

強化学習による 合目的かつ接地した一方向コミュニケーションの創発

○柴田克成 笹原冬月* (大分大学)

Emergence of Purposive and Grounded One-Way Communication by Reinforcement Learning

*Katsunari Shibata and Kazuki Sasahara (Oita University)

Abstract— Communication is not just the manipulation of words, but needs to decide what is communicated considering the surrounding situations and to understand the communicated signals considering how to reflect it on the actions. In this paper, aiming to the emergence of purposive and grounded communication, communication is seamlessly involved in the entire process consisted of one neural network in each agent, a transmitter or a receiver, and no special learning for communication but reinforcement learning is used to train it. A real robot control task was done in which a transmitter agent generates two sounds from 1,785 real camera image signals representing the state of the robot field, and a receiver agent controls the robot according to the received sounds. After learning, appropriate communication was established to lead the robot to the goal. It was found that, for the learning, the experience of controlling the robot by the transmitter is useful, and the correlation between the communication signals and robot motion is important.

Key Words: Grounded Communication, Emergence, Reinforcement Learning, Neural Network, Robot

1 はじめに

現在、喋ることができるロボットは多数存在し、中でも、相互にやりとりするロボットは知的存在に見える。人間と機械に対し、相手が見えない状態で別々にやりとりをし、機械が審判員に「人間が相手だ」と思わせることができるかを競うチューリングテスト¹⁾において、近年、短時間のテキストのやりとりで、審判員が機械との会話を人間との会話と間違えるケースもあると言う²⁾。これは大きな驚きであり、知性とは何かという問いに一石投じているように見える。しかしながら、人間のような知性を持っていると感じるロボットが見当たらないという現実から考えると、この現象は、特定の問題に特化し、人間のような知性を持った存在のように振る舞ういわゆる「弱い AI」³⁾の優れた一面を示していると考えられる。しかしこのようなアプローチでは、人間によるプログラミング、もしくは、人間が実際に行った会話の真似に大きく依存するため、その能力にはどこかに限界があると予想される。したがって、単なる言葉のやりとりを越え、状況や相手の気持ちを理解したより人間的で知的な会話の実現はこのアプローチの延長線上にはなく、“Artificial General Intelligence”とも呼ばれ、思考をし、心を持つと言われるいわゆる「強い AI」³⁾を指向したアプローチへの質的な変換が必要であると筆者らは考えている。

求められる大きな質的な変換の一つとして、長年コミュニケーションにおいて問題となって来た「シンボルグラウンディング(記号接地)問題」⁴⁾がある。会話が接地されていないということは、ロボットが言葉の意味を理解しないまま、手順にしたがって言葉を発していることを意味するので、これでは強い AI を指向したコミュニケーションとは言いがたい。中野ら⁵⁾や Steels ら⁶⁾の語彙の創発モデルでは、提示された物体から抽出された特徴と言葉や記号を結びつけることによって、パターンとシンボルを融合させることが提案

されている。そのモデルは、基本的には、2 エージェントが同じものを眺めているという仮定の下に、聞き手の言葉を話し手の言葉に近づけていくことで両者間に共通の意味を持つ言葉(シンボル)が形成されていく。

ここで問題となるのは、パターンとシンボルを元々別のものとして扱い、両者の間を橋渡しするために、他の学習とは独立した特別な学習が必要となることである。Steels は⁶⁾の中で、「この記事で議論されている実験では、エージェントは言語ゲームをプレーすることが前提になっているが、いったいゲームそれ自身はどうやって創発するのか？」と自ら問いかけている。この問いかけは問題の核心を突いている。

そもそも動物や原始人が用いる原始的なコミュニケーションは、食物の場所を知らせたり、危険が近づいていることを知らせるといったように、生きることに直結した目的を持っているように見える。つまり、敵から逃れ、食物にありつくために、状況を的確に認識して、必要な情報を記憶し、適切な行動やコミュニケーションを行うのだと捉えれば、コミュニケーションも、行動と同様に強化学習などによって合目的に学習することができる。さらに、ニューラルネットを用いれば、認識から行動までをトータルに学習できる⁷⁾のと同じように、コミュニケーションも単なるシンボル操作ではなく、認識や記憶などの他の機能と密接に関係し合って生成されるようになると期待される。また、シンボルは、相手への伝達、もしくは、自分自身の内部での論理的な思考のために創発する信号の一形態であると考えれば、パターンとシンボルをつなぐための特別な学習をわざわざ用意しなくても、強化学習のような合目的な学習の中でシンボルは創発する可能性もある⁹⁾¹⁰⁾。そうすれば、自ずとパターンとシンボル間の密で自然なつながりが形成され、シンボル接地問題の解決にもつながると期待される。現代人のコミュニケーションは必ずしも生きることに直結していないように見えるが、何らかの形で生きることに間接的に関

