

報酬予測に基づく初期コミュニケーションの獲得

Acquiring the early communication based on the reward prediction model

大井手 友美 (JST Erato, 阪大) 渡辺 絢子 (JST Erato, 阪大)
正 荻野 正樹 (JST Erato, 阪大) 正 浅田 稔 (JST Erato, 阪大)
Tomomi OOIDE, Osaka University, 2-1, Yamadaoka, Suita, Osaka
Ayako WATANABE, Osaka University, 2-1, Yamadaoka, Suita, Osaka
Masaki OGINO, Osaka University, JST Erato Asada Project
Minoru ASADA, Osaka University, JST Erato Asada Project

This paper proposes a learning model which enables a baby robot to acquire the early communication in human development. The robot stores the time sequence of sensor information in its memory when the internal state rises up by the sudden sensor change such as big sound or face detection. The memory helps the robot to predict the response of the caregiver. The experimental result shows that the robot can acquire one of the early communications, peekaboo, by the proposed system.

Key Words: Formation of memory, Internal state, Reward prediction

1 はじめに

現在、様々な観点からロボットに人間らしさを取り入れたコミュニケーションロボットの研究が行われている。

本研究では、ロボットに内部状態をもたせ、人間同士が行っているようなコミュニケーションを行うことを目指す。そのため、本研究では乳児期のコミュニケーションに焦点を当てる。

乳児は生後4ヶ月までには、養育者が与える規則性のある行動に敏感になる¹⁾。そして、同時期に乳児は自分の母親のタイミングや相互作用における相対的な随伴性への調律を発達させ始め、またこの調律を乳児自身に向けられる行動が母親と一致している見知らぬ女性へも般化させる傾向があることが示唆されている²⁾³⁾。そして、乳児が対面的やりとりにおいて相互交換をするようになった時から、社会的パートナーに関する予測を発達させる²⁾。以上から、乳児はコミュニケーションを行う際に、養育者のタイミングに同調し、それをを用いて記憶したことを予測するようになると考えられる。

そこで、本研究では、認知科学や心理学そして、脳科学に基づき、ロボットの内部状態を構成する。そして、乳児でみられるコミュニケーション能力をロボットに学習させ、自然で適応的なコミュニケーションを生み出すことを目的とする。

脳科学では、外界の予測のためにドーパミンニューロンが重要な役割を果たしていると考えられている。大脳基底核にあるドーパミンニューロンは報酬を受けると発火する。そして、どういう出来事で報酬が受けられるのかを学習すると、その一連の出来事の開始を認識した際に発火し、報酬の予測がはずれて報酬が受けられないと活動が抑制されるという反応をする⁴⁾。また、脳科学に

おける、記憶に関する知見として、人間の脳にある海馬は扁桃体から大きな情動刺激を受けるとその刺激の前後の出来事を鮮明に記憶する⁵⁾がある。

そこで、本研究ではロボットが情動に基づいて養育者の行動を記憶し、その記憶の報酬に基づいて養育者の行動の予測を行うモデルを提案する。また、養育者の行動を記憶していない場合は、センサ入力に対して情動変化を起こし、記憶がある場合には、記憶と予測に基づいて情動変化を起こす情動モデルを提案する。そして、乳児と養育者が行う典型的な遊びとして「いないいないばあ」を用いて以上のモデルが妥当であるかを検証する。

2 提案するコミュニケーションモデルとシステムのメカニズム

2.1 乳児と養育者のコミュニケーション「いないいないばあ」

本研究の目的は、養育者のタイミングに基づく規則性の学習と報酬予測に関するモデルを提案することである。実証のために、乳児のコミュニケーションの実例として、「いないいないばあ」を用いる。

「いないいないばあ」は乳児が養育者の感情、情緒、情動に同調するようになるやりとりとして挙げられている⁶⁾。又、社会的相互作用における規則性を検出する能力と、これらの規則性をもとに特定の予測を発達させる能力を捉える研究¹⁾の際に、実験手段として用いられている。

