

# 大脳小脳の機能に基づいた人型エージェントの表情動作に対する異常検知を表す定性微分方程式の構築

Qualitative Differential Equation of Cortico-Cerebellar Function Explaining for Human Error Detection toward Facial Expression of a Human-like Agent

田和辻 可昌 \*1  
TAWATSUJI Yoshimasa

松居 辰則 \*2  
MATSUI Tatsunori

\*1 早稲田大学 大学院人間科学研究科  
Graduate School of Human Sciences, Waseda University

\*2 早稲田大学 人間科学学術院  
Faculty of Human Sciences, Waseda University

Recently, uncanny valley, which refers to human negative emotional response towards humanlike agents, has been investigated in various scientific fields while there lacks a comprehensive model providing an explanation for the results of these researches. Our aim is to provide such qualitative model based on brain functional connectivity, which allows us to explain the cause-and-result of the uncanny valley phenomena. In this study, we proposed a qualitative model representing how error was detected during the observation of the humanlike agent's smile, based on the cortico-cerebellar function. The model consists of two systems; the one is for expecting facial movement, and the other is for emotional processing. Our model suggests the rapid emotional processing to calculate the expectation of facial movement. Thus it is proposed that thalamus should play a central role in error detection toward facial expression of a humanlike agent.

## 1. はじめに

### 1.1 背景

汎用人工知能をエージェントに実装し、人間との相互創発的関係を築くためには、人間とエージェントの間で親和的なインタラクションを構築することが必要不可欠である。この上で近年の3DCGおよびアンドロイドに関する技術的發展が相まって、ヒトに酷似したエージェントが今後重要な役割を担うと考えられる。なぜならば人型であることの大きな特徴として、人間にとって親和性の高い非言語情報の表出が可能であるという点が挙げられるためである。特に、非言語情報の中でも自身の情動状態を表す表情は、相手の心的状態を直感的に理解する上で極めて重要な役割を果たす。

ところが、設計と帰属の観点から考えると、必ずしも意図して設計された表情を、観察者がその意図通り受け取るとは限らない。例えば、人型エージェントに笑顔を実装する場面を考える。このとき、設計者は人型エージェントに笑顔を表出させることを目的として、ヒトの笑顔表出動作に対して形態的にエージェントの表情動作を近づけるを試みる。しかし、わずかなタイミングのずれによって、ユーザには不気味な笑顔に映ることが経験的に考えられる (ref. 「不気味の谷」[1])。このような課題はヒトの表情生成が複雑な (表情筋動作) システムから構成されていることから、技術的な再現が極めて難しいことが背景にあると考えられる。

このような人型エージェントに対する否定的感情については、様々な分野で研究が進められてきている。一方で、これらの研究はある特定の条件下での人型エージェントに対する否定的な感情を実験的に捉えるにとどまっており、これらを包括的に説明するモデルは現在提案されていない。したがって、否定的な感情が生起する現象を理解するためには、その感情生起に至る原理を踏まえた説明モデルが必要不可欠である。

### 1.2 先行研究: 「期待との不一致」

これまで既往研究では、否定的感情生起の要因が「期待との不一致」という観点から説明されてきた。具体的には、以下の二つの情報処理の齟齬である。一方では、対象 (ここでは人型エージェント) に対する情報過程において「知覚対象が人間である」という期待が得られる。他方では、「知覚対象は人間ではない」という期待に反する情報処理がなされる。これらの情報処理の齟齬が、不気味さを形成する要因となっているというものである。

Seyama ら [2] は、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を実験的に確認した。これはヒトが暗黙的にも人間の目の大きさに比べて大きい目をエージェントが有していることが異常として検知されていることを示唆している。また、Saygin ら [3] 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という動作を被験者に観測させ、エージェントを観察している被験者の脳血流量の変化を fMRI によって取得し、それぞれ比較した。その結果、人間と機械的なロボットの動作に比べ、アンドロイドの動作を観察している際は、脳領域の高い活動性が認められた。ここで得られた結果に対し、Saygin らは以下のように考察している。一般的に、脳は predictive coding を行っており、うまく予想ができなかった刺激に対して脳の活動は強くなると考えられる。このことから、人間のような外見を持つアンドロイドは、人のような動作を行うであろうという被験者の予想に反して機械的な動作を生成することから prediction error が発生し、脳の活動が強くなったと考察している。

