

人 - ロボットインタラクションにおける「ためらう」ロボットの 実験的評価

An experimental valuation of a robot who “hesitates” in human-robot interaction

田中 一晶^{1*} 岡 夏樹¹
TANAKA Kazuaki¹ OKA Natsuki¹

¹ 京都工芸繊維大学 大学院工芸科学研究科

¹ Graduate School of Science and Technology, Kyoto Institute of Technology

Abstract: If robots learn new actions through human-robot interaction, it is desirable that the robots can utilize rewards as well as direct teachings to reduce humans' efforts. However, in the process of reinforcement learning, which is based on rewards, we found that humans had difficulty in understanding the intention of a robot who dared to take an action of a low expected reward, and that the robot did not given proper instructions. We hence thought that it must be useful if the robot can show his intention of daring to take a low-reward action by hesitation. In this study, we designed AIBO to express hesitation by the delay of his actions. We conducted an experiment in which AIBO learned shaking hands through human-robot interaction. The experiment demonstrated that a robot who hesitates learned more quickly, and made better impression on humans about teachability.

1 はじめに

将来、人の日常生活の場で、ロボットが人の仕事をサポートすることが予想される。そのようなロボットに予め必要な行動を全てプログラムしておくことは現実的ではない。よって、人と触れ合うロボットは人や環境とのインタラクションを通して新たな行動を獲得する能力が必要不可欠である。

ロボットが人とのインタラクションを通して学習する場合、人から与えられる直接教示のみをあてにして学習するだけでは人の負担が大きいため、報酬に基づく学習(強化学習)を併用できることが望ましい。また、人が人とのインタラクションを通じて学習する場合には、学習者の能力に応じて、足場として簡単な学習課題から徐々に難しい学習課題を与えること(Scaffolding)が知られている[8]が、先行研究では、ロボットとのインタラクションでも、人は足場を与えることを明らかにした[5]。人から与えられる報酬を利用してロボットが学習する場合、人から新しい課題(足場)が与えられると報酬は増加し、ロボットが課題を達成するにつれて報酬は減少する。そして、そのようなインタラクションでは、強化学習における探索行動(探索のために報

酬の期待値が低い行動をあえて実行すること)の意図は人に伝わりにくく、ある程度学習しているにも関わらず、まだ学習が足りないという誤解を人に与えてしまい、Scaffoldingの妨げになるという問題が浮かび上がった[5]。そこで、我々は、ロボットが探索行動の際に「ためらう」ことによって、最適行動(報酬の期待値が高い行動)と異なる行動を取ろうとしていることが人に伝われば、人から適切な教示が与えられると考えた。

ロボットの感情や意図を表現し、人とのインタラクションに利用した研究は[1, 2]が挙げられる。これらの研究では、ロボットに多くの表情や振る舞いを実装しているが、「ためらい」に対応する表情や振る舞いは実装されておらず、また、ロボットの表情や振る舞いは研究者が作りこんで実装したものである。これに対し、本研究では、ロボットの「ためらい」を特定の振る舞いとして実装するのではなく、ロボットが次の行動を実行するまでの時間を利用して表現する。具体的には、報酬の期待値が高い行動は即座に実行するが、期待値の低い探索行動は間をおいて(ためらって)から実行するように、動的に時間を決定する手段を用いる。

本論文では、ロボットが人とのインタラクションを通して学習する実験を実施し、人-ロボットインタラクションにおける「ためらい」の有用性を「学習効率」と「人の印象」の両面から明らかにする。

*連絡先: 京都工芸繊維大学 大学院工芸科学研究科
〒606-8585 京都市左京区松ヶ崎橋上町
E-mail: d8821007 @ edu.kit.ac.jp

2 学習アルゴリズムと「ためらい」の表現方法

ここでは、本研究で採用した学習アルゴリズムと、ロボットの「ためらい」の表現方法について説明する。

2.1 Q-Learning

本研究では、学習アルゴリズムとして Q-Learning[6]を採用する。Q-learning では、状態 s における行動 a_n の価値 $Q(s, a_n)$ を行動を行うたびに報酬 r に基づいて更新し、最適行動を学習する。状態 s で行動 a_n を実行した結果、状態 s' に遷移した際に、以下の更新式に従って実行した行動 a_n の Q 値を更新する。

$$Q(s, a) \leftarrow Q(s, a) + \alpha \{r + \gamma \max_a Q(s', a) - Q(s, a)\} \quad (1)$$