2.2 いけないいないばあによるコミュニケーションモデルとメカニズムの概要説明

モデルのシステムは、センサ入力部、報酬予測モジュール・記憶モジュール・内部状態モジュールの3つのモジュール、そして、表情表出部から成っている。各モジュール

の関わりの概要をいらないばあをコミュニケーションモデルとして挙げながら説明する。

入力と同じ記憶をもっている場合ともっていない場合でコミュニケーションの内容が異なるがロボット内部にもっているシステムは同じなので、主に入力と同じ記憶をもっている場合について解説する。入力と同じ記憶がある場合のコミュニケーションの流れを Fig.1 を用いて示し、次に、そのときのロボット内部の概要を Fig.2 を用いて示す。

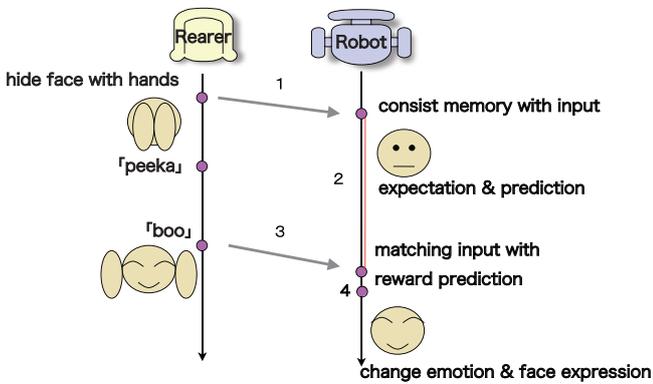


Fig.1 The communication model of memory

Fig.1 に書かれた番号毎に、コミュニケーションの流れを説明する。

1. 養育者が手で顔を隠したことを、ロボットは「いないいないばあ」の記憶と結びつけそれが始まること予測し始める。
2. 養育者が手で顔を隠している間、養育者の顔がいつ現れるかを予測そして期待している。
3. 養育者が手を降ろし、「ばあ」と言ったときに予測が合っていたのかを比較し、そのタイミングのずれで情動が変化する。
4. その情動変化に応じて、表情表出をする。

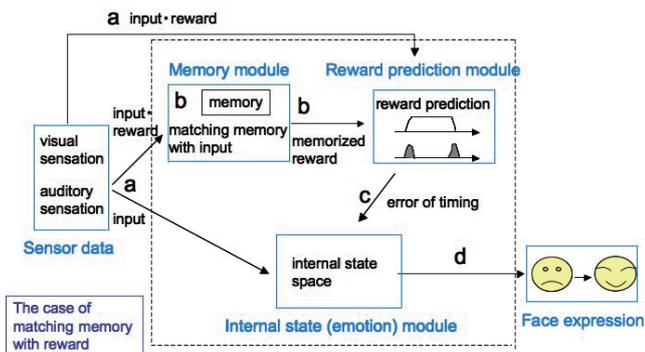


Fig.2 The case of having memory

Fig.2 をアルファベット順に解説する。なお、働いている機能を青の枠で囲っている。

- a センサ情報とその時の報酬を各モジュールに入力する。
- b そのときのセンサ情報と報酬を記憶し、また記憶と入力の比較で一致した記憶の報酬を報酬予測モジュールに送る。そして、記憶モジュールでは一致した記憶と入力を随時比較する。
- c 報酬予測モジュールに入力として報酬が入ってきた時、予測したタイミングとの差を内部状態モジュールに渡す。
- d c に対する情動変化に基づいて、表情表出する。

入力と同じ記憶がない場合は、入力情報に対して直接情動変化を示し、表情表出する。

3 システム

4 記憶モジュール

記憶モジュールの役割は2つあり、一つは入力情報の記憶、もう一つは記憶と入力の比較である。記憶の内容はセンサ情報(音と画像)・内部状態・報酬である。

記憶と入力の比較は、記憶の先頭部分の比較と記憶と入力の先頭部分が一致した後の比較に分けられる。記憶の先頭部分の比較とは、ある時刻 t の入力と時刻 $t-1$ の入力の二つを時系列で繋がった入力として、記憶と比較するというものである。そして、記憶と入力の先頭部分が一致した後は、その一致した記憶と各時刻で入力に対して随時比較が行われる。また、記憶の最後の部分が現れるタイミングにはある程度のズレが許容されると考えられるので許容範囲を設けて記憶と入力の比較を行う。

