

# 脳を模倣した学習器群による鏡像自己認識機構の獲得

南宇人<sup>1</sup> 本谷康平<sup>1</sup> 水山遼<sup>1</sup> 西田圭吾<sup>1</sup> 内田貴久<sup>2</sup> 廣芝和之<sup>3</sup> 天野薫<sup>4</sup>

大阪大学 生命機能研究科<sup>1</sup> 大阪大学 基礎工学研究科<sup>2</sup>

Dwango Media Village<sup>3</sup> 情報通信研究機構 脳情報通信融合研究センター<sup>4</sup>

## 1 はじめに

認知科学の分野において、鏡に映った自己像を自分だと判別できる能力である「鏡像自己認識」の理解を試みた研究が始まり、霊長類やヒトの幼児を対象として、これまでに様々なアプローチがとられてきた。特にマークテストなど、行動評価ベースの心理実験が行われ、「鏡像自己認識」を定量化する動きがあった。さらにfMRIを用いた脳活動計測によって、鏡像自己認識と関連した脳機能の解明も進められており、概念モデルで取り上げられた「鏡像自己認識」の構成要素が、脳内のどの領域と対応するのかが明らかになりつつある。一方で時間分解能の限界から、情報処理の順番や分岐など、各要素の詳細な対応関係は明らかにされていない。本研究では、鏡像自己認識と関連する脳領域についての神経科学的な知見に基づき[1]、各脳機能と情報のフローを簡略化することで、運動情報の連動性から自他判別を行う鏡像自己認識モデルを構築すると共に、実際にそのモデルによる自他判別が可能であることを確認する。また、脳に由来する予測モデルを採用した場合の、自他判別性能に対する影響を調査する。

## 2 鏡像自己認識のモデル化

本章では、鏡像自己認識と関連する脳領域と鏡像自己認識の処理経路について我々の仮説を提案する。その上で、本仮説に基づいたネットワークモデルについて説明する。

### 2.1 脳内メカニズムの仮説

通常、鏡像自己認識は運動情報の連動性と顔などの形状認識から構成されると考えられているが、本研究では前者に注目する。視覚系についての先行研究より、初期視覚野からMT野を経て頭頂葉に至る背側経路で、視覚情報から運動成分を抽出し、処理を行なっていることが明らかになっている。一方、人間が筋骨格系を動かす際の運動制御パラメータは、運動前野や一次運動野を経て処理されている。さらに頭頂葉(AIP/PFG)は、腹側運動前野(F5)から送られてくる遠心性コピー(感覚野へ送られる、運動指令信号のコピー)を元に、自身の運動に伴う視覚フィードバックを予

測し、実際の視覚フィードバックと照合していることが示唆されている[1]。我々は、この照合が上手く行った時のみ、「視覚情報中の運動成分」が、「自分の動き」として知覚されると考える。これらの考えに基づき、鏡像自己認識の情報処理フローを模式化したものが図1である。

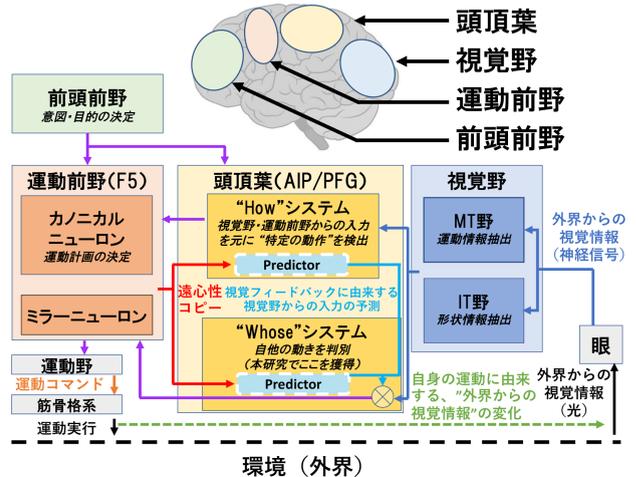


図 1: 鏡像自己認識の脳内メカニズム仮説

青色の矢印は「外界からの視覚情報に由来する求心性の神経信号」、赤色の矢印は「遠心性コピー」、橙色の矢印は「運動コマンド」、紫色の矢印は「その他の神経信号」、黒色の矢印は「外界との相互作用」、緑色の矢印(点線)は「自身の運動に由来する、外界からの視覚情報の変化」を示している。

### 2.2 自己運動感判別モデル

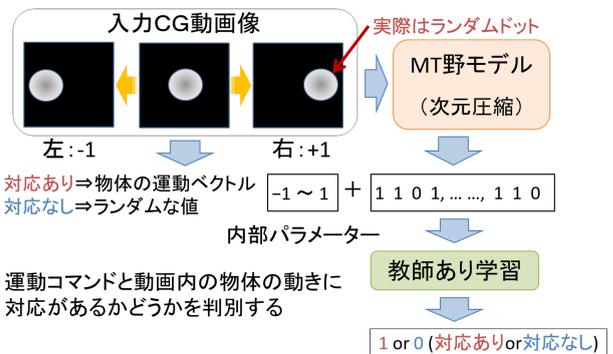


図 2: 自己運動感判別モデルの学習過程

2.1の仮説に基づき鏡像自己認識を再現するため、我々は各脳機能を簡略化し、段階的にモデルを構築した。最初のモデルには、「予測」を組み込まず、自身の運動コマンドと、視覚経路で抽出された入力動画中の運動成分との連動性から、ある瞬間の「運動成分」を「自分の動き」として判別するように学習させた(自己運動感モデル)。この

