 光毅ほか訳, 東京大学出版会 (2003): *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, Simon & Schuster Paperbacks (1996)
- [6] 飛松 省三: 事象関連電位を用いた顔認知機構の解明, *Brain and Nerve*, Vol.64, No.7, pp.717-726 (2012)
- [7] Palermo, R., Rhodes, G.: Are you always on my mind? A review of how face perception and attention interact, *Neuropsychologia*, No.45, pp.75-92 (2007)
- [8] 川人 光男: 『脳の計算理論』, 産業図書 (1996)
- [9] Ahmed, A., Simmons, Z.: Pseudobulbar affect: prevalence and management, *Therapeutics and Clinical Risk Management*, Vol.2013, No.9, pp.483-489 (2013)
- [10] 山尾幸広, 芝田純也, 宇佐美清英, 松本理器編訳: 「脳幹の感覚機能, 運動機能, 反射機能」. 『カンデル神経科学 Fifth Edition』, メディカル・サイエンス・インターナショナル, 999-1015 (2014)
- [11] Meltzoff, A.N., Moore, K.M.: Imitation of facial and manual gestures by human neonates, *Science*, Vol.198, No.4312, pp.75-78 (1996)