

特集 「人工知能と Emotion」

# 脳機能から見た人型エージェントに対する否定的感情生起プロセス

## Functional Processing of Brain Generating Negative Emotional Response toward Human-like Agents

田和辻 可昌  
Yoshimasa Tawatsuji  
早稲田大学大学院人間科学研究科  
Graduate School of Human Sciences, Waseda University.  
wats-kkoreverfay@akane.waseda.jp

松居 辰則  
Tatsunori Matsui  
早稲田大学人間科学学術院  
Faculty of Human Sciences, Waseda University.  
matsui-t@waseda.jp

**Keywords:** negative emotion, uncanny valley, brain functional model.

### 1. はじめに

本章では、まず不気味の谷現象がどのような現象であるかを説明したのち、この現象理解に取り組む意義について述べる。

#### 1.1 不気味の谷現象

外的表現を有し知的に振る舞うソフトウェアエージェントや物理的身体を有するロボット（以下、合わせて単にエージェント）と人間の共生の在り方を考えるうえで、エージェントを介したインタラクションにおける人間の感情特性を理解することは重要である。特に、近年の映像技術や工業技術の発展に伴うエージェントの人型化において、「不気味の谷」[森 70]と呼ばれる現象が注目されている。一般的に、エージェントの外見が工業ロボットのような機械的なものよりも、手足や顔の表現が与えられ、人間の外見に近づいたもののほうがエージェントに対する親和度は高いと考えられる。このことから図1（破線）のように、横軸に人間に対するエージェントの類似度、縦軸にエージェントに対する人間の親和度をとると、両者の関係は単調増加の関係が予想される。ところが、一般的には図1（実線）のように、人間に対する類似度が極めて高いある地点において、急激に親和度が負の方向へと転落すると考えられている。また、さらに類似度が高くなり、人間とほとんど区別がつかなくなると、親和度が正の方向へと回復し、完全に人間との類似度が等しくなったときに高い親和度が得られると考えられている。このような類似度と親和度の描くグラフの形状から、高い類似度における急激な親和度の下落を不気味の谷と呼ぶ[森 70]。

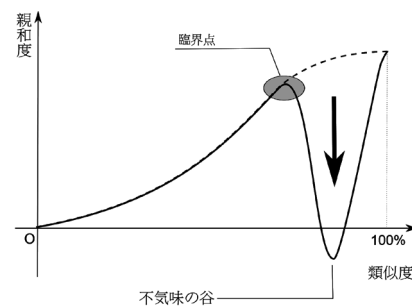


図1 不気味の谷の概念図  
([森 70]を参考に作成)

#### 1.2 不気味の谷現象理解に取り組む意義

これまで多くの研究が、不気味の谷現象の理解を目指して多角的なアプローチで取り組まれてきた。この取り組みの意義は大きく二つの視点で与えることができる。一つ目の意義は、図1に見られるように、谷を越えることで高い親和度をエージェントが得られるという点である。このような高い親和度を得ることができれば、より人間とエージェントとが共生する社会を実現することができると考えられる。ところが、この谷を越えることの困難さは森[森 70]の提唱に始まり、数々のメディアによって共有されてきた。そこで、不気味の谷を克服することを目指すのではなく、むしろ、いかに谷の手前で最大限の親和度を得るか、ということを目指すべきであるとも指摘される[森 70]。このためにも、親和度が落ちる直前の「臨界点」にあたる箇所と、その下落の要因を明確化する必要がある。

もう一つ目の意義は、不気味の谷現象の理解を通して、人間が有する同種認識の機能を理解することができるという点である。言い換えれば、ヒトがヒトとそうでないものをどのようにして見分けているのか、ひいて

はヒトはヒトをどのように認識しているのか、ヒトとは何かという根源的な問いに対して貢献することができる [Hanson 05]。石黒は、アンドロイドを人間に近づけていく中で、「不気味の谷をどのように越えるか、越えたならばなぜ越えられたかを理解することが、人間とロボットの相互作用、さらには、人間のコミュニケーションの原理に迫ることにつながる」(p. 10) と述べている [石黒 05]。このように、不気味の谷現象を基点として、ヒトがどのようにしてヒトとそうでないものを弁別しているのか、つまり、ヒトの同種弁別の機能を解明する一助となるといえる。このような観点を強調するため、以降本稿で考えるヒト型指向のエージェントを特に「ヒト型エージェント」と呼ぶ。

まとめると、不気味の谷現象を理解する意義は、大きく二点存在するといえる。一点目は、人間と円滑なコミュニケーションを取ることができる親和性が高いエージェントを構築するという社会的意義である。二点目は、ヒトがいかにしてほかの同種個体を認識しているかについて理解を進めることができるという学術的意義である。

### 1.3 本稿の構成

2章では、これまでの不気味の谷に関する先行知見について俯瞰的にまとめる。このうえで、3章では、我々がこれまで取り組んできた研究のうち、脳機能に着目した不気味さ生起過程を説明するモデルについて紹介する。さらに4章では、そのほか我々が取り組んできた研究に関して、不気味の谷と動作に関する研究、視聴覚に関する複数モダリティと不気味さに関する研究について紹介する。最後に5章では、不気味の谷に関する総合的な考察を行い、不気味の谷を考えるうえで重要であると考えられる課題について取り上げる。