The mirror self-recognition mechanism acquired by groups of learners imitating the brain: Sorato Minami<sup>1</sup>, Kouhei Motoya<sup>1</sup>, Ryo Mizuyama<sup>1</sup>, Keigo Nishida<sup>1</sup>, Takahisa Uchida<sup>2</sup>, Kazuyuki Hiroshiba<sup>3</sup>, Kaoru Amano<sup>4</sup> (1: Osaka University, Graduate School of Frontier Biosciences, 2: Osaka University, Graduate School of Engineering Science, 3: Dwango Media Village, 4: National Institute of Information and Communications Technology, Center for Information and Neural Networks)

モデルは、入力動画像からMT野モデル[2]を介して運動成分を抽出する一方、入力動画生成時のパラメータ(運動コマンド)を内部パラメータとして保持している。モデル内では、これらを連結したものを入力、その対応の有無を教師データとし、教師あり学習を行う(図2参照)。

その後、先述したモデルを予測モデルに拡張する(図3参照)。予測モデルは動画特徴量と運動コマンドから予測特徴量を計算し、これと1フレーム先の動画特徴量の差分をニューラルネット(NN)への入力とする。学習は入力動画像と運動コマンドの対応を元に行う。

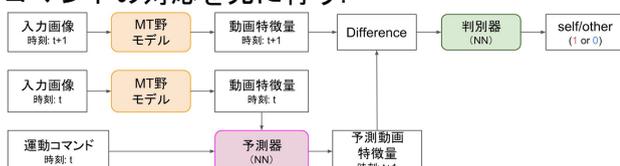


図3: 予測モデルの学習過程

### 3 実験

2.2で提案した自己運動感判別モデルにおいて、簡単な条件下で入力と運動コマンドの対応付けが可能か検証する。その際、入力動画像としてCG画像を用意し、各学習モデルの学習推移と汎化能力を評価する。

#### 3.1 入力データセット

本実験では、二次元のランダムドットが横方向へランダムに移動する動画像を入力データセットとして与えた。1フレームの静止画像は $100 \times 100$  pixelである。そして、この静止画像が1フレームごとに $-5 \sim 5$  pixel(整数値, 11パターン)分だけ移動する。このように生成された画像100フレームを1動画とし、データセットとして10000個の動画を用意する。本研究ではそのうち、9910は学習データに、残りの90はtestデータに使用した。

#### 3.2 実験設定

各動画セットに対して、1フレームごとの移動距離に応じた運動ベクトル(運動コマンド)が $-5 \sim 5$ の整数として与えられる。両者に対応関係があるペアを正例、1つの動画とペアになる運動コマンドが、対応のない他動画の運動コマンドからランダムに選択されたものを負例とする。学習用、テスト用の入力データセットとして、正例データと負例データを1対1の頻度で使用した。ネットワークの実装フレームワークとしてChainerを使用し、最適化アルゴリズムにはSGDを用いた。学習係数は0.01とした。学習は予測値(self:1 or other:-1)と正解値(self:1 or other:-1)の差の2乗平均を最小化するように行った。

#### 3.3 結果

モデル出力が正解値と一致した割合を算出した。各モデルに対する学習時、テスト時のaccuracy推移を図4に示す。初期モデルに比べ、予測モ

デルは学習速度が明らかに速く、テスト時の精度も高かった(iteration = 100000, first: accuracy = 0.9843, predict: accuracy = 1.0000)。このため予測モデルでは途中で計算を打ち切った。本結果から、予測を取り入れたより脳に近いモデルで、自己運動感の判別能力が向上することが明らかになった。

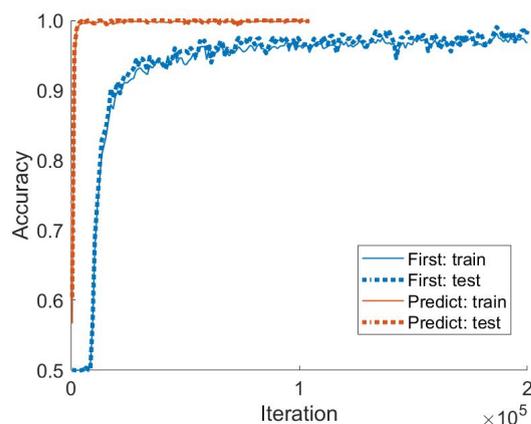


図4: 学習時のaccuracyの変化

近年では脳の知見を取り入れ「運動予測」を行う鏡像自己認識モデルを、ロボットに適用した報告もあり[3]、鏡像自己認識に対する予測の寄与が考察されている。このことを踏まえ、今後は予測モデルの有用性を詳細に検討する必要がある。

### 5 おわりに

本研究では、鏡像自己認識の神経科学的な知見に基づき、運動情報ベースの自己運動感判別モデルを構築し、単純なCG動画像を入力データとして使用してその性能を検証した。今後は、予測の有無によるモデルの性能差を詳細に検討するため、多様な入力動画像に対する汎化性能や、ノイズへのロバスト性などを比較する予定である。

#### 謝辞

本研究は、WBAIハッカソンで取り組んだテーマに基づいており、ドワンゴ人工知能研究所の山川宏氏には多大なアドバイスを頂いた。また、文部科学省博士課程教育リーディングプログラムの補助を受けている。ここに記して謝意を表す。

#### 参考文献

- [1] Murata, A., Wen, W., Asama, H.: "The body and objects represented in the ventral stream of the parieto-premotor network", *Neurosci Res*, Vol.104, pp.4-15, 2016.
- [2] Newsome, W. T., Paré, E. B.: "A selective impairment of motion perception following lesions of the middle temporal visual area (MT)", *J Neurosci*, Vol.8, No.6, pp.2201-2211, 1988.
- [3] Zeng, Y., et al.: "Toward Robot Self-Consciousness (II): Brain-Inspired Robot Bodily Self Model for Self-Recognition", *Cognitive Computation*, pp.1-14, 2017.