以上から、ヒトが経験的に獲得し暗黙的に期待する笑顔に関する表情動作と、実際の人型エージェントが表出する表情動作との間に齟齬がおこることで異常が検知されると考えられる。なぜならば、既往研究が示唆する「期待との不一致」を考えると、不気味の谷はこの「期待との不一致」の枠組みで包括的に説明されると考えられ、したがって、表情動作に対する否定的な評価 (ex. 「笑顔が不気味である」) の生起プロセスも同様の枠組みで説明されることが期待されるためである。

連絡先: 田和辻可昌, 早稲田大学 大学院人間科学研究科, 〒359-1165 埼玉県所沢市堀之内 135-1 フロンティアリサーチセンター 213 実験室, E-mail: watsk-koreverfay@akane.waseda.jp

### 1.3 目的と本研究の意義

本研究の目的は、人型エージェントの笑顔に対する異常検知プロセスを説明する脳機能モデルを構築することである。具体的には笑顔に関する期待動作知覚と知覚動作知覚間で齟齬を計算するプロセスを説明する脳機能モデルの構築を試みる。

また、本研究では脳機能、つまり脳部位の機能的接続性に着目してモデル化を行う。ここでは一つの機能を有する脳の系を機能単位とし、それらの接合によって一つの現象を明らかにするアプローチである。このアプローチでは、それらの機能がどの順番で作用するか、という機序が最も重要となる。したがって、機能単位を支える神経基盤の知見に着目し、それらの機序が成り立つかどうかについて制約をかけた定性シミュレーションを行う。これまで感情に関する既往研究は以下の二つのアプローチに分けることができる。一つは、神経生理学などに代表されるように、機能を構成する部品（感情に関係する部位の細胞の構造や機能）を明らかにする手法である。もう一つは、認知科学に代表されるように、トップダウンなプロセスとしてのモデルを構築するものである。これらはそれぞれ重要な立場ではあるが、それぞれ固有の課題が存在する。前者は、その構成単位の仕組みが分かったとしても、それが果たす機能の観点を導入しない限り、その構成単位の集合がなす現象を説明することは難しいという問題がある。後者は、プロセスとして記述されたモデルが実際にどのように脳というシステムの上で実現されているのかという点にまでは記述されない。したがって、双方の課題を相補的に解決する本アプローチは、今後の汎用人工知能分野において感情というシステムを考える上で重要な役割を果たすと考えられる。

## 2. これまでに構築したモデル

### 2.1 従来モデルの概要

これまで我々は、ヒトが人型エージェントを観察した際に生じる否定的評価は、二つの情動情報処理の結果生じる齟齬に起因すると考えてきた [4] (see also: (Shimada 2007) [5])。まず、一つの情報処理系では、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「評価」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を行い、人型エージェントを「非人間」と「評価」する系である。このモデルは従来研究の「期待不一致」の枠組みを、神経科学的知見から説明する有用なモデルと考えられる。

### 2.2 従来モデルの説明限界

これまでのモデルでは、入力を入型エージェントの目鼻口における形態的異常度として定義していた。そして、この予め与えられた異常度に従って、人型エージェントに対する「評価」が時系列的にどのように変化するかを検討されていた。したがって、このモデルにおける大きな課題として、異常度が予め与えられており、この異常度がどのように検出されるのか、という枠組みについては説明がなされていない、ということが挙げられる。

我々は人間の異常検知に関するモデルを神経科学的知見に基づいて構築し、従来のモデルとの統合を試みてきた。図 1 に本研究における異常検知メカニズムの概略図を示した。次節以降より、この提案モデルについて説明する。

## 3. 内部モデルを用いた異常検知メカニズム

### 3.1 異常度の計算

ヒトが知覚した対象に対して異常を検出する方法としては、期待との不一致という点から考え、知覚対象の期待動作に対し

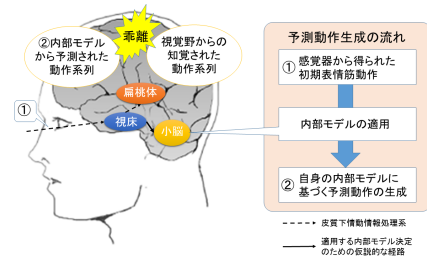


図 1: 予測動作生成と異常検知に関わる脳部位。

て、知覚された情報が大きく外れている際に異常が検出されるというものであると考えられる。そこで、表情表出における計算では、期待される表情筋動作  $m^*(t)$  と、実際に知覚され視覚野から伝達された表情筋動作  $m(t)$  との間で起こる乖離で異常が計算されると考えた。また、この乖離  $d(m^*, m)(t)$  は  $m^*(t)$  と  $m(t)$  の差のノルムで計算され、異常度はこの乖離に定性的に比例すると考えた。

