

ロボットの随伴性行動を生成するための振る舞い統合手法

Integration Method for Generating robot's contingent behavior

滝本 佑介 長谷川 孔明 今井 倫太
Yusuke Takimoto Komei Hasegawa Michita Imai

慶應義塾大学理工学部情報工学科

Department of Information and Computer Science, Faculty of Science and Technology, Keio University

In order for a robot to communicate smoothly with humans, it is important to respond appropriately to the surrounding environment and events. In this paper, we propose a simple bi-layered architecture (SB architecture) to integrate behaviors in two stages of prioritization and weighted average. SB architecture integrates voluntary movements by prioritization and involuntary movements and reflex movements by weighted averaging, thereby the robot can respond to surrounding environment and events and promote the communication with humans by performing necessary behaviors. The advantage of SB architecture is that it can easily generate various actions by combining multiple actions and it is possible to propose a simple design for robot behavior.

1. はじめに

ロボットが人間と円滑なコミュニケーションを行うためには、音声対話の機能の他に、視線・ジェスチャ・体の向き・立ち位置の移動といった振る舞いも人間から見て違和感のない形で生成できる必要がある。しかし現在普及しているコミュニケーションロボットは主に、事前に設計された動きしか行わないシナリオベースで設計されている。シナリオベースのロボットは対話シナリオに沿って行動するため、予め想定していない出来事がロボットの周囲で起きたとしても反応することができず、ロボットが周囲の状況を共有できていないと人間に感じさせてしまう原因となる。状況が共有できない感覚は、コミュニケーションにおける文脈が共有できていないと人間に思わせてしまうことに直結しており、人間とロボットがコミュニケーションを継続することを著しく阻害する。ロボットが人間と円滑なコミュニケーションを行うには、周囲の環境やイベントに対して適切に反応し、目の前にある状況を人間と共有していると感じさせるライブネスを示すことが求められる。本論文では、人間との円滑なコミュニケーションを目的としたロボットのライブネスのある振る舞いの生成に取り組む。

外界のイベントに反応しながら人間とインタラクションが可能なロボットの研究は複数行われている。発達心理学の分野では周囲の環境の変化に対して即座に反応する行動の性質を随伴性と呼び、幼児は随伴性を持つロボットに積極的に関わろうとする傾向がある [Movellan 05]。ロボットの随伴性行動を生成するモデルとして、Sumioka らは振る舞いの中から適切なものを選択するモデルを提案した [Sumioka 10]。また人間の動作を模倣した振る舞いをロボットに取り入れることで人間とロボットのインタラクションを後押しする研究も行われている。Schulz らは目、頭、まぶたに着目し人間の生理的な動作を模倣する Humotion フレームワークを提案した [Schulz 16]。実世界の環境に反応して振る舞うロボットに関するアーキテクチャには、古くはサブサンクション・アーキテクチャがある [Brooks 86]。サブサンクション・アーキテクチャは振る舞いのモジュールを目的ごとに階層構造として持ち、上位層のモ

ジュールが下位層を抑制することで動作する。

しかし Sumioka らが提案した部位ごとに単一の振る舞いのみを選択するモデルでは、単純過ぎるため表現力が足りず、複雑な振る舞いを生成することが出来ない。一方、Schulz らの提案した Humotion フレームワークといった複雑なモデルでは、新しい振る舞いの追加や、設計者が意図しない振る舞いをロボットが行った際に修正するにはモデル自体を変更しなければならないという問題点があり、ロボットを開発するコストが高くなる。またサブサンクション・アーキテクチャも同様にデータフローが複雑になりやすく、振る舞いの追加や修正が困難である。