*現在、スズキ(株)勤務

わっている、このアプローチの延長線上に存在する可能性は高いのではないかと考えている。

われわれの脳の中には言語野があると言われているが、言語野はその他の領野から孤立しているわけではなく、また、その構造も他の領野とそれほど違っているようにも見えない。コミュニケーションは、脳の言語野だけで作られているのではなく、超並列で柔軟な処理システムとしての脳全体が作り出したものであろう。このように考えても、コミュニケーションは日々の生活の中で創発すべきものであり、その他の機能の学習と切っても切れないものと考えられる。そもそも、シンボルとパターンを別々のものとして扱うために「シンボル接地問題」が生じてしまうと考えられる。

合目的なコミュニケーションの創発は、進化的なアプローチによるもの¹¹⁾や強化学習によるもの¹²⁾がすでに示されている。著者らのグループも強化学習を通じてのコミュニケーションの獲得を目指して来た¹³⁾¹⁴⁾⁹⁾¹⁰⁾。その中で、ノイズのある環境で学習させると、コミュニケーション信号が離散化することも示し、シンボル創発との関連性を指摘した⁹⁾¹⁰⁾。しかし、これらでは環境がとても単純で、コンピュータシミュレーションによってのみ示されて来たため、特にコミュニケーションの「接地」という点で説得力がなかった。

そこで本論文では、接地された合目的なコミュニケーションの獲得を示すため、実際のカメラ、スピーカー、マイクロフォン、ロボットを用いたタスクを用意し、送信者が1,000個以上のカラーの画像信号をニューラルネットに入力し、適切な周波数の音を2つ生成して出力することを学習し、受信者は受信した音から、ロボットへの適切な動作信号を生成してコントロールすることを学習する。それぞれのエージェントはニューラルネットを使って、ロボットがゴールに着いたときの報酬と壁に近づいた時の小さな罰だけから強化学習によって並行して学習していく。そして、単に報酬や罰による学習を通して、本当にセンサ信号に基づいた接地したコミュニケーション信号の合目的な生成ができるかどうかを確認する¹⁵⁾。ただしここでは、入出力ともに連続値とし、シンボルの創発は次の課題とした。

カメラを使ったコミュニケーションロボットの研究もすでにいくつかある¹⁶⁾¹⁷⁾¹⁸⁾。しかし、カメラはコミュニケーションの相手を見つけたり、相手に見つめられているという感覚を与えるために使われており、本研究で目指すように、カメラ画像をコミュニケーションに直接反映させ、両者の有機的な統合を学習によって獲得した例は見当たらない。

2 ニューラルネットを用いた強化学習⁷⁾

強化学習は試行錯誤に基づいた自律的かつ合目的な学習であり、ニューラルネットは通常、次元の呪いのための状態の爆発を防ぐための非線形関数近似器として用いられる。著者らのグループでは、この組み合わせによって、さまざまなことを同時に考慮することができる並列処理が合目的的、シームレスに、そして調和的に学習され、その結果、報酬を得て罰を避けるために、認識や記憶(リカレントネットを用いた場合)などの必要な機能が内部に創発することを提唱して来た。この柔軟で並列な処理によって、各機能ごとに独立に優れたプログラムを開発し、それを統合して知能

ロボットを構成する「機能モジュール」アプローチに別れを告げ、「フレーム問題」を根本的に解決することに有効であると期待される。

そのシステムは、センサ信号を入力し、アクチュエータコマンドを出力するニューラルネットにより構成される。強化学習に基づいて教師信号を自律的に生成して教師あり学習を行うため、外部からは報酬と罰を与えるだけで、逐一教師信号を与える必要はない。この論文では、連続値入出力を扱うため、強化学習の手法として、Actor-Critic¹⁹⁾を用いる。したがって、ニューラルネットの出力はCriticの出力 P とActorの出力 \mathbf{a} に分けられる。Actorの出力 \mathbf{a} はベクトルであり、探索要素としての乱数ベクトル \mathbf{rnd} を加えた後、アクチュエータの動作信号として用いられる。TD誤差は

$$\hat{r}_{t-1} = r_t + \gamma P(\mathbf{s}_t) - P(\mathbf{s}_{t-1}) \quad (1)$$

と計算される。ここで、 r_t は時刻 t に与えられる報酬、 γ は割引率、 \mathbf{s}_t は時刻 t でのセンサ信号で、これがニューラルネットの入力となる。 $P(\mathbf{s}_t)$ は、 \mathbf{s}_t が入力の場合のCriticの出力である。Criticの教師信号は、

$$P_{d,t-1} = P(\mathbf{s}_{t-1}) + \hat{r}_{t-1} = r_t + \gamma P(\mathbf{s}_t) \quad (2)$$

