公 開 用

尾川 順子，阪口 豊，並木 明夫，石川 正俊，``ダイナミクス整合にもとづく感覚運動統合---ターゲットトラッキングにおける戦略の学習---,"電子情報通信学会ニューロコンピューティング研究会（2001.3.14, 東京）/ 電子情報通信学会技術研究報告 NC2000-109, Vol. 100, No. 686, pp. 47-54, Mar. 2001.

ダイナミクス整合にもとづく感覚運動統合 ―ターゲットトラッキングにおける戦略の学習―

尾川 順子* 阪口 豊† 並木 明夫‡* 石川 正俊*

*東京大学 大学院工学系研究科 計数工学専攻
†電気通信大学大学院 情報システム学研究科 情報ネットワーク学専攻
‡科学技術振興事業団 戦略的基礎研究推進事業（CREST）

*〒 113-8656 東京都文京区本郷７－３－１
TEL: 03-5841-6937 e-mail: naoko@k2.t.u-tokyo.ac.jp

あ ら ま し 人間やロボットなどにおける感覚運動統合において，ダイナミクス整合という概念を提起する．これは有限の計算資源やハードウェアの物理的特性や制御系の制約，タスクの内容といった拘束条件の下で，情報処理系や制御系の時間的な特性，すなわちダイナミクスを適応的・能動的に調整することによって，システム全体としてのパフォーマンスの最大化を実現するという考えである．本稿ではこの問題に対する一つのアプローチとして，パフォーマンスを報酬とした強化学習を用いることにより具体的なアルゴリズムを構成し，ターゲットトラッキングを例題とした数値実験によりその効果を検証した．

キーワード ダイナミクス整合，感覚運動統合，強化学習，ターゲットトラッキング

Sensory-Motor Fusion Based on Dynamics Matching —Learning of Strategies in Target Tracking Task—

Naoko OGAWA* Yutaka SAKAGUCHI† Akio NAMIKI‡* and Masatoshi ISHIKAWA*

*Graduate School of Engineering, University of Tokyo
†Graduate School of Information Systems, University of Electro-Communications
‡CREST, Japan Science and Technology Corporation

*7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN
TEL: +81-3-5841-6937 e-mail: naoko@k2.t.u-tokyo.ac.jp

Abstract The authors propose a novel concept named “dynamics matching” for designing sensory-motor fusion system. This means a strategy to adjust adaptively the contents of the information processing and the control to maximize overall performance of the system under constraints such as restricted computational resources, physical properties of the sensor and motor systems and requirements by given tasks. The authors construct an algorithm utilizing reinforcement learning where the system performance is regarded as the “reward.” The effect of the proposed concept is illustrated by a numerical experiment of a target-tracking task.

key words dynamics matching, sensory-motor fusion, reinforcement learning, target-tracking

1 はじめに

人間やロボットなどの感覚運動統合システムには，有限の計算資源やハードウェアのもつ物理的制約などといった制約条件が存在する．このような制約の下では，与えられたタスクの内容に応じて自らの特性や情報処理の内容を能動的・適応的に調整し，タスクとシステムの動的な特性を整合させることによって，システム全体として最大のパフォーマンスを達成することができると予想される． 本稿ではこのような問題意識を「ダイナミクス整合」として提起する．さらにこの問題を解決する一つのアプローチとして，パフォーマンスを報酬とした強化学習を用いたアルゴリズムを構成し，ターゲットトラッキングタスクを例題として数値実験によりその効果を例証する．

2 ダイナミクス整合とは

2.1 本研究の問題意識

近年，ロボティクスの分野においては，感覚運動統合を工学的に実現することによって [1]，ロボットによる高速で複雑な動作を可能にしたり，人間の高次脳機能の解明を試みるという研究が多く行なわれている．

ロボットをはじめとする感覚運動統合システムを設計する場合，通常は，与えられたタスクに対して適切に動作するよう，設計者がケースバイケースでハードウェアとソフトウェアを調整している．しかし，実世界で機能するロボットの特性，例えばセンシングの物理的限界や計算処理による時間遅れ，アクチュエータの物理的特性などを事前に完全に記述することが現実的に困難であることから，設計者の経験や抽象的なモデルに基づいた設計がなされている．また，研究レベルにおいても，このような複雑に絡み合った要因を制御理論的な形で定式化することが難しく，情報処理の計算の負荷を無視した数値実験や実機実験が数多く見られる．このような従来の最適設計法には限界があり，感覚運動統合システムがもつ潜在的な能力を最大限に引き出せていない可能性が高い．

一方，人間は，未知の制約が数多く存在する状況下でも，自らのもつ計算能力や運動能力をタスクの内容に整合させることで，システム全体として合理的なパフォーマンスを実現していると予想される．

本稿では，人間がもつこのような性質を工学的に実現することを目指して，処理系や制御系のもつ動特性（ダイナミクス）を外界の制約やタスクに応じて適応的に調整する手法を「ダイナミクス整合」として提起し，考察する．