ここで、 $\max_a Q(s', a)$ は、遷移後の状態 s' で最大の Q 値を持つ行動 a の Q 値であり、将来得られる報酬の期待値を表している。 α は学習率、 γ は割引率であり、本研究では、それぞれ 0.1、0.3 とした。また、報酬 r については次節にて説明する。

また、行動の選択方法は Boltzmann 選択を採用し、Boltzmann 温度は 0.03 とした。

2.2 評価教示と直接教示

本研究では、ロボットは人から与えられる評価教示と直接教示を併用して学習する。

人がロボットに与えることができる評価教示は、「撫でる」という正の報酬 r と「叩く」という負の報酬 $-r$ の二種類とし、 $r = 0.1$ とした。これらの刺激の識別と、それに対する感情表出には、人が AIBO の頭のタッチセンサに触れた瞬間から、刺激に応じてその時点での尤もらしい表出（喜ぶ、悲しむ）を行う随時表出システム [4] を利用する。

人から a_n を実行するように直接教示 I が与えられた際には、必ず行動 a_n を実行し、報酬 r に $I = 0.1$ を加算して $Q(s, a_n)$ を更新する。強化学習に直接教示を併用する研究は [7] が挙げられる。この研究では、ポテンシャル関数によって動的に I を決定しているが、本研究では簡単のために定数を使用する。

また、人とロボットとのインタラクションでは、学習が進み、ロボットが望ましい行動を選択するようになると、それに伴って人が与える評価教示は減少することがわかっている [5]。一般的な Q-Learning では、報酬が与えられなくなると Q 値はゼロに漸近してしまうが、この報酬の減少には、NNC(No News 規準)[3]による報

酬を利用することで対応できることが明らかとなっている [5] ため、本研究でも同様の方法を利用する。NNC は教示が与えられないことを肯定的な評価と捉える暗黙的な評価規準である。これにより、人から報酬が与えられなかった場合は、NNC による報酬 $r = Q(s, a_n)$ に基づいて学習し、現在の Q 値を維持する。

2.3 「ためらい」の表現方法

本研究では、ロボットの「ためらい」を次の行動を実行するまでの時間で表現する。状態 s において行動 a_n を実行するまでの時間 $T(s, a_n)$ は、行動 a_n の選択確率 $p(s, a_n)$ に従って、以下の式で動的に決定する。

$$T(s, a_n) = t_0 + t_1 \{1 - p(s, a_n)\} \quad (2)$$

ここで、 t_0 は最短の時間であり、 t_1 は t_0 から延長する最大の時間である。これにより、行動の選択確率が高い行動ほど即座に実行し、行動の選択確率が低い探索行動は間をおいてから（ためらってから）実行する。本研究では、 $t_0 = 0.1[s]$ 、 $t_1 = 3.0[s]$ とした。

3 評価実験

本節では、我々が提案する「ためらい」の人 - ロボットインタラクションにおける有用性の評価実験について説明する。

3.1 実験設定

本実験では、図 1 に示すように、AIBO が「おて」という音声を認識すると、立った状態から座って右手を出すという動作を、人から与えられる評価教示や直接教示を利用して学習するタスクを設定する。AIBO には 2 節で説明した学習アルゴリズムと「ためらい」の表現方法を AIBO の開発環境である RemoteFramework を使用して実装する。

Q-Learning で使用する各状態は、以下の入力 $i_0 \sim i_7$ に従って定義する。

$i_0 \sim i_2$: それぞれ、立っている状態、座っている状態、伏せている状態であれば (1)、それ以外は (0)。

$i_3 \sim i_6$: それぞれ、左手、右手、左足、右足を出している状態であれば (1)、それ以外は (0)。

i_7 : 「おて」という音声を認識すると (1)、初期状態 s_0 に遷移すると (0) に初期化する。

例えば、図 1 内の 4 つの状態 ($s_0 \sim s_3$) は、初期状態では $s_0 : 10000000$ 、初期状態で「おて」を認識すると $s_1 : 10000001$ 、「おて」を認識し、座って右手を出す

