筋骨格ロボットを用いた跳躍運動の学習 *

柿谷慧 *1, 新山龍馬 *1, 國吉康夫 *2

Learning of Jumping Motion with Muscloskeletal Robot

Kei KAKITANI*3,Ryuma NIIYAMA and Yasuo KUNIYOSHI

*3 The University of Tokyo, Graduate School of Information Science and Technology
Eng. 2nd bldg. , 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, JAPAN

Learning Multi-DOF motion is a challenging task. With previous learning algorithm, the number of required trials has been too large for real robot. Recently we proposed a new method called “Exploration of Virtual Goal Switching Pattern” which is a quick method to discover semi-optimal solution. In the present paper, we propose “Muscle Command Switching Pattern”, which is an extension of our previous method for use with musculoskeletal robots. We describe a series of jumping motion learning experiments with a musculoskeletal robot called “Mowgli2”. The result shows that vertical jumping motion is acquired with only 100 trials (about 25 minutes).

Key Words : Motion Learning, Muscloskeletal Robot, Jumping Motion, Dynamic Multi-DOF Motion, Muscle Command Switching Pattern

1. 序　　論

運動はロボットの重要な機能であり，運動を生成するアプローチとして学習による運動の獲得がある．運動を学習することにより，設計者が想定していない未知の環境や変動する環境において環境に適した運動が獲得されることが期待されており，またロボットの力学的解析が困難な場合においても適用できると考えられている．

ロボットが様々なタスクに対応する場合，多自由度が必要であると考えられる．また，ロボットの運動性能を活かして素早く動くためには，ダイナミックな運動を扱う必要がある．運動の学習については多くの研究が行われてきたが，そのなかでダイナミックで多自由度を持つ運動を学習した例はほとんどない．そこで，本研究では「ダイナミック多自由度運動」[(1)]の学習を扱う.「ダイナミック多自由度運動」とは，多自由度を持つロボットにおける運動で，かつ速度や加速度の影響が無視できない運動である．

運動学習の最終的な目的は実機においてうまく機能する運動を獲得することである．しかし，ロボットや環境を完全にモデル化するには非常に多くの労力を必要とするために，シミュレーションと実世界を完全に一致させることは困難であり，実機での学習が必要となる．

ダイナミックな運動としては跳躍・着地や走行，投球など様々なものがあるが，中でも跳躍は人間や生物にとって基本的な運動である．この運動を脚ロボットにおいて実現した例として文献[(2)]がある．この研究では跳躍・着地のために空気圧人工筋を用いており，この人工筋は質量に対する出力が大きいために高い跳躍が可能となっているが，このようなロボットの跳躍を解析的アプローチにより行うことは困難である．

以上より，本研究は実ロボットにおいて「ダイナミック多自由度運動」の学習・獲得を行うことを目的とし，そのタスクとして，解析的アプローチによる解決が困難な筋骨格ロボットによる跳躍動作を扱う．

2. 実機によるダイナミック多自由度運動の学習

本研究においては,「実機」による「ダイナミック多自由度運動」の学習を扱う．ダイナミック多自由度運動は，ダイナミックなことにより速度を状態変数として扱う必要があり，また多自由度なため多くの状態変数や駆動自由度を持つ．このため，静的な場合や少自由度の場合と比較して，扱うべきパラメータや変数が非常に多くなる．そして，実機による学習はシミュレーション上とは異なる点が存在し，その中でも重要なのが運動のばらつきやセンサの誤差といった「不確かさが存在する」という点と，環境や身体などが「時間により変動する」という点である．特に評価値へのノイズの混入や変動は学習に大きな影響を及ぼす．このため探索手法がこ

* 原稿受付 2008 年 10 月 24 日
*1 学生員, 東京大学 (東京都文京区本郷 7-3-1)
*2 正員, 東京大学
Email: kakitani@isi.imi.i.u-tokyo.ac.jp

れらノイズや変動に対応している必要がある．さらに，実機で学習が行える現実的な試行回数であるかという点も大きな制約となる．

運動学習について様々な研究が行われてきたが，手法から大きく分類すると，強化学習を用いるものと GA 等の最適化手法を用いるものがある．この中でも「ダイナミック多自由度運動」の実機による学習に関連した研究として，以下のようなものがある．強化学習を用いた研究として，3 自由度 4 リンクのロボットでダイナミックな立ち上がり動作を学習させた例[3]や 4 脚 8 自由度のロボットで準静的な歩行動作を学習させた例[4]などがある．また最適化手法を用いたものとして，2 自由度 3 リンクのロボットで山登り法による探索を行い，ダイナミックな動作を含む様々な前進運動を獲得させたもの[5]や，Genetic Algorithm(GA) を用いてシリアルリンク構造のロボットの跳躍運動学習をシミュレーション上で行ったもの[6]などが挙げられる．GA と強化学習を組み合わせた研究[7]では，強化学習における状態遷移を GA により絞ることにより，静的な運動ではあるが非常に多くの自由度を扱っている．しかし，これらはいずれも自由度が少ない，運動が準静的を仮定している，実機での学習でない，という理由から，4 自由度を超えるような「多自由度」のロボットにおいて「ダイナミック」な運動を,「実機」で学習した研究はほとんどない．