2.2 ダイナミクス整合にもとづく感覚運動統合の構成

図 1 は，ダイナミクス整合にもとづく感覚運動統合システムをモデル化したものである．

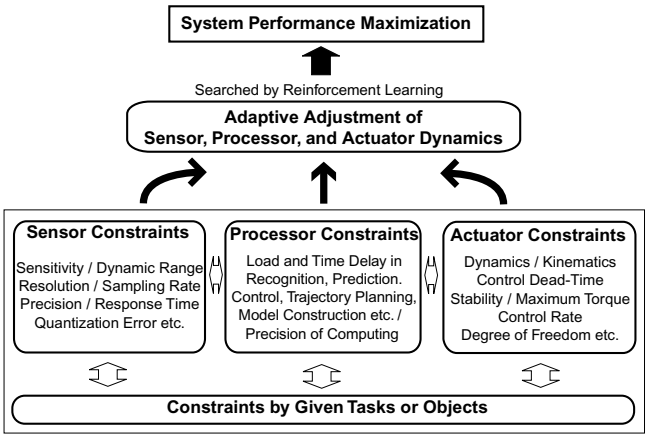

図 1. ダイナミクス整合にもとづく感覚運動統合のコンセプト．

アクチュエータ系やセンサ系，そして処理系には，外界の状況やタスクに応じてさまざまな物理的制約および計算的制約が存在し，互いに影響しあっている．このもとで，各系の時間的特性（ダイナミクス）を適応的に調整することによって，システム全体としてのパフォーマンスの最大化を図るというのが基本的なコンセプトである．

このような機能をシステムが持つためには，外界の内部モデルを保持して外界の状態を予測するとともに，システム自身のモデルも保持して，システム自身に起因する制約を考慮した観測戦略，計算戦略，運動戦略を能動的に獲得できるような構成にすることが望ましい．その際，物理的制約や計算的制約の間には複雑なトレードオフ関係が存在するため，システムはこれらを考慮しつつ，個々のサブシステムの動作の最適化ではなく，システム全体としての動作を保証する必要がある．

また，先に述べたように，外界，センサ，制御系がもつ物理的な特性や計算処理の時間遅れなどを事前に記述することは現実的に困難である．したがって，環境との相互作用の過程でそれらを学習によって適応的に獲得することが求められる．

感覚統合システムが適切に動作しているかどうかは，最終的には，システムが与えられたタスクを適切に実行しているかどうかによって判断されるべきである．そこで，このタスク遂行の善し悪しを評価関数としてシステム内部のパラメータを試行錯誤的に決定すれば，最終的にシステムとそのタスクの組み合わせに対して最高のパフォーマンスを達成するパラメータを獲得するようなアルゴリズムを構築することができる．筆者らは今回，試行錯誤的に解を求めていく能動的な学習法である強化学習 [2] を採用し，タスク遂行の評価関数を報酬として用いることにより，ダイナミクス整合を獲得するというアルゴリズムを提案する．

3 ターゲットトラッキングにおけるダイナミクス整合の実現

本稿では，ターゲットトラッキング（カメラによる運動物体への追従）を例題として，ダイナミクス整合のアルゴリズムを具体化する．本章では，この例題に即してダイナミクス整合のコンセプトを論じる．

このコンセプトの概要を図 2 に示す．システムは，可動範囲という物理的制約，および予測・軌道計画の時間遅れという計算的制約のもとで，強化学習により予測戦略を能動的に調整して，対象の推定位置の平均二乗誤差を状況に応じて最小化する．

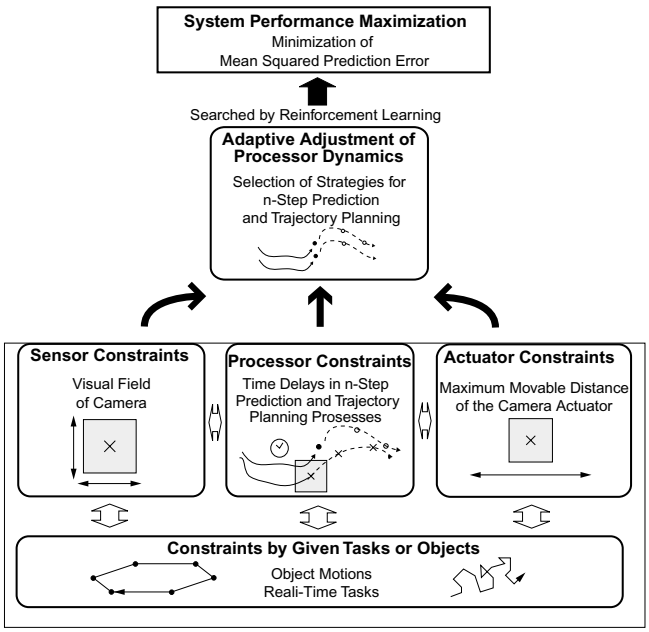

図 2. ターゲットトラッキングにおけるダイナミクス整合の獲得のコンセプト．アクチュエータの可動範囲という物理的制約，および予測・軌道計画処理の時間遅れという計算的制約のもとで，強化学習により予測戦略を能動的に調整することで，システム全体にとってのパフォーマンス（報酬）を最大化，すなわち平均二乗予測誤差を状況に応じて最小化する．