と計算される。そして、Actorの教師信号は

$$\mathbf{a}_{d,t-1} = \mathbf{a}(\mathbf{s}_{t-1}) + \hat{r}_{t-1} \mathbf{rnd}_{t-1} \quad (3)$$

と計算される。ただし、 $\mathbf{a}(\mathbf{s}_{t-1})$ は \mathbf{s}_{t-1} が入力のときのActorの出力であり、 \mathbf{rnd}_{t-1} は探索のために $\mathbf{a}(\mathbf{s}_{t-1})$ に加えられた乱数ベクトルである。そして、この $P_{d,t-1}$ と $\mathbf{a}_{d,t-1}$ を教師信号として、 \mathbf{s}_{t-1} を入力としたニューラルネットを誤差逆伝搬(Back Propagation, BP)法で1回だけ教師あり学習させる。ここでは、ニューロンの出力関数に-0.5から0.5の値域のシグモイド関数を用いたため、ニューラルネットの出力の値域と実際のCriticの出力の値域を調整するために上記の教師信号からは0.5を引き、ニューラルネットの出力には0.5を足して上記の式で用いた。Actorについてはタスクに依存するので、後述する。学習は非常に簡単で汎用的であり、コミュニケーションの学習のためやこのタスクのための特別な学習は行っていないことがわかる。

3 ロボット制御タスクと学習に用いた手法

3.1 システムアーキテクチャとロボット制御タスク

Fig. 1にシステムアーキテクチャと行ったタスクの様子を示す。30cm×30cmのフィールドに置かれた小型の移動ロボット(e-puck)と送信者と受信者からなる。送信者は固定されたカメラを持ち、上からフィールドを見下ろす。そして、カメラ画像のRGBの画素値を入力ベクトル \mathbf{s} として、ニューラルネットに入力する。さらに、スピーカーを有し、2つの音を出力する。2つの音の周波数はActorの出力 \mathbf{a} と試行錯誤成分の \mathbf{rnd} の和のそれぞれの成分を1,000Hzから1,300Hzの範囲に線形変換して決定した。2つの音はそれぞれ1秒間の正弦波とし、少しの間を空けて続けて出力した。プログラムのバグのため、発信した信号の周波数は、実際には約20Hz小さい音であった。受信者はマイクを持ち、送信者からの2つの音を受け取る。受信者もニューラルネットを有し、受信した2つの音をそれぞれFFTにか

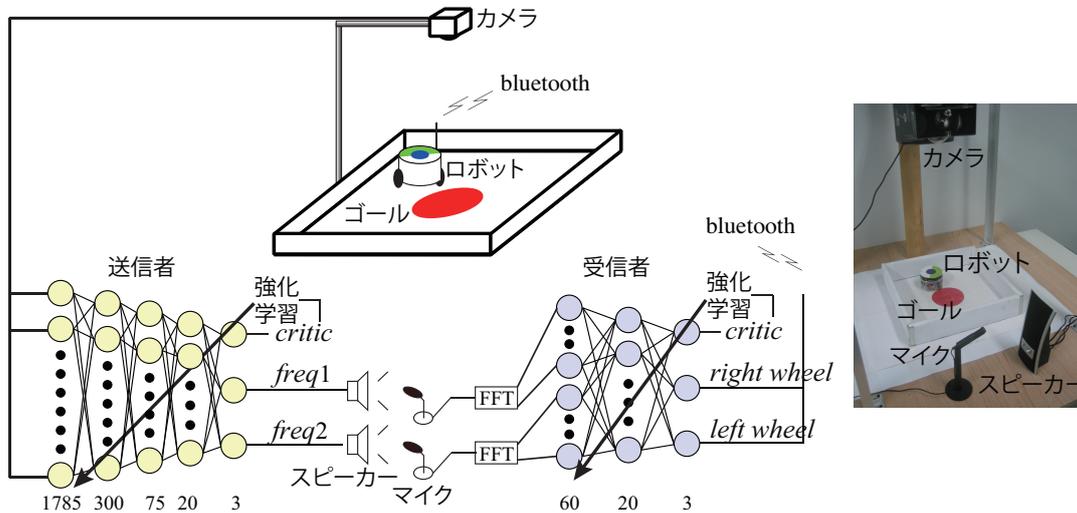


Fig. 1: システム構成とロボット制御タスク。図では2つのスピーカーと2つのマイクがあるが、実際は、右の写真のような1つのスピーカーから2つの音が少しの間隔を置いて発生し、1つのマイクでそれを受信した。

けて、1,000Hz から 1,300Hz の範囲で 10Hz 幅でとった平均スペクトル 30 個を最大値で割って正規化した合計 60 個の信号を入力ベクトル s とした。受信者は、Actor の出力 a と試行錯誤成分の rnd の和の2つの成分をロボットの右と左の車輪の制御信号とし、Bluetooth を経由してロボットに送り、ロボットを制御した。

