

# ポピュレーション符号化を利用した 自他の動き等価性の早期発見による共同注意の学習

長井志江\*

## Joint Attention Learning based on Early Detection of Self-Other Motion Equivalence with Population Codes

Yukie Nagai\*

This paper presents a robotic learning model for joint attention based on self-other motion equivalence. Joint attention is a type of imitation, by which a robot looks at the object that another person is looking at by producing an eye-head movement equivalent to the person's. It means that this ability can be acquired by detecting an equivalent relationship between the robot's movement and the person's. The model presented here enables a robot to detect the eye-head movement of a person as optical flow in the vision and the movement of its eyes and head as a motion vector in the somatic sense. Because both of the movements are represented with population codes, the robot can acquire the motion equivalence as simultaneous activations of homogeneous neurons that are responsible to a same motion direction in the two senses. Experimental results show that the model enables a robot to learn to establish joint attention based on the early detection of the self-other motion equivalence and that the equivalence is acquired in a well-structured visuomotor map. The results moreover provide analogies with the development of human infants, which indicates that the model might help to understand infant development.

**Key Words:** Joint attention, Imitation, Self-other motion equivalence, Population coding

### 1. はじめに

身体運動に基づくコミュニケーションでは、自己の運動と他者の運動との関係をいかに表現するかが問題となる。人間とロボットとのインタラクションにおいて、ロボットが人間の運動を視覚センサを通して観察しているとすると、ロボットはその視覚情報と、自己の運動を知覚、制御している体性感覚情報との間に、何らかの合目的な関係を成立させなければならない。このとき、一般に視覚情報と体性感覚情報は異なる様式で表現されるため、ロボットにとって自己の運動と他者の運動との関係を統一的に扱うことが困難になる。そして、このような問題は人間側でも同様に発生しており、工学的視点から人間がどのような感覚情報表現を用いることで効率的なコミュニケーションを実現しているか議論することは、認知発達ロボティクス [1] の視点から見ても意義深い。

上記のような問題を含むタスクとして、ここでは共同注意 [2] を取り上げる。共同注意は自己が他者の注視している対象物を同時に注視する行為であり、人間のコミュニケーション能力の

発達の礎として、認知科学や発達心理学の分野 [3] ~ [6]、さらにはヒューマン - ロボットインタラクションの分野 [7] ~ [10] で盛んに研究されている。ここでの身体運動とは、他者の眼球や頭部の回転による視線変化運動に対して、自己も同じ方向に視線を向けるために、他者とほぼ等しい眼球、頭部運動を生成することである。つまり、視覚情報として検出される他者の運動と、体性感覚情報としての自己の運動との間に、等価な関係を成立させることが求められる。このような身体運動の関係は模倣にも共通しており、自他の運動間の等価性を異なる感覚上のどの特徴量を用いて評価するか、またそれらをいかに表現するかが、これらの能力を効率的に獲得する鍵と考えられる。

これまでに、工学的視点からいくつかの共同注意の学習モデルが提案されてきたが、自他の運動間の等価性についてはほとんど考慮されていなかった。例えば、他者の視線方向が既知であると仮定して、報酬に基づく学習により自己の視線変化運動との対応関係を獲得するモデル [11] [12] や、他者の頭部画像から得られる輝度値情報を入力として、自己の運動との関係をニューラルネットワーク (NN) で学習するモデル [13] [14] が提案されている。またこのモデルの発展として、あらかじめ入力画像を他者の視線方向に対応した自己組織化マップで分類しておく方法 [15] もある。これらのモデルの共通点は、対象物を発見したときの報酬を引金として、知覚 - 運動対の随伴性を検出するこ

原稿受付

\* (独) 情報通信研究機構

\* National Institute of Information and Communications Technology

とで共同注意能力を獲得するもので、自他の運動間の等価性を積極的に利用していないことから、学習に膨大な時間を費やしたり、学習の加速のために前処理を行うといった工夫を必要としていた。これに対して、共同注意を行うときの自他の注視姿勢はほぼ等価であるという仮定のもとに、他者の頭部画像から確率モデルで推定した頭部姿勢を、そのまま自己の目標姿勢として利用する方法 [16] も提案されている。しかし、他者の頭部姿勢は本来知ることのできない変数であり、その空間での等価性は外部からの教示を受けない限り学習者が自律的に発見することは難しい。

一方、発達心理学の分野では、人間の幼児の共同注意が単なる対象物の発見による随伴性学習に基づくものではなく、他者の視線変化運動の追従から始まること [6]、そして、そのときの他者の頭部回転の動きは幼児の視線方向の理解を助けること [5] [17] が指摘されている。また、新生児でも他者の頭部回転運動を模倣できること [18] や、運動中の動きの呈示によってさまざまな模倣が誘発されること [19] [20] が発見されている。つまり、共同注意能力の基礎には頭部回転運動の模倣があり、それらはいずれも視覚上の動きによって発達が促進されていると考えることができる。この動きに注目したロボットの模倣モデルとして、視覚上の自他の動き等価性を利用したモデル [21] [22] が提案されている。これは、模倣対象となる自他の身体部位（手や腕）の視覚上での動きベクトルを等しくすることで大まかな運動を真似るもので、学習者が自律的に等価性を発見できる点で有効である。しかし、共同注意のように自他の身体部位（頭部）が同じ感覚上に知覚されない場合には、動きの等価性を異なる感覚様式間で評価しなければならないという問題を残す。

そこで本稿では、知覚 - 運動情報を感覚様式に依存しないポピュレーション符号化 [23] で表現することで、ロボットが自律的に自他の動き等価性を発見し、共同注意能力を獲得できるモデルを提案する。提案モデルの特徴を以下に示す。

- (1) ロボットは他者の視線変化時の動きを、それを観察したときのカメラ画像上のオプティカルフローとして検出し、自己の動きを頭部姿勢の遷移ベクトルとして表現する。ロボットは視覚上の動きベクトルと体性感覚上の動きベクトルの間の等価性を早期に発見することで、効率的に共同注意能力を獲得する。
- (2) ロボットは視覚と体性感覚の動きベクトルを、感覚様式に依存しないポピュレーション符号化で表現する。動きベクトルを特定の運動方向に対して反応選択性をもつ、複数のニューロン活動の集合として表すことで、動き等価性を異なる感覚間での同質なニューロンの同時活動として自律的に発見する。
- (3) ロボットは視覚入力として、オプティカルフローの他に他者の頭部画像のエッジ特徴を用いる。オプティカルフローが他者の大まかな運動方向を示すのに対して、エッジは正確な注視姿勢を表すことから、共同注意の学習精度を向上させることが期待できる。

以下では、まず 2 章で、共同注意における自他の動き等価性と、知覚 - 運動情報の表現方法としてのポピュレーション符号化について説明する。次に 3 章で、それらのアイデアに基づいた

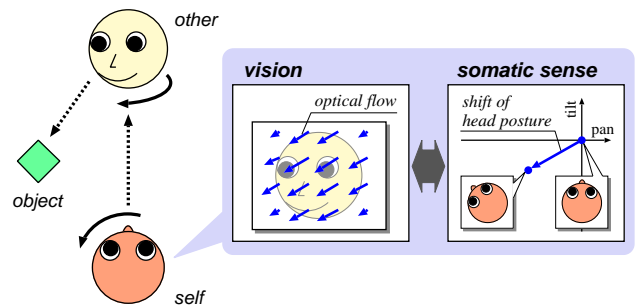


Fig. 1 Self-other motion equivalence in joint attention

共同注意の学習モデルを提案し、4 章で、実ロボットを用いた実験によりモデルの有効性を示す。最後に、5 章でまとめを述べ、認知発達ロボティクスの観点から見たモデルの評価と、今後の課題を示す。