まず，トラッキングタスクの概要を図 3 に示す．ある 2 次元平面上（以下物体面と呼ぶ）を運動する対象の位置を推定・計測する問題を考える．すなわち，視野の限られたカメラを物体面に平行なある平面（カメラ面と呼ぶ）内で動かして対象を追跡し観測する．ここでは教師あり学習によって対象の運動の内部順モデルを構築すると同時に，そのモデルを用いて対象の軌道を予測する．このシステムのパフォーマンスは，対象の推定位置の平均二乗誤差によって評価することとする．

カメラを動かすアクチュエータには，可動範囲という物理的制約が存在する．また，カメラのセンサにも，視野角という物理的制約がある．したがって，システム全体のパフォーマンスを保証するためには，対象の運動を数ステップ先まで先読みし，これらの制約条件の下で対象を見失わ

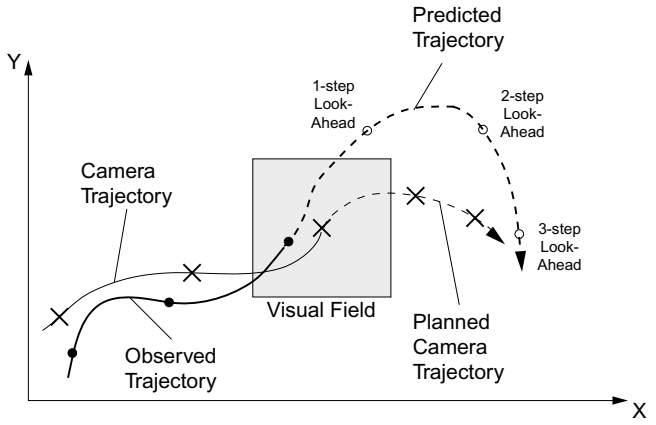

図 3. ターゲットトラッキングの概要．

ないような軌道計画をおこなう必要がある．しかし，一般に予測や軌道計画の計算処理は，先読み量が多くなるほど負荷が増大するため，実時間で動作させるにはいたずらに先読み量を増やすことができない．つまり，パフォーマンスを評価基準として見ると，センサやアクチュエータの物理的制約と処理系の計算的制約は相互にトレードオフの関係にある．この例題では，予測戦略（先読み量）を調整することで，最適なパフォーマンスを得ることを考える．よって，ダイナミクス整合の見地からすると，この両者を天秤にかけて，最適な予測戦略を選択する必要がある．

しかし，これらの制約条件は相互に依存しており，また状況にも影響されるため，このトレードオフ条件をあらかじめ陽に求めることは困難であるので，強化学習によって試行錯誤的に解を求める手法が有効になる．このアルゴリズムを次節で示す．

4 学習のアルゴリズム

以下では前節で述べたコンセプトにもとづいて図 4 のようなモデルを提案し，そのアルゴリズムを示す．アルゴリズムの大まかな流れは図 5 のようになる．

4.1 トラッキング

まず，トラッキングのアルゴリズムについて説明する．運動する対象の位置を $\boldsymbol{x}(t) = (x(t), y(t))^T$ とし，これを位置 $\boldsymbol{X}(t) = (X(t), Y(t))^T$ にあるカメラで観測する．ここでは観測に伴う誤差は考えない．

次に観測値を用いて，状態空間表現された対象の軌道の内部順モデル $\{A_i, \boldsymbol{B}_i\}$ を更新する．更新則は，予測位置 $\boldsymbol{x}_p(t)$ と観測された位置 $\boldsymbol{x}(t)$ との誤差 $\boldsymbol{e}(t) = \boldsymbol{x}_p(t) - \boldsymbol{x}(t)$ を教師信号として，勾配法にもとづく教師あり学習

$$\Delta A = -\beta_s \boldsymbol{e}(t)\boldsymbol{x}^T(t) \tag{1}$$

$$\Delta \boldsymbol{B} = -\beta_s \boldsymbol{e}(t) \tag{2}$$

にしたがう．ここで β_s は学習係数である．そして，更新

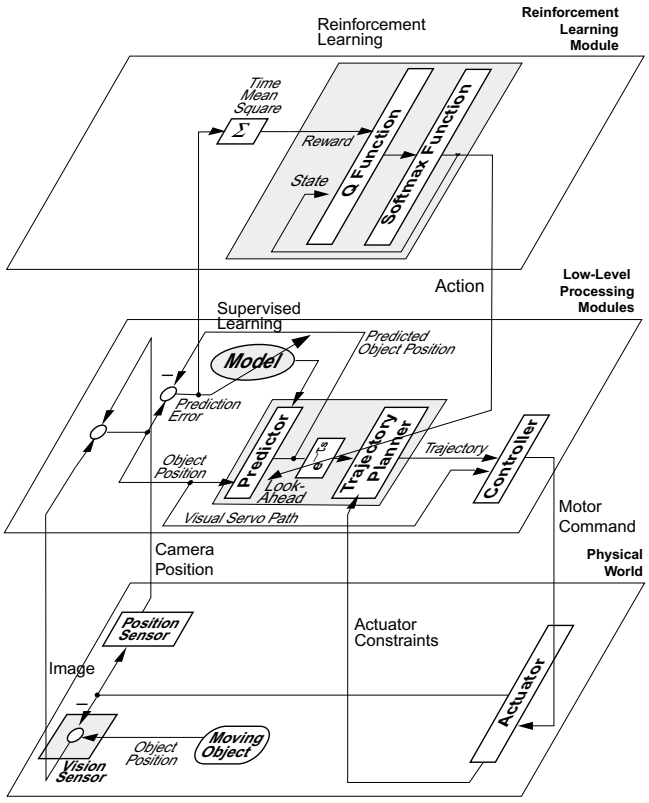

図 4. モデルの概要.