○ 次のような「おて」が出来るように教えて下さい



※「おて」と言ってから
教えて下さい

○ 使える教示は次の6つです



図 1: 実験協力者に提示した「おて」の手順とインストラクション

実験協力者には () 内の文字を消したものを提示した。

s_3 : 01001001 というように定義する。以降、これらの状態は $s_0 \sim s_3$ で参照する。

また、「おて」の音声認識は、AIBO の耳に付属するマイクを介し、RemoteFramework のライブラリを利用して行う。

AIBO が各状態で実行する行動 $a_0 \sim a_6$ は、それぞれ、立つ、座る、伏せる、左手を出す、右手を出す、左足を出す、右足を出すの 7 種類とした¹。

3.2 実験方法

我々が提案する「ためらい」の効果を評価するため、次の行動を実行するまでの時間 T を以下のように設定し、3 つの条件での実験を実施する。

条件 1(ためらい条件): T を行動の選択確率 $p(s, a_n)$ に応じ、式 (2) によって動的に決定する。

条件 2(せっかち条件): 式 (2) で、行動の選択確率 $p(s, a_n)$ が 0.9 のときと同じ時間に固定する。つまり、常に $T = 0.4[s]$ の時間で次の行動を実行する。

条件 3(おっとり条件): 式 (2) で、行動の選択確率 $p(s, a_n)$ が 0.1 のときと同じ時間に固定する。つまり、常に $T = 2.8[s]$ の時間で次の行動を実行する。

一つの実験につき 20 分間、各実験協力者で実施する順番を変えて合計 60 分の実験を実施した。実験の前には、 $p(s, a_n)$ が 0.5 のときと同じ時間に固定した状態

¹但し、立った状態で行動 a_0 (立つ) は実行できないので、その際は状態遷移に関係しない遊び行動 (足を上げておしっこをするような振る舞いなど) を代わりに実行する。座った状態における a_1 (座る) や、伏せた状態における a_2 (伏せる) も同様である。また、立った状態では後足を出すことはできないので、 a_5 と a_6 の代わりに足を蹴り上げる動作を行う。その際も状態遷移は起こらないものとした。

(つまり、常に $T = 1.6[s]$ の時間で次の行動を実行する) の AIBO と 10 分程度、練習を行ってもらった。

実験協力者には、予め図 1 を A4 の紙 1 枚で提示し、口頭でも紙面の内容と同様の説明を行った。

実験は 6 人の実験協力者に行ってもらった。実験協力者の属性と実験 (条件 1~3) を実施した順番を表 1 に示す。

表 1: 実験協力者の属性と実験 (条件 1~3) を実施した順番

実験協力者	属性	年齢	条件の順番
A	社会人	20 代	1 2 3
B	本学学生 (工学系)	20 代	1 3 2
C	他大学学生 (医学系)	20 代	2 1 3
D	社会人	20 代	2 3 1
E	主婦	50 代	3 1 2
F	本学学生 (工学系)	20 代	3 2 1

実験では、プログラム上で実験開始から 20 分間の各状態での Q 値と与えられた教示を時間ごとに記録した。実験で評価するポイントは以下の 3 点である。

- 我々が提案する「ためらい」の表現方法が学習速度にどのような影響を与えるのか。
- 人が感じるロボットの「教えやすさ」と「生き物らしさ」が「ためらい」によって変化するか。
- 人は実際にロボットがためらっているように感じるか。

3.3 実験結果

図 2 に a) 状態 s_1 と b) 状態 s_2 における正しい行動 (それぞれ a_1 (座る)、 a_4 (右手を出す)) の選択確率の各条件ごとの推移を示す。グラフの横軸は Q 値の更新回数、縦軸は行動の選択確率を表し、各点は実験協力者 6 名の結果を平均したものである。

表 2: アンケートの結果

実験協力者	教えやすかった順番	生き物らしいと思った順番
A	3 1 2	1 3 2
B	1 3 2	2 1 3
C	1 3 2	3 1 2
D	1 2 3	2 1 3
E	1 2 3	2 1 3
F	1 3 2	2 1 3