その理由として，強化学習においては状態空間が爆発的に大きくなることで，状態ー行動間の写像や状態価値関数の学習が困難になり，また最適化手法を用いた場合においても探索空間が非常に大きくなるため，学習回数が実機で可能な回数を超えると考えられる．このため，多くの研究では人間がタスク依存の事前知識を与えることでこの問題を回避している．しかし，タスク依存の事前知識を用いると未知の環境や身体における学習は困難となり，根本的に問題を解決しているとは言いがたい．

これに対し著者らは，タスクに依存しない運動表現として「仮想目標切替パタン」を用いた手法を提案し，シミュレーション上において手法の有効性を確認した[1]．この「仮想目標切替パタン」は，目標へ近付く方向へ作用するコントローラを仮定し，そのコントローラへの目標値を時変化させて与えるという運動表現をとることで，学習の探索対象となるパラメータを少なくしている．次章ではこの「仮想目標切替パタン」を筋骨格ロボットへ拡張した「筋指令切替パタン」を提案する．

また実機において学習を行うために，次章では，それまでの試行で最も評価値の高かった運動を再現することで，評価値のノイズや変動に対応する方法を示す．

3. 跳躍運動の学習手法

ダイナミックな運動学習の研究のさきがけとして，文献[5]がある．この中で銅谷は，運動学習を構成する要素として,「運動パタンの表現」「運動の評価関数」「評価関数の最適化手法」の 3 つを挙げている．基本的な学習の流れとしては，この「運動パタンの表現」により運動をパラメータで表し，あるパラメータで運動を実行する．実行された運動を「運動の評価関数」によって評価し，その評価値を元に「運動の評価関数」が最も高くなるパラメータを「最適化手法」により探索する．

この枠組みを受けて，本研究では「運動パタンの表現」として「仮想目標切替パタン」[1]を筋骨格系に拡張した「筋指令切替パタン」を用いる．また「評価関数の最適化手法」ではランダム探索と，最大評価値更新を行う山登り探索を組み合わせて用いる．

3·1 仮想目標切替パタン　できるだけタスクに依存する前提知識を含まない運動表現として，著者らは「仮想目標切替パタン」[1]を提案し，実機で学習を行うことでその有効性を確認した[8]．この「仮想目標切替パタン」では，まず，目標姿勢に向け制御する目標値コントローラ C を用意する．次に，運動が行われる時間を N 個のフェーズに区切り，それぞれのフェーズ i の長さを時間幅 T_i とする．このフェーズに対応する目標姿勢を $\boldsymbol{g}_i$ とする．そして，時刻 t に対応するフェーズ i の目標姿勢 $\boldsymbol{g}_i$ を目標値コントローラ C に与えることで，時刻によって目標姿勢が切り替わり，これによって運動を制御する．この時間幅 T_i と目標姿勢 $\boldsymbol{g}_i$ を運動を表現するパラメータとし，これらのパラメータを学習の対象とする．よって，この運動表現において運動を学習するということは，評価値を最大にする N 個の時間間隔 T_i・目標姿勢 $\boldsymbol{g}_i$ のパラメータセットを求めることである．

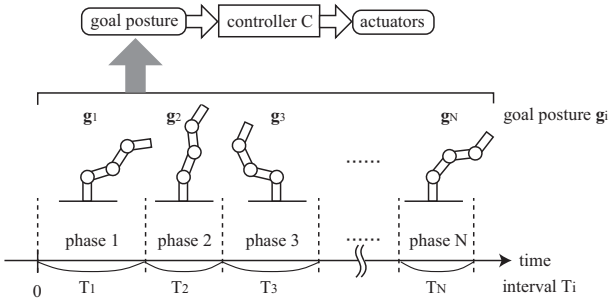


Fig. 1　Virtual Goal Switching Pattern.