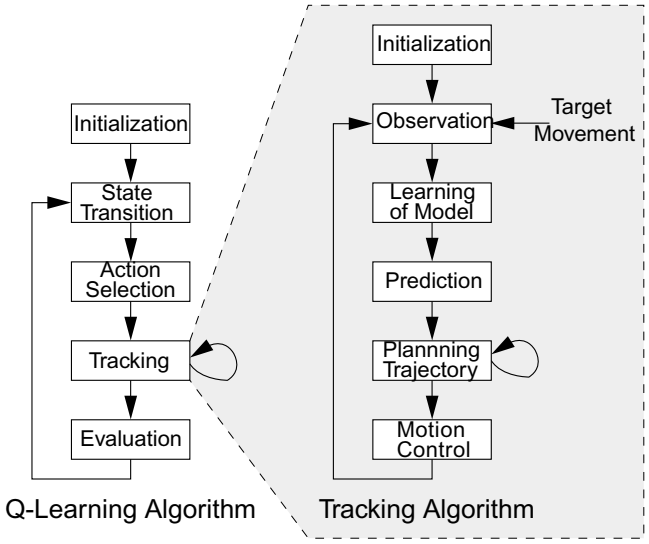

図 5. 数値実験のアルゴリズム.

した内部モデル $\{A_i, \boldsymbol{B}_i\}$ を用いて,

$$\boldsymbol{x}_p(t+1) = A_i \boldsymbol{x} + \boldsymbol{B}_i \tag{3}$$

にしたがって次の時刻での対象の位置 $\boldsymbol{x}_p(t+1)$ を予測する. $\boldsymbol{x}_p(t+k)$ $(k = 1, 2, 3, \ldots)$ は, 式 (3) を逐次適用していくことで求めることができる. よって, k 手先読みの場合, 予測値の時系列 $\{\boldsymbol{x}_p(t+1), \ldots, \boldsymbol{x}_p(t+k)\}$ が得られる. 予測が完全に正しいと仮定した場合, これをそのままカメラの目標軌道とできれば, 誤差 0 で追従できることになる.

また, アクチュエータの可動範囲の上限としては, 簡単のために 1 時刻の間に動ける最大変位を一定値とおき, X, Y 方向それぞれ $X_{\max}$, $Y_{\max}$ として定めた.

次に軌道計画を行なうにあたっては, 逐次的に精度を上げていく方法を用いた. 初期値とする暫定的な軌道を

$$\begin{aligned} X(t+i) &= X(t) + i s_x \min\{|x_p(t+k) - X(t)|/k, X_{\max}\}, \\ Y(t+i) &= Y(t) + i s_y \min\{|y_p(t+k) - Y(t)|/k, Y_{\max}\}, \\ \text{where} \quad & s_x = \mathrm{sgn}(x_p(t+k) - X(t)), \\ & s_y = \mathrm{sgn}(y_p(t+k) - Y(t)), \\ & \qquad \text{for } i = 1, \ldots, k \end{aligned} \tag{4}$$

というように最遠予測値方向への直線軌道として定める. この暫定軌道から, $|X(t+i) - X(t+i-1)| \leq X_{\max}$, $|Y(t+i) - Y(t+i-1)| \leq Y_{\max}$ の条件のもとで, 最急降下法

$$\Delta \boldsymbol{X}(t+i) = \mu(\boldsymbol{x}_p(t+i) - \boldsymbol{X}(t+i)) \text{ for } i = 1, \ldots, k \tag{5}$$

により, カメラの軌道が対象の軌道 $\{\boldsymbol{x}_p(t+i)\}$ に近づくように反復計算する.

ここで, 前節で示したようなダイナミクス整合の概念より, 予測および軌道計画の計算時間を陽に考える必要がある. 処理にかけられる時間は対象軌道の一定のサンプリング周期値 T に制限されているため, 先読み量 k が多くなるほど予測のための処理時間は増大し, 逆にカメラの軌道計画にかけられる時間は減少する. ここでは, 簡単のためこの制約条件を, 式 (5) の計算の反復回数 n を k に比例した回数 νk だけ減らすという形で実現した.

以上のような一連の処理によって, トラッキングを N 時刻の間おこなう. 対象を途中で見失ってしまった場合には探索モードに移行し, カメラ面内を一定速度でランダムに移動して対象を探すものとする. この間は, 上述の一連の処理は行なわれない. 対象が視野に入ったら, 探索モードから抜け, 再びトラッキングを始めるものとした.

4.2 Q 学習

前節で述べたトラッキングアルゴリズムは, 強化学習の 1 試行中に組み込まれている. 強化学習は Q 学習 [2] によって行なわれ, システムは, 選択される行動が収束するまで試行を繰り返す.