## 2. 自他の動き等価性とポピュレーション符号化

### 2.1 共同注意における動き等価性

共同注意 [2] は、自己が他者の視線を追従することで他者と同じ対象物を注視することであり、このときの自他の視線変化の動きには等価な関係が存在する。Fig. 1 のように自己が他者と対面し、その間に対象物が置かれている状況を想定する。ここで両者が互いを正視した状態から、他者が頭部を回転させて対象物の方向へ視線を向けると（頭部と眼球の向きは一致すると仮定する）、他者の動きは自己の視覚上でオプティカルフローとして検出することができる（図中央）。ここでは、左下方向のフローベクトルが得られる。一方、自己が正面を向いた状態から、自発的に対象物の方向へ視線を向けると、自己の動きは体性感覚上の頭部姿勢の遷移ベクトルとして検出される（図右）。このベクトルを頭部姿勢のパン、チルトの二次元空間で表現し、視覚上のオプティカルフローと比較すると、両者はほぼ同じ方向性をもっていることが分かる。このベクトルの方向性の等しさが共同注意における自他の動き等価性であり、ロボットは視覚情報と体性感覚情報を知覚 - 運動の対として等価性を検出することで、共同注意能力を獲得することができる。

### 2.2 知覚 - 運動情報のポピュレーション符号化

上述のように、等価性の検出が必要な知覚 - 運動情報は、視覚と体性感覚という異なる感覚様式で知覚される。そこで、ロボットがこれらを統一的に扱い、自律的に等価性を発見できるように、知覚 - 運動情報の表現として感覚様式に依存しないポピュレーション符号化を採用する。ポピュレーション符号化 [23] とは、脳の神経細胞における情報表現の一種であり、一つの情報を特定の刺激に対して反応選択性をもつ複数のニューロン活動の集合として表現する方法である。個々のニューロンはある適刺激を中心にガウス関数で近似される非線形の応答特性をもち、呈示された刺激に対して、異なる反応選択性をもつ複数のニューロンがそれぞれの応答特性に応じた活動を示すことで、その集合として元の情報を表現する。例えば、ネコの V1 のニューロンにおける視覚刺激の傾きや動き方向の情報表現 [24] や、サル MT 野における視覚刺激の動き方向の表現 [25]、またサルの

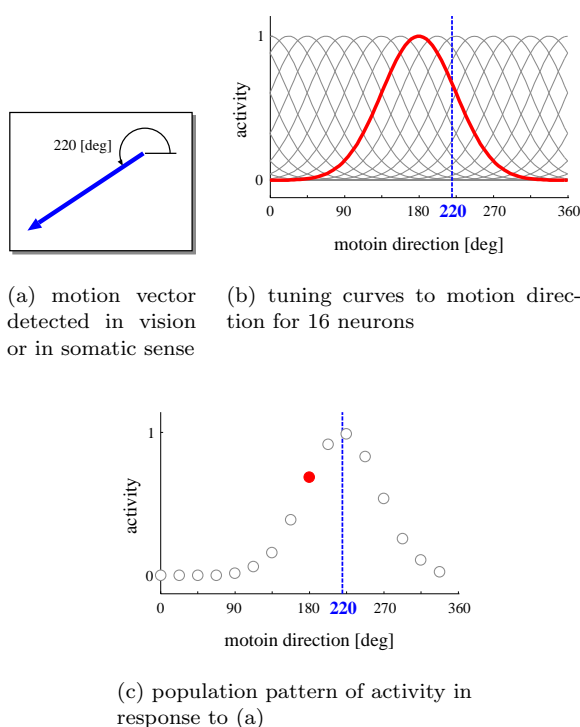


Fig. 2 Population coding of motion vector

運動野における手先の運動方向に応じた応答 [26] [27] がポピュレーション符号化であることが知られている。

Fig. 2 にポピュレーション符号化の例を示す。(a) は符号化する情報の例として、視覚もしくは体性感覚上で検出された動きベクトル (220 [deg])、(b) は符号化に用いるニューロンの応答特性のグラフである。ここでは、22.5 [deg] おきの動き方向に対して反応選択性をもつ 16 個のニューロンが存在すると仮定する。各ニューロンは適方向を中心としたガウス状の応答特性をもち、呈示された刺激に対してその関数値に応じた応答を示す。例えば、グラフ中央の太線で描かれたガウス曲線は、180 [deg] の動き方向に対して反応選択性をもつニューロンの応答特性であり、このニューロンは (a) の刺激に対して約 0.7 (最大値で正規化) の活性度を示すことになる。このようにして、16 個全てのニューロンの活動パターンを描いたのが (c) である。入力刺激の動き方向に近い方向に反応選択性のあるニューロンほど、大きな活性度を示しているのが分かる。このようなニューロン集団の活動によって、元の情報を脳内で分散的に表現している。また反対に、このニューロン活動から情報を復号化して運動制御などに利用する場合は、各ニューロンの適方向を活性度に応じた割合で合成したり、最尤法により集団の活動を最もよく説明する応答関数を推定するなどして、一意の運動方向を求める手法が用いられる [28]。

### 2.3 ポピュレーション符号化を用いた動き等価性の発見

以上の符号化法を共同注意の学習モデルに利用することで、知覚 - 運動情報が分散表現され、その関係がより構造化されて獲得されることが期待できる。特に、2.1 節で説明した自他の

動き等価性は、視覚と体性感覚上で同じ動き方向に反応選択性をもつ同質なニューロンの同時活動として検出することができ、ロボットは自律的かつ早期にこの関係性を発見できると考えられる。また一般に、ポピュレーション符号化は刺激のずれやノイズに対して頑健であり [29]、人間の情報認識の不確かさまでも含めて表現できることが指摘されており [25]、認知発達ロボティクスの視点から人間の認知モデルを考察する上でも有効であると考えられる。

### 3. 動き等価性の早期発見に基づく共同注意の学習モデル

前章のアイデアに基づいた共同注意の学習モデルを Fig. 3 に示す。ロボットは人間と対面した状態で、人間が対象物の方向へ視線を変化させたときのカメラ画像から、視覚情報としてオプティカルフローとエッジを抽出し、自己の運動である頭部姿勢の遷移ベクトルとの対応関係を学習することで、共同注意能力を獲得する。視覚 - 運動対の表現には階層型 NN を用い、ポピュレーション符号化された自他の動き情報間で等価性を早期に発見することで、学習の高速化を実現する。また、静的な視覚情報であるエッジ特徴も利用することで人間の正確な注視姿勢を推定し、学習の精度向上を図る。以下で、各機能について順に説明する。

#### 3.1 画像特徴の抽出

##### 3.1.1 オプティカルフローの検出

ロボットはまず、カメラ画像  $I$  からオプティカルフロー  $F$  を検出する。2 枚の連続する画像をそれぞれ微小領域 (受容野と呼ぶ) に分割し、受容野ごとにテンプレートマッチングを行って対応点を検出する。ある時刻における  $k$  番目の受容野の左上座標を  $(x_k, y_k)$ 、一時刻前の対応点の座標を  $(px_k, py_k)$  とすると、この受容野におけるフローベクトル  $F^k$  は、対応点の移動距離の累積和として、

$$F^k = \left[ \begin{array}{c} \sum_t (x_k - px_k) \\ \sum_t (y_k - py_k) \end{array} \right] \quad (1)$$

と計算される。 $t$  は一定期間である。

実際の実験場面での検出例を Fig. 4 に示す。ロボットは同軸上に固定された (a) 周辺視カメラと (b) 中心視カメラをもち、(a) の画像を用いてテンプレートマッチングで人間の顔を画像中心に捕らえた後、(b) の中央矩形領域内から画像特徴を抽出する。テンプレートには人間が正面と上下左右各 45 [deg] の計 5 方向を向いたときの画像を使用している。(c) に示すのが、人間がカメラ正面を向いた状態から対象物の方向 ((a) 内左下) へ視線を変化させたときに検出したオプティカルフロー<sup>†</sup>である。この検出例から、オプティカルフローは人間の視線変化の動き方向を大まかではあるが、すでにベクトル表現された中間解釈を必要としない状態で獲得できていることが分かる。また、ロボットはこのフローの量変化から人間の視線変化運動のタイミングを推定しており、フロー量が上方閾値を越えた直後に下方閾値を下回ったとき、人間の視線変化運動が終了したものと見て、その直後に自身の視線変化運動を生成する。