次に、実験後に実施したアンケートの結果を表 2 に示す。アンケートの内容は表中に示す通り、条件 1~3 において 1) 教えやすかった順番、2) 生き物らしいと思った順番の 2 つである。また、アンケート結果を踏まえてインタビューを実施した。その内容を表 3 に示す。

表 3: インタビューの結果

質問:	3つの条件(AIBO)の違い
対象:	全ての実験協力者
回答:	(A,C,E,F): 学習の速さが違う (B,D): 行動の早さが違う
質問:	行動の早さに対する印象
対象:	実験協力者 B,D
回答:	(B,D): 行動が遅いときは指示を待っていると思った。 (B,D): 条件1(ためらい条件)は適度な早さだった。 (B,D): 条件2(せっかち条件)は早すぎ、条件3(おっとり条件)は遅すぎる。
質問:	条件1が一番教えやすいと思った理由
対象:	実験協力者 B,C,D,E,F
回答:	(C,D,F): 条件3と違いは無いように思ったが、なぜか教えやすかった。 (B,D): 速すぎず、遅すぎず、行動のテンポが良く教えられた。 (B,E): 自発的である。
質問:	条件3が一番教えやすいと思った理由
対象:	実験協力者 A
回答:	(A): 一旦覚えた後、余計な行動をあまりしなかった。
質問:	条件2が一番教えにくいと思った理由
対象:	実験協力者 A,B,C,F
回答:	(A,C,F): すぐ覚えたと思ったのに余計な行動をし過ぎる。 (C,F): わがままで、教えた通りにしてくれない。 (A): やる気が無い。 (B): 行動が速過ぎる。
質問:	条件3が一番教えにくいと思った理由
対象:	実験協力者 D,E
回答:	(D): ずっと指示待ちで、イライラした。 (E): 自発的で無い。
質問:	条件1が一番生き物らしいと思った理由
対象:	実験協力者 A
回答:	(A): 余計な行動もするが、きちんと学習もする。
質問:	条件2が一番生き物らしいと思った理由
対象:	実験協力者 B,D,E,F
回答:	(B,D,E,F): 生き物の犬のように好き勝手に行動する。 (B,D): 反応が早い。 (E): 突然違う行動をするのが犬らしい。 (F): 気まぐれであるため。
質問:	条件3が一番生き物らしいと思った理由
対象:	実験協力者 C
回答:	(C): 行動の種類が多い気がした。
質問:	条件2が一番生き物らしくないと思った理由
対象:	実験協力者 A,C
回答:	(A,C): 生き物にしては、ちゃんと学習していないと思った。
質問:	条件3が一番生き物らしくないと思った理由
対象:	実験協力者 B,D,E,F
回答:	(D,E,F): 機械的な印象がある(ロボットらしい)。 (B,D): ずっと指示待ちであり、犬らしくない。

4 考察

4.1 「ためらい」が学習効率に与える影響

実験では、実験協力者は、学習の初期段階では直接指示と評価指示を併用し、AIBO が正しい行動を選択するにつれて徐々に直接指示は与えなくなっていく。そして、さらに学習が進むと評価指示も徐々に与えなくなっていくという指示の与え方が観察された。これは、先行

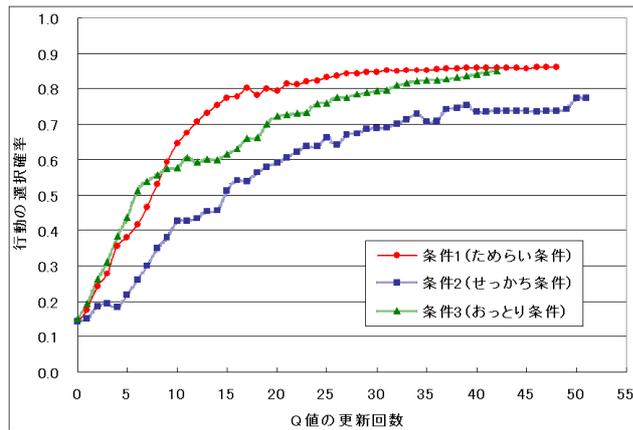
研究 [5] でも観察された報酬の増加と減少 (Scaffolding) が生じたためである。