1 試行が終わるごとに先読み戦略に対して Q 値の更新が行なわれる. もともとの先読み量 k_t を Q 学習における状態 s, 選択した先読み量 k_{t+1} を行動 a とすれば,

$$\Delta Q(s,a) = \beta \left\{ r(t) + \gamma \left(\max_{a'} Q(s,a') - Q(s,a) \right) \right\} \tag{6}$$

によって Q 値が更新される. ここで $r(t)$ は報酬であり, 1 試行を通じての二乗予測誤差の和 (これを $1/N$ 倍すると

平均二乗予測誤差となる）の逆数

$$r(t) = \frac{1}{\sum_{t=1}^{N} |\boldsymbol{e}(t)|^2} \tag{7}$$

とした．そして，このQ値に対し，Boltzmann選択法により行動選択が行なわれる．行動 a が選択される確率 $P(a)$ は

$$P(a) = \frac{\exp(\alpha T Q(s,a))}{\sum_{a=a'} \exp(\alpha T Q(s,a'))} \tag{8}$$

となる．ここで α は定数，T は試行数である．

5 数値実験

5.1 実験の設定

まず，対象の軌道として，決定論的・周期的に運動する軌道（対象の運動モデルを構築可能）と，ランダムに運動する軌道（運動モデルを構築不可能）の2種類を考えた．

(a) 周期軌道

対象の軌道 $\boldsymbol{x}(t)$ は

$$\boldsymbol{x}(t+1) = A\boldsymbol{x}(t) + \boldsymbol{B} \tag{9}$$

$$A = \begin{pmatrix} \cos\omega & -\frac{a}{b}\sin\omega \\ \frac{b}{a}\sin\omega & \cos\omega \end{pmatrix} \tag{10}$$

$$\boldsymbol{B} = \begin{pmatrix} \frac{a}{b}q\sin\omega + p(1-\cos\omega) \\ -\frac{b}{a}p\sin\omega + q(1-\cos\omega) \end{pmatrix} \tag{11}$$

$$\boldsymbol{x}(0) = \begin{pmatrix} a\cos(\theta-\omega)+p \\ b\sin(\theta-\omega)+q \end{pmatrix} \tag{12}$$

で与えた．このとき，対象は楕円軌道 $a^2(x-p)^2 + b^2(x-q)^2 = 1$ を描く．今回は $a = 400$ [mm]，$b = 100$ [mm]，$p = 0$ [mm]，$q = 0$ [mm]，$\omega = \pi/3$ [rad]，初期位相 $\theta = \pi/12$ [rad] としたので，結果的に軌道は図6のような歪んだ6角形となる．

なお，システムが保持する対象の軌道の内部モデルの初期値としては，上に挙げた $\{A_i, \boldsymbol{B}_i\}$ に平均0，分散0.1の正規白色雑音を加算したものとした．

(b) ランダム軌道

一定速度100 [mm/s]で擬似的にBrown運動するような対象に対しても，同様にトラッキング実験を行なった．

以下ではカメラの視野の大きさを300 [mm]×300 [mm]とした．1回のトラッキングは $N = 30$ 時刻かけて行なうものとした．また，学習のパラメータは，$\beta_s = 10^{-9}$，$\beta = 0.01$，$\gamma = 0.999$，$\alpha = 0.5$，$\mu = 0.1$，$n = 9$，$\nu = 1$ とした．

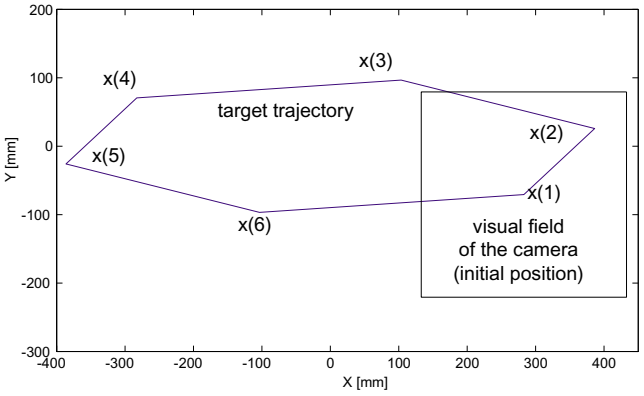

図6. 対象の軌道とカメラ画角．

先読み量の候補として0ステップ，1ステップ，2ステップ，3ステップの4種類を用意し，最大可動距離を $X_{\max} = Y_{\max} = D_{\max}$ として，これを155 [mm]から355 [mm]まで25 [mm]ごとに変化させて挙動を調べた．なお，先読み量が0ステップの場合は予測を行なわないため，追従誤差を予測誤差とみなして評価に用いた．

5.2 実験結果

まず，最大可動距離 $D_{\max}$ がそれぞれ255 [mm]，355 [mm]の場合に，今回の実験でパフォーマンスの評価基準として用いた，平均二乗予測誤差（報酬の逆数）がどのように推移していったかを図7および図8に示す．強化学習により，パフォーマンス最大化を達成できていることがわかる．他の $D_{\max}$ についても同様な結果が得られた．