学習は非常に簡単であり、送信者と受信者はそれぞれ独立に通常の強化学習によって前の節で述べたように学習する。ロボットのフィールドの中央に直径 20cm の赤い円があり、ロボットの中心が赤い円の中に入ると 0.9 の報酬が与えられ、1 試行が終了する。ロボットが壁に近づくと前のステップの場所に戻され、-0.01 の小さな罰が与えられる。

Fig. 2(a) に生のカメラ画像の例を示す。計算時間を短縮するため、画像サイズを 26×20 に縮小した。Fig. 3 にロボットの前、後および相対的、絶対的向きの定義を示す。緑の部分はロボットの前を表し、画像の垂直な軸からの角度を絶対角度 θ とし、ゴールの中心とを結ぶ線からの角度を相対角 α とする。

強化学習に先立ち、 26×20 の画像を入力して、ロボットの位置と向きを変化させてロボットの相対位置と向き ($\cos\alpha, \sin\alpha$) を教師あり学習させたところ、ロボットの向きに対する誤差はほとんど減少しなかった。画像入力からロボットのすべての位置における相対的な向きを認識することが困難であると考え、Fig. 2(b) のように、中心にロボットが来るように画像を平行移動させてからニューラルネットに入力したところ、誤差は減少した。その際、画像を平行移動させて空いたスペースにはグレーの色を敷き詰めた。これは、カメラを移動させたと考えられることもできる。しかし、事前知識を与えない自律学習の観点からは、その移動も学習によって獲得させることが望ましいが、これも次の課題とした。さらに、精度を上げるために、画像の中心部の 5×5 の画素の部分は 10×10 とし、中心部と周辺で解像度を変化させた。その結果、画像の画素数は 595 となり、各画素で RGB の 3 つの信号があることから全部で入力は 1,785 個となった。個々の信号は入力前に -0.5 から 0.5 の間に線形変換した。

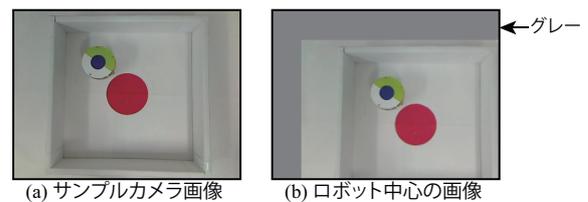


Fig. 2: ロボット中心の画像

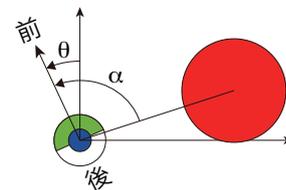


Fig. 3: ロボットの前およびロボットの絶対角度 θ と相対角度 α

3.2 事前学習の効果

このタスクでは、ロボットは回転することで向きを変え、その後前進または後退することでゴールにたどり着くことができる。回転方向は時計回り、反時計回りのどちらでも良いが、無駄な動きをしないためには、 $\alpha \leq 90^\circ$ または $180^\circ < \alpha \leq 270^\circ$ の場合には時計回り、その他の場合は反時計回りをすることが望ましい。 $\alpha = 90^\circ$ または $\alpha = 270^\circ$ のあたりでは、最適方向は α の微小な違いで急激に変化する。学習後、ロボットはゴールにたどり着くことができるようになったが、回転方向は上述のような最適なものではなく、どのような場合でも同じ方向にしか回転しなかった。これは、送信者にとって、コミュニケーション信号がロボット動作に直接影響を与えず、受信者を通して影響するため、微妙な学習ができなかったためではないかと考えた。

そこで、コミュニケーションの学習の前に、送信者は受信者を介さずに直接ロボットを制御することを強化学習で学習させてみた。その場合には、ロボットの回転方向は相対角度 α にしたがって適切に選ぶことができるようになった。その後、学習したニューラルネットの内部表現を用い、つまり、入力層と中間層の結合

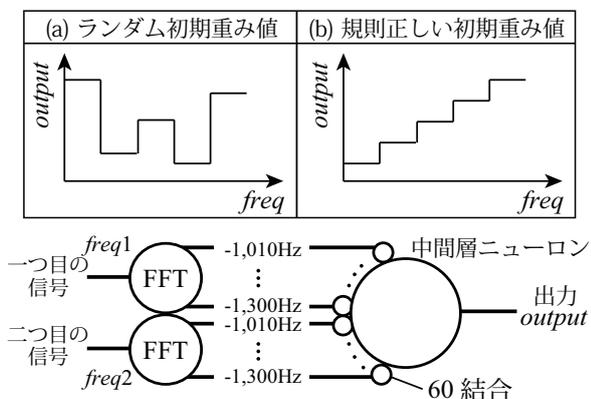


Fig. 4: 受信者エージェントの規則的な初期重み値によるコミュニケーション信号(音)の周波数と各中間層ニューロンの出力との間の相関の確保

はそのままにし、中間層と出力層の結合をすべて0にして受信者とともにコミュニケーションの学習を行った。その結果、次の3.3に示す方法を導入後、相対角度 α にしたがって正しく回転方向を選ぶことができるようになった。別の学習によって得られた内部表現が、その後の学習に利用されてより良いコミュニケーションが得られるようになるという大変面白い結果となった。