### 3.2 表情筋動作系列の生成モデル

本小節では、先の小節で紹介した乖離を求める上で必要な予測表情筋動作  $m^*(t)$  が、どのような情報処理を経て得られるかについて述べる。ここでは、動作予測には小脳における内部モデルが必要であること、それは表情動作生成のメカニズムから理解されることの二点に関して述べる。

まず、表情認知において、表情に関する予測動作がどのように計算されるかを考える。任意の時刻の予測動作を生成することは、顔の形態（ここでは表情筋の状態を表す）における初期の変化量から任意時刻の動作を計算するということである。このためには、時間に対する後ろ向き計算をすることは間に合わず、対象の前向き計算が求められる [6]。このような前向き計算として、小脳の内部モデルに基づく動作生成の知見は重要である。

小脳と表情表出の関係性について、Ahmed ら [8] によると、情動調節障害において大脳-橋小脳経路が感情表出に重要な役割を果たしていると考えられている。具体的には、大脳皮質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ものである。また、表情は表情筋によって構成されているが、これらの表情筋が属する顔面神経核は「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 [9]。これらを踏まえると、小脳は大脳（運動野）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、実行系である橋の調整を行っているかと推察される。

そこで、表情筋の状態を制御するといった小脳の機能を定性的に表すと、以下の定性微分方程式が考えられる。

$$\frac{d^2 m}{dt^2} + \lambda_1(x) \frac{dm}{dt} + \lambda_2(x) m = 0 \quad (1)$$

ただし、ここで与えられたパラメータ  $\lambda_1, \lambda_2$  はそれぞれ、速度および加速度を制御する変数で、情動状態  $x$  から定まるパラメータである。この二階微分までの項を考えた点としては、五味 [7] の研究において、小脳（腹側傍片葉）のプルキンエ細胞の発火パターンに対して、視線の位置に関する二階微分方程式までを考えた逆ダイナミクスモデルでよく近似できているという知見に基づいた。

式 (1) に対して, 変数  $y = dm/dt, \xi = \begin{pmatrix} m & y \end{pmatrix}^T$  (ただし,  $x^T$  はベクトル  $x$  の転置を表す.) を導入すると, 次の表情筋に関する状態方程式を得ることができる.

$$\frac{d\xi}{dt} = \Lambda \xi \quad \text{ただし} \quad \Lambda = \begin{pmatrix} 0 & 1 \\ -\lambda_2(x) & -\lambda_1(x) \end{pmatrix} \quad (2)$$

ここで,  $z$  を表情筋の状態の出力として定義すると, 式 (2) と  $z = m$  によって制御系の状態方程式のみなすことができる. このことから,  $\Lambda$  は表情筋動作生成に関する内部モデルになっていると考えられる. 以上より, 広範な部位からまとめあげられた情動状態に即した内部モデルに従って表情筋を制御するという, 小脳の機能が定性的に記述された.

### 3.3 内部モデルに基づいた表情筋動作予測プロセス

先に構築したモデル (式 (2)) を用いて, ヒトがどのように対象の表情筋動作予測を行うのかについて述べる. ヒトは対象が表情筋動作を行う際に, 初期動作情報を知覚する. すなわち, これは時刻  $t = 0$  における他者の表情筋動作  $\xi^*(0) = \begin{pmatrix} m(0) & y(0) \end{pmatrix}^T$  を得ることに等しい. ここで, 次の仮定を置くと, 式 (2) で定義した表出のモデルを用いて他者の表情筋動作予測に適用できる.

仮定: 他者の表情筋動作生成過程は自身の表情筋動作生成過程と一致するとみなす. 言い換えると, 他者の動作生成の内部モデル  $\Lambda^*$  を自身の内部モデル  $\Lambda$  に対して,  $\Lambda^* = \Lambda$  が成り立つと考える.

これによって, 式 (2) に  $\xi^*(0)$  を初期値として与えることで, 任意時刻における他者の表情筋の状態, つまり予測動作を得ることができる.

## 4. 情動状態を表す定性モデルの構築

### 4.1 基本モデル

これまで情動状態をモデル化する試みは数多くなされてきた. これらの中でも情動状態を一つの心理空間に位置づけ, その状態変化をモデル化する試み [10] がなされてきた. そこで, 三輪らの情動方程式 [10] を参考に, 情動状態の時間発展に関して, 次のような基本的な定性的力学系モデルを考える.