<sup>†</sup> フローベクトルは 8 方向 (←, ↖, ↑, ..., ↘) にそれぞれ (赤, シアン, 青, ..., 緑) と色分けしている。逆方向は同色である

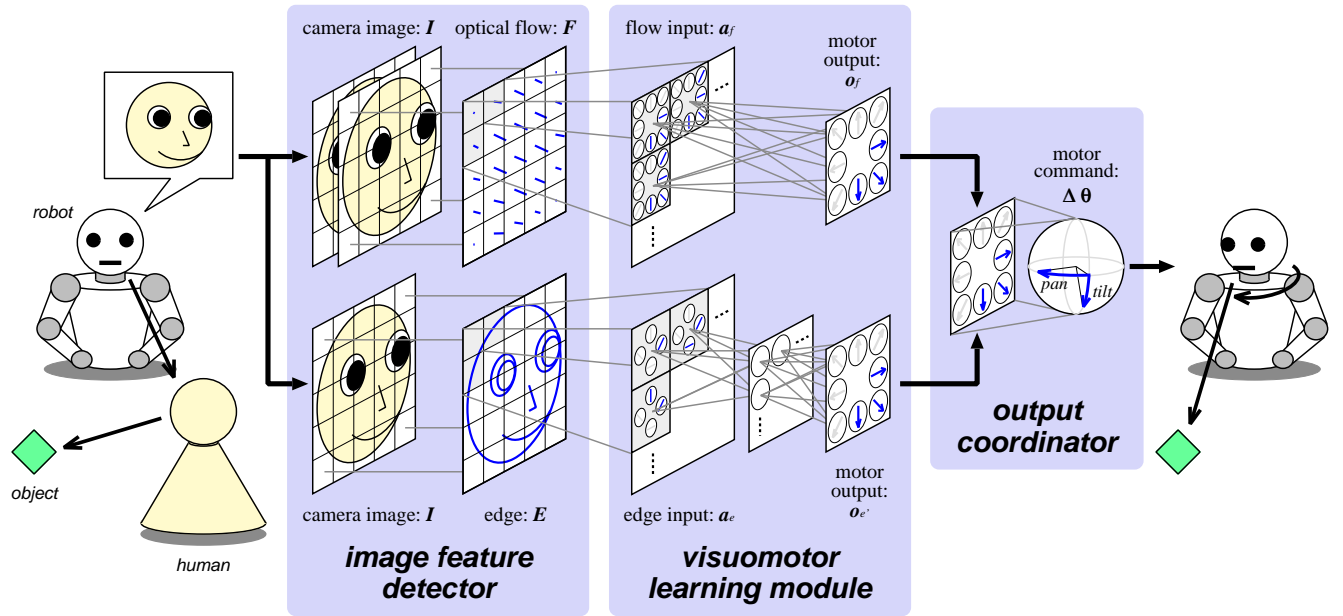
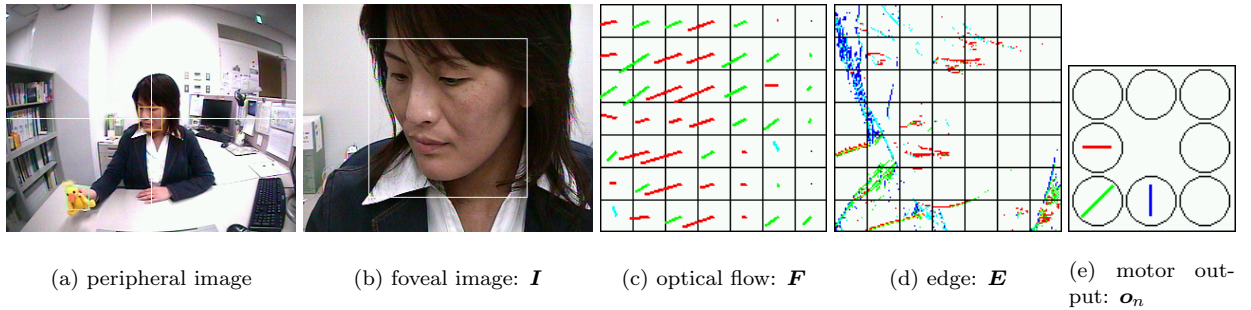


Fig. 3 Learning model for joint attention based on self-other motion equivalence



(a) peripheral image (b) foveal image:  $I$  (c) optical flow:  $F$  (d) edge:  $E$  (e) motor output:  $o_n$

Fig. 4 An example of input-output datasets

### 3.1.2 エッジの抽出

ロボットはオプティカルフローの検出と同時に、カメラ画像  $I$  からエッジ特徴  $E$  を抽出する。画像に対して4方位  $(e_1, e_2, e_3, e_4) = (-, \backslash, |, /)$  にそれぞれ選択性をもつ4つの微分フィルタを施し、エッジ画像を生成する。例えば、水平エッジ  $e_1$  に関して、画素  $(x, y)$  でのエッジ特徴の有無を示す値  $E_1(x, y)$  は、画像の輝度値  $I(x, y)$  から

$$E_1(x, y) = \begin{cases} 1 & \text{if } \epsilon_1(x, y) > \epsilon_{\text{thr}} \\ 0 & \text{else} \end{cases} \quad (2)$$

$$\epsilon_1(x, y) = \sum_{i=-1}^1 |I(x+i, y-1) - I(x+i, y+1)| - \sum_{j=-1}^1 |I(x-1, y+j) - I(x+1, y+j)| \quad (3)$$

と計算される。  $\epsilon_{\text{thr}}$  は定数の閾値である。同様にして、他の方位のエッジについても  $E_n(x, y)$  が求められる。

Fig. 4 (d) に (b) の矩形領域内から抽出したエッジ特徴<sup>†</sup>を示す。これは、オプティカルフローの量変化から人間の視線変化運動が終了したと判断されたときに検出したエッジ特徴である。この検出例から、エッジはオプティカルフローと比較すると、人間の視線方向としての解釈に段階的な処理を必要とするが、より正確な方向を推定できる情報を提供していることが分かる。

### 3.2 視覚 - 運動対の表現

次に、ロボットはNNを用いて、視覚情報から自己の運動への対応関係を学習する。オプティカルフローとエッジを認識するNNをそれぞれフローNNとエッジNNとし、前者を2層、後者を3層とした独立したモデルを構成する (Fig. 3 中央参照)。これは、オプティカルフローがそのベクトル表現として、すでに人間の動き方向を直接的に表しているのに対し、エッジ特徴は人間の視線方向としての解釈に段階的な処理を必要とすると思われることによる。このようにそれぞれの視覚情報の特徴を活かし、異構造のNNを用いることで、フローNNでは学習の

<sup>†</sup> フローベクトルと同様に、4方位のエッジ  $(-, \backslash, |, /)$  について (赤, シアン, 青, 緑) と色分けしている。