用いている記憶の種類には、「短期記憶と長期記憶」がある。短期記憶は、常に行われており、各時刻の入力を保存している。また、短期記憶には限界があるので、古い記憶から順に上書きされていく。長期記憶は次の二つの場合に形成される。一つは、入力と先頭部分が一致する記憶が無い場合に、一定以上の内部状態が入力されたときである。この場合、そのときより前に一定以上の内部状態が入力された出来事までを記憶する。そして、時系列として繋がっている長期記憶は一つの長期記憶として認識する。もう一つは、入力と先頭部分が一致する記憶が存在した場合に、ある時点で記憶と異なる出来事が起きた場合である。この場合、記憶と一致した部分から一致しなかった部分までを長期記憶とする。長期記憶にも限界があるが、限界に達したら使われていない記憶から消えるようになっている。

5 報酬予測モジュール

Schultz et al.⁴⁾ は次に示すようなドーパミンニューロンの働きを発見した。ドーパミンニューロンは報酬を受けると発火する。そして、報酬を受けられる行動を学習すると、その一連の出来事の開始を認識した際に発火し、報酬の予測が外れて報酬がうけられないと活動が抑制さ

れるという反応をするものである。この Shultz et al. が示したドーパミンニューロンの発火パターンは強化学習で用いられる TD 誤差の振舞として説明できる⁷⁾。

以上から、このモジュールでは TD 誤差の式 $\delta(t)$ を用いてドーパミンの発火の有無を示す。

$$\delta(t) = r(t) + \gamma V(t) - V(t-1) \quad (1)$$

$r(t)$ は報酬を表す。報酬の種類は 2 種類あり、1 つは顔が現れたこと、もう一つは音が入力されたことである。 V は期待報酬である。 V は TD 学習を用いて計算する。また、 γ は割引率である。 γ は 1 とする。

6 内部状態モジュール

人間の情動モデルとして Russell の 2 次元情動モデルを用いる。本研究では、表情と内部状態は結びついていないが、将来渡辺ら⁸⁾が行った、表情と内部状態の結びつけを行い、他者の表情から他者の情動状態がわかるというロボットの内部環境モデルを本研究に取り入れるために、同様の情動モデルを用いた。Russell の 2 次元情動モデルは「快-不快」「覚醒-睡眠」を最も基本的な感情意味次元と考え、様々な情動のカテゴリーがこの 2 つの次元で構成される空間上に位置付けられている。

Russell の 2 次元情動モデルから、ロボットの内部状態空間を作成した。(Fig.3 参照) これは、「快-平常」「覚醒-平常」の 2 次元からなる。本研究の実験で用いる報酬が快のもののみなので、記憶に対する予測を行う際に、睡眠と不快の次元は重要でないと考えたためである、

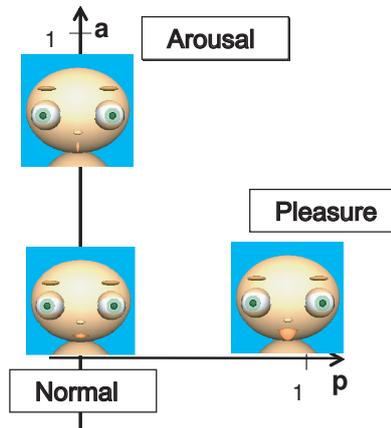


Fig.3 Facial expression

ロボットの内部状態は Plesasure(快),Arousal(覚醒) それぞれに対応した

$$\mathbf{S} = (p, a) \begin{cases} (0 < p < 1) \\ (0 < a < 1) \end{cases} \quad (2)$$

で表す。 $\mathbf{S} = (p, a)$ は Fig.3 で示した空間上を動く、Pleasure,Arousal の度合いを示す状態変数である。ロボットはこの内部状態変数に従い表情を表出させる。