本論文では振る舞いの統合部分に着目し、優先度付けと重み付き平均化の2段階で振る舞いを統合する SB (Simple Bi-layered) アーキテクチャを提案する。SB アーキテクチャは随意運動を優先度付け、不随意運動・反射運動を重み付き平均化で統合することでライブネスのあるロボットの振る舞いを生成する。SB アーキテクチャは振る舞いモジュールを優先度や重みによって統合している点でサブサンクション・アーキテクチャに代表される多くのアーキテクチャと似ている。一方で振る舞いの統合の仕方を、優先度付けされる随意運動の層と、重み付き平均化される不随意運動・反射運動層という、2層のシンプルな構造で設計する点で異なる。随意運動と不随意運動・反射運動を別々に扱い、2つのレイヤでシンプルに統合したとしても、多様さとライブネスを損なうことなくロボットの振る舞いを生成することができる。さらに、シンプルな統合方法を取ることで行動設計の全体的な難易度を下げ、簡潔で表現力の高いライブネスのあるロボットの振る舞い設計が可能になる。

2. 背景

2.1 コミュニケーションにおけるライブネス

コミュニケーションにおいて、相手とのやりとりの時間的な要素が重要な効果を持つことが知られ始めている。例えば、人間同士のインタラクションでは、相手がテレビの向こうなのか、目の前に居るライブな存在なのかで脳の状態が異なることが知られている [Shimada 06]。さらに、テレビを介した母親が幼児に動きを教えるコミュニケーションにおいて、幼児に見せる母親の映像側だけに2秒程度のほんの僅かな遅延を入れると、幼児が教えてもらった動きをしづらくなる知見もある

連絡先: 滝本佑介, 慶應義塾大学理工学部情報工学科,
神奈川県横浜市港北区日吉 3-14-1, Tel: 045-566-1515,
takimoto@ailab.ics.keio.ac.jp

[Yamamoto 16]. ビデオ映像ではなく、目の前にいるライブな存在であることの効果、またビデオ映像であっても時間的な遅延の存在がコミュニケーションに影響が与えることが分かる。

また、目の前にいる相手とのコミュニケーションでは相手の反応が重要であり、反応性が重要であることはロボットが相手であっても同様であることが知られている [Movellan 05]. 発達心理学の分野では、相手が自分の行動や周囲の状況に反応することを随伴性と呼ぶ。随伴性は人間とロボットのコミュニケーションにおいても重要な役割を果たす。

Movellan の研究では、随伴性が幼児に与える影響について比較実験が行われた [Movellan 05]. 実験には音と光を発するロボットを使用し、ロボットが幼児の声に反応する場合と、ランダムなタイミングで音を発する場合の 2 条件で行われ、実験の結果幼児の声に反応するロボットにおいてのみ幼児が興味を示した。極めて単純な随伴性を持つロボットであっても幼児が興味を示したことから、インタラクションにおいて随伴性が重要であると言える。

随伴性を持つ行動を生成するコミュニケーションロボットの研究も行われており、Sumioka らは人間とのインタラクションにおいて、注視に関する随伴性の検出と学習を行うロボットを作成した [Sumioka 10]. また、人間とロボットの円滑なコミュニケーションを目的に、人間の動きを模倣した振る舞いをロボットに取り入れる研究も存在する。Schulz らは人間の目、頭、まぶたに関する生理的な動きを模倣する Humotion フレームワークを提案した [Schulz 16]. Humotion フレームワークは目、頭、まぶたが取るべき目標値を入力すると、生理的な動きを考慮したモータ値を出力する。

上記の研究が示しているのは、ロボットが物理的な体を持って目の前に存在するだけでは、人間とのコミュニケーションを成立させる上では不十分であるということである。人間の動きや周囲のイベントに随伴性のある形で反応し、ライブネスを示すことが重要である。

2.2 サブサンプリング・アーキテクチャと Behavior-Based アプローチ

実世界の環境やイベントに反応したロボットの行動を生成するアーキテクチャとして、サブサンプリング・アーキテクチャが存在する [Brooks 86]. サブサンプリング・アーキテクチャは目的ごとの振る舞いモジュールから成る階層構造アーキテクチャで、下位層は単純な振る舞いを生成し、上位層に行くにしたがって高度な振る舞いが実装される。上位層の振る舞いが下位層の振る舞いを抑制し、環境に対する適用性の高いロボットが実現される。