## 2. 先行研究

不気味の谷に関する研究はこれまでに多数報告されてきた。そこで本章では、これまで不気味の谷に関してどのような研究が進められてきたかについて俯瞰的に述べる。なお、体系的なレビューについては、Kätsyriら [Kätsyri 15] に詳しい。

### 2.1 多彩な分野からのアプローチ

主に不気味の谷に関する研究は、ヒト型エージェントの外見に着目した研究が見られる。例えば、目の形 [Seyama 07] や、肌のテクスチャなどの形態的特徴 [MacDorman 09]、また、人間の目に対する感受性と不気味さの関係性に関する研究 [Schein 15] がある。そのほかには、動作を伴うヒト型エージェントを対象とした研究 [Chaminade 07, Saygin 12]、視聴覚のマルチモーダルに関する研究 [Mitchell 11]、表情表出における顔の上半分が不気味さに与える影響を調べた研究 [Tinwell

11] があげられる。さらに、人間の情報処理を計算モデルで構築し、このうえで不気味の谷を解釈する計算論的アプローチ [Moore 12, Shimada 07] も数少ないが見受けられる。これに加え、幼児の認知機能に着目した研究も見受けられる。幼児の例では母親と母親ではない女性、両者をモーフィングした画像に対する幼児の視覚選好性を調べたもの [Matsuda 12] や、発達段階に合わせてヒト型エージェントをどのように知覚するようになるかを実験的に検討する発達心理学的アプローチ [Shimada 07] がある。これらに加えて、Lewkowiczらは、生後6, 8, 10, 12か月の乳児に人間の顔、ヒト型エージェントの顔、形態的に目が大きなヒト型エージェントの顔を見せ、発達段階と不気味の谷との関連性を調査した [Lewkowicz 12]。また、人間以外の種 (e.g. マカクザル) で不気味の谷が見受けられるかを検証する比較行動学的アプローチ [Steckenfinger 09] によって、種を超えて共有されている防衛本能システムが否定的反応を形成していることが示唆されている。

### 2.2 不気味の谷現象に関する説明パラダイム

それでは、これらの研究から得られた結果について、統一的に説明する枠組みは存在するだろうか。不気味の谷現象を説明するうえで、知覚的齟齬 (perceptual mismatch) とカテゴリー曖昧性 (categorical ambiguity) の枠組み [Kätsyri 15]、あるいは期待誤差 (prediction error) [Saygin 12] の枠組みが重要である。まず、(知覚)心理学によって提唱された知覚的齟齬・カテゴリー曖昧性と不気味の谷の関係性について述べる。これらは言い換えると、人のカテゴリー知覚の特質から、カテゴリーが曖昧な刺激に対して否定的な評価がなされるというものである。Cheethamらによると、人はヒト型エージェントに対して、「人間」と「非人間」に関するカテゴリー知覚を行っていることが示唆されている [Cheetham 13]。このため、「人間」と「非人間」のカテゴリーの境界 (Category Boundary) においては、各カテゴリーのいずれに帰属させるかという知覚過程において競合が起こると考えられる。この結果、ヒト型エージェントに対する評価を否定的なものにする、と考えられている [Yamada 12]。

これに対して、期待誤差の枠組みは、Sayginらによって提唱された [Saygin 12]。一般的にヒトの脳は外界に対して predictive coding (予測) を常に行っている。したがって、うまく予測ができなかった刺激に対して脳の活動は強く活性化されると考えられる。このことから、人間のような外見をもつアンドロイドは、人のような動作を行うであろうというヒトの予測に反して機械的な動作を生成することから期待誤差が発生する [Saygin 12]。

### 2.3 動作を伴わない外見的特徴に関する研究

森の示した不気味の谷グラフ (図1) に対して、類似

度をいかに定義するかは重要な課題である。多くは（一方の端点を特定のCGエージェントやロボット、もう一方の端点を特定の人物とした）モーフィング画像を用いて刺激を作成し、刺激の段階的な変化を類似度として定義している。このように作成された一連の刺激に対する知覚心理学的実験から、人はヒト型エージェントに対して、「人間」と「非人間」に関するカテゴリ知覚を行っていることが示唆されている [Cheetham 13]。

#### 2.4 計算論的アプローチに基づく不気味の谷研究

これまでに紹介した心理学的実験をもとにしたアプローチに対して、人間の認知的情報処理をモデル化することで、不気味さがどのように表出されるかを説明するアプローチも見られる。ここでは重要と考えられる2編の文献をあげる。具体的には、人間がもつ側抑制に着目したモデル [Shimada 07] と知覚心理学における知覚のマグネット効果 (perceptual magnet effect) に着目したモデル [Moore 12] である。人間の認知情報処理の各段階には側抑制の機能があると考えられている。この考えを用いて、人間に対する類似度が高い刺激、言い換えれば、類似度が最大となる箇所近傍における刺激に対して応答感度は抑制されやすくなることが考察された [Shimada 07]。このような応答感度の抑制は情報処理が高次になるにつれて顕著となる。これらの段階的な側抑制を重ねていくと、不気味の谷のグラフが得られる。一方で、知覚のマグネット効果の説明モデルを用いた研究については、先に見た知覚的齟齬に当てはまるモデルである。ここでは、対象が人間であると知覚する働きと、非人間であると知覚する働きとの二つの知覚が働き、それらによって生じた知覚のひずみが不気味さを生起させると考えられている。具体的には、刺激に対して人間であると知覚(ターゲット知覚)する確率分布と、人間でないと知覚(バックグラウンド知覚)する確率分布を事前確率分布として設け、知覚された刺激が人間・非人間のうち、いずれのカテゴリに帰属されるかをベイズ推定の枠組みで計算すると、不気味の谷のグラフの形状が得られるというものである。