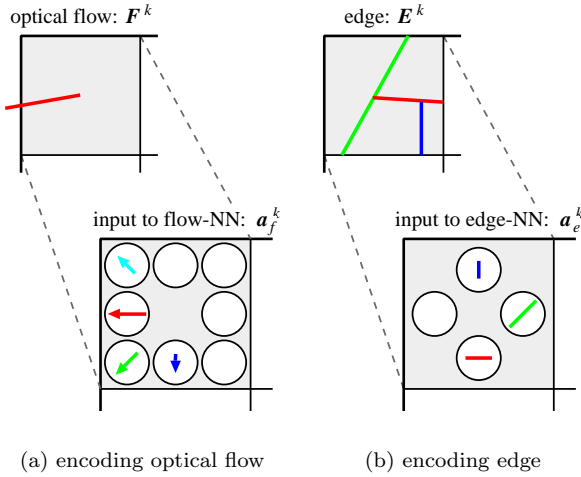


Fig. 5 Population coding of visual information

高速化を実現し、エッジ NN では学習の精度を向上させることが期待できる。以下で、ポピュレーション符号化による視覚-運動情報の表現について説明する。

### 3.2.1 オプティカルフローのポピュレーション符号化

フロー情報は、受容野ごとに 8 方向 ( $f_1, f_2, \dots, f_8$ ) = ( $\leftarrow, \swarrow, \dots, \searrow$ ) に反応選択性をもつ 8 種類のニューロン活動の集合として NN に入力される。Fig. 5 (a) に符号化の例を示す。上図が受容野  $k$  で検出されたオプティカルフロー、下図が符号化された結果である。円がニューロンを、円内の矢印がそのニューロンの選択的方向を示しており、矢印の長さでニューロンの活性度を表している。矢印のないニューロンは活性度ゼロである。各ニューロンの活性度  $a_{f_n}^k$  ( $n = 1, \dots, 8$ ) は、その受容野におけるフローベクトル  $F^k$  と 8 方向への単位ベクトル  $u_n$  との内積から、

$$a_{f_n}^k = \begin{cases} F^k \cdot u_n / \max_k \|F^k\| & \text{if } F^k \cdot u_n \geq 0 \\ 0 & \text{else} \end{cases} \quad (4)$$

と計算される。分母は正規化のためである。このような符号化によって、Fig. 5 (a) のように、フローベクトルに近い方向 ( $\swarrow, \leftarrow, \searrow, \downarrow$ ) に反応選択性をもつニューロンがその類似度に応じた強さで活性化され、逆方向 ( $\searrow, \rightarrow, \swarrow, \uparrow$ ) に選択性をもつニューロンの活動がゼロに抑えられるようなパターンが得られる。そして、全ての受容野について同様の処理を行うことで、フローベクトルをその方向性の情報を分散表現したニューロン活動として NN へ入力する。

### 3.2.2 エッジのポピュレーション符号化

エッジ情報は、受容野ごとに 4 方位 ( $e_1, \dots, e_4$ ) = ( $-$ ,  $\dots$ ,  $\swarrow$ ) に反応選択性をもつ 4 種類のニューロン活動の集合として NN に入力される。Fig. 5 (b) にエッジの符号化例を示す。上図が受容野  $k$  で抽出されたエッジ、下図が符号化した結果である。オプティカルフローの場合と同様、円がニューロンを、円内の線分がニューロンの選択的方向を示しており、線分の長さでニューロンの活性度を表している。活性度  $a_{e_n}^k$  ( $n = 1, \dots, 4$ )

は、その受容野に含まれるエッジ量  $E_n^k$  と全受容野での最大エッジ量との比から、

$$a_{e_n}^k = E_n^k / \max_k \sum_{m=1}^4 E_m^k \quad (5)$$

$$E_n^k = \sum_{x=x_k}^{x_k+\delta x} \sum_{y=y_k}^{y_k+\delta y} E_n(x, y) \quad (6)$$

と計算される。ここで、 $E_n(x, y)$  は式 (2) で与えられるエッジの有無を示す値であり、 $(x_k, y_k)$  は受容野  $k$  の左上座標、 $(\delta x, \delta y)$  は受容野の幅と高さである。Fig. 5 (b) の例では、( $-$ ,  $\swarrow$ ,  $\downarrow$ ) の 3 方位に反応選択性のあるニューロンがそれぞれのエッジの長さに応じた割合で活性化され、( $\swarrow$ ) に選択性のあるニューロンの活動がゼロに抑えられるようなパターンが生成される。同様にして全受容野について符号化を行うことで、エッジをその方位性の特徴と、画像全体に対するエッジ量の空間的な比を情報としたニューロン活動として NN に入力する。これは、オプティカルフローがその方向性の情報によって人間の動き方向を表現していたのに対し、エッジは方位性だけでなく、むしろ空間的な偏りによって人間の視線方向を表していることによる。例えば Fig. 4 (d) の場合、人間が左方向を注視しているときには顔の要素 (目や鼻、口) はその方向に寄り、結果としてエッジが画像の左側に多く抽出される。この性質を利用して、エッジ量の空間的な比をニューロン活動として表現することで、人間の注視方向をより正確に推定できるようにする。

### 3.2.3 運動出力のポピュレーション符号化

運動出力は、フロー NN とエッジ NN とともに 8 つの運動方向 ( $f_1, \dots, f_8$ ) = ( $e'_1, \dots, e'_8$ ) = ( $\leftarrow, \dots, \swarrow$ ) に反応選択性をもつ 8 個のニューロン活動として表現される。これらのニューロンはオプティカルフローを符号化したニューロン (Fig. 5 (a) 参照) と同じ性質をもち、各ニューロンが特定の方向に対する運動成分を担うことで、その活動の集合として一意の頭部回転運動を生成する。ニューロン活動から頭部回転運動への復号化については次節で説明する。

### 3.3 運動出力の調停と頭部回転運動の生成

フロー NN とエッジ NN からの出力は、それぞれの NN の学習収束度に応じて一つの最終出力へと統合される (Fig. 3 右部参照)。統合後の運動出力の例を Fig. 4 (e) に示す。フロー NN の出力層ニューロンの活性度を  $o_{f_n}$  ( $n = 1, \dots, 8$ )、エッジ NN の出力層ニューロンの活性度を  $o_{e'_n}$  とすると、最終出力  $o_n$  は、

$$o_n = o_{f_n} / \max_n o_{f_n} \quad (7)$$

$$o_{f_n} = o_{f_n} / \text{range } o_{f'} + o_{e'_n} / \text{range } o_f \quad (8)$$

$$\text{range } o_f = \max_n o_{f_n} - \min_n o_{f_n} \quad (9)$$

$$\text{range } o_{e'} = \max_n o_{e'_n} - \min_n o_{e'_n} \quad (10)$$

と計算される。出力層ニューロンは、学習が不十分な場合には 8 個全てが平均的に活性化され、学習が収束すると Fig. 4 (e) のようにある特定の方向のニューロンだけが強く活性化されるようになることから、各出力ごとに活性度の最大値と最小値の差をとることで NN の学習収束度を推定することができる。式

(7) は正規化のためである．

生成された出力  $o_n$  は、最終的に次式によりロボットの頭部回転運動  $\Delta\theta$  へと変換される．

$$\Delta\theta = \begin{bmatrix} \Delta\theta_{pan} \\ \Delta\theta_{tilt} \end{bmatrix} = \begin{bmatrix} g_{pan} \sum_n u_{nx} o_n \\ g_{tilt} \sum_n u_{ny} o_n \end{bmatrix} \quad (11)$$

ここでは、ロボットの頭部がパンとチルトの二軸の回転自由度をもつと仮定している． $(g_{pan}, g_{tilt})$  は設計者によって与えられる回転速度を定義するゲイン、 $(u_{nx}, u_{ny})$  は8方向への単位ベクトル  $u_n$  の水平、垂直成分である．この計算によって、8個の出力層ニューロンの運動成分を各ニューロンの活性度に応じて合成した頭部回転運動が生成される．Fig. 4 (e) の例では、ロボットは左下の3個のニューロンの運動ベクトルを合成することで、左下方向への頭部回転運動を生成する．ここでは、ちょうど人間の視線方向を追従する運動となり、共同注意が実現される．また、このときの出力層ニューロンの活動を (c) のオプティカルフローと比較すると、どちらも左下方向のベクトルが現れており、視覚上の動きと体性感覚上の動きとの間で等価な関係が成立していることが分かる．特に、ポピュレーション符号化した後のフロー情報は出力層ニューロンと同じ方向に反応選択性をもつニューロンの活動として表現され、動き等価性は同質なニューロンの同時活動として自律的に検出することができるようになる．