$$\frac{dx}{dt} = f(x) + u(t) \quad (3)$$

このモデルでは, 情動状態  $x$  自身に関する時間的な変化に加えて外乱 (入力項)  $u(t)$  が与えられ, 時々刻々と情動状態が変化の様子が表されている. このようなモデルを考える上で, ヒトにおける情動情報処理の観点から見ると, 以下の二点に関して検討が必要であると考えられる. 一点は, 情動状態がどのような脳機能によって定まるかという点である. もう一点は, 情動状態を表す  $x$  が有する次元の神経科学的な妥当性である. そこで, この二点に関して次節で検討する

### 4.2 情動状態空間の神経科学的知見との対応

式 (3) に関して, 情動空間の次元と情動情報処理を支える機能ネットワークについて述べる. 本研究では, 先行研究の知見に基づき, 二つの神経系の機能として, 情動状態  $x$  が定義されると考える. すなわち,  $x \in \mathbb{R}^2$  であり,  $x = (x_1, x_2)^T$  と表すことができると考える. Russell が提案した感情の円環

モデル (感情は快-不快, 覚醒の程度を基底とする心理空間上にマッピングされるというもの) に関して Gerber らはヒトの表情を観察している際の被験者の脳血流を fMRI によって計測し, 各軸に係る機能的部位の同定を試みている [11]. この結果, 快-不快 (特に, 不快) に関する神経系として背側前帯状回皮質および頭頂皮質が, 覚醒 (の低さ) に関する神経系として扁桃核, 小脳, 背側橋核, 右内側前頭前野が関わっているとされた. これによって, 一方の系が他方の系に与える影響など神経科学的知見を基に考えることで, 式 (3) の関数  $f$  の具体的な性質を考察することができると考えられる. ただし, この結果には他者の情動評価に関する系に加え, 観察による自身の情動変化に係る神経系が関与していることが考えられるため, 他者の情動評価に関する系と, 自身の情動変化に関する系にさらに分離する必要があると考えられる.

ここで, 自身の情動変化に関して, ヒトが刺激 (ここでは人型エージェントの表情動作) を観測する前後での情動状態の変化について考える. 表情動作が表出される前, 観察者の情動状態はある定常状態, 力学系において固定点  $x^{\#*1}$  上にあると考える. そして, 表情表出を観察することで, 入力  $u$  が与えられ, 情動状態  $x$  が別の点に移される. その後, 速度における程度の差はあるものの, 情動状態は元の定常状態に戻ると考えられる. これらのことから, 式 (3) における固定点周りの挙動を考えると, 少なくとも固定点  $x^{\#}$  は沈点であると推察される. これは, 関数  $f$  の  $x = x^{\#}$  におけるヤコビアン  $J^{\#}$  の固有値の実部がすべて負であることを表す. 今, 式 (3) の固定点  $x = x^{\#}$  周りに摂動  $y$  を与えテラー展開を行うことで, 次の線型化方程式を得る.

$$\frac{dy}{dt} = J^{\#} y \quad (4)$$

固定点が沈点であり, 情動空間を張る基底ベクトルが二次元であるとすると, この  $J^{\#}$  は少なくとも以下の条件式を満足すると考えられる.

$$J^{\#} = P^{-1} \begin{pmatrix} -a_1 & 0 \\ 0 & -a_2 \end{pmatrix} P, \quad a_1, a_2 > 0 \quad (5)$$

ただし,  $P$  は正則行列. これによって, 刺激によって変化した情動状態は時間を有して元の情動状態に戻ると考えられる.

## 5. 総合考察

### 5.1 視床を中心とした広範ネットワーク

これまで, ヒトは対象の表情動作の予測を自身の内部モデルを用いて行っていることについて述べ, その定性的計算モデルを示した. ここで重要な点として, 対象から得られた初期知覚動作  $\xi(0)$  に自身の内部モデル  $\Lambda = \Lambda(x)$  を適用させて他者の表情筋動作予測を行っているのであれば, 対象がどの情動状態  $x^*$  にあるか, ということが事前に明確でないと予測ができない, ということが挙げられる. これは  $\Lambda$  が  $x$  の関数であることから明らかであると考えられる. したがって, 対象の表情筋動作の予測を行うためには, 迅速に情動情報が処理される必要がある

ここでの迅速な情報処理に関する系は, 感覚器から視床を通じて扁桃核に投射される系 (皮質下経路) である. このとき, この迅速な情報処理に対して, どのような経路を経て小脳

\*1 固定点とは, 力学系の理論において,  $\frac{dx}{dt} = 0$  となる点のことである.