まず、図 2(状態 s_1 と状態 s_2 における正しい行動の選択確率の推移) を参照されたい。状態 s_1 、 s_2 共に、1~3 全ての条件で選択確率が上昇しているが、条件 1(ためらい条件) と条件 2(せっかち条件) を比較すると、条件 1 の方が学習が早いことがわかる。条件 2 では AIBO の行動と行動の間隔 T (以下、行動間隔) が短く ($T = 0.4[s]$)、学習の初期段階において誤ったタイミングで直接指示が与えられたり、誤った行動に評価指示が与えられることが度々起こっていた。これに対し、条件 1 では、全ての行動の実行確率が低い学習の初期段階では、行動間隔も長くなるため、適切なタイミングで直接指示や評価指示が与えられていた(行動の種類は 7 つであり、初期状態での実行確率は 0.14 であるため、式 (2) より $T = 2.68[s]$ となる)。この違いが学習速度の差となって表れたのだと考えられる。

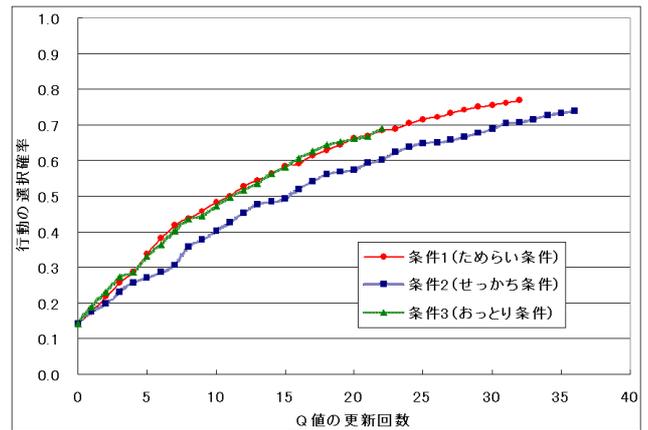
一方、条件 3(おっとり条件) では、行動間隔が常に長い ($T = 2.8[s]$)、条件 1(ためらい条件) と同様に適切なタイミングで直接指示や評価指示が与えられていた。そのため、図 2 では、学習速度に大きな差は見受けられない。ここで、表 4 に、各状態における 20 分経過した時点での正しい行動の選択確率を実験協力者ごとに示す。条件 1 と条件 3 を対応ありの t 検定で比較した結果、状態 s_1 では $t(5) = 0.66, n.s.$ 、状態 s_2 では $t(5) = 3.22, p < .05$ となり、状態 s_1 では統計的な有意差は無いが、状態 s_2 では条件 1 の方が有意に学習が進んでいることがわかった。20 分経過した時点では、状態 s_1 では十分に学習が進んでおり、既に Q 値が収束しているため、実行確率に差は無かったのだと考えられる。しかし、状態 s_2 では 20 分経過時点ではまだ学習途中であり、条件 3 は条件 1 ほど学習が進んでいなかったためにこの差が表れたのである。この原因として、20 分間に状態 s_2 で Q 値が更新された回数が条件 1 より条件 3 の方が少なかったことが挙げられる(表 5 参照)。条件 2 では実行確率に関わらず、2.68[s] の行動間隔で行動を実行するが、条件 1 では行動の実行確率が高くなるにつれて、その行動を実行するまでの行動間隔は徐々に短くなるため、20 分間に状態 s_2 で Q 値が更新された回数が増加したのだと考えられる。

表 4: 20 分経過時点での正しい行動の選択確率

条件協力者	状態 s_1			状態 s_2		
	条件 1	条件 2	条件 3	条件 1	条件 2	条件 3
A	0.87	0.60	0.89	0.80	0.72	0.81
B	0.89	0.86	0.86	0.82	0.77	0.77
C	0.89	0.69	0.87	0.81	0.77	0.77
D	0.85	0.81	0.89	0.85	0.79	0.77
E	0.95	0.89	0.95	0.82	0.80	0.77
F	0.75	0.84	0.65	0.83	0.82	0.79
平均	0.867	0.782	0.851	0.819	0.779	0.779