渡辺ら⁸⁾の内部状態変化ダイナミクスの式を基に、ロボットの内部状態のダイナミクスを表す。式(3)にその式を示す。

$$\begin{cases} \tau \dot{\mathbf{S}} = -\mathbf{S} + \Sigma \mathbf{r}_j + \mathbf{S}_0 \\ \Sigma \mathbf{r}_j = \mathbf{r}_i + \mathbf{r}_e \end{cases} \quad (3)$$

\mathbf{S}_0 は定常時の内部状態である。 \mathbf{r}_i は入力と一致した記憶がない場合のセンサ情報に対して内部状態を変化させる変数ベクトルである。 \mathbf{r}_e は入力と一致した記憶がある場合に、報酬を受け取ったときのタイミング誤差によって内部状態を変化させる変数ベクトルである。 τ の要素 τ_p, τ_a は、 $\mathbf{S} = (p, a)$ の時定数である。

Rochat et al.¹⁾の実験指標において、乳児の注視が指標となっていたため、乳児の馴化を表現するために、渡辺ら⁸⁾の内部状態変化ダイナミクスに興味度を新たに加えた。乳児にとって、養育者の顔は特別なものである⁹⁾、顔に対して興味度を設けた。興味度 \mathbf{I} を表す式を以下に示す。

$$\mathbf{I} = 1.0 - hab_m \times \frac{time}{at} \quad (4)$$

$time$ はロボットが養育者に興味を持ち始めてからの時間を表す。 hab_m は同じ記憶が呼び出された回数を表す。 at は養育者の顔に依存した定数を表す。

$$at = \begin{cases} 100 & (face\ appear) \\ 150 & (face\ hide \rightarrow face\ appear) \\ 80 & (face\ appear \rightarrow face\ hide) \end{cases} \quad (5)$$

$face\ appear$ は顔が現れている状態を表す。 $face\ hide$ は顔が隠れている状態を表す。

7 実験設定

7.1 実験内容

本研究では Rochat et al.¹⁾が行った実験を基に、モデルの検証を行う。2ヶ月児を記憶を持たない時期、4・6ヶ月児を記憶を持つ時期と捉え、両方の状態に対して規則的ないないいいいばあ和不規則ないないいいいばあを行い、ロボットの内部状態、表情を変化させる実験を行った。規則的ないないいいいばあとは、養育者が手で自分の顔を隠し、「いないいいい」と言って、ある時間が過ぎたら、手を降ろして「ばあ」と言う。といった一般的なものである。これを、続けて3回繰り返す。不規則ないないいいいばあとは、声だけで「いないいいいばあ」として、手で顔を隠して、何も言わずに顔を出すと言うものである。以上2つを不規則ないないいいいばあとして、規則的ないないいいいばあ後に、連続で行う。

7.2 実験条件

Rochat et al.¹⁾の実験によると、各いいいいいいばあに対して2ヶ月児は笑顔の量と注視について同様な反応を示した。我々は、これを記憶がまだ出来ておらず、他

者の行動を予測できない段階と定義し、記憶モジュール、報酬予測モジュールを取り除いて、入力に対する情動表出だけを見た。

また、4・6ヶ月児は規則的ないないないばあよりも不規則ないないないばあにおいて注視の量が増え、笑顔の量が減った。これは、記憶に基づいて、他者の行動を予測しているため、記憶と同じことが起きた場合に喜び、記憶と異なることが起きた場合に、驚き注視する、と考えた。そのため、4・6ヶ月児を想定した場合には、全モジュールを用いた。

8 実験結果

各実験に対して、時系列でのロボットの興味度・内部状態・顔の出没・報酬・音の入力について示し考察を行う。また、出来事を記憶出来る場合において、形成された記憶に対して述べる。

以下に示す図の見方を説明する。図には、上から順に、実験時の興味度・内部状態・報酬・顔の出没・音の入力についての時間的変化を示した。上部の画像はその時刻の入力画像と入力画像に対するロボットの表情である。図中のロボットは音に対して反応した際のロボットの表情である。縦軸に各データの内容、横軸に時間を示した。図上のカラーは、青に近いほど値は小さくピンク色に近いほど値は大きい。また、興味度・内部状態・音は連続な値であり、報酬と顔の出没は0または1で表す。なお、顔の出没については、顔が出ている場合が1であり、隠れている場合が0である。図中の赤い線がいないないばあの開始を表している。そして、緑の線がいないないばあの終了を表している。また、不規則ないないないばあの開始を黄色、終了を青色で示した。