3·2 筋指令切替パタン　筋骨格ロボットにおいては，関節を拮抗駆動とすると，屈筋と伸筋へ与える指令によって関節角の平衡点と剛性の両方を制御することができる．この拮抗駆動は，指令を与えると平衡点へ向かうトルクを出すという点で「仮想目標切替パタン」

におけるコントローラの役割をはたし，さらにこのコントローラは剛性という制御パラメータを変化させることが可能である．これらより拮抗駆動において屈筋と伸筋への指令を与えることは「仮想目標切替パタン」における仮想目標とコントローラの制御パラメータを与えることとほぼ等価と考えられる．

以上のような理由から，本研究においては筋骨格ロボットを扱うにあたって「仮想目標」を筋骨格系に拡張した「筋指令」を用いる．また，これを用いた運動表現を「筋指令切替パタン」と呼ぶ．

「筋指令切替パタン」では,「仮想目標切替パタン」と同様に運動が行われる時間を N 個のフェーズに分ける．各フェーズの時間幅を T_i とする．この各フェーズ i に対応する筋指令を $\boldsymbol{m}_i$ とし，時刻がフェーズ i の間は人工筋に筋指令 $\boldsymbol{m}_i$ が与えられる．時刻により筋指令がステップ状に切り替わることで，運動が生成される．時間幅 T_i と筋指令 $\boldsymbol{m}_i$ を「筋指令切替パタン」のパラメータとし，これらが学習の対象となる．

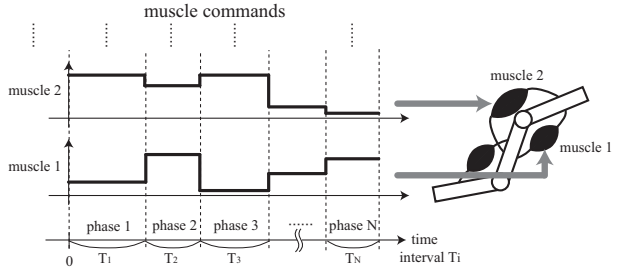


Fig. 2　Muscle Command Switching Pattern.

一般的に強化学習においては，状態から行動へのフィードバック関数を求めるが，関数は全ての状態から全ての行動への写像であるため，一般には非常に多くのパラメータから構成されている．これが学習の回数を増大させている原因のひとつであると考えられ，また同じ状態を何度も通過しつつも異なる動作を行う手順的な運動を表現するのが難しい．これに対し著者らの提案する「仮想目標切替パタン」や「筋指令切替パタン」は，フィードバックはあらかじめ与えたローカルなフィードバック器にまかせ，これらへの指令を手順的に与えることで運動を生成する．この際に与える指令も，各微小時間毎に指令を与えるような詳細な与え方ではなく，運動する時間を少数のフェーズに分け，同じフェーズの間は同じ指令値を与えるという非常に粗い与え方をする．このため，フィードバック関数そのものを求める場合に比べパラメータが少なく，学習に有利であると予想される．また，手順的な運動を容易に表現できるという特徴がある．

3·3　ランダム探索と，最大評価値更新を行う山登り探索

本研究では，学習に用いる最適化手法として，ランダム探索と山登り探索を用いる．まずランダム探索を複数回行い，この中で最も評価値の高いパタンを抽出する．つぎにこのパタンから山登り探索を行う．探索においては，探索すべき各パラメータの値域を $[0,1]$ に正規化する．

ランダム探索に使用する乱数は，筋指令では 0 から各筋肉の指令値の上限値までの範囲，時間間隔では 0 から上限値までの範囲に，一様に分布するものを使用する．

山登り探索では，それまで試行した中で最も評価値の高いパタンのパラメータ $\boldsymbol{p}_{best}$ に微小変位 $\Delta\boldsymbol{x}$ を加えた近傍のサンプル $\boldsymbol{p}_{best}+\Delta\boldsymbol{x}$ を生成する．このサンプルを用いて運動を実行し，評価値がより大きくなればこのパラメータを採用，そうでなければ棄却する．この 1 回を 1 試行と呼ぶ．加える微小変位 $\Delta\boldsymbol{x}$ は，パラメータの全次元で変化させる微小変位と，１次元のみ変化させる微小変位の二種類を確率的に選択して用いる．前者は，微小変位のスケールパラメータを δ としたとき，変位のすべての要素 $(\Delta\boldsymbol{x})_i$ を $[-\delta,\delta]$ の一様乱数とすることにより生成する．後者は，微小変位のうち確率的に選択されたある一つの要素 i のみ $[-\delta,\delta]$ の一様乱数とし，他の要素を 0 とする．これは，前者により学習の初期で大まかな探索を行い，後者により極値付近で細かな探索を行うことで探索回数を少なくすることを狙いとしている．この二種類の微小変位を 1:1 の割合で確率的に選択し，サンプルを生成する．