ここで重要な点は、Shimadaらは、不気味さがどのようなプロセスで表出するか、に関しても言及している点である。彼らは感情の二重経路 [LeDoux 96] に着目し、処理が速い系と処理が遅い系との二つの情報処理結果における齟齬によって、不気味さが形成されたと考えた。速い系では、皮質の高次認知を介さず、皮質下の迅速な(粗雑な)処理によって刺激が「人間」とであると評価される。これに対して遅い系では、刺激に対して皮質を介した(精密な)情報処理が行われ、「非人間」とであると評価される。この二つの結果における齟齬がシグナルとなり、不気味さが形成されるというものである。

#### 2.5 先行研究から導かれる仮説モデル

先行研究から、ヒト型エージェントに対する不気味さ

を形成する情報処理として、二つの情報処理が関与していることが示唆される。一つ目は、迅速で自動的に、知覚対象を「人間」に帰属させる情報処理である。注意されるべきは、この処理において、知覚対象は「人間」であることを言語的に認識されているわけではないという点である。むしろ、これまでの学習経験に基づいて、顔の大まかな形態的パターンから、人間の顔を見た場合と同様の反応を返すことを意味している。この迅速な情報処理は、観察者に「知覚対象は『人間』である」という予測を構築し、これにやや遅れる情報処理に制約をかけているとも考えられる。ここで顔の学習経験による専門化 (Expertise) という点についてはLewkowiczら [Lewkowicz 12] も言及している。彼らは発達仮説 (development hypothesis) として、早期の発達段階での学習経験の副産物として不気味の谷が形成されていることも十分考えられると述べている [Lewkowicz 12]。もう一つ目は、一つ目にやや遅れて、刺激を「非人間」に帰属させる情報処理である。ただし、この情報処理は必ずしも刺激を「非人間」に帰属させるわけではない。特に、不気味に感じられるヒト型エージェントの場合は、「非人間」に帰属されていることを表す。以上のように考えると、情報処理のモデルを図2のように与えることができる。

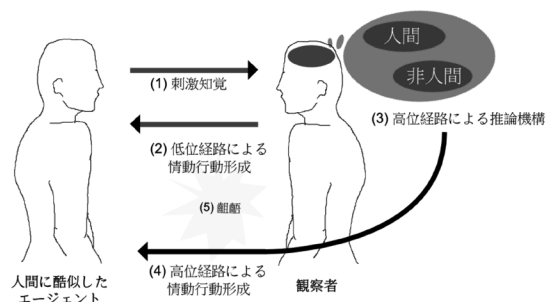


図2 否定的情動反応形成モデル

### 3. 脳機能モデル概説

本章では、これまで我々が構築してきたヒト型エージェントに対する否定的感情形成を表す脳機能モデル [田和辻 15] に関して概説する。2.5節で述べた仮説モデルがどのように脳の機能的結合の中で達成されているかを明らかにすることを目的としている。

#### 3.1 本研究のアプローチと意義

我々は、ヒト型エージェントの顔を知覚するプロセスにおいてどのようにして否定的な反応が表出するかという点について説明を与える脳機能モデルを構築することを目的としている。これまで、脳機能、つまり脳部位の機能的接続に着目してモデル化を行ってきた。言い換えれば、一つの機能を有する脳の系を機能単位とし、それ

らの結合による脳の機能ネットワークを考えた。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、という作用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った。以下、人間の情報処理モデルを構築するにあたって、どのような立場でモデルを構築するかについて述べる。

脳神経科学的研究は、着目する対象の微視性（巨視性）に関して多段階に分けることができる。このような中で、感情を理解するためには、どのレベルで脳機能をモデル化する必要があるかを十分検討しなければならない。まず、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の仕組み（感情に関係する部位の細胞の構造や機能）を明らかにするという研究立場がある。この立場はさらに細かく段階に分類され、脳機能の中でも最小単位と考えられる細胞内のイオン交換を対象とした研究、そのイオン交換によって達成される膜電位の電位変化を対象とした研究、あるいはこれらを巨視的な視点で捉えて、現象（もしくは現象を実現する機能）をモデル化する計算論的な研究である。これらの立場の研究は、神経系を理解するうえで極めて重要な立場である。ところが、複数の脳部位によって達成されると考えられる感情を、これらの立場のレベルから捉えることは極めて難しい。これは記述粒度がミクロになればなるほど、それらの組合せによって達成されるマクロ現象の振舞いが予測困難であること、あるいは立式ができては原理的に方程式が解けないという課題があるためである。

このような粒度の細かい対象を扱う立場からやや巨視的な対象を扱う立場として、脳の部位に着目し、それらの部位がどのような解剖学的構造を有しているか、さらには、そのうえで実現すると考えられる機能を数学的なモデルで記述するという立場が存在する。ここでの機能記述は、主に機械学習におけるニューラルネットワークをベースとした微分方程式を用いて記述されることが多い。実験的な方法論としては、ある認知課題遂行時や感情誘発時に賦活している部位を特定しようとするものがある。このような立場はそれぞれが互いを補い合い、一つの部位（あるいは二つの脳部位の機能的結合）における機能をモデル化するうえで重要な立場であるといえる。