3.3 コミュニケーション信号と動作の間の相関

うまく学習が進まなかった理由をさらに調査したところ、送信者が発するコミュニケーション信号と受信者が発する動作信号の間の相関が小さかったことにも問題があることがわかった。受信者の個々の中間層ニューロンのFFT後の各入力に対する初期重み値はランダムに決定していた。したがって、ニューロンの出力はFig. 4(a)のように、コミュニケーション信号の周波数に対して単調に変化しない。そのため、受信者のActorの出力である動作信号も、入力信号の周波数に対して相関が小さくなる。相関がなければ、送信者がロボット動作をより適切にするために、周波数を増やすべきか減らすべきかを知ることが困難になる。そこで、ここでは、個々の中間層ニューロンにおいて、1つのコミュニケーション信号に対する30個の入力に対する初期重み値を、Fig. 4(b)のように、周波数の増加とともに徐々に増加または減少するようにした。また同じく受信者の入出力間で相関を保つために、受信者のActorの出力に加える試行錯誤のための探索要素 \mathbf{rnd} を ± 0.1 と、送信者の探索要素の ± 1.8 より大幅に小さくした。このことは、文献¹⁴⁾の報告とも合致する。

4 実験

前節の手法を導入し、実機を使って学習を行った。学習に用いたパラメータをTable 1に示す。送信者のニューラルネットは入力が高次元であるため5層とし、受信者は3層とした。そして、6,000試行の学習を行った。ロボットの初期位置の範囲は学習の進展とともに徐々に広くしていった。Fig. 5に、学習後に探索要素を切ったときの2つの試行例を示す。一つの試行(a)では、ロボットは画像の左上の領域に置き、絶対角度 $\theta = 0^\circ$ 、つまり、ロボットの緑の部分が白い部分より上に来るようにした。試行(b)では、ロボットを左下に置き、同じく絶対角度 $\theta = 0^\circ$ とした。それぞれの試行について、カメラ画像、送信者のCriticとActor(コ

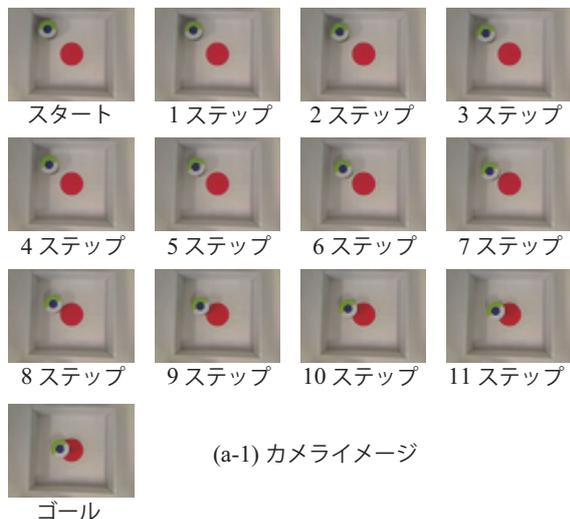
ミュニケーション信号である音の周波数 $freq1, freq2$), 受信者のCriticとActor(ロボットの両輪の動作信号 $left_wheel, right_wheel$)の出力の時系列を示す。試行(a)では、送信者は1回目に高い周波数の音、2回目に低い周波数の音を出している。そして、ロボットは反時計回りに回転しながら後方に移動している。その後、送信者は、2回目に前よりも少し高い周波数の音を出すようになり、ロボットは回転せずに後退をするようになっていく。試行(b)では、まず、低い周波数の音の後に高い周波数の音を伝達して、ロボットは時計回りに回りながら前進し、その後、2つ目の音が中間的な周波数になり、ロボットはゴールに着くまで前進していることがわかる。