へと伝達されるかに関して、我々は視床が重要な働きをしていると考えた。理由は以下の三点からなる。一点目は、扁桃体から小脳への投射が著者が調べた範囲では認められないことから、扁桃体から小脳への迂回では潜時を要してしまうと考えられるためである。二点目は、視床が扁桃体との結合をもつと同時に、小脳や運動認知に関わる大脳基底核や大脳皮質との接続を持つためである。三点目は、近年の研究において、人間の視覚系情動情報処理に視床枕が重要な役割を果たしていることが示唆されているためである [12]。以上のことから、視床においてある程度の情動情報が処理され、小脳へと送られると考えられる。すなわち、視床を中心とした広範なネットワーク構造を脳が有していると考えた。

## 5.2 各情報処理の機序

本小節では、これまでに構築したモデルから考えられる視覚情報と情動情報の処理機序について考察する。視覚情報処理において、網膜で受容された情報(図2:①)は視床を経て、情動情報は橋-小脳へ(図2:②)、知覚情報は視覚皮質へと伝達される(図2:③)。情動情報が小脳へと初めに渡される段階においては、形態的な情報が得られていないため、初期動作の方向が得られていない(図2:②時点)。そこで、皮質で知覚処理された方向に関する情報が視床へと戻され、小脳へと入力されると考えられる(図2:④)。この情報が入力されたうえで、小脳は内部モデルと初期動作を用いて、任意時刻の動作予測を生成すると考えられる(図2:⑤)。また、これとほぼ並列的に視覚情報処理が続けて行われ、視覚皮質を介して初期動作に続く表情動作が知覚される(図2:⑥)。この二つの処理結果が統合的に比較され、乖離があるかどうか計算される。

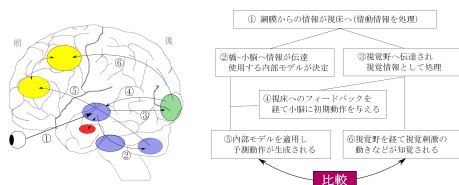


図 2: 予測動作と知覚動作の比較がなされるまでの処理過程。青色の部分は、上側から視床、橋、小脳を、緑色は視覚皮質を示している。また、赤色は扁桃体、黄色は前頭前野を示す。

## 6. まとめと今後の課題

本研究では、小脳の内部モデルに基づいた人型エージェントの表情表出に対する異常検知メカニズムのモデルを提案した。この結果、表情予測を行うためには、初期の表情動作に基づく情動状態を迅速に処理する必要があることが示唆された。我々はこの迅速な処理が低位経路によって与えられること、さらにその上で、視床が重要な働きを行っていると考えた。

今後は、以下の二点について神経科学的な知見を踏まえて検討し、モデルを精緻にする必要があると考えられる。一点目は、視床が情動情報処理を行うかどうかに関する文献を調査する必要がある。二点目は、情動状態に関する方程式における関数  $f$  について検討する必要がある。

## 謝辞

本研究は、早稲田大学 2015 年度特定課題研究助成費(特定課題 B)(課題番号:2015B-390)、および日本学術振興会科学研

究費補助金(課題番号:25540165)の支援を受けている。

## 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/>, 参照日: 2013/2/16
- [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, H. 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] Shimada, M., Minato, T., Itakura, S., Ishiguro, H.: Uncanny valley of androids and its lateral inhibition hypothesis, *Robot and Human Interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, pp. 374-379 (2007)
- [6] 川人 光男: 脳の計算理論, 産業図書 (1996)
- [7] 五味 裕章: 反射性眼球運動における小脳の役割と計算モデル, 日本視覚学会論文誌, Vol. 6, No. 4, pp. 483-489 (2013)
- [8] Ahmed, A., Simmons, Z.: Pseudobulbar affect: prevalence and management, *Therapeutics and Clinical Risk Management*, Vol. 2013, No. 9, pp. 483-489 (2013)
- [9] 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, In *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*, pp. 999-1015 (2014)
- [10] 三輪 洋靖, 伊藤 加寿子, 高信 英明, 高西 淳夫: 人間との情緒的コミュニケーションを目的とした人間形頭部ロボットの開発(第3報, 情動方程式・ロボットパーソナリティの導入), 日本機械学会論文集 C 編, Vol. 70, No. 699, pp. 3244-3251 (2004)
- [11] Gerber, A.J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., Kangarlu, A., Zhu, H., Russell, J., Peterson, B.S.: An affective circumplex model of neural systems subserving valence, arousal, & cognitive overlaying during the appraisal of emotional faces, *Neuropsychologia*, Vol. 46, No. 8, pp. 2129-2139 (2008)
- [12] Pessoa, L., Adolphs, R.: Emotion processing and the amygdala: from a 'low road' to 'many roads' of evaluating biological significance, *Nature Reviews. Neuroscience*, Vol. 11, No. 11, pp. 773-783 (2010)