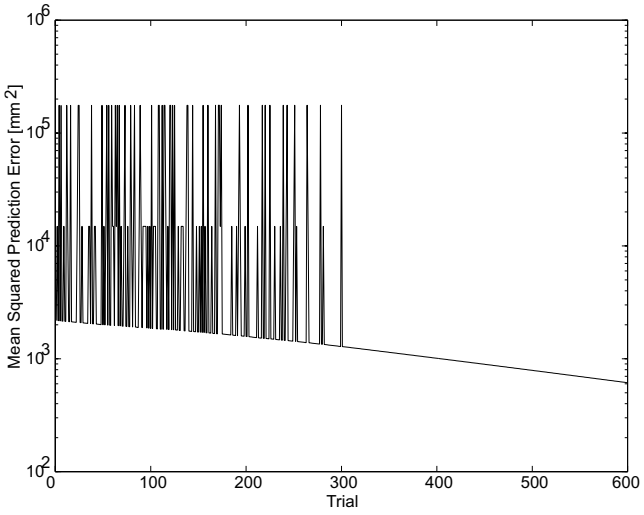

図7. $D_{\max} = 255$ [mm] の場合の，平均二乗予測誤差の推移．

次に，同様に最大可動距離 $D_{\max}$ がそれぞれ255 [mm]，355 [mm]の場合に，各先読み量に対し式(8)で表される

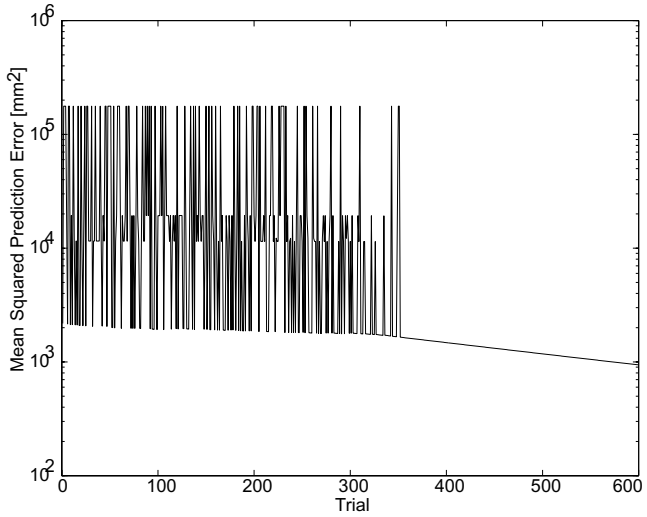

図 8. $D_{\max} = 355$ [mm] の場合の，平均二乗予測誤差の推移．

先読み量の選択確率 $P(a)$ が学習によってどのように推移していくかを 図 9 ，図 10 に示す．数百回の反復で収束し，ただ一つの最適な先読み量だけを選ぶようになることがわかる．他の $D_{\max}$ についても同様な結果が得られた．

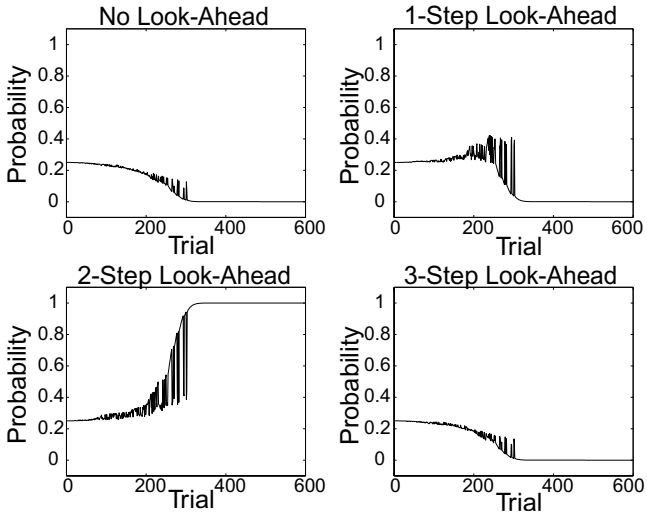

図 9. $D_{\max} = 255$[mm] の場合の，各先読み量の選択確率の推移．2 手先読みに収束している．

さらに，それぞれの最大可動距離について，600 試行が終了した時点で，各先読み量に対する Q 値を正規化したものを図 11 に示す．割合が多い先読み量ほど，式 (8) に従って高い確率で選択される．数値実験の結果，最大可動距離が小さくなる，すなわちアクチュエータの制約が厳しくなるにつれ，先読み量のより多い戦略が選択されるようになった．また，対象が Brown 運動の場合には，先読み量 0（ビジュアルサーボ）という戦略が多く選択されるようになった．

これは，以下のように解釈できる．

まず，アクチュエータの制約が緩い場合について考えてみる．この場合には，計算的制約がパフォーマンスに与える影響が相対的に大きくなる．つまり，先読み量を増やすと計算負荷により軌道計画が初期で打ち切られ，得られる軌道が最適軌道からずれてしまうため，最終的なパフォーマンスが低下する．その結果，先読み量が少ない戦略が多く選択される．