Table 1: 学習で用いたパラメータ

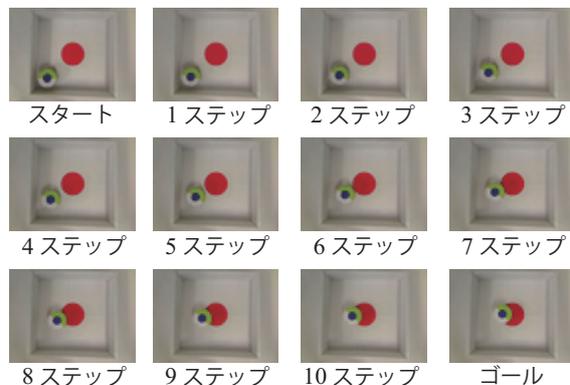
	送信者	受信者
ニューロン数	1785-300-75-20-3	60-20-3
学習係数	0.5	0.3
初期重み値 (入力→中間層)	事前学習で獲得した重み値	規則的 (-2.0 - 2.0)
初期重み値 (中間層→出力層)	ランダム [-0.5 - 0.5]	ランダム [-2.0 - 2.0]
探索要素	ランダム [-1.8 - 1.8]	ランダム [-0.1 - 0.1]
報酬	0.9	
罰	0.01	
割引率 γ	0.96	

Fig. 6(a)に、ロボットの位置と向きの絶対角度 θ の組み合わせに対する、2つの音の周波数、つまり、送信者の2つのActorの出力を示す。周波数は実際にキャプチャーされた画像から送信者が生成したものである。周波数はロボットの位置や向きによって異なっていることがわかる。しかし、ロボットから見たゴールの相対位置が等しい場合(例えば、(a-1)の左上の場合と(a-2)の左下の場合)は入力される画像は大きく異なるが、2つの音の周波数は似ている。Fig. 6(b)に、2つの音の周波数の組み合わせに対する動作信号、つまり、受信者の2つのActorの出力を示す。この図は、実際に2つの正弦波音声をスピーカーから発生させ、マイクでそれを捉え、FFTを掛けた後に受信者のニューラルネットに入力してデータを取った。動作信号は、2つの音の周波数の変化に対して徐々に滑らかに変化していることがわかる。Fig. 6(c)には、ロボットの状態と動作信号の関係を示す。動作信号は、実際にキャプチャーされた画像から送信者のニューラルネット、スピーカー、マイク、FFT、受信者のニューラルネットを通して生成されたものである。適切なコミュニケーションを通して、ロボットは、完全に最適とまではいかないものの、状態に応じて適切な動作をしていることがわかる。

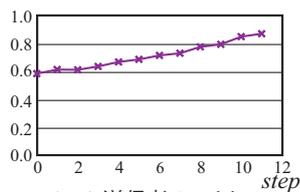
コミュニケーション信号はロボットが実行すべき動作を表現しているだけであり、ロボットの状態や状態価値の値は表現していない。したがって、受信者はロボットの状態を考慮した適切なCriticを表現することができない。しかし、音に合った行動をすれば報酬をもらったり、報酬を早くもらうことに結びつき、逆に音に合わない行動をして罰を受けたり、報酬をもらうのが遅くなることを通じてコミュニケーション信号とロボットの動作信号の間の適切なマッピングを獲得することができたと考えられる。



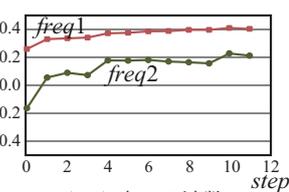
(a-1) カメライメージ



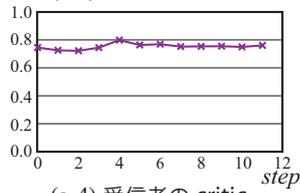
(b-1) カメライメージ



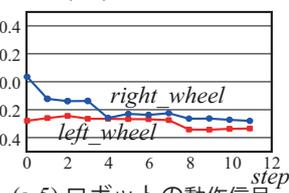
(a-2) 送信者の critic



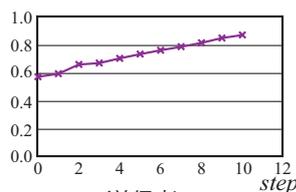
(a-3) 音の周波数



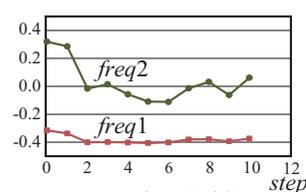
(a-4) 受信者の critic



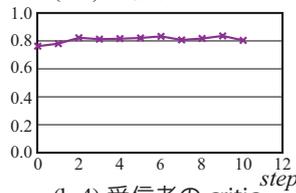
(a-5) ロボットの動作信号



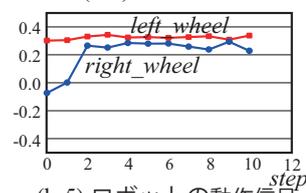
(b-2) 送信者の critic



(b-3) 音の周波数



(b-4) 受信者の critic



(b-5) ロボットの動作信号

Fig. 5: 2つのサンプル試行におけるロボットの行動, 送信者と受信者の出力の変化。コミュニケーション信号(音)は適切な行動を表現し, 状態や状態価値を表現しないので, 受信者の Critic の出力は時間とともに増加しない。