他方で、これらよりもさらに巨視的な観点から人間の情報処理機能を捉える立場がある。言い換えれば、脳部位の結合状態は別として、トップダウンに抽象的な機能をモデルとして与えることでモデル上での振舞いを観察する立場である。

我々はこれらの立場のうち、部位や部位間の機能的結合に着目することで、神経科学的な妥当性は担保しつつ、人間の情報処理モデルを構築することを目指す。このとき、各機能を定量的な微分方程式で書き下すことはせず、入力-出力の関係の枠組みで機能を捉え、それらの機能

の作用機序として、トップダウン的に情報処理プロセスを捉えることを考える。この方針によって、数値計算による解析の困難さを避けることができると考えられる。以上のような巨視的観点からのトップダウンアプローチは、微視的観点からボトムアップ的に感情を理解するアプローチと組み合わせる必要がある [Chi 16]、この二つの並行的アプローチによって、より感情理解の深化が進むと考えられる。

### 3.2 視線実験

仮説モデルから、ヒト型エージェントを見ている場合と人間を見ている場合で、行動に違いが現れると考えられる。そこで、人間がヒト型エージェントと人間を観察する際、どのような知覚過程が伴うのかを実験的に検討する。具体的には、明らかに人間であると判断されるエージェントに対する知覚処理と、明らかに人間ではないと判断されるエージェントに対する知覚処理においてどのような差が見られるかを、エージェントの人間・非人間判断中の被験者の視線情報を計測し、時系列的に分析することで明らかにする。

PCに刺激画像を提示し、それが人間か非人間のいずれであるかを判断する課題を被験者に与えた。刺激として、(a) 人形、(b) CG 描写された男性、(c) 女性の人型アンドロイド、(d) CG で作成された女性、(e) 人間の男性である（(a) から (e) に向かうにつれて見掛けの人間に対する類似度が上昇するように主観的に選択した）。分析の対象となった被験者は、CG 画像で作成された女性および人間を「人間」、CG で描写された男性を「非人間」と判断していた。次に、視線遷移の結果について述べる。人間を見ている場合 [Yarbus 67] と同様、CG 描写された顔を見ている場合でも視線は目や鼻、口に集中することが定性的に読み取れた。これは顔に対する特有の遷移パターンであり、刺激が人間であってもヒト型エージェントであっても誘発し得るものであると考えられる。

一方、右目に対する視線停留時間に着目すると、ほかの顔特徴部位と比較して刺激間で特徴的な変化が確認された。具体的には、CG 画像で作成された女性の右目に

Eye Fixation Duration on the Right Eye

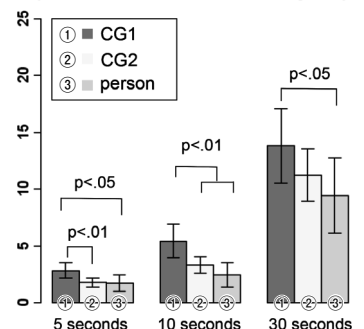


図3 観測開始から5秒、10秒、30秒間の被験者の右目に対する視線停留時間 [秒] の平均

対する視線停留時間は、すべての時間長に対して人間の右目に対する視線停留時間よりも長くなる傾向が認められた(図3)。また、CG画像で作成された男性の右目に対する視線停留時間は、観測時間が経過するに従って、人間の右目に対する視線停留時間と比較して定性的に長くなることが示唆された。ここから、間違いなく人間でないと判断された刺激は、間違いなく人間であると判断された刺激に比べて、右目に対する視線停留時間が長くなる傾向があることが考えられる。このような行動上の違いが現れるのは、高次の情報処理プロセスで人型エージェントと判断された場合と、人間と判断された場合で違いがあることに起因すると考えられる。

### 3.3 脳機能でどのように達成されているか

それでは、仮説モデルは脳機能の組合せとしてどのように達成されているかについて考える。まず、Shimadaらの提案[Shimada 07]に基づき、感情の二重経路に関わる扁桃体を中心として、視床や大脳皮質が関与していると考えた。また、視線実験で得られた行動上の違いを表現するにあたって、大脳基底核における線条体が重要な働きをしていると考えた。モデルの詳細は[田和辻 15]を参照されたい。

#### §1 モデル1

本モデルで取り上げた脳部位は、視床、扁桃体、大脳皮質、海馬、線条体である。そこでまず、これらの脳部位の各機能を脳科学的知見をもとに抽出し、それらの神経生理学的結合関係から機能の結合を試みる。次に、シミュレータを用いてこの機能的結合関係を定性的に記述する。最後に、シミュレータ上で構築されたモデルに対して入力情報を与え、得られた出力から不気味さがどのように表現されるかを考える。ただし、シミュレーションにおける重要な点として、本研究ではその具体的な定量値に着目せず、境界標(0やしきい値)との大小関係のみを扱う。

視床は感覚器官から得られた情報の中継地点であり、得られた情報を扁桃体および感覚皮質へと伝達する[LeDoux 96]。次に、扁桃体は得られた刺激の情動評価を行い、刺激が個体にとって価値があるか(快)、価値がないか(不快)を判断する[小野 12]。さらに、扁桃体は、顔、特に表情に反応すること[小野 12]、また同じ顔の繰り返し呈示によってその活動度が抑制されること[西条 05]が知られている。また、線条体を含めた大脳基底核は強化学習との関連性が指摘されている[Wichmann 14]ことから、期待報酬量と報酬の差分だけ行動を強化するものと考えた。