(a) 状態 s_1 (初期状態 s_0 で「おて」を認識した状態) での行動 a_1 (座る) の選択確率



(b) 状態 s_2 (状態 s_1 で行動 a_1 を選択し座った状態) での行動 a_4 (右手を出す) の選択確率

図 2: 各状態における正しい行動の選択確率の推移

表 5: 20 分間に各状態で Q 値が更新された回数

実験協力者	状態 s_1			状態 s_2		
	条件 1	条件 2	条件 3	条件 1	条件 2	条件 3
A	57	70	56	36	41	34
B	48	64	43	54	42	38
C	50	56	48	44	43	41
D	59	51	42	51	59	53
E	62	68	60	32	36	22
F	48	52	50	60	50	42
平均	54.0	60.2	49.8	46.2	45.2	38.3

ここで、条件 1(ためらい条件) と条件 2(せっかち条件) においても、20 分経過時点での選択確率の比較を対応ありの t 検定で行ったところ、状態 s_1 では $t(5) = 1.56, n.s.$ 、状態 s_2 では $t(5) = 4.14, p < .01$ となり、条件 3 との比較と同様に、状態 s_1 では有意差は無かったが、状態 s_2 では条件 1 の方が有意に学習が進んでいることがわかった。これは先に考察したように、条件 1 の方が条件 2 よりも学習が早いいため、最終的な実行確率も条件 1 の方が高くなったと考えられる。しかし、行動間隔は条件 2 が最も短いため、状態 s_2 に遷移する回数は条件 2 が最も多くなるはずであるが、表 5 の通り、平均では条件 1 が最も多いという結果となった。これは、そもそも条件 2 の学習速度の方が遅いため、状態 s_2 に遷移する回数も増加しなかったのであろう。

まとめると、1) 学習の初期段階では、人から直接教示・評価教示を待つように、行動間隔は長い方が良いが、2) ある程度学習が進んだ後は、一人一人の教示を待つのではなく、選択確率が高い行動は素早く実行することで、学習が促進されることがわかった。我々が提案した「ためらい」の手段(条件 1)はこの両方の要求をを満たしているため、最も学習効率が良かったのだと考えられる。

4.2 「ためらい」に対する人の印象

本研究で実施した 3 つの実験では、学習率や行動の選択確率を決める値である Boltzmann 温度は全て同じであり、異なるのは行動間隔(次の行動を実行するまでの時間)のみとした。それにもかかわらず、3 つの条件(AIBO)の違いについて、6 人の実験協力者の内 4 人が「学習の速さが違う」と回答している(表 3 を参照)。そして、アンケート結果(表 2)の通り、6 人中 5 人が「ためらい」の表現方法を実装した条件 1(ためらい条件)が最も教えやすかったと解答している。これは、4.1 節で考察したように、条件 1 が最も学習効率が良かったことが起因しているのかもしれないが、条件 1(ためらい条件)と条件 3(おっとり条件)では最終的な学習状態(正しい行動の実行確率)に差はあるが、学習速度に差は無いため、より大きな要因として、我々が提案する「ためらい」の表現方法が、暗黙的に人の印象に良い影響を与えていると考えられる。条件 1 が一番教えやすかった理由として、「条件 3 と違いは無いように思ったが、なぜか教えやすかった」という曖昧な回答をした実験協力者が多かった(3 人)ことや、条件 3 が一番教えにくかったと感じた実験協力者が「イライラした」「自発的でない」と回答をしていることから、行動間隔が暗黙的な情報として作用している可能性がある。

一方、一番教えにくかった条件として、条件 2(せっかち条件)が最も多く挙げられていた(5 人)が、一番生き物らしいと思った条件でも、条件 2 が最も多く挙げられていた(4 人)。教えにくかった直接的な原因は、4.1 節で考察した通り、行動間隔が短かった(行動が早過ぎた)ために教示が与えにくかったからだと考えられるが、実際にそう回答した実験協力者は少なく(1 人)、なかなか学習しないことを理由として述べていた(3 名)。図 2 からわかるように、条件 2 は条件 1 や 3 と比べる