また、サブサンプリング・アーキテクチャに代表される多数の振る舞いを集めてロボットやエージェントの行動を生成するデザインは Behavior-Based アプローチとも呼ばれる。Behavior-Based アプローチで作成されたロボットは、サブサンプリング・アーキテクチャで採用されていた明確な階層性による優先順位を持たないことも多く、様々な振る舞いモジュール同士が、協調や抑制を行いながら行動を生成する。例えば、Breazeal が構築した Kismet は、ロボットの感情や欲求の影響および、外界のセンサ情報に応じて活性化する振る舞い群がロボットの動きを作り出し、ライブネスの豊かなコミュニケーションロボットとなっている [Breazeal 04].

2.3 従来研究の問題点

従来研究で提案されてきた随伴性のある振る舞いや人間の動きを模倣した振る舞いを生成する研究では、実際に振る舞い同士をどのように連携させて行動生成させるのかについて言及

した研究は少ない。また振る舞いを生成するためのアーキテクチャやフレームワークについて言及している研究やサブサンプリング・アーキテクチャにおいても、次に示す何点かの課題が存在する。

ある時刻においてロボットがとりうる振る舞いが複数存在する場合、実際に出力する振る舞いを決定する必要がある。特に随伴性を持った振る舞いは外界からの刺激に反応する行動であるため、同一時刻において体の同じ箇所を動かす振る舞いが生じ競合しやすい。Sumioka らの提案したモデルでは、1 つのみの振る舞いを選択するという単純な操作により振る舞いの統合が簡潔に表現される [Sumioka 10]. 一方で振る舞い同士が混合されて出力される事がないため、組み合わせによって生じる複雑な振る舞いが表現ができないという問題点が存在する。

また Schulz らの Humotion フレームワークでは目、頭、まぶたを動かすモータの目標値を決定するために値の計算や選択をモデル中の各所で行っている [Schulz 16]. しかし、モデル中に値を決定する処理が分散しているため全容を把握することは難しく、新たな振る舞いの追加や既存の振る舞いの修正を行うことは困難であり、ロボットを開発する上で問題となる。

アーキテクチャ内のモジュール構成が複雑で開発や修正が困難であることはサブサンプリング・アーキテクチャや多くの Behavior-Based アプローチでも同様であり、開発には職人技が必要となる。

2.4 随意運動・不随意運動・反射運動

理想的なコミュニケーションロボットは、コミュニケーションに必要な行動をおこなったり、コミュニケーションの結果与えられる目的に従って行動したりする必要がある。これらはコミュニケーションにおける主要な行動であり、随意行動に相当する。一方でライブネスを感じさせるためには、生理的な運動である不随意運動や反射運動を持たせる必要がある。

随意運動とは意思を伴った行動のことであり、例として興味のある物体を表現するために注視したり、発言を聞いていることを表現するために頷いたりといった振る舞いが挙げられる。一方で不随意運動とは意思を伴わない行動のことであり、周期的な瞬きや呼吸が例に挙げられる。また反射運動は不随意運動の中でも、眼球運動に伴ったまばたきといった感覚刺激に起因した行動のことであり、

随意運動、不随意運動、反射運動をロボットの振る舞いとして統合する際の課題として、目的をもって生成される随意運動に対してどのように不随意運動や反射運動を組み合わせ、コミュニケーションを行いながらもライブネスを感じさせるかという点が挙げられる。従来はコミュニケーションを扱うことがメインであるばかりに、不随意運動や反射運動は完全に抑制して、コミュニケーションのシナリオに沿った随意運動の実行に主眼が置かれていたが、コミュニケーションにおけるライブネスの重要性を考えると、これら三つの行動を統合するアーキテクチャを考案することは大変重要な取り組みである。

