博士論文要旨

ヒト型エージェントに対する否定的感情表出過程を表現する定性的脳機能モデルの構築

Qualitative Brain Function Model Explaining the Mechanism of Negative Emotion towards Humanlike Agent

田和辻　可昌（Yoshimasa Tawatsuji）　　指導：松居　辰則

1．はじめに

技術の進展に伴い，人間とエージェント（e.g.ロボットやアバター）とがインタラクションをとる機会が増えてきている．このとき，エージェントの振る舞いの意図を人間にとって可読性が高くなるように設計することは重要な課題である．人間の擬人化作用を利用し，エージェントの外的表現を人間に近づけることが試みられている一方，「不気味の谷」［1］と呼ばれる現象は重要な課題である．本現象はヒト型エージェントに対する単なる否定的感情という現象のみならず，人間が他者をどのように認識するかを検討するうえで重要な現象であり，心理学や神経科学，情報科学を含む多領域にわたって本現象の実態や形成メカニズムを説明する試みがなされてきた．この中で，「予測誤差」は不気味の谷の形成において重要な役割を果たすことが示唆されている［2］．ところが，どのような一連の情報処理過程によって不気味の谷が形成されているのかについては統一的な理解はなされていない．本研究では，「作ることによる理解」（Analysis-by-Synthesis）の構成論的手法に則り，不気味の谷の形成メカニズムを説明する定性的脳機能モデルを構築することを目的とする．本手法では，まず脳の各領域の機能的側面をモデル化し，それらを機能的に結合することで対象となる脳全体の振る舞いを表現する．

本論文の第II部では，ヒト型エージェントの静的側面（静止画）に対する人間の知覚処理の特性とその神経基盤に関するモデル構築，第III 部では，ヒト型エージェントの動的側面（動画）に対する人間の知覚処理の特性とその神経基盤に関するモデル構築を行った．第IV 部では，これらのモデルを体系的かつ統一的に記述する枠組みとしてデバイスオントロジーに着目し，各モデルを統一的な観点から位置づけるためのアプローチの有効性について論じた．

2．顔の静的情報に対する神経系情報処理

2.1．顔の生物性判断における視線遷移

ヒト型エージェントと人間の顔画像をそれぞれ刺激として用いて，ヒト型エージェントの「顔」に対する特有の知覚処理を，観察中の視線を計測することで抽出を試みた．顔刺激の提示時間長の観点から顔刺激の各顔特徴領域に対する視線停留時間の長さを分析した．この結果，人間とは異なると判断されたヒト型エージェントの右目に対する停留時間は，刺激が提示される時間が長くなるにしたがって，人間の右目に対する停留時間よりも定性的に長くなる傾向があることが示唆された．このことから，ヒト型エージェントに対しては，その観察初期段階において人間と同等の処理を行う過程と，それに続く人型エージェント特有の処理を行う過程の二段階からなると考えられた．

2.2．顔認知における否定的情動形成モデル

前小節で抽出されたヒト型エージェントに対する情報処理過程を説明するための定性的脳機能モデルの構築を試みた．このモデルでは，ヒト型エージェントに対する知覚情報処理は観察対象に対する雑多な情報処理過程と観察対象に対する精密な情報処理過程からなり，この二つの情報処理の齟齬の結果否定的な感情が形成されると考えた（図1）．本モデルは，扁桃体や線条体などによる快-不快評価機構，大脳皮質による高次認知処理，海馬による記憶や評価の一貫性評価の機構からなる．シミュレーションの結果，形態的特徴が人間から離れるに従い，エージェントに対する評価が振動することが示唆された．この評価の振動が不気味さの様相を表現していると考えられた．

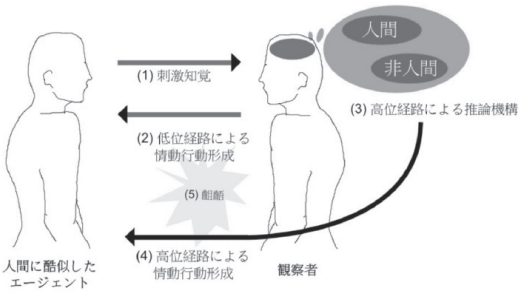

図１．本研究で提案する否定的情動反応形成モデル

3．顔の動的情報に対する神経系情報処理

3.1．表情動作の典型性が人間の印象に与える影響

ヒト型エージェントにおける表情動作速度が，人間の典型的な表情動作速度から逸脱するにしたがって否定的な印象を形成させる，と仮説を立て心理学的実験を実施した．人間の自発的な表情表出を行う際の速度を典型的な動作速度と定義し，典型的な動作速度および非典型的な動作速度の笑顔・怒り顔をヒト型エージェントに実装した．実験参加者にそれぞれの表情動作の観察をしてもらい，違和感，快-不快評価，表情強度，表情速度の適度についてそれぞれ評価を

