1E3-1

脳機能の定性的記述を用いた人型エージェントに対する情動評価の時系列的変化モデルの提案

Proposal of Generation Model for Time Series Transition of Emotional Value with a Humanlike Agent Using Qualitative Description Based on Brain Function

田和辻 可昌 *1　村松 慶一 *2　松居 辰則 *2
TAWATSUJI Yoshimasa　MURAMATSU Keiichi　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

In the research field of human agent interaction, it is a critical issue that human can feel repulsive against an agent when it looks considerably humanlike, as the unnncay valley. We hypothesized that when human observes a humanlike agent, the observer can perceive it as both human and non-human, and that the contradiction between the two kinds of perception causes negative response toward it. In the experiment, the participants were asked to judge whether faces of humanlike agents or a person was human or not with their eye tracked and their gaze direction estimated. The results indicated that observers had two-steps information processing to the agent, and we proposed a model providing an explanation for how the human negative response emerges with the concept of the dual pathway of emotion. In addition, we proposed the advanced model with the functions of the hippocampus and the striatum added. To verify the model, the transition of emotional value was simulated using the qualitative description for the model.

1. はじめに

ヒューマンエージェントインタラクションの分野において，不気味の谷は重要な課題である．図 1 に示すように，ロボットやコンピュータエージェント（以下，合わせてエージェントと呼ぶ）の外見が人間に近づくにつれ人間のエージェントに対する親和度は上昇するが，人間との類似度がかなり高くなったある地点において親和度が急激に下落することがあると考えられており，これが不気味の谷である [Mori 70].

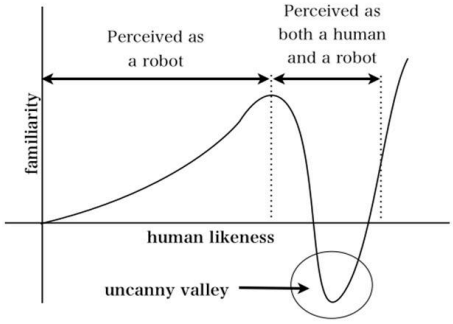

図 1: 不気味の谷の概念図 ([Mori 70] を参考に作成)

不気味の谷は人間に酷似したエージェントに対する否定的な反応としてとらえられ，様々な研究が進められてきたが，どのようなメカニズムによってこの否定的反応が形成されるかに関しては統一的な見解が得られていない．そこで本研究では，人間がエージェントを観察した際に生じる否定的反応がどのようなメカニズムで形成されるかを統一的に説明する基盤モデルの構築を目指す.

連絡先: 田和辻 可昌，早稲田大学 大学院人間科学研究科，埼玉県所沢市三ヶ島 2-579-15，090-5880-0631，watskkoreverfay@akane.waseda.jp

2. 仮説

Noma らによると，人間に酷似したロボットを 1 秒間観察するにあたり，人間が呼吸する際に生じる微小な胸の動きやまばたきをロボットに与えると，動きの伴わない静止したロボットを観察する場合と比較して，大多数がロボットを人間であると認識することを実験的に明らかにした [Noma et al. 06]．また Minato らは，人間が人間に酷似したロボットと会話を行うときは，人間と会話する場合と比較して有意に会話相手の右目に対する視線停留頻度が多いことを明らかにした [Minato et al. 04]. これらのことから，人間は人型エージェントに対して，人間を観察している場合と同等の処理と，人間でないものを観察している場合の処理を行っていることが示唆される．そこで，本研究では，人間に酷似したエージェントを観察している際に人間はエージェントを人間／非人間の両観点から知覚しており，この二つの知覚による情報間の齟齬が否定的反応を生起させると考えた.

3. 仮説検証実験

人間に似たエージェントを人間はどのように知覚しているのかを明らかにするため，図 2 の 5 種類の「顔」画像を観察中の被験者の視線を非接触型視線計測器 EMR-AT VOXER を用いて計測し，人間と判断した場合と非人間と判断した場合でどのように視線の性質に違いがあるかを検証した.