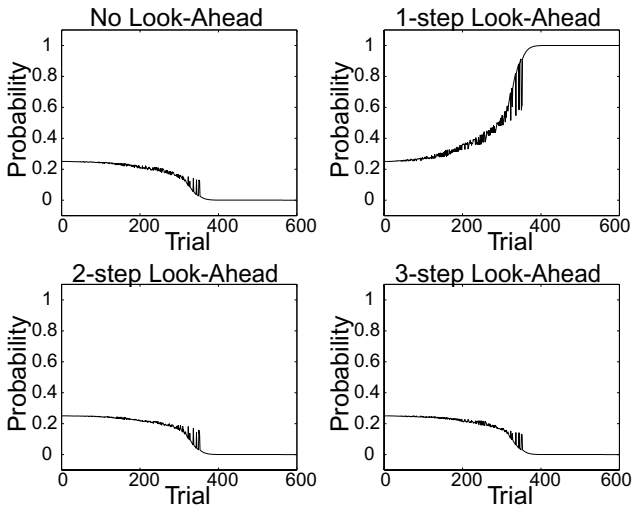

図 10. $D_{\max} = 355$ [mm] の場合の，各先読み量の選択確率の推移．1 手先読みに収束している．

次に，アクチュエータの制約が厳しい場合について考える．この場合には計算的制約に比べて物理的制約の影響が無視できない．例えば，図 12 のように最大可動距離が対象の運動距離と同程度あるいはそれ以下の場合には，1 手先読みだと，図 12(a) に見られるように途中で対象に追従しきれなくなることがある．この場合は図 12(b) に示したように，2 手先読みのほうが，たとえ計算負荷は大きくても追従能力は高い．つまり，計算的制約よりも物理的制約が効いてくるような状況では先読み量を増やして，物理的制約を補償していると考えられる．この傾向は，最大可動距離が小さくなるにつれて顕著になっていることが，実験結果からわかる．

また，対象がランダム運動をしている場合は，もはや予測をすることに全く意味がない．よって，負荷のかかる予測処理を行なわないという，ビジュアルサーボという戦略を選択したと見ることができる．

最後に，最大可動距離 $D_{\max}$ がそれぞれ 255 [mm]，355 [mm] の場合の，600 試行終了時点でのトラッキングの様子をそれぞれ図 13 および図 14 に示す．図 13 では先読み量は 2 であり，カメラがあまり動かずに視野の広さをフルに使ってトラッキングしていることがわかる．図 14 では先読み量は 1 であり，対象の軌道にほぼ沿いながらトラッキングしている様子がわかる．

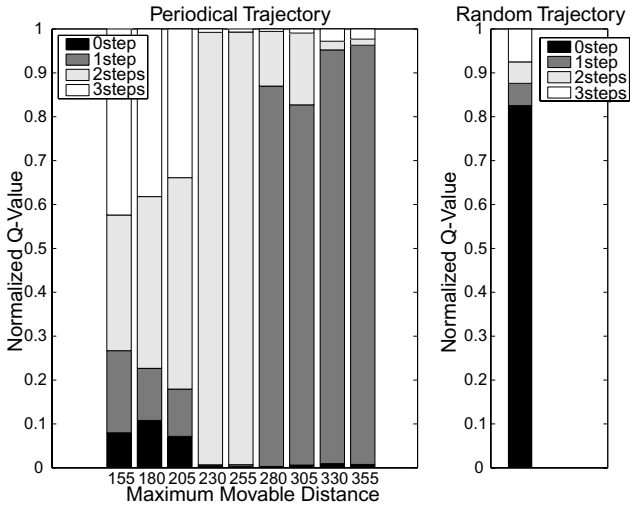

図 11. アクチュエータ制約（最大可動距離）を変化させたときの，各先読み量に対する正規化 Q 値．割合が多い先読み量ほど高い確率で選択されることになる．アクチュエータの制約が厳しくなるにつれ，先読み量のより多い戦略が選択されるようになった．また，対象が Brown 運動の場合には，先読み量 0 という戦略が多く選択されるようになった．