### 3.4 視覚注視経験に基づく NN の学習

以上のモデルを用いて、ロボットは視覚注視経験に基づき NN を学習することで、共同注意能力を獲得する [14]．視覚注視とは、他者の視線方向とは無関係に、環境中の特徴的な視覚刺激を注視することである．ここでは、テンプレートマッチングを用いて人間の顔を視覚追従する機能と、色情報を用いて対象物を注視する機能がロボットに備わっている．また、ロボットはオプティカルフローの量変化から人間の視線変化運動のタイミングを検出しており、人間の運動生成をトリガとしてロボット自身も頭部回転運動を生成し、注視対象を人間の顔から対象物へと切替える．これは、人間の幼児が他者の頭部回転の動きに誘発されて、自身も頭部を回転させるという知見 [6] に基づいている．ロボットはこのときに検出されるオプティカルフローとエッジ特徴、そして自身の頭部回転運動の組を学習データとして、誤差逆伝播法により NN を学習することで共同注意能力を獲得する．

ここで学習の収束性に関して、環境中に複数の対象物が存在すると、ロボットは必ずしも共同注意を実現する正しい視覚・運動対だけを学習するのではないという問題を生じる．例えば環境に対象物が  $m$  個存在し、人間とロボットが一樣な確率でこれらを注視するとすると、ロボットが視覚注視機能によって人間と同じ対象物を注視する確率は  $1/m$  となる．つまり、学習データのうち共同注意が成立したときの正しいデータは  $1/m$  となり、残りの  $(m-1)/m$  は共同注意が成立していないとき、つまりロボットが人間と異なる対象物を注視したときのデータとなる．しかし、この後者のデータでは視覚入力と運動出力の間には何の一貫性もなく、組合せはほぼランダムである．これに対し、前者のデータでは一つの視覚入力に対して唯一の運動

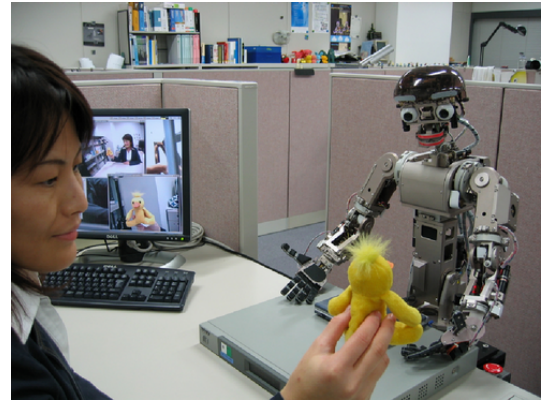


Fig. 6 Experimental environment for joint attention

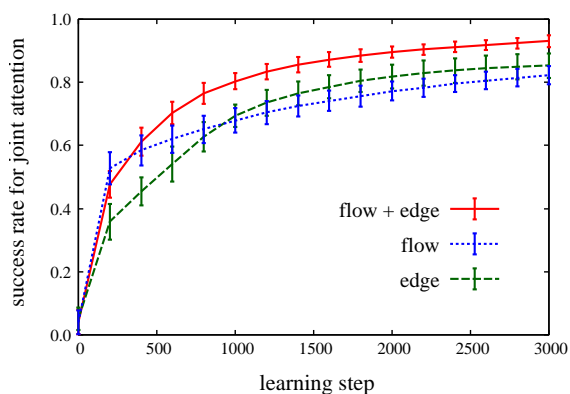
出力が対応し、入出力間に一意の相関関係が存在する．この性質を利用し、さらに NN を必要十分な大きさに設定しておくことで、ロボットは共同注意の成否にかかわらず全ての視覚・運動経験を学習することを通して、相対的に共同注意が成立したときの相関関係を強化することが期待できる．この学習収束性は、前研究 [14] において対象物が 10 個存在する環境でも、十分に共同注意能力が獲得されることが示されている．

## 4. 実験

### 4.1 実験設定

学習モデルを Fig. 6 に示す幼児型ロボット *Infanoid* [30] に実装して、共同注意の学習実験を行った．*Infanoid* はコミュニケーション能力の発達メカニズムを研究することを目的として開発された上半身型ヒューマノイドロボットで、頭部や腕に多数のセンサとアクチュエータをもつ．本実験では視覚情報の取得のために、左右目にそれぞれ組にして固定された周辺視と中心視カメラのうち左目のカメラを用いた．カメラ画像の大きさは周辺視、中心視ともに  $320 \times 240$  画素で、中心視カメラ画像の中央  $168 \times 168$  画素の領域を入力画像として、 $49 (7 \times 7)$  個の受容野に分割して使用した (Fig. 4 参照)．視線方向の制御には目の 3 自由度 (左目パン 1 自由度、右目パン 1 自由度、チルト 1 自由度) と首の 3 自由度 (パン 1 自由度、チルト 2 自由度) の計 6 自由度を使用し、まず眼球の方向を運動出力に従って制御した後、それに追従するように首の角度を調整することで自然な運動を生成した．首のチルト 2 自由度には、目標運動を等分した運動出力を割り当てた．学習モデルは、フロー NN のニューロン数を入力層 392 ( $8 \times 49$ ) 個、出力層 8 個とし、エッジ NN のニューロン数を入力層 196 ( $4 \times 49$ ) 個、中間層 3 個、出力層 8 個としたものを用いた．エッジ NN の中間層のニューロン数は予備実験の結果から決定した．

これに対して、人間側は学習実験を通して同一の人物が、テーブルを挟んでロボットと正対するように座った．そして、各試行でまずロボットを正視し、その後、対象物の方向へ視線を変化させることで、顔の正面で対象物を捕らえるようにした．対象物はあらかじめ色情報で定義した 2 個を使用し、試行ごとにテーブル上でランダムな位置と高さに設置した．人間は各試行で注視する対象物をランダムに選択し、実験を通してその注視



(a) change in success rate for joint attention

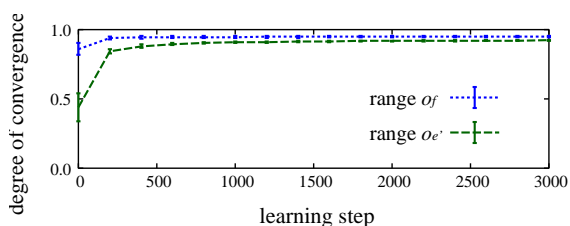
(b) change in range  $o_f$  and range  $o_{e'}$ 

Fig. 7 Results of learning experiment

確率がほぼ一樣になるようにした．また、ロボットも視覚注視機能によって各対象物をほぼ同確率で注視するものとし、学習時には約 50%の確率で人間と同じ対象物を注視できるようにした．

#### 4.2 学習実験

始めに、提案モデルの学習性能と学習における視覚情報の役割を評価する実験を行った．実験に際して、まず実環境で Infanoid に学習に用いる入出力データを 200 組獲得させ、その後、獲得したデータを用いてオフラインでの学習実験を行った．200 組のデータのうち、約半数は共同注意が成立したときに獲得されたデータ、残りの半数は失敗したとき、つまりロボットが人間と異なる対象物を注視したときに獲得されたデータである．

##### 4.2.1 共同注意の成功率の変化と各入力情報の役割

Fig. 7 (a) に学習回数に対する共同注意の成功率の変化を示す．共同注意の成否は、ロボットの運動出力後の視線方向と対象物の方向との角度誤差が  $\pm 8$  [deg] 以下（中心視カメラで対象物を捕らえられる範囲）であるときを成功として、それ以外を失敗として判断した．3 本の曲線は、視覚入力としてオプティカルフローとエッジの両方を使用した場合と、どちらか一方のみを使用した場合の結果を示している．どちらか一方の入力を用いたときは運動出力の統合は行わずに、NN からの出力をそのままロボットの視線変化運動へと変換して評価した．グラフは学習の初期条件を変化させて 50 回行った実験結果の、平均と分散を示している．