図 2: 実験で用いた「顔」画像

これらエージェントのうち，CG1 および人間は人間の顔画像，CG2 は人間でない「顔」画像と判断され，人間／非人間判断の困難さにおいては，CG1 の画像は他の画像に比べて判断が難しかったという回答が得られた．このため，分析では女性の CG 画像 (CG1)，男性の CG 画像 (CG2)，人間の画像 (人間) を用いた．図 3 に計測開始から 5 秒間におけるある被験者の視線の動きと，各エージェントの右目に対する視線停留時間の時系列的変化を示す．各画像に対する被験者の視線停留時間に関して，CG1 に対しては開始 5 秒の段階から他の画像に比べて有意に右目に対する視線停留時間が長かった (CG1 と CG2：$p < .0053$，CG1 と人間 $p < .00433$)．これは判断の困難さが視線停留時間に影響を与えたと考えられる．一方で，CG2 の右目に対する視線停留時間は，観察における初めの 5 秒間においては人間のものと有意な差は認められなかったが，観察時間が経過するにしたがって人間のものと比較して有意に長くなることが認められた．これより，「人間に似た顔画像を見ている場合は人間の顔画像と同等の処理を行った後に人間と異なる情報処理を行う」ことが示唆された．

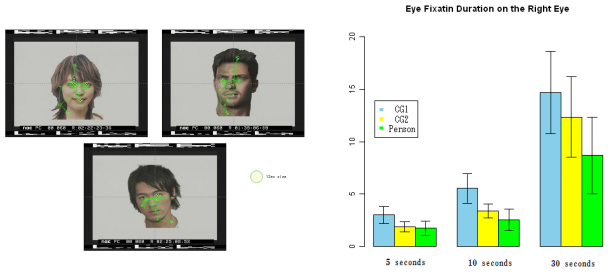

図 3: 計測開始から 5 秒間の視線停留時間の変化（左）と右目に対する視線停留時間の時系列的変化（右）

4. 不気味の谷発生メカニズムモデル

この節では前節の実験結果を踏まえて脳科学的見地からモデルを構築する．さらに，先行研究のモデルと統合することでより汎用的なモデルへと発展させることを考える．その後，実験の結果にあたる人間観察時と比較して，エージェント観察時に右目に対する視線停留時間が増加する現象に関しての説明を行う．

4.1 感情の二重経路に着目したモデル

森は，不気味の谷を人間の対象に対する本能的な自己防衛反応として位置付けている [Mori 70]．マカクザルを用いた実験から，サルであっても不気味の谷で考えられているような否定的な反応が形成されることが確認されている [Steckenfinger et al. 09]．このことから否定的反応が形成されるメカニズムを考える上では，種に共通したシステムである脳，またその中でも系統発生学的に古い部位に着目することは重要であると考えられる．そこで，刺激に対する情動反応を形成する比較的に古い部位である大脳辺縁系に属する扁桃体に着目した．ここで，情動とは動物が刺激に対する本能的な評価として位置付けられるものとして定義する [小野 12]．LeDoux によると，人間の情動情報処理は，迅速だが雑多な情報処理からなる低位経路 (皮質下経路) と，これにやや遅れる形で詳細な情報処理を行う高位経路 (皮質経路) から成り立っている [LeDoux 96]．

前節の二段階の知覚プロセスは，この感情の二重経路から低位経路によっておいて人間を知覚している場合と同等の情動処理および情動行動形成がなされ，高位経路によって人間ではないということに対する情動処理および情動行動形成がなされると考えられる．ここで，情動行動は目に対して視線を向けるということであると考える．本研究ではこの低位経路と高位経路の間の情動情報処理の齟齬が否定的反応を形成すると考えた．このモデルを図 4 に示す．

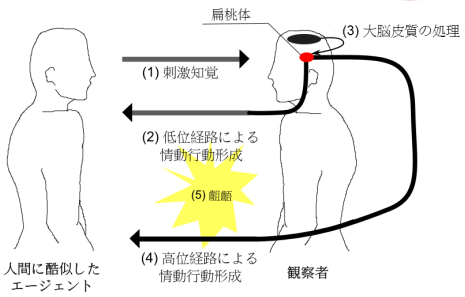

図 4: 感情の二重経路を用いた否定的情動反応形成モデル

4.2 先行研究モデルとの統合

Moore は不気味の谷の発生メカニズムに関して知覚のマグネット効果によるモデルの適用可能性を示した [Moore 12]．このモデル (以下，Moore モデル) ではカテゴリ知覚の観点から，人間が誤ってエージェントを人間と知覚することによって知覚の齟齬が発生し，不気味の谷が発生すると解釈している．しかし，本研究のモデルと Moore モデルとの統合を行うことでより詳細なモデルを構築するにあたり，両モデルの概念粒度が異なるため，Moore モデルを脳機能の観点から再解釈することを考える．