と学習速度は明らかに遅いため、人の印象にも強く影響を与えたのであろう。しかし、学習が遅く、ランダムに様々な行動を実行する振る舞いに対し、人は「わがまま」や「やる気がない」といった印象を持っており、これが生き物らしいと感じた理由となったのだと推測する。これは、実験に使用した AIBO が「犬型」のロボットであるということが大きく影響しているため、人型のロボットで実験した場合は、違う結果となるかもしれない。

また、行動間隔の違いに気づいた実験協力者 2 名に行動の早さに対する印象を尋ねたところ、「行動が遅いときは、指示を待っていると思った」と回答しており、「ためらい」という直接的な表現ではないが、行動間隔に対して、我々が意図した捉え方をしている。しかし、それぞれの条件に対する印象では、「条件 2 は早すぎ、条件 3 は遅すぎる」と的を得た回答しているが、条件 1 に対しては「適度な早さだった」と回答しており、行動間隔が動的に変わっていたことには気づいていなかった。学習開始時は「ためらって」行動するが、徐々にためらわずに行動できるようになることは、生き物として自然な振る舞いであるため、早さの変化を意識することなく、その時々で「適度な早さ」であると感じたのであろう。

5 まとめと今後の展開

本研究では、ロボットの内部状態(学習状態)を人にフィードバックする手段として、「ためらい」に着目し、その振る舞いを明示的に表現するのではなく、「時間」による暗黙的な情報を利用してロボットがためらう方法を提案した。そして、人-ロボットインタラクションを通して、我々が提案する「ためらい」が学習効率・人の印象の両面において良い影響を与えることを明らかにした。

我々の提案方法では、行動の実行確率に応じてその行動を実行するまでの時間を動的に決定し、実行確率が高い行動は即座に実行するようになる。よって、実験協力者が毎回直接教示を与えていると、学習が進むにつれてロボットの行動は徐々に早くなるため、直接教示を与える前に行動するようになり、実験協力者はそれ以降直接教示を与えなくなるという場面が多く観察された。このように、ロボットの学習状態が適切にフィードバックできており、実験協力者の負担の軽減に繋がっている。しかし、ロボットが探索行動(あえて実行確率の低い行動を実行すること)の際に、ロボットの行動が遅くなった場合には、実際にロボットがためらっていると感じた実験協力者は少なかった(6人中2人)。探索行動が「まだ学習が足りない」という誤解を与えるという、先行研究で浮かび上がった問題は生じ

ていなかったが、それが「ためらい」の効果によるものかどうかは明らかとなっていない。

今後の展開として、我々は「ためらい」がより有効にはたらく条件について検討していく。

謝辞

本研究の一部は科学研究費補助金 基盤研究(C)「3項間インタラクションを通じた認知発達メカニズムの解明」の支援を受けた。

参考文献

- [1] C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. *IEEE/RSJ International Conference*, Vol. 2, pp. 858–863, 1999.
- [2] A. Brooks, J. Gray, G. Hoffman, A. Lockerd, H. Lee, and C. Breazeal. Robot's play: interactive games with sociable machines. *Computers in Entertainment*, Vol. 2, No. 3, pp. 1–18, 2004.
- [3] K. Tanaka, X. Zuo, Y. Sagano, and N. Oka. Learning the meaning of action commands based on no news is good news criterion. *Workshop on Multimodal Interfaces in Semantic Interaction*, pp. 9–16, 2007.
- [4] 田中一晶, 岡夏樹. ペットロボットによる感情表出のタイミングがユーザとのインタラクションに与える影響. HAI シンポジウム 2006, 1B-1, 2006.
- [5] 田中一晶, 岡夏樹. Scaffolding(足場づくり)を利用した学習系の構築. FIT2008 第7回情報科学技術フォーラム, RJ-006, 2008.
- [6] C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, Vol. 8, No. 3-4, pp. 279–292, 1992.
- [7] E. Wiewiora, G. Cottrell, and C. Elkan. Principled methods for advising reinforcement learning agents. *Proceedings of the Twentieth International Conference on the Machine Learning*, pp. 792–799, 2003.
- [8] D. Wood, J. S. Bruner, and G. Ross. The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry*, Vol. 17, pp. 89–100, 1976.