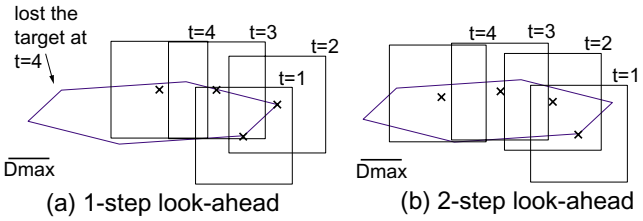

図 12. アクチュエータの制約が厳しい場合のトラッキング．(a) 1 手先読みの場合，途中で対象を見失う．(b)2 手先読みなら，対象を見失わずに追従できる．

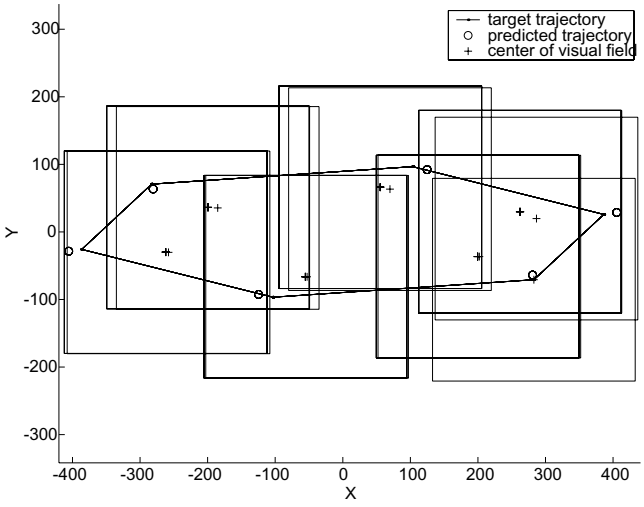

図 13. $D_{\max} = 255$ [mm] の場合のトラッキングの様子．カメラがあまり動かずに視野の広さをフルに使ってトラッキングしている．

5.3 考察

実験の結果，システムは強化学習によって，アクチュエータの物理的制約に応じて異なる予測戦略を選択するようになることがわかった．予測誤差のプロットを見ると，先読み量を試行錯誤している間は誤差が大きいが，先読み量が 1 種類に収束するにつれて誤差が最小値に近づく．つまり，先読み量を適切に選択することで，状況に応じたシステムパフォーマンスの最適化が達成されている．

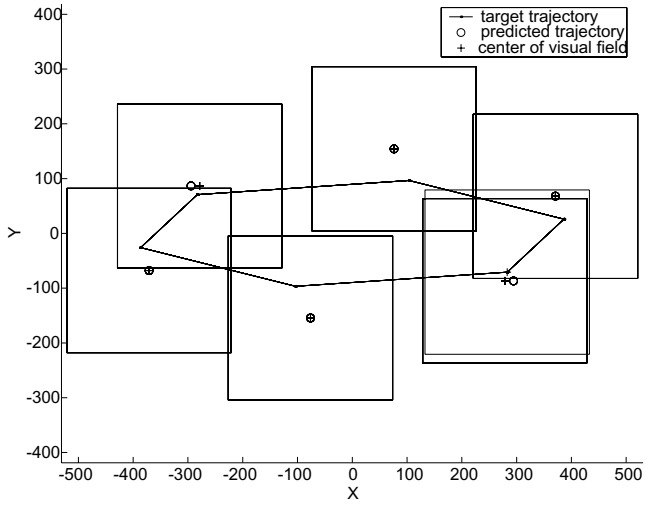

図 14. $D_{\max} = 355$ [mm] の場合のトラッキングの様子．対象の軌道にほぼ沿いながらトラッキングしている．

このように，このシステムは物理的制約や計算的制約というトレードオフ条件のもとで，状況に応じたパフォーマンス最大化を学習によって達成できた．よって，本稿で提案したモデルおよびアルゴリズムはダイナミクス整合の獲得に有効であったといえよう．

さらに，振舞いやタスクといったトップダウン的な視点から結果を眺めてみると，アクチュエータの制約が緩い場合には対象に忠実に追従し，制約が厳しい場合には画面の広さを利用して制約範囲内でトラッキングしようとしている．また，対象がランダムな運動をしている場合に最終的にビジュアルサーボ戦略が獲得されたことは，もはやタスク遂行にとって意味のない先読み処理を避けるようになったことを意味している．このように，振舞いレベル，タスクレベルからこのシステムの挙動をみても，きわめて合理的なトラッキングをしている．

なお，今回はアクチュエータの可動範囲をパラメータとして変化させたが，これを一定としてセンサの視野角，あるいは対象の運動の激しさを変化させても，同様の結果が得られるであろう．この例に限らず，一般にダイナミクス整合の考慮においては最適解は状況やタスクに依存すると考えられるため，今回のように強化学習による試行錯誤的な最適化は有効な方法であると思われる．

この実験で重要なことは，従来はあまり顧みられてこなかった物理的制約や計算的制約の相互作用を陽に考慮したということである．今回の例題の設定は実際のシステムに即したものではないが，このような問題意識は他のシステムを実際に動かす際にもやはり重要になってくると思われ

る．特に，ロボットにその高速性を生かした運動をさせる場合には，物理的・計算的制約がクリティカルになり，時間方向のパフォーマンスに大きく影響することが予想される．今後，実機に即した設定でこの問題を検証していくことが求められる．

6 むすび

本稿では，さまざまな物理的・計算的拘束条件の下で，情報処理系や制御系の時間的な特性，すなわちダイナミクスを適応的・能動的に調整することで，システム全体としてのパフォーマンスの最大化を実現する，という考えをダイナミクス整合として提起した．そしてこの問題に対するひとつのアプローチとして，パフォーマンスを報酬とした強化学習を用いて具体的なアルゴリズムを構成し，ターゲットトラッキングを例題とした数値実験によりその効果を検証した．

今回紹介したような例題はダイナミクス整合に対する一つのアプローチに過ぎないが，同様の考え方はロボットアームの運動制御やロボットハンドによるマニピュレーションなどにも応用できると予想される．今後は人間のダイナミクス整合機能について理論面・実験面から検証をすすめてその機構を解明するとともに，ロボティクスへの応用によってその有用性を確かめていきたい．

参考文献

[1] A. Namiki, Y. Nakabo, I. Ishii and I. Ishikawa, “1ms Sensory-Motor Fusion System,” in *IEEE/ASME Transactions on Mechatronics*, Vol. 5, No. 3, pp. 244–252, 2000.

[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, 1998.