Guenther らは，自己組織化マップを用いた視床-大脳皮質のモデルを構築し，知覚のマグネット効果を説明するモデルを構築した [Guenther et al. 02]．このことから，知覚のマグネット効果のモデルを転用した Moore モデルは，視床-大脳皮質における情報処理モデルとして解釈できる．そこで，Moore モデルによって観察したエージェントが人間であるか非人間であるかの判断が大脳皮質と海馬における記憶との照合がなされ，知覚の齟齬によって生じた否定的反応が扁桃体において負の情動反応として生起するとして，両モデルの統合的説明を試みた．図 5 に統合モデルを示す．

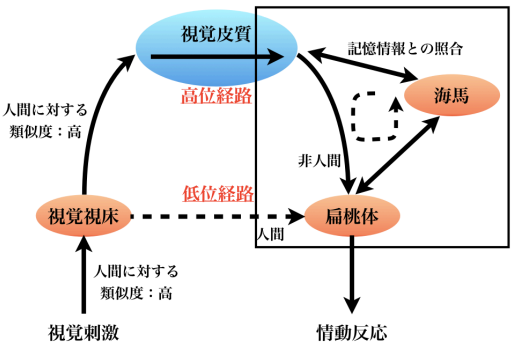

図 5: 不気味の谷発生メカニズムモデル

4.3 強化学習と視線行動形成モデル

得られた刺激を運動へと変換する系として大脳基底核は重要な役割を担っており，この中でも辺縁系と大脳基底核の間での連絡は，辺縁系から得られた情動情報を行動へと変換すると考えられている [小野 12]．鮫島らは大脳基底核が強化学習の性質を有していると考え，線条体を中心とした大脳基底核の強化学習モデルを提案している [鮫島ら 01]．強化学習は得られた報酬と期待報酬値の差分だけ行動を強化する枠組みである．ここで，報酬は扁桃体の情動値（快／不快），期待報酬値は大脳皮質において計算されると考え，情動値を観察対象が人間であることを保証する情報を知覚した際は正，観察対象が人間であることを保証しない情報を知覚した際は負となるように定義する．また，このときの情動反応は視線を対象の目に向けるという行動によって表出されると考える．このとき，人間に酷似したエージェントであっても人間であっても対象を知覚した際，低位経路によってまず，対象から人間であることを保証する情報が知覚され，対象に対して視線を向ける．このとき，対象が人間であるといった期待報酬値が形成される．対象が人間であれば，その後知覚される情報は人間を保証する情報が知覚されるので，報酬値は期待報酬値へと近づき，視線を向けるという行動は弱められる．一方で，対象が人間に酷似したエージェントの場合は，その後知覚される情報は人間を保証しない情報であり，報酬値は期待報酬値へと近づかず，視線を向けるという行動は強化される．この結果実験結果に見られたように，時間が経過するにしたがって対象の右目に対する視線停留時間に差が生じたと考えられる．

5. シミュレーション

前節で提案したモデルに対して定性推論の手法を用いて，各脳部位の結合と，情動評価および情動行動に関する脳部位の機能を定性的記述によってモデル化した．さらに，定性シミュレータ STELLA を用いて,「顔」画像を観察中の被験者の情動状態のシミュレーションを行った．

5.1 モデルの定性表現

本モデルを記述するにあたり，(1) 脳部位間の結合に関する記述と，(2) 各脳部位の機能に関する記述を行った．前者は活動度と呼ばれる変数を各脳部位に設定し，これが閾値を超えると結合された脳部位の活動度が上昇することとした．つまり，n を本研究で用いる脳部位の数として，$i, j \in \{1, \cdots, n\}$ に対して，ある脳部位 X_i における活動度を $a(X_i)$，$\omega_{ij} (i \neq j)$ を脳部位 X_i, X_j の接続強度，θ_i を X_i の活動閾値として，

$$a(X_j) \quad = \quad \mathrm{sign}\left(\sum_{i \in \{1,\cdots,n\}} \omega_{ji} a(X_i) - \theta_j \right) \qquad (1)$$