3. SB アーキテクチャ

3.1 SB アーキテクチャの構成

本論文では随意運動と不随意運動・反射運動を、優先度付けと重み付き平均化の 2 段階で統合を行うことでライブネスのある振る舞いを生成する SB (Simple Bi-layered) アーキテクチャを提案する。SB アーキテクチャは随意運動を優先度に従って選択し、複数の不随意運動・反射運動と共に重み付き平均化により組み合わせられるため、コミュニケーションに必要な動きをしつつも、周囲の環境やイベントに反応した多様な振る

舞いを生成することが可能である。また SB アーキテクチャは独立に設計された振る舞いを 2 層構造で統合する簡潔な構造であるため、ロボットの使用環境や目的に応じた振る舞いの追加、修正、削除が容易であり、ロボットの開発コストを下げるができる。図 1 に SB アーキテクチャの構成図を示す。

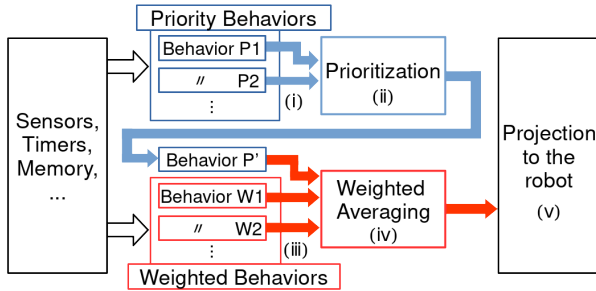


図 1: SB アーキテクチャの構成図

SB アーキテクチャには優先度付き振る舞い (Priority Behavior) と重み付き振る舞い (Weighted Behavior) をそれぞれ必要なだけ登録することができる。またロボットにより搭載するアクチュエータの数や機構が異なり、ロボットごとに固有なデータ形式を用いることは汎用化の面から望ましくないため、頭、目、まぶた、腰、右手、左手といった体の部位ごとに抽象化されたデータを扱う。

各振る舞いはセンサで取得されるイベントや、タイマーから得られた時刻、他の情報処理モジュールによる処理結果に起因して任意のタイミングで生起し (図 1 (i) (iii)), 各時刻において生のモータ値や、モータ値を抽象的に表現する 3 次元空間中の座標といった体の動かし方を部位ごとに出力する。次に振る舞い群から出力された体の部位ごとの動かし方の競合を解消し、単一の動かし方への統合を行う。SB アーキテクチャは、優先度付け (図 1 (ii)) と重み付き平均化 (図 1 (iv)) の 2 段階で統合を行う。優先度付けとは部位の動かし方の中から単一の動かし方を選択する操作のことであり、体の部位の動かし方に対応した静的な実数である優先度 p が最も大きい動かし方を選択する。優先度付けは優先度付き振る舞いに対して適用し、1 つの統合された振る舞いを得る。一方、重み付き平均化とは部位の動かし方の中から単一の動かし方を選択するのではなく、重みによって平均化を行う操作のことであり、体の部位の動かし方に対応した非負の実数である重み w を用いた重心を求める。重み付き平均化は重み付き振る舞い、および優先度付けによって統合された優先度付き振る舞いに対して適用し、最終的に統合された振る舞いを得る。また優先度 p と重み w は振る舞いの設計と合わせて設定する。アーキテクチャの最終段階では、体の部位ごとに抽象化されたデータ形式をロボットごとに固有な値に変換し出力する (図 1 (v))。

3.2 振る舞いの分類

各振る舞いは SB アーキテクチャに登録する際に優先度付き振る舞いと重み付き振る舞いのいずれか一方に分類する必要がある。分類方法として、随意運動を優先度付き振る舞いに、不随意運動・反射運動を重み付き振る舞いに対応付けを行う。

振る舞いの優先度付けによる統合は、動かし方のいずれか一つのみを選択するため、離散的である一方、単純で振る舞いの設計が容易であるという特徴がある。ロボットの行動目的を達成するための随意運動は、他の随意運動と混ざると行動が指し示す意味が人に伝わらなくなってしまう可能性があり、コミュニケーションのための行動生成の観点からは致命的である。随