5 結論

ロボットの状態をカメラ画像として観測できる送信者から, ロボットに動作コマンドを送って制御できる受信者へのコミュニケーションが, ロボットがゴールに到達した際の報酬と壁に近づいた際の罰による強化学習を送信者と受信者が独立に行うだけで, コミュニケーションのための特別な学習をすることなく形成されることを, 実際のロボット, カメラ, スピーカー, マイクを用いて示した。また, 送信者が実際にロボットを制御する経験をすることで学習が加速されること, さらに, 受信者において, 受信する信号と出力する動作信号の間に相関を持たせることが学習には重要であることを示した。本論文では, コミュニケーション信号が連続であり, 「シンボル接地問題」の解決とは言えないが, たくさんのセンサ信号の中から, 何をコミュニケーション信号として伝えるべきか, そして, コミュニケーション信号をどう動作に反映させるべきかも含めて, 合目的かつ接地したコミュニケーションをコミュニケーションに特化したものではない汎用的な学習を通して獲得できることを示すことができた。

謝辞

本研究は日本学術振興会科学研究費補助金基盤研究(B)19300070 および基盤研究(C)23500245 の補助の下で行われた。

参考文献

- 1) Turing, A.M.: Computing Machinery and Intelligence, *Mind*, **59**, pp. 433-460 (1950)
- 2) Can a machine think? - results from the 18th Loebner Prize contest, [http://www.reading.ac.uk/research/Highlights-News/featuresnews/research/research/featureloebner.aspx](http://www.reading.ac.uk/research/Highlights-News/featuresnews/research/research/Highlights-News/featuresnews/research/research/featureloebner.aspx)
- 3) Searle, J.: "Minds, Brains and Programs", *Behavioral and Brain Sciences*, **3** (3), pp.417-457 (1980)
- 4) Harnad, H.: Symbol Grounding Problem. *Physica D*, **42**, pp.335-346 (1990)
- 5) Nakano, K. et al.: Self-Organizing System Obtaining Communication Ability. *Biological Cybernetics*, **58**, pp.417-425 (1988)
- 6) Steels, L.: Evolving grounded communication for robots. *Trends in Cognitive Science*, **7**(7), pp. 308-312 (2003)
- 7) Shibata, K.: Emergence of Intelligence through Reinforcement Learning with a Neural Network. *Advances in Reinforcement Learning*, InTech, pp.99-120 (2011)
- 8) 柴田克成: 強化学習とニューラルネットによる知能創発, 計測と制御, **48**(1), pp. 106-111 (2009)
- 9) Shibata, K.: Discretization of Series of Communication Signals in Noisy Environment by Reinforcement Learning. *Adaptive and Natural Computing Algorithms*, pp. 486-489 (2005)
- 10) 柴田克成: コミュニケーションの強化学習におけるノイズ付加による連続値信号の離散化, 電子情報通信学会技術研究報告, **103** (734), NC2003-203, pp. 55-60, 2004
- 11) Werner, G.M. & Dyer M.G.: Evolution of Communication in Artificial Organisms. *Proc. of Artificial Life II*, pp.1-47 (1991)

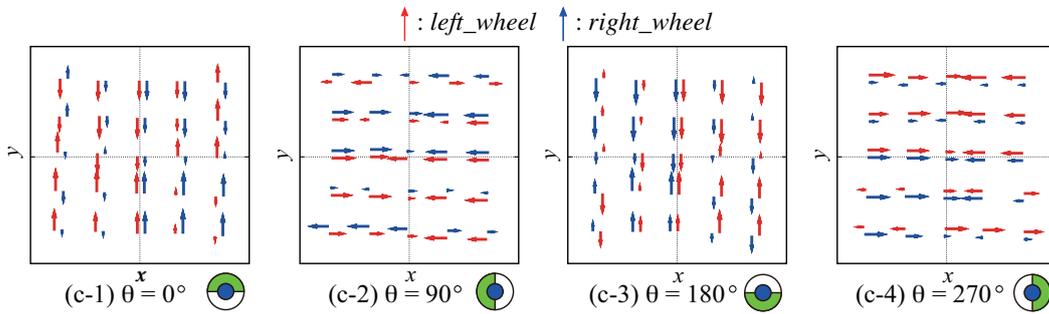
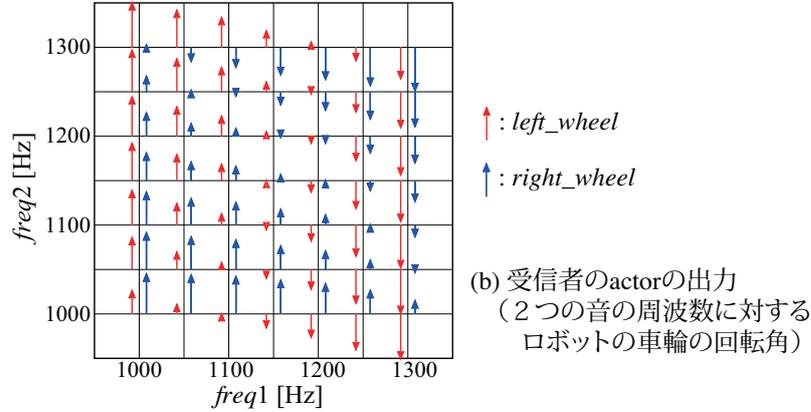
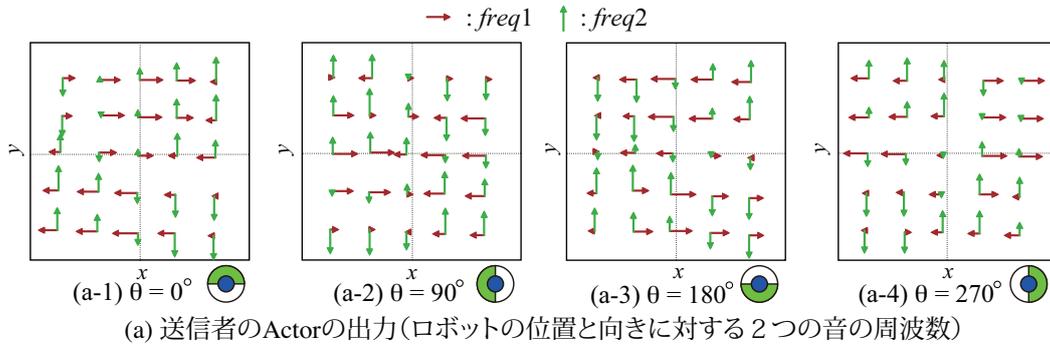