この結果より、まず一種類の入力情報だけを用いたときの学習曲線を比較すると、フロー入力を用いた場合は学習初期にお

いて学習の立上りが速くなっていること、これに対して、エッジ入力を用いた場合は学習の立上りは遅いが、学習後期にタスク精度がより向上していることが分かる．これは 3.1 節で指摘したように、人間の注視方向を示す情報として、オプティカルフローは大まかではあるが直接的にその方向を示しているのに対して、エッジは段階的な解釈を必要とするが、より正確な注視方向を推定できる情報を提供しているためと考えられる．そして、これらの結果とフローとエッジ両方の入力を用いた場合の結果を比較すると、後者は学習の立上りも速く、学習後期には精度の高いパフォーマンスを得ていることが確認できる．このことから、提案したモデルは二種類の入力の両方の長所を活かすことができたといえる．また、前研究 [13] [14] で入力情報として画像の輝度値を用いたときの実験結果（学習回数  $10^5$  回で同パフォーマンスに到達）と比較しても学習が加速しており、動き等価性とポピュレーション符号化の利用によって、学習が効率的になったといえる．

##### 4.2.2 NN の学習収束率の変化

次に、学習過程でのフロー NN とエッジ NN の収束率の変化を調べた．Fig. 7 (b) に式 (9) (10) における range  $o_f$  と range  $o_{e'}$  の値の変化を示す．これらが、式 (8) で最終的な運動出力への貢献度として用いられる．

この結果から、学習初期ではフロー NN からの出力がより高い割合で採用され、学習後期ではフロー NN とエッジ NN からの出力がほぼ均等に採用されていることが分かる．特に、学習後期においてフロー NN とエッジ NN を共に活用することで、個々の NN を用いたときよりも高い精度を得ることができた (Fig. 7 (a) 参照)．これは、フロー入力とエッジ入力が情報のずれに対する頑健性と情報の正確性の点で相補的な関係にあり、両 NN からの運動出力を合成することで互いの短所が補われたためと考えられる．つまり、本来正確な注視方向を推定できるエッジ NN が、その正確さゆえにテンプレートマッチングの大きなずれによって生じる入力情報の変化に対応できず、いくつかの状況において正確な運動出力を生成することができなかった．これに対して、フロー NN の運動出力はエッジ NN のそれに比べて元来正確性に欠けているが、その分上記のような問題に対して頑健であり、画像入力のずれに対しても常にある程度の正確性をもった運動出力を生成することができた．この結果、両 NN からの運動出力をその収束率に応じて合成することで、両者が相補的に機能し、より精度の高いパフォーマンスが得られたものと推測される．

#### 4.3 NN の結合荷重の解析

知覚 - 運動情報のポピュレーション符号化によって NN がどのように構造化されたのかを確認するため、学習後の NN の結合荷重を解析した．

##### 4.3.1 フロー NN の解析

まず、フロー NN の入力層と出力層の間の結合荷重を Fig. 8 に図示する．(a) が学習前、(b) が学習後である．縦軸と横軸がそれぞれ入力層ニューロンと出力層ニューロンの選択的方向を示しており、対応するマスの色の濃度でそのニューロン間の結合荷重を表している．解析に際して、フロー入力は特にその方向性の情報が重要であると考えられることから、ここでは入



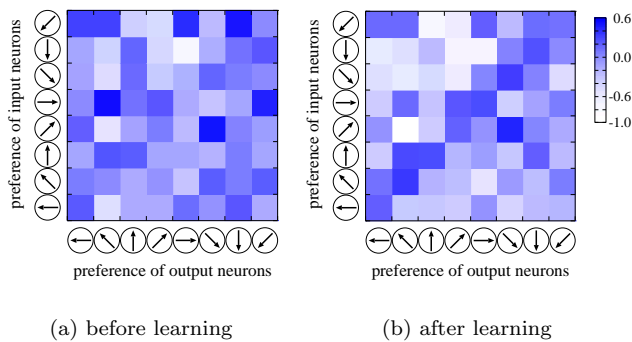


Fig. 8 Connecting weights of flow-NN

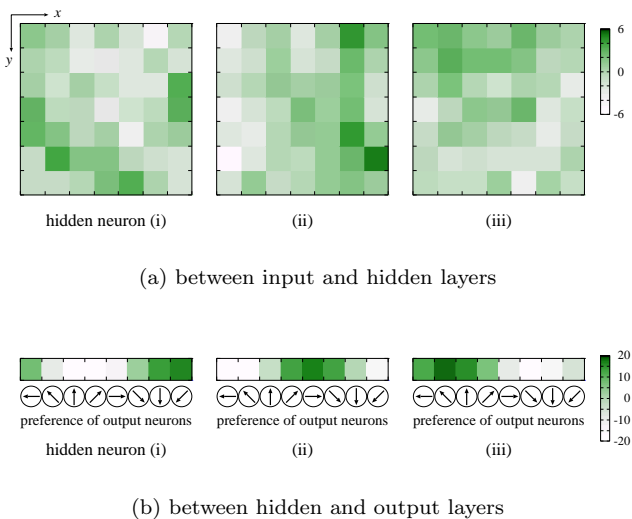


Fig. 9 Connecting weights of edge-NN after learning

力層ニューロンの属性のうち受容野の位置は無視し、全受容野にわたった選択的方向ごとの平均値を代表結合荷重として表示した。

この結果より、学習前にはランダムに設定されていた結合荷重が、学習後には左下から右上への対角線上で、相対的に高い結合荷重を有するような構造化が起きていることが確認できる。この対角線はちょうど入力層ニューロンと出力層ニューロンの選択的方向が一致するところに相当しており、このことから、入力層と出力層の同質なニューロン間の強い結合として自他の動き等価性が獲得できていることが分かる。ここで、学習データのうち共同注意が成立し、オプティカルフローの方向性と運動出力の方向性とが一致していたのは約半数であったことに注意されたい。尚、入力層ニューロンの受容野の位置を考慮した解析も行ったが、位置に関してはこのような構造化は見られなかった。

#### 4.3.2 エッジ NN の解析

次に、エッジ NN の学習後の結合荷重の解析結果を Fig. 9 に示す。(a) が入力層から中間層への結合荷重、(b) が中間層から出力層への結合荷重である。それぞれ 3 つの図が 3 個の中間

層ニューロン (i), (ii), (iii) に対応しており、(a) が各中間層ニューロンと受容野ごとの入力層ニューロンとの平均結合荷重を、(b) が出力層ニューロンの選択的方向ごとの結合荷重を表している。つまり、(a) では入力層ニューロンの方位性が無視され、 $7 \times 7$  の受容野の画像空間に対応した結合加重が表示されていることになる。これは、エッジ入力のもつ情報として、エッジの方位性よりも画像内での空間的な位置の方が、人間の視線方向を推定する上でより重要と考えたためである。

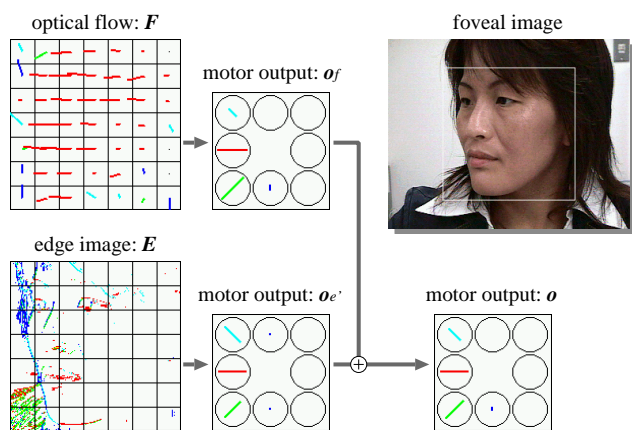
これらの結果についてまず (b) を解析すると、3 個の中間層ニューロンがそれぞれ特定の運動方向に反応選択性をもつ出力層ニューロンと強く結び付いていることが確認できる。(i) は左下方向、(ii) は右方向、(iii) は左上方向の運動に対応した出力層ニューロンと最も強く結合し、さらにその近傍のニューロンともガウス関数で近似される結合荷重を有している。このことから、中間層ニューロンでは出力の運動方向に対しての役割分担が行われ、3 個のニューロンの組合せで全ての運動方向に対応していることが分かる。ここで、中間層でどのような役割分担が行われるかは NN の初期設定に依存し、他の設定では異なる分担 (例えば、上、左下、右下) が獲得されることが確認されている。