このようにして構築された情報処理モデルに対して、定性シミュレータSTELLA<sup>\*1</sup>を用いて表現した。このうえで、入力(強膜の広さと目の大きさ)に対する情動

状態(ポジティブな評価あるいはネガティブな評価の二値)を評価し、時間経過によって観察対象に対する評価値が時系列的にどのように変化するかを、シミュレーションを行うことで分析した。

シミュレーションによって得られた評価および脳部位の活動の時間的遷移に関して述べる。まず、人間の目を見ている場合であれば、ある時刻以降は一定の評価値(期待評価値)に収束および扁桃体の活動も非活動状態(活動度=0)となることが示唆された。一方で、人間の目と異なる構造のもの(形態的に大きな目など)を見ている場合は、評価値は0へと収束し、扁桃体の活動状態も不安定になる(活動度が振動する)ことが示唆された。グラフの結果から、人間の目を観察している際は、扁桃体はある時刻まで活動が続けられるが、一定時刻を過ぎると非活動状態となることが示された。これに対して、人間の目と異なる構造をもつエージェントを観察している場合は、時間が経過しても扁桃体の活動状態は持続していることが示唆された。

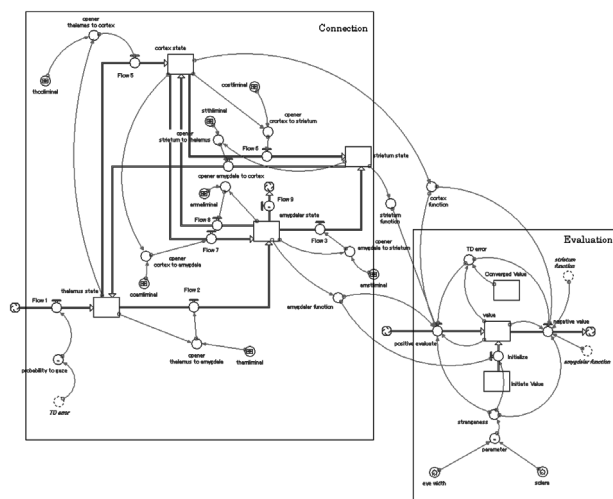


図4 STELLAにおいて構築した本研究のモデル

#### §2 モデル2

実装されたモデル1では、入力の初期値によって正の情動評価になるか負の情動評価になるかが決定的であるという課題があった。この入力はその後の系の挙動すべてに影響を与えることから、この箇所を適切な表現に改める必要がある。また、エージェントを見る際には、目以外にも鼻や口などに視線が移ることが考えられるため、それらの視線移動を考慮に入れた実装を行った。そこで、実装面では図5のような情報の流れを想定した。まず、エージェントの顔の特徴部位から異常度が高い部分を検知し、異常度に比例して扁桃体が活動する。ただし、この扁桃体の活動は感情の二重経路による二つの処理に分けられ、時間的に早い段階における雑多な情報処理による活動と、これにやや遅れる詳細な情報処理による活動に分けられる。ここで、早い段階の処理においては、ある一定値に満たない異常度は0として扱われる。

\*1 isee systems : STELLA R. Systems Thinking for Education and Research, <http://www.iseesystems.com/software/Education/StellaSoftware.aspx>

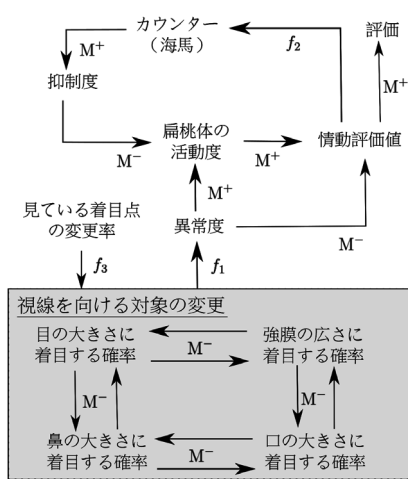


図5 各パラメータの定性的関係

この一定値は感情の二重経路における低位経路の情報処理に関する雑多さを表している。

各変数の定性的関係を、統計処理言語および処理環境であるフリーソフトウェア R<sup>\*2</sup> によって記述し、各部位に対する異常度、低位経路における情報処理の雑多さ、着目点の変更率をパラメータとして変化させ、対象に対する評価値がどのように変化するかに関してシミュレーションを行った。この結果、顔の特徴部位の中に異常が多く存在していると、扁桃体は抑制されず、活動が維持されることが示唆された。一方で、人間の顔に近いエージェントを観察している場合は、人間のエージェントに対する評価が増加と減少を繰り返すことが示唆された。これは、対象となるヒト型エージェントに対して抱く不気味さの様相を表していると考えられ、不気味さとは、快情動と不快情動の混合状態であることが示唆された。

#### 4. 応用事例

本章では、前章に加え、これまで我々が取り組んできた不気味の谷とエージェントの動作、不気味の谷とマルチモーダルに関する研究についてそれぞれ述べる。

##### 4.1 不気味の谷とエージェントの動作

ヒト型エージェントの設計を考えるうえで、外見と動作が人間の印象に与える影響は大きい。石黒は、外見と動作が最適に組み合わせられることで、相乗効果が得られることを提案している [石黒 05]。このことから、外見と動作の最適な組合せを検討することは、今後のヒト型エージェントの設計において重要な指針になるといえる。萩原らは、エージェント動作の滑らかさに着目し、エージェントの外見における人間との類似度と、動作の滑らかさとの最適な組合せを検討した [萩原 16]。ヒト型