4・6ヶ月を想定した規則的ないないないばあを行った場合の実験結果の図を Fig.4 に示す。

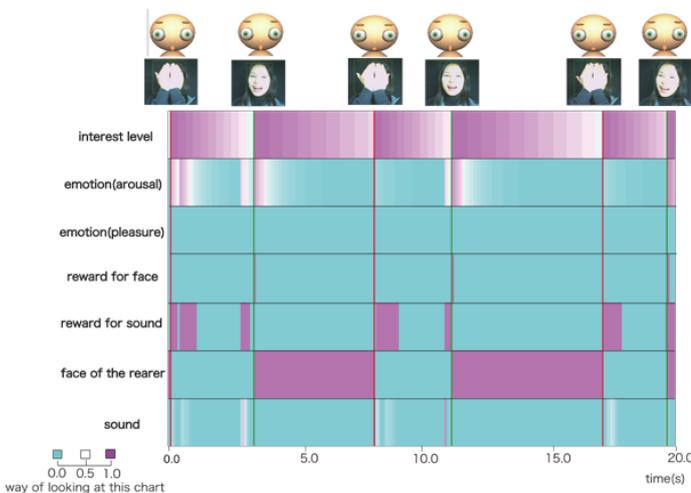


Fig.4 The internal state of a robot for order peekaboo. a robot is 4 months older

各入力に対するロボットの表情を見てみると、二回目以降のいないないばあの開始時にロボットは驚いてい

ない。これは、一回目のいないないばあで記憶ができてからは、それ以降は手で顔を隠すことでいないないばあが行われることを予期するためである。このように、出来事を記憶をし、入力と比較する機能は正常に働いている。

4・6ヶ月を想定した不規則ないないないばあを行った場合は、顔を手で隠していないないばあを言わない場合に、ロボットの情動は余り変化しなかった。これは最初のいないないばあが記憶されているため、開始部分では規則的ないないないばあだと判断したからである。そして、記憶と異なり音の入力がなかったため、不規則ないないないばあ終了時に驚きの感情が現れた。このように、記憶を行う機能が正常に働いていた。

また、2ヶ月児を想定した実験においては、Rocha et al.¹⁾と同じように、両方のいないないばあに対して同程度の注視が見られた。

参考文献

- [1] Philippe Rochat, Jane G. Querido, and Tricia Striano. Emerging Sensitivity to the Timing and Structure of Protoconversation in Early Infancy. *Developmental Psychology*, 35(4):950-957, 1999.
- [2] philippe Rochat. *THE INFANT'S WORLD*, chapter 4. Harverd University Press, 2001.
- [3] Ann E. Bigelow. Infants' sensitivity to imperfect contingency in social interaction. *Early social cognition*, pages 137-154, 1999.
- [4] W. Schultz, P. Dayan, and P. F. Strick. A neural substrate of prediction and reward. *Science*, 275:236-250, 1997.
- [5] James L. McGaugh. *MEMORY AND EMOTION*. Orion Publishing Group Ltd, 2003.
- [6] Alan Fogel. *Developing through Relationships*. Chicago: University of Chicago Press, 1993.
- [7] P. Read Montague, Peter Dayan, and Rerrence J. Sejnowski. A Framework for Mesencephalic Dopamine Systems Based on Predictive Hebbian Learning. *Neuroscience*, 16(5):1936-1947, March 1996.
- [8] 渡辺 絢子, 荻野 正樹, and 浅田 稔. 他者観測による内部状態と表情の範疇獲得. In *ロボティクス・メカトロニクス講演会 '06 予稿集*, pages 1P1-E15, 2006.
- [9] M.H. Johnson and J. Morton. Biology and cognitive development. *Blackwell*, 1991.