また、この解析結果を踏まえて (a) を観察すると、左下方向の運動出力を担当する中間層ニューロン (i) は入力画像の左下の受容野に属するニューロンと強い結合をもち、同様に右方向、左上方向の運動出力を担当する (ii) と (iii) は、それぞれ画像の右半分、左上の受容野に属するニューロンと強い結合をもっていることが分かる。これは、エッジ入力とその画像上での位置を特に重要な情報としていることを意味し、エッジの画像空間での偏りによって人間の注視方向が推定されていることを示している。さらにエッジの方位性の役割も調べるため、(a) の結合荷重を入力層ニューロンの選択的方向も考慮して解析したところ、目や口の位置に対応する受容野 (それぞれ、 $2 \cdot 3$  行目と  $5 \cdot 6$  行目の受容野) では、特に水平方位に反応選択性をもつ入力層ニューロンと各中間層ニューロンとが強く結合されていることが確認できた。これは、人間の目や口の要素が水平方位のエッジとして検出されることからきており、エッジ NN ではエッジの位置情報だけでなく、方位性に関する情報も同時に利用されていることが分かる。以上の解析結果から、ポピュレーション符号化によって入出力情報の方向性や方位性、また空間的位置を考慮した構造化された NN が獲得できたといえる。

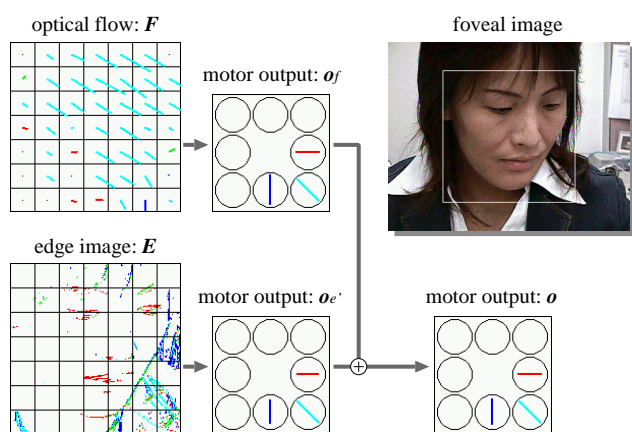
#### 4.4 共同注意の実現

学習で獲得した NN を Infanoid に実装して、人間との共同注意実験を行った。実験条件は学習時とほぼ同じで、同一の人物が対応した。Fig. 10 に試行ごとのロボットの入出力結果を示す。(a) は人間が画面左方向の対象物を注視した場合、(b) は右下方向の対象物を注視した場合である。ロボットは、図右上の中心視カメラ画像からオプティカルフローとエッジを抽出し (図左)、各 NN で出力を計算した後 (図中央)、それらを学習の収束度に応じて統合した最終出力 (図右下) から眼球と頭部の回転運動を生成する。それぞれの結果から、人間の注視方向と運動出力の方向とがきれいに対応しており、NN が適切な視覚・運動対を獲得できていることが確認できる。実際にこの出





(a) when the person is looking at the left



(b) when the person is looking at the lower right

Fig. 10 Joint attention experiments after learning

力でロボットの視線方向を変化させたところ、正しく対象物を発見して共同注意を成立させることができた。対象物を人間とロボットの間テーブル上でさまざまな位置と高さで設置して実験した結果、共同注意の成功率は90% (18/20 試行)であった。失敗した2試行ではロボットの視線変化方向は対象物の方向から約15 [deg] ずれており、許容誤差 ( $\pm 8$  [deg]) を上回っていた。これは、テンプレートマッチングのずれによって生じたエッジ NN からの運動出力の誤差が、フロー NN からの出力によって補正しきれなかったことと、それを含めて、学習データがあらゆる状況を網羅できていなかったことが原因と考えられる。これについては、顔検出方法を改善することで対処することが望まれる。

尚、本実験では学習時と学習後の評価時とで同一の人物が対応したが、異なる人物でも同じ NN で共同注意を実現できるものと期待される。前研究 [13] [14] で画像の全画素の輝度値を入力として用いた際は、学習モデルは画像の変化に対して敏感であり、人物の違いに対応することは困難であった。これに対し

て、本モデルは4.3節の解析実験で確認したとおり、フロー NN では人間の頭部回転時の動き方向を、エッジ NN では人間の顔特徴の偏りを利用して人間の視線方向を推定している。前者は人物の違いによらず不変的な特徴であり、後者もエッジを受容野ごとに扱ったことで個人の顔形状の違いは無視され、偏りという不変的な特徴にのみ注目していると考えられる。このことから、共同注意能力の汎化という点でもポピュレーション符号化が有効に機能しているといえる。

### 5. おわりに

本稿では、共同注意の学習モデルとして自他の動き等価性を利用したモデルを提案した。共同注意は自己と他者の視線変化運動の間に等価な関係を見出すことで発達すると考えられる [6] が、等価性の発見に際し、自他の動きが視覚と体性感覚という異なる感覚様式で検出されることが問題となる。提案モデルでは NN の情報表現としてポピュレーション符号化を採用し、視覚と体性感覚の動き情報を感覚様式に依存しない方向選択性ニューロンの活動として表現することで、自他の動き等価性を両感覚上の同質なニューロンの同時活動として検出できる枠組を実現した。学習実験の結果、提案モデルによって自他の動き等価性が自律的かつ早期に発見できること、そして、それによって共同注意の学習が加速されることを示した。また、ポピュレーション符号化によって視覚 - 運動情報の方向性、方位性、空間的位置といった属性が適切に表現され、NN の構造化が促進されることを確認した。

提案モデルはロボットの自律的学習モデルとしてだけではなく、認知発達ロボティクス [1] の視点から人間の認知機能を解明する上でも重要な役割をもつ。まず実験結果に類似した発達学的知見として、視線追従を行う際に、発達初期の幼児だけが他者の動きを重要な手がかりとして使用し、発達後期の幼児は他者の静止した注視姿勢だけからその視線方向を正しく追従できることが知られている [5] [17]。これは模倣能力にも共通して見られる特徴である [19] [20] が、なぜそのような情報利用の発達の变化が起きるのかは明らかではない。これに対して本稿では、動き情報と静的情報としてオプティカルフローとエッジという二種類の視覚入力を用いることで、前者が学習初期における学習の高速化に貢献し、後者が学習後期の高精度化に貢献するという相補的な役割を担うことを実験的に示した。そして、その理由としてフローの解釈のしやすさとエッジの情報としての正確性を挙げ、それらが学習の各段階で適切に利用されていることを指摘した。学習実験で見られた視覚入力の貢献度の変化と幼児の発達の变化とは現象面において一致しており、実験で得られた見解も幼児の認知発達を説明する理由として有効であると考えられる。また、発達初期の幼児が他者の頭部回転の動きに誘発されて反射的に頭部を回転させるといった現象 [6] [18] も、提案モデルで採用した自他の動き間の直接的な結合によって引き起こされるものと説明できる。

フローやエッジ、頭部の運動ベクトルといった特徴量の利用とそのポピュレーション符号化によって、NN の構造化が促進され共同注意能力の汎化につながったことも、人間の認知機能、特に神経機構を理解する上で重要な発見と考えられる。動物の

脳ではフローやエッジといった視覚特徴や身体の運動ベクトルという体性感覚情報が抽出され、それらが複数のニューロン活動の集合として符号化されていることが発見されている [24] ~ [27]。この特徴量の抽出と符号化のメカニズムは動物が様々な視覚 - 運動経験をとおして獲得したものであり、共同注意タスクに特化されたものではない。それにもかかわらず、本稿ではその情報処理メカニズムの利用によって、共同注意の学習が効率化されることを示した。ここで採用したメカニズムは実際のニューロンのそれに比べて非常に単純ではあるが、タスク学習の効率化を確認したことで、ニューロンにおける情報処理機構の発達の解明にもつながっていくと期待される。