とする．今回は，$\omega_{ij} = 0.3, \theta_i = 0.5$ で固定した．図 6 に視床と大脳皮質の接続を STELLA に実装した様子を表す．これは，視床の活動度が閾値を超えると，伝達の on/off が切り替わり，大脳皮質の活動度を上昇させるということを模式化している．

次に各脳部位の機能に関して述べる．扁桃体は，知覚した情報の情動評価値 $v \in \mathbb{Q}$（$\mathbb{Q}$ は定性値を表す）を決定し，その定性的な値に沿って情動行動を起そうとする．また，大脳皮質は対象に対する期待値 $\kappa \in \mathbb{Q}$ を計算し，海馬は扁桃体の情動評価を大脳皮質が算出した期待値へと収束させるように情動評価を行うように働くと考えた．これは対象に対する情動評価値 v

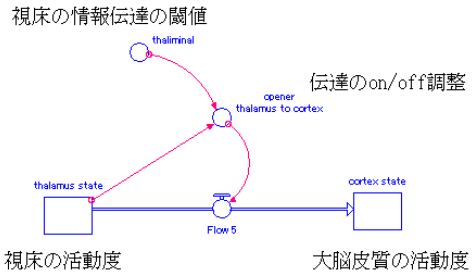

図 6: 視床 (Thalamus) と大脳皮質 (Cortex) との接続を表した模式図

は時間 t に対して，以下の力学系を用いて表した．

$$\frac{dv}{dt} \quad = \quad (\kappa - v)\, v \qquad (2)$$

大脳基底核として線条体を辺縁系の情動評価値と大脳皮質で計算された期待値との差分の大きさに比例して，情動行動を強化する系とした．これらは先の活動度が 0 以上の場合において機能するようにした．

構築したモデル全体の概要を図 7 に示す．大きく Evaluation とした評価機能の機構と，それ以外の Connection とした各脳部位の接続記述からなるモデルである．また，入力としてエージェントの目の形態的特徴（強膜の広さと目の広さ）の人間の目に対する類似度を用いた．これらの積が 0.5 を超えると情動評価がポジティブに，0.5 を下回ると情動評価がネガティブに働くように設定した．

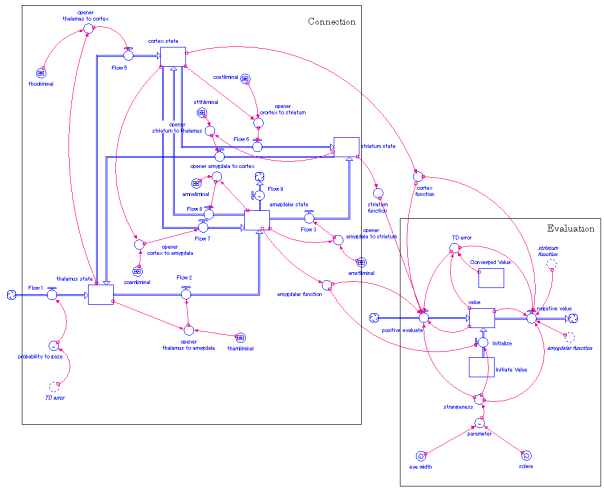

図 7: STELLA において構築した本研究のモデル

入力として目の構造が人間のそれとどの程度似ているか，また，人間の入力に対する情動状態（ポジティブな評価あるいはネガティブな評価の二値）を出力とし，時間経過によって系全体の挙動および観察対象に対する情動評価が時系列的にどのように変化するかをシミュレートした．この結果を図 8 に示す．情動評価および各脳部位の活動は，人間は人間の目を見ている場合であれば安定した情動状態および活動状態へと収束するのに対し，人間の目と異なる構造のものを見ている場合は情動状