意運動を離散的な優先度選択にすることで、他の随意運動を気にせず設計することができ、設計の難易度を下げることができる。随意運動とは対照的に、不随意運動・反射運動は人間の場合でも他の行動と同時に行われる場合がほとんどであり、他の振る舞いと混ざることができる。また振る舞い自体が複雑ではないため、優先度よりも複雑な表現が可能な重み付き振る舞いとすることで柔軟な振る舞いの設計を行うことができる。

3.3 優先度と重みの決定

SB アーキテクチャに登録された振る舞いの個数が少ない場合は、優先度付けと重み付き平均化に必要なパラメータである優先度 p と重み w は手作業により決定することができる。しかし登録された振る舞いの個数が増加すると優先度 p と重み w の個数は膨大になり、手作業でパラメータを決定するのは困難になる。自動的にパラメータを決定するために、任意の個数のセンサ情報とその時に期待される振る舞いのペアを記述し、全てを満たすように非線形最小二乗法でパラメータの最適化を行う。最適化には実装のしやすさから準ニュートン法の一つである BFGS 法 [Broyden 70] を用いる。

また SB アーキテクチャは単純な 2 段階構造なため優先度 p と重み w の数値から意味合いを把握しやすく、最適化により決定されたパラメータをベースにしつつ手作業によりパラメータの調整を行うことができる。機械学習を行った結果に人間が介入することで、意図しない振る舞いが出力された場合にも修正を行うことが可能になり、安全性の高いロボットの開発が可能になる。

4. 実装

本論文では卓上コミュニケーションロボット Robovie-mR2 [Matsumura 16] を用いて SB アーキテクチャの実装を行った。Robovie-mR2 は頭の稼働に 3 自由度、片目にまぶたを含めて 3 自由度、腰に 1 自由度、片腕に 4 自由度、合計 18 自由度の可動軸をもつ。Robovie-mR2 の外観を図 2 に示す。

センサ情報として Web カメラからは人間の顔の 3 次元位置と向き、手のひらの検出、物体の 3 次元位置、人間の注視対象の推定を行った結果を使用した。顔の 3 次元的位置と方向は、CLNF [Baltrusaitis 13] により推定された顔の 2 次元ランドマークと、別途計測された 3 次元形状とのマッチングを行うことで推定した。手のひらと物体の検出には HOG 特徴量 [Felzenszwalb 10] を使用し、別途計測された 3 次元的な大きさを用いることで 3 次元位置を推定した。またマイクロフォンからは人間の発話を取得し発話区間認識 (VAD) [Ramirez 07] を行った。

本論文においては、優先度付き振る舞いとして動いている顔や物体の注視、アイコンタクト、共同注視、発話に対応した頷き、人物に対する腕振りといった 18 振る舞いを実装した。また重み付き振る舞いとして周期的および眼球運動に伴う瞬きや呼吸を表現する動き、サッカーボール眼球運動を表現する動きといった 5 種類を実装した。優先度 p および重み w はセンサ情報とその時に期待される振る舞いのペアを用いて最適化により決定した。

実装された振る舞いを SB アーキテクチャが統合しロボットが動作する様子を図 2、図 3 に示す。図 2 では青リングを注視していたロボットが赤リングが動くのを検出し注視対象を変更している。注視対象の変更の際は眼球が頭よりも先に動作し、素早い眼球の動きであるサッカーボール眼球運動を表現している。図 3 は人間を注視していたロボットが、人間の視線変化を検出し、視線追跡を行う様子である。注視対象の変更と同様に眼