エージェントには、東京で営業の仕事をしている真面目で礼儀正しい性格の 20 代後半の男性という文脈を与えたうえで、外見の肌のテクスチャと動作（「お辞儀」動作）の滑らかさ（fps を調節したもの）をそれぞれ段階的に与えて被験者に提示した。この結果、必ずしもテクスチャの写実性が高いものと動作の滑らかさが高いものの組合せが最適であるとは限らないことが示唆された。

##### 4.2 不気味の谷とマルチモーダル

ヒト型エージェントが今後社会的に活用される状況を考えると身体的表現に加え、音声による表現表出を行うことが考えられる。これにあたり、単一モーダルの場合と複数モーダルで表現を行うヒト型エージェントでは、どのような違いが見られるかを検討することは重要である。Tanaka らは、顔を光学的妥当性と形態的妥当性の二軸で段階付けたヒト型エージェントに対して、人間の音声と合成音声とをそれぞれ付与し、天気予報の内容を読み上げたものを被験者に聞かせた [Tanaka 11]。まず、音声の伴わない顔のみの提示であれば、光学的妥当性と形態学的妥当性が高いエージェントが最も人間らしさ（similarity）が高いと評価された。ところが、音声を付与した場合、必ずしも光学的妥当性と形態的妥当性がともに高いエージェントが人間らしいと評価されることは限らないことが示唆された。Mitchell ら [Mitchell 11] によると、人間の音声と合成音声を、人間とロボットの映像にそれぞれ合わせて提示したところ、それぞれ不一致条件（人間音声-ロボット映像、合成音声-人間映像）のときに不気味さが高くなることが示唆されている。これらのことから、単一モーダル（e.g. 外見）においては不気味さが表出されなくとも、異なるモーダルを付与することで、（特にそれらが不一致を起こす場合において）不気味さが表出する可能性が推察される。

#### 5. Open Question

最後に不気味の谷現象を考えるうえで、既往研究が着手できていない課題について述べる。この課題は大きく三点ある。一点目は、「不気味の谷」現象のグラフを考えるうえで、描写されたグラフがインタラクションのどの瞬間を切り出して構築されたグラフであるかを明らかにすることである。二点目は、不気味さとは何かについて明確な定義を与えることである。三点目は、ヒト型エージェントがどのような役割で用いられているかに関する状況設定を明確にし、そのうえで不気味さが表出されるかについて検討する必要がある。

一点目は、ヒト型エージェントと対面してからのインタラクション経過時間を考慮することを表している。ここで、経過時間は具体的に、長期経過と短期経過の二つある。まず、長期経過を踏まえたインタラクションに関して、Zlotowski らは、エージェントとの接触頻度とエー

\*2 The R Project for Statistical Computing : <http://www.r-project.org/index.html>

エージェントに対する好感度、不気味さがどのように変化するかを実験的に検討している。エージェントとインタラクションを数回行うと、エージェントの外見に依存せず、エージェントに対する好感度(Likability)が変化すること、また、その変化の方向は、エージェントの態度(ポジティブな態度をとったかネガティブな態度をとったか)に依存することが示唆された[Zlotowski 15]。このため、数回の接触頻度を視野に入れた、言い換えれば長期経過を踏まえたインタラクションを考える場合は、エージェントの態度がエージェントに対する評価を決定するうえで重要な要素になる可能性がある。一方で、エージェントとの短期的なインタラクションにおいては、どの瞬間において否定的な感情が表出されるかも検討することが必要である。具体的には、一目見た瞬間的に否定的な感情が表出されるのか、それとも、これよりも長く、1回あるいは数回程度のやり取りがあるうえで、否定的な感情が表出されるのかを考える必要がある。これは、どの瞬間の否定的な感情を不気味さと呼ぶか、という二点目の課題にも関わり、重要である。したがって、インタラクションにおける経過時間とそのときの感情状態の変化の関係性を十分に検討する必要があると考えられる。

二点目は、不気味の谷現象を考えるうえで急務な課題であるといえる。なぜならば、この不気味さとは何かについて、既往研究で統一的な定義が与えられていないという現状があるためである。この課題は大きく次の二つの問いに答えることに当たる。すなわち、(1) 不気味さとはどのような感情か、(2) 不気味さとはどのようにして形成されているのか、についてである。Hoらは、ロボットに対する不気味さが、恐れを主要因として、嫌悪感、動揺、不安が関与していることを実験的に明らかにしている[Ho 08]。要するに、不気味さは単一の感情から説明されるというわけではなく、複雑な複合感情であることが推察される。そこで、不気味さがどのような神経基盤によって形成されるのか、というアプローチは、形成過程を明確にするという点において、今後重要な位置を占めると考えられる。Urgenらは、期待誤差と不気味の谷の関係を脳波を用いた実験を通して調査した[Urgen 15]。この結果、期待誤差に関与する信号N400の発生源は、中側頭・上側頭(middle and superior temporal area)、側頭頭頂接合部(temporal parietal junction)、前頭前野であることが示唆された。したがって、今後は信号の発生源における部位の機能を踏まえた検討が必要である。