さらに、この感覚情報表現を用いたモデルが、共同注意だけではなく幼児の様々な認知機能を説明するモデルとして発展していくことが期待できる。本稿の実験からポピュレーション符号化が異種感覚間の等価性の早期発見に役立つことが確認されたが、幼児はそのような等価性に敏感であることが指摘されている。新生児模倣 [18] [31] をはじめとして、視覚と触覚の感覚間協応 [32] や、視覚と聴覚の感覚間協応 [33] [34] など、幼児が生後まもなくから異種感覚間の等価性を検出できるという証拠が多く発見されている。そして、これらの能力は言語使用などの高次の認知機能の発達の礎ともされており、これらを提案モデルと同じ枠組で統合的に扱うことで、高次認知機能の発達メカニズムの解明にもつながっていくと期待される。

以上の課題に取り組む際に議論しなければならないのは、タスクに応じた特徴量の選択と、高次のニューロンにおける情報表現の問題である。本稿で提案したモデルでは、共同注意タスクに有効とされる感覚情報を設計者があらかじめ選択し、そしてそれらを方向性や方位性といった低次の特徴量でのみ取り扱っていた。これに対して、人間は種々の感覚から常に膨大な情報を得ており、状況に応じて注目すべき特徴量を適切に切替えている。また、脳内での情報処理も、低次の特徴量だけではなくタスクに応じた高次の特徴量を抽出することで、タスクをより効率的に遂行していると考えられる。前者についてはすでに関連研究 [35] もあり、ロボットがタスク学習と同時に、そのタスクや置かれた環境に応じて重要と思われる特徴量を抽出するフィルタを学習するモデルが提案されている。今後は、このような特徴量選択の仕組みとその情報表現が、タスク学習をとおしていかに同時に発達していくかも検討していく。

## 参考文献

- [1] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi. Cognitive developmental robotics as a new paradigm for the design of humanoid robots. *Robotics and Autonomous Systems*, 37:185–193, 2001.
- [2] C. Moore and P. J. Dunham, editors. *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum Associates, 1995.
- [3] M. Scaife and J. S. Bruner. The capacity for joint visual attention in the infant. *Nature*, 253:265–266, 1975.
- [4] G. Butterworth and N. Jarrett. What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9:55–72, 1991.
- [5] C. Moore, M. Angelopoulos, and P. Bennett. The role of movement in the development of joint visual attention. *Infant Behavior and Development*, 20(1):83–92, 1997.
- [6] V. Corkum and C. Moore. The origins of joint visual attention in infants. *Developmental Psychology*, 34(1):28–38, 1998.
- [7] C. Breazeal and B. Scassellati. Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, 8(1):49–74, 2000.
- [8] B. Scassellati. Theory of mind for a humanoid robot. *Autonomous Robots*, 12:13–24, 2002.
- [9] M. Imai, T. Ono, and H. Ishiguro. Physical relation and expression: Joint attention for human-robot interaction. In *Proceedings of the 10th IEEE International Workshop on Robot and Human Communication*, pages 512–517, 2001.
- [10] H. Kozima and H. Yano. A robot that learns to communicate with human caregivers. In *Proceedings of the First International Workshop on Epigenetic Robotics*, pages 47–52, 2001.
- [11] E. Carlson and J. Triesch. A computational model of the emergence of gaze following. In *Proceedings of the 8th Neural Computation and Psychology Workshop*, 2003.
- [12] B. Lau and J. Triesch. Learning gaze following in space: a computational model. In *Proceedings of the Third International Conference on Development and Learning*, 2004.
- [13] 長井, 浅田, and 細田. ロボットと養育者の相互作用に基づく発達の学習モデルによる共同注意の獲得. *人工知能学会論文誌*, 18(2):122–130, 2003.
- [14] 長井, 細田, 森田, and 浅田. 視覚注視と自己評価型学習の機能に基づくブートストラップ学習を通じた共同注意の創発. *人工知能学会論文誌*, 19(1):10–19, 2004.
- [15] A. Morita, Y. Yoshikawa, K. Hosoda, and M. Asada. Joint attention with strangers based on generalization through joint attention with caregivers. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3744–3749, 2004.
- [16] A. Shon, D. Grimes, C. Baker, M. Hoffman, S. Zhou, and R. Rao. Probabilistic gaze imitation and saliency learning in a robotic head. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2876–2881, 2005.
- [17] J. D. Lempers. Young children's production and comprehension of nonverbal deictic behaviors. *The Journal of Genetic Psychology*, 135:93–102, 1979.
- [18] A. N. Meltzoff and M. K. Moore. Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology*, 25(6):954–962, 1989.
- [19] S. W. Jacobson. Matching behavior in the young infant. *Child Development*, 50:425–430, 1979.
- [20] A. Vinter. The role of movement in eliciting early imitations. *Child Development*, 57:66–71, 1986.
- [21] Y. Kuniyoshi, Y. Yorozu, M. Inaba, and H. Inoue. From visuomotor self learning to early imitation - a neural architecture for humanoid learning. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 3132–3139, 2003.
- [22] M. Ogino, S. Matsuyama, J. Ooga, and M. Asada. Motion recognition and generation for humanoid based on visual-somatic field mapping. In *Proceedings of the Third International Conference on Development and Learning*, 2004.
- [23] A. Pouget, P. Dayan, and R. Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1:125–132, 2000.
- [24] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148:574–591, 1959.
- [25] S. Treue, K. Hol, and H. Rauber. Seeing multiple directions of motion — physiology and psychophysics. *Nature Neuroscience*, 3(3):270–276, 2000.
- [26] A. P. Georgopoulos, A. B. Schwartz, and R. E. Kettner. Neuronal population coding of movement direction. *Science*,

- 233:1416–1419, 1986.
- [27] J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K. Chapin, J. Kim, S. J. Biggs and M. A. Srinivasan, and M. A. L. Nicolelis. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408:361–365, 2000.
- [28] S. Deneve, P. E. Latham, and A. Pouget. Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, 2(8):740–745, 1999.
- [29] S. Deneve, P. E. Latham, and A. Pouget. Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, 4(8):826–831, 2001.
- [30] H. Kozima. Infanoid: A babybot that explores the social environment. In K. Dautenhahn, A. H. Bond, L. Canamero, and B. Edmonds, editors, *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, chapter 19, pages 157–164. Amsterdam: Kluwer Academic Publishers, 2002.
- [31] A. N. Meltzoff and M. K. Moore. Imitation of facial and manual gestures by human neonates. *Science*, 198:75–78, 1977.
- [32] A. N. Meltzoff and R. W. Borton. Intermodal matching by human neonates. *Nature*, 282(5737):403–404, 1979.
- [33] P. K. Kuhl and A. N. Meltzoff. The bimodal perception of speech in infancy. *Science*, 218(10):1138–1141, 1982.
- [34] D. J. Lewkowicz and G. Turkewitz. Cross-modal equivalence in early infancy: Auditory-visual intensity matching. *Developmental Psychology*, 16(6):597–607, 1980.
- [35] 港 and 浅田. 注視機構実現に向けた視覚 - 行動学習による画像特徴と状態空間の構成. 日本ロボット学会誌, 21(1):87–93, 2003.

---

長井 志江 (Yukie Nagai)

1974年11月9日生。1999年青山学院大学大学院理工学研究科機械工学専攻博士前期課程修了。2002年10月大阪大学大学院工学研究科知能・機能創成工学専攻博士後期課程単位取得退学。同年11月同大学大学院工学研究科産学官連携研究員。2004年(独)情報通信研究機構専攻研究員。2006年ドイツビーレフェルト大学研究員となり現在に至る。共同注意や模倣能力の獲得など認知発達ロボティクスの研究に従事。博士(工学)。(日本ロボット学会正会員)