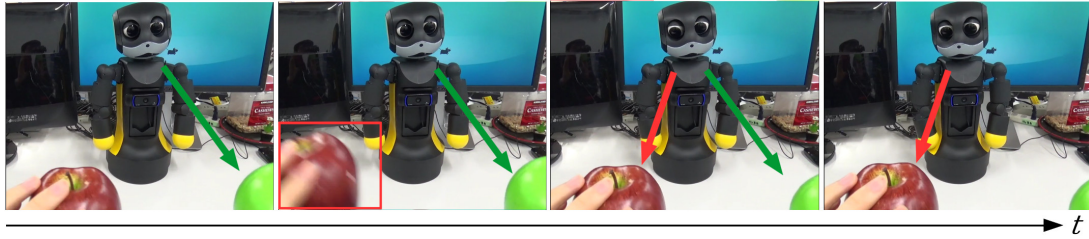


図 2: 動いた赤リンゴを素早く眼球のみで注視しサッカード眼球運動を表現, 次のいで頭を向ける様子



図 3: 人間の視線変化を検出し素早く眼球のみで視線追従を行いサッカード眼球運動を表現, 同時に眼球運動に伴うまばたきが発生, 眼球運動を追うように頭も伴って視線追跡を行う様子

球が素早く動くことでサッカード眼球運動を表現し, さらに眼球運動に伴ってまばたきが発生している。

5. 結論

本論文では, 随意運動と不随意運動・反射運動を分類し, 優先度付けと重み付き平均化の 2 段階で統合することでライブネスのある振る舞いを生成する SB アーキテクチャを提案した. SB アーキテクチャは周囲の環境やイベントに対して即座に反応し, 複数の振る舞いを混合することで多様な振る舞いを生成する. また単純な 2 段階構造により, 振る舞いの追加, 修正, 削除が容易になり, ライブネスのあるロボットの開発コストを下げることができる。

参考文献

- [Baltrusaitis 13] Baltrusaitis, T., Robinson, P., and Morency, L.-P.: Constrained local neural fields for robust facial landmark detection in the wild, in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 354–361 (2013)
- [Breazeal 04] Breazeal, C. L.: *Designing sociable robots*, MIT press (2004)
- [Brooks 86] Brooks, R.: A robust layered control system for a mobile robot, *IEEE journal on robotics and automation*, Vol. 2, No. 1, pp. 14–23 (1986)
- [Broyden 70] Broyden, C. G.: The convergence of a class of double-rank minimization algorithms 1. general considerations, *IMA Journal of Applied Mathematics*, Vol. 6, No. 1, pp. 76–90 (1970)
- [Felzenszwalb 10] Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D.: Object detection with discriminatively trained part-based models, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 32, No. 9, pp. 1627–1645 (2010)
- [Matsumura 16] Matsumura, R., Shiomi, M., Nakagawa, K., Shinozawa, K., and Miyashita, T.: A Desktop-Sized Communication Robot: “robovie-mR2”, *Journal of Robotics and Mechatronics*, Vol. Vol.28, No. No.1 (2016)
- [Movellan 05] Movellan, J. R.: An infomax controller for real time detection of social contingency, in *Development and Learning, 2005. Proceedings. The 4th International Conference on*, pp. 19–24 IEEE (2005)
- [Ramirez 07] Ramirez, J., Górriz, J. M., and Segura, J. C.: *Voice activity detection. fundamentals and speech recognition system robustness*, INTECH Open Access Publisher NewYork (2007)
- [Schulz 16] Schulz, S., Lier, F., Kipp, A., and Wachsmuth, S.: Humotion: A Human Inspired Gaze Control Framework for Anthropomorphic Robot Heads, in *Proceedings of the Fourth International Conference on Human Agent Interaction*, pp. 207–214 ACM (2016)
- [Shimada 06] Shimada, S. and Hiraki, K.: Infant’s brain responses to live and televised action, *Neuroimage*, Vol. 32, No. 2, pp. 930–939 (2006)
- [Sumioka 10] Sumioka, H., Yoshikawa, Y., and Asada, M.: Reproducing interaction contingency toward open-ended development of social actions: Case study on joint attention, *IEEE Transactions on autonomous mental development*, Vol. 2, No. 1, pp. 40–50 (2010)
- [Yamamoto 16] Yamamoto, E., Nagata, K., and Hiraki, K.: Subtle temporal delays in infants’ responses affect mothers’ infant-directed actions: A Study of Mother-child interactions, *International Journal of Psychology*, Vol. 51, p. 234 (2016)