三点目は、実応用志向の立場からの課題である。この課題は、ヒト型エージェントが使用される状況によって不気味さが増加あるいは低減されるのか、また、状況に依存した不気味さが存在するかについて検討することである。今後、ヒト型エージェントやアンドロイドは、さまざまな場面で使用されることが期待される。例えば、医療現場や介護現場、娯楽施設に加え、教育現場への導

入が考えられる。一方で、既往研究の実験では、ヒト型エージェントを提示するのみで、固定された状況の中(使用文脈を与えること)で提示する試みは余りなされてこなかった。このような特定の場面においては、ある特定の文脈のもとで、導入されたヒト型エージェントに対して感情が表出される。例えば、教育場面において用いられるペダゴジカルエージェントは、生徒の学習における心的状態に大きな影響を与える。設計者の意図として、生徒の学習意欲を高めようと笑顔を提示したとしても、かえってその笑顔動作生成が非人間的で、生徒によって否定的に受け取られる可能性がある。したがって、ある特定の文脈において、ヒト型エージェントの外見や行動が、観察者にどのような影響を与えるかは実応用志向を考えるうえで、今後重要な課題になると推察される。

## 6. まとめと今後の課題・展望

本概説では、不気味の谷に係るさまざまな先行研究を俯瞰的にまとめた。この中で、不気味の谷現象の生起プロセスを理解するにあたり、知覚的齟齬・カテゴリー曖昧性と、期待誤差の説明パラダイムが重要であることについて述べた。そして、この説明パラダイムのうえで考えられる情報処理モデルが脳機能としてどのように表現されるか、という点について、我々がこれまで取り組んできた脳機能モデルの概説を行った。これに加えて、静止した写真などに対する否定的感情だけでなく、動作を加えたエージェントを用いた研究や、音声情報を加えたマルチモーダルと否定的感情の関係性に関する研究についても簡単に紹介を行った。最後に不気味の谷において検討されるべき課題として、Open Questionという形で課題の概略を三点まとめた。

不気味の谷を検討することで、不気味さという感情とも考えられる心的状態が、どのようにして生起しているのかという情報処理プロセスの解明が今後期待される。特に、心的状態の表出過程が脳機能の作用機序においてどのように達成されているか、という本アプローチは、そのほかのさまざまな感情を理解するうえでも重要な役割を果たすと期待される。このような脳機能からのアプローチは、さまざまな神経科的科学的知見に加え、実験を通じた心理学的知見、さらに、それらを体系的かつ包括的に説明する計算論的アプローチが融合されることで達成される。脳機能から心的状態の変化過程を検討する試みは、このように人間の感情理解を学際的になすうえで大きな役割を果たすといえる。また、冒頭でも述べたとおり、我々は不気味の谷はそれ自体が「ヒトがいかにして同種を他種と弁別しているか」という機能の根源的な表れであると考えている。ヒトとは何か、に関する哲学的な問いに答えるにあたり、不気味の谷が果たす役割は大きいと考えられる。