Fig. 6: (a) ロボットの位置 (x, y) に対する送信者の actor 出力である音の周波数 $(freq1, freq2)$ 。矢印の位置がフィールド上のロボットの場所を表す。(a-1,2,3,4) で、ロボットの向き θ が異なる。ロボットの向きは、各図の脇にある小さいロボットの図で示す。水平の茶色の矢印が最初の信号の周波数 $freq1$ を、垂直の緑の矢印が2つ目の信号の周波数 $freq2$ を表す。(たとえば、1,000Hz: 上または右方向に最大の長さ、1,150Hz: 長さ 0、1,300Hz: 下または左方向に最大の長さ) (b) 2つのコミュニケーション信号の周波数 $(freq1, freq2)$ に対する受信者の Actor 出力であるロボットの動作信号 ($left_wheel, right_wheel$)。 (c) ロボットの位置 (x, y) と向き θ に対するロボットの動作信号 ($left_wheel, right_wheel$)。各状態に対する動作信号は赤と青の矢印で表している。赤い矢印は左車輪に対する動作信号、青い矢印は右車輪の動作信号を表す。矢印の向きはロボットの向きに応じて変化させた。

12) Ono, N. Ohira, T. & Rahmani, A.T.: Emergent Organization of Interspecies Communication in Q-Learning Artificial Organs. *Advances in Artificial Life*, pp.396-405 (1995)

13) Shibata, K. & Ito, K.: Emergence of Communication for Negotiation By a Recurrent Neural Network. *Proc. of ISADS (Int'l Symposium on Autonomous Decentralized System) '99*, pp.294-301 (1999)

14) Nakanishi, M. & Shibata, K.: Effect of Action Selection on Emergence of One-way Communication Using Q-learning. *Proc. of the 10th Int'l Symp. on Artificial Life and Robotics (AROB)*, CD-ROM, GS7-3 (2005)

15) Shibata, K. & Sasahara, K: Emergence of Purposeful and Grounded Communication through Reinforcement Learning. *LNC3(Lecture Notes in Computer Science)*, **7064**, pp.66-75 (2011)

16) Mitsunaga, N., et al.: Robovie-IV: A Communication Robot Interacting with People Daily in an Office. *Proc. of IROS'06 (IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems)*, pp. 5066-5072 (2006)

17) Suga, Y., et al.: Development of Emotional Communication Robot, WAMOBEA-3. *Proc. of ICAM'04 (The 4th Int'l Conf. on the Adv. Mechatronics)*, pp.413-418 (2004)

18) Bennewitz, M. et al.: Fritz - A Humanoid Communication Robot. *Proc. of RO-MAN'07 (16th IEEE Int'l Conf. on Robot & Human Interactive Communication)*, pp. 1072-1077 (2007)

19) Barto, A.G. et al.: Neuronlike Adaptive Elements That can Solve Difficult Learning Control Problems'. *IEEE Trans. of SMC*, **13**, pp.835-846 (1983)

20) Rumelhart, D.E. et al.: Learning Internal Representation by Error Propagation. *in Parallel Distributed Processing* (1986)