求めた．この結果，外見で否定的な評価がなされたエージェントについては，表出速度が速い場合の非典型的笑顔は，典型的な速度と比較して否定的印象を与えることが示唆された．また，外見の評価あるいは表出する感情に関わらず，遅い速度での表情表出は観察者に否定的な印象を与えることが示唆された．さらに，重要な点として，表情表出速度を遅くすると人間はその背後に文脈を読み取り，何らかの理由があってその表情を表出していると感じることが示唆された．これはヒト型エージェントの表情動作であっても社会的な存在として認知させることができるということを示唆しており，学習教育文脈など非言語コミュニケーションが必要な場面においても多様な表現をヒト型エージェントが表情で伝達できることを示唆している．

3.2．表情認知における否定的情動形成モデル

人間の表情認知が動作予測の情報処理と動作知覚の情報処理の二つの情報処理によって実現されていると仮説を立て，動作予測を行う神経基盤として小脳を検討した．本モデルでは，初期動作が受容されるとその情報が小脳に連絡され，小脳の内部モデルによってそのあとの動作系列が計算され予測されることがモデル化されている．シミュレーションの結果，典型的な表情動作を行うヒト型エージェントについては否定的な印象は形成されないが，非典型的な動作を与えた場合は，否定的な評価がなされることが示唆され，非典型的動作に対する否定的評価を説明する脳機能モデルの構築が達成された．

4．脳機能モデルの統一的記述体系

これまでに構築された定性的脳機能モデルの記述観点が統一されていないという点を受け，それらを体系的に記述する枠組みとして拡張デバイスオントロジー［3］の枠組みに着目した．拡張デバイスオントロジーは人工物における機能を体系的に記述するための概念を提供するが，本枠組みが生物器官としての脳における機能に対しても適用可能であることを，I-goals とNI-goals［4］の観点から検討した．さらに，この記述の一環として，神経科学的知識に基づく神経細胞群間の活性伝播に関するオントロジーをロール概念に基づいて記述し，具体的なドメインとして衝動性眼球運動に関する活性伝播オントロジーを構築した（図2）．この結果，活性伝播に関するオントロジーはデバイスオントロジーに基づいた対象の捉え方ではないことが明らかとなった．このため，デバイスオントロジーの観点で脳機能を捉えるための脳機能概念を今後明示する必要がある．

また，このデバイスオントロジーの考えに基づくことで，不気味の谷にみられる否定的感情形成過程における部分機能に見られる，刺激の生物学的重要性検出に関する機能や典型性逸脱検出機能などに共通する「重要性の評価」を明示的に区分し，また，それらの部分機能を達成する役割をもつ構造を神経系のどこに対応させるのか，という点を明示する必要性が指摘された．このような観点は，今後感情を含む高次認知機能を説明する脳機能モデルを検討する上で重要な視点になると考えられる．

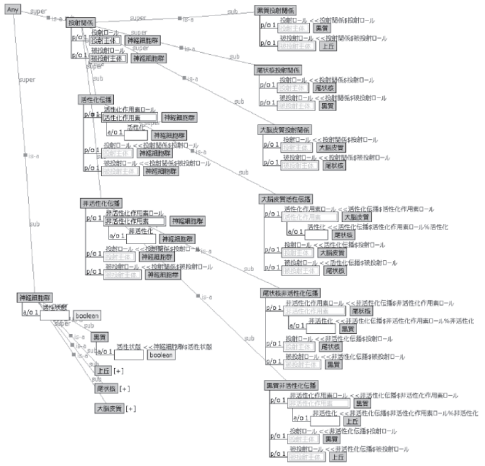

図2．衝動性眼球運動に関する活性伝播オントロジー

5．まとめと今後の課題

本研究では，(1) 顔の静的情報におけるヒト型エージェントに対する特有の知覚情報処理過程の抽出とその基盤となる脳機能モデルの提案，(2) 顔の動的情報（i.e.動作速度）における典型性・非典型性が観察者の印象に与える影響とその基盤となる脳機能モデルの提案を行い，(3) それらを統一的な観点から説明する枠組みとして，脳の機能をデバイスオントロジーの観点から整理することを試みた．

今後は，心理実験で用いられた顔画像の性差を含む刺激の詳細な検討を行う必要がある．また，定性シミュレータの構築にあたり，定性的な情報処理を検討する上での時間概念の取り扱いが重要な課題となる．さらに，これまでに構築した脳機能モデルを体系的に記述するためにデバイスオントロジー的視点についてさらに検討し，脳機能を体系的に記述可能な枠組みの詳細化を行うことが挙げられる．

参考文献

［1］ 森 政弘:不気味の谷，エナジー誌，7 (4)，33-35 http://www.getrobo.com/（2013年２月16日参照）

［2］ Saygin, A.P. et al.: The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions, Social Cognitive and Affective Neuroscience, 7 (4)，413-422（2012）

［3］ 來村 徳信ら:オントロジー工学に基づく機能的知識体系化の枠組み，人工知能学会論文誌，17（1），61-72 (2002)

［4］ Mizoguchi, R. et al.: A functional ontology of artifacts, The Monist, Vol.92, No.3, pp.387-402 (2009)