## ◇ 参考文献 ◇

- [Chaminade 07] Chaminade, T., Hodgins, J. and Kawato, M.: Anthropomorphism influences perception of computer-animated characters' actions, *Social Cognitive and Affective Neuroscience*, Vol. 2, No. 3, pp. 206-213 (2007)
- [Cheetham 13] Cheetham, M., Pavlovic, I., Jordan, N., Suter, P. and Jancke, L.: Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data, *Frontiers in Psychology*, Vol. 4, No. 108, pp. 1-12 (2013)
- [Chi 16] Chi, K. R.: Neural modelling: Abstractions of the mind, *Nature*, Vol. 531, pp. S16-S17 (2016)
- [萩原 16] 萩原 愛, 田和辻可昌, 村松慶一, 松居辰則: 動作の滑らかさに着目した擬人化エージェントの最適な外見と動作に関する実験的検討, 第11回日本感性工学会春季大会, G3-5 (2016)
- [Hanson 05] Hanson, D., Olney, A., Pereira, I. A. and Zielke, M.: *Upending the uncanny valley*, *Proc. National Conf. on Artificial Intelligence*, Vol. 20, No. 4, Menlo Park, CA; Cambridge, MA; London; AAAI Press (2005)
- [Ho 08] Ho, C. C., MacDorman, K. F. and Pramono, Z. D.: Human emotion and the uncanny valley: A GLM, MDS, and isomap analysis of robot video ratings, *Proc. 3rd ACM/IEEE Int. Conf. on Human Robot Interaction*, pp.169-176 (2008)
- [石黒 05] 石黒 浩: アンドロイドサイエンス, システム制御情報学会誌, Vol. 49, No. 2, pp. 47-52 (2005)
- [Kätysri 15] Kätysri, J., Förger, K., Mäkäräinen, M. and Takala, T.: A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness, *Frontiers in Psychology*, Vol. 6, pp. 390-406 (2015)
- [LeDoux 96] LeDoux, J.: *The Emotional Brain - The Mysterious Underpinnings of Emotional Life*, Simon & Schuster Paperbacks (1996), 松本 元, 川村光毅 ほか 訳: エモーション・ブレイン—情動の脳科学, 東京大学出版会 (2003)
- [Lewkowicz 12] Lewkowicz, D. J. and Ghazanfar, A. A.: The development of the uncanny valley in infants, *Developmental Psychobiology*, Vol. 54, No. 2, pp. 124-132 (2012)
- [MacDorman 09] MacDorman, K. F., Green, R. D., Ho, C. C. and Koch, C. T.: Too real for comfort? Uncanny responses to computer generated faces, *Computers in Human Behavior*, Vol. 25, No. 3, pp. 695-710 (2009)
- [Matsuda 12] Matsuda, Y., Okamoto, Y., Ida, M., Okanoya, K. and Myowa-Yamakoshi, M.: Infants prefer the faces of strangers or mothers to morphed faces: An uncanny valley between social novelty and familiarity, *Biology Lett.*, Vol. 8, No. 5, pp. 725-728 (2012)
- [Mitchell 11] Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermehorn, P. W., Scheutz, M. and MacDorman, K. F.: A mismatch in the human realism of face and voice produces an uncanny valley, *i-Perception*, Vol. 2, pp. 10-12 (2011)
- [Moore 12] Moore, R. K.: A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena, *Scientific Reports*, Vol. 864, No. 2 pp. 1-5 (2012)
- [森 70] 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)
- [西条 05] 西条寿夫, 堀悦郎, 田積徹, 小野武年: 表情認知における扁桃体の神経機構, 日本薬理学雑誌, Vol. 125, No. 2, pp. 68-70 (2005)
- [小野 12] 小野武年: 脳科学ライブラリー 3 脳と情動—ニューロンから行動まで—, 朝倉書店 (2012)
- [Saygin 12] Saygin, A. P., Chaminade, T., Ishiguro, H. Driver, J. and Frith, C.: The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions, *Social Cognitive and Affective Neuroscience*, Vol. 7, No. 4, pp. 413-422 (2012)
- [Schein 15] Schein, C. and Gray, K.: The eyes are the window to the uncanny valley: Mind perception, autism and missing souls, *Interaction Studies*, Vol. 16, No. 2, pp. 173-179 (2015)
- [Seyama 07] Seyama, J. and Negayama, R. S.: The uncanny valley: Effect of realism on the impression of artificial human faces, *Presence: Teleoperators and Virtual Environments*, Vol. 16, No. 4, pp. 337-351 (2007)
- [Shimada 07] Shimada, M., Minato, T., Itakura, S. and Ishiguro, H.: Uncanny valley of androids and its lateral inhibition hypothesis, *Proc. IEEE Int. Symp. on Robot and Human Interactive Communication*, pp. 374-379 (2007)
- [Steckenfinger 09] Steckenfinger, S. A. and Ghazanfar, A. A.: Monkey visual behavior falls into the uncanny valley, *Proc. National Academy of Sciences*, Vol. 106, No. 43, pp. 18362-18366 (2009)
- [Tanaka 11] Tanaka, K., Matsui, T. and Kojima, K.: Experimental study on appropriate reality of agents as a multi-modal interface for human-computer interaction, *Int. Conf. on Human-Computer Interaction*, pp.613-622, Springer Berlin Heidelberg (2011)
- [田和辻 15] 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)
- [Tinwell 11] Tinwell, A., Grimshaw, M., Nabi, D. A. and Williams, A.: Facial expression of emotion and perception of the uncanny valley in virtual characters, *Computers in Human Behavior*, Vol. 27, pp. 741-749 (2011)
- [Urgen 15] Urgen, B. A., Li, A. X., Berka, C., Kutas, M., Ishiguro, H. and Saygin, A. P.: Predictive coding and the uncanny valley hypothesis: Evidence from electrical brain activity, *Proc. Workshop: Cognition: A Bridge between Robotics and Interaction in 10th Int. Conf. on Human Robnot Interaction*, pp. 15-21 (2015)
- [Wichmann 14] Wichmann, T. and DeLong, R. M.: 大脳基底核 (南部 篤 訳), In Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A. and Hudspeth, A. J. (Eds.), *カンデル神経科学, Principles of Neural Science, Fifth Edition*, (日本語版第一版), 金澤一郎, 宮下保司 (日本語版監修), pp. 963-978, メディカル・サイエンス・インターナショナル (2014)
- [Yamada 12] Yamada, Y., Kawabe, T. and Ihaya, K.: Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon, *Japanese Psychological Research*, Vol. 55, No. 1, pp. 20-32 (2012)
- [Yarbus 67] Yarbus, A. L.: Eye movements during perception of complex objects, In Riggs, L. A. (Trans., ed.), *Eye Movements and Vision* (Translated by Haigh, B.), pp. 171-211, Springer, US (1967)
- [Zlotowski 15] Zlotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C. and Ishiguro, H.: Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception, *Frontiers in Psychology*, Vol. 6, pp. 883-895 (2015)

2016年8月2日 受理

## 著者紹介



田和辻 可昌 (学生会員)

2013年早稲田大学大学院人間科学研究科修士課程修了。現在、同研究科博士後期課程に在学中。関心は不気味の谷を通した人間の他者認知を実現する機構の記述など。日本感性工学会、認知科学会各会員。



松居 辰則 (正会員)

1994年早稲田大学大学院理工学研究科博士後期課程修了。博士(理学)。早稲田大学理工学部助手、東京芸文大学教育学部助手、電気通信大学大学院情報システム学研究科助教授を経て、現在、早稲田大学人間科学学術院教授。専門分野は、感性情報処理、知識科学、数理統計学、e-learningなど。IEEE、ACM、電子情報通信学会、日本感性工学会、教育システム情報学会、日本認知科学会各会員。