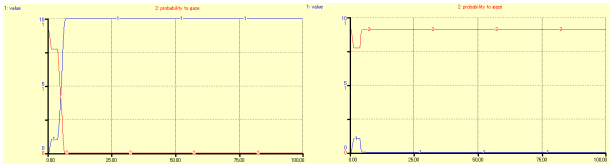

図 8: 人間の目 (左) および人間の目と異なる構造のものを観測中の情動評価と対象を見るという行為の選択確率の時系列的変化

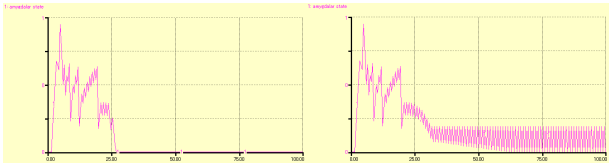

図 9: 人間の目 (左) および人間の目と異なる構造のものを観測中の扁桃体の活動度の時系列的変化

態は負の方向へと変化し，活動状態も不安定に維持されている状態が示唆された．

また，人間の目あるいは人間の目と異なる構造を持つエージェントを観察している際の，扁桃体の活動度 $a(X)$ の時系列的変化の結果を図 9 に示す．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持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，対象に対する情動評価が行われることが示された．これは，人間の目と異なる構造を持つエージェントを観察している際は，継続して対象に対して視線を向け続けることから，絶えず情報の入力が続いている状態であり，この結果，対象に対する情動評価が継続していることを示していると考えられる．

人間が人間に酷似しているエージェントを観察しているときは，人間を見ている場合や見かけが明らかに機械的なロボットを見ている場合と比較して，脳全体が広く活動していることが知られており [Saygin et al. 12]，本研究の結果はこれを説明しているという点で妥当なモデルであると考えられる．このモデルによって，人間に酷似したエージェントに対する人間の初期の情動反応，つまり，低位経路と高位経路の応答による情報処理結果の齟齬までを踏まえた短期的な情動状態の記述が可能であることが示唆された．

6. まとめ

本研究では，系統発生学的に古い脳部位に着目し，人間に酷似したエージェントに対する人間の否定的反応がどのように形成されるかに関して，定性的記述法を用いたモデルを構築した．人間が人間に酷似したエージェントを観察した場合は，まず迅速な情報処理によってエージェントを人間として知覚し，その後高次認知処理を経て非人間として知覚する．このとき，先の情報処理, に対して一貫性を持たない後続の情報は，先の情報処理の結果と一貫性を持つように処理されようとするが，それが情報の齟齬として扁桃体によって知覚される．この齟齬を低減させるために，脳システム全体が活動することで不安定な状態になり，エージェントに対する否定的反応が形成されることが示唆された．

今後の課題としては，今回構築したモデルの妥当性に関するより詳細な検証を脳科学的見地に基づいて考察する必要がある．また今回のシミュレーションでは各パラメータを固定して行ったが，それらのパラメータの値と情動評価値や系全体の振る舞いとの関係性をより深く考察していく必要がある．さらに，否定的反応が形成された後，エージェントがどのような行動をとれば，否定的な反応が解消されるかの検討を行うことも重要であると考えられる．

参考文献

[Mori 70] Mori, M.: The Uncanny Valley, K.F. MacDorman & Minato Takashi trans., *Energy*, Vol. 7, No.4, pp.33-35 (1970)

[Noma et al. 06] Noma, M., Saiwaki, N., Itakura, S., Ishiguro, H.: Composition and Evaluation of the Humanlike Motions of an Android, *Proceedings of International Conference Humanoid Robots*, pp. 163-138 (2006)

[Minato et al. 04] Minato, T., Shimada, M., Ishiguro, H., Itakura, S.: Development of an Android Robot for Studying Human-Robot Interaction, *Proceedings of IEA/AIE Conference 2004*, pp. 424-434 (2004)

[Steckenfinger et al. 09] Steckenfinger, S. A., Ghazanfar, A. A.: Monkey visual behavior falls into the uncanny valley, *Proceedings of the National Academy of Sciences*, Vol. 106, No. 43, pp. 18362-18366 (2009)

[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)

[LeDoux 96] LeDoux, J. E.: *The Emotional Brain - The Mysterious Underpinnings of Emotional Life*, Simon & Schuster Paperbacks (1996)

[Moore 12] Moore, R. K. : A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena, *Scientific Reports*, Vol. 2, No. 864, pp. 1-5 (2012)

[Guenther et al. 02] Guenther, F. H., Bohland, J. W.: 音カテゴリの学習 - ニューラルモデルとそれを支持する実験結果-, 日本音響学会誌, Vol.58, No. 7, pp. 441-449 (2002)

[鮫島ら 01] 鮫島 和行, 銅谷 賢治: 強化学習と大脳基底核 (<特集>運動学習) , バイオメカニズム学会誌, Vol. 25, No.4, pp. 167-171 (2001)

[Saygin et al. 12] 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, *Social cognitive and affective neuroscience*, Vol. 7, No. 4, pp. 413-422 (2012)