しかし，実世界で学習する場合は，この山登り探索だけでは容易に破綻する．この山登り探索は，新たなサンプルを採用するか棄却するかの基準として，今までの探索の中で最も高い評価値（「最大の評価値」）を用いている．そのため，この評価値が不当に高いとサンプルが全て棄却され，学習が停滞するからである．この「不当に高い」というのは，ノイズによって評価値が真の値よりも大幅に大きくなる場合や，評価関数の変動によって，以前は正しい評価値だったものがもはや正しくなくなる場合である．この問題に対し文献[1]や本研究ではある確率で，それまで試行した中で最も評価の高いパラメータで運動を再現し，その結果得られる評価値を「最大の評価値」として更新することを行う．これにより，ノイズにより不当に高くなっていた「最大の評価値」を正しい値に修正する効果が期待される．また，「最良」とされた運動の評価値を更新することで，環境の変動にも対応することができる．

4. 筋骨格ロボットによる跳躍運動の学習実験

著者らが開発した筋骨格ロボット “Mowgli2” を用いて，出来るだけ高く飛ぶことを目標とした跳躍運動の学習を行った．

4·1 筋骨格ロボット Mowgli2 の概要 運動学習に用いた筋骨格ロボット “Mowgli2” の外観を図 3 に示す．股関節・膝関節・足関節を備えた多自由度 2 脚ロボットで，総自由度数は 6 DOF，重量は 3 kg，脚を伸ばした状態での全長は 0.84 m，関節の可動範囲はすべて 150 deg である．筋配置，関節可動角およびリンク長を図 4 に示す．

片脚を駆動する筋の本数は 6 本で自由度に対して冗長な構成となっており，そのうち 4 本は圧力比例弁で制御される McKibben 型人工筋で，残りの拮抗筋 2 本は受動バネで代用されている．また，単関節筋と二関節筋が混在している．状態を計測するため，関節角度センサ，体幹の位置と姿勢を計測する姿勢センサ，空気圧筋の内圧を計測する圧力センサを搭載している．多様な運動を実行するため，通常使われる電磁弁（ON/OFF 弁）ではなく，空気圧筋の内圧を連続値で制御できる圧力比例弁を採用している．圧力比例弁と制御回路を本体に搭載し，空気圧と電源は外部から供給する．外部とつながるケーブルは最小限とすることで本体の運動を妨げない．

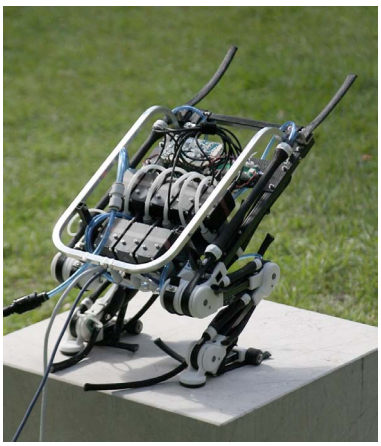


Fig. 3 Musculoskeletal robot : Mowgli2.

跳躍の鉛直高さを測るため，腰関節の付近に MicroStrain 社の姿勢センサ 3DM-GX1 を搭載した．RS-232C により Mowgli2 上の CPU ボードで加速度，角速度を取得している．データの取得はおよそ 77fps で行われる．取得した角速度から姿勢のクォータニオンを積算，この姿勢を利用して加速度データから重力加速度を差し引き，二階積分することで位置を算出している．この姿勢センサの位置における鉛直方向の変位を跳躍高さとする．

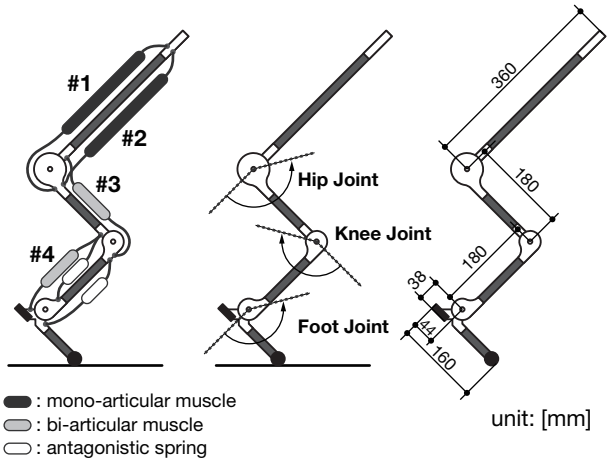


Fig. 4 Muscles, joints, and link length of Mowgli2.

4·2 学習の実験設定 学習における運動表現としては「筋指令切替パタン」を用い，圧力比例弁への入力電圧を筋指令とした．フェーズ数は 2 とした．また，運動は左右対称であるとして，駆動自由度は 4 つとした．１つのフェーズあたりの最大時間幅は 0.5s とした．

前半のランダム探索 50 試行，後半の山登り探索 50 試行を 1 探索とし，この 100 試行からなる探索を 3 回行った．山登り探索における局所探索範囲の大きさは $\delta = 0.1$ とした．最良運動の再現と最良評価値の更新は，10% の確率で行った．

学習の評価関数は，時刻 t における鉛直方向の位置 $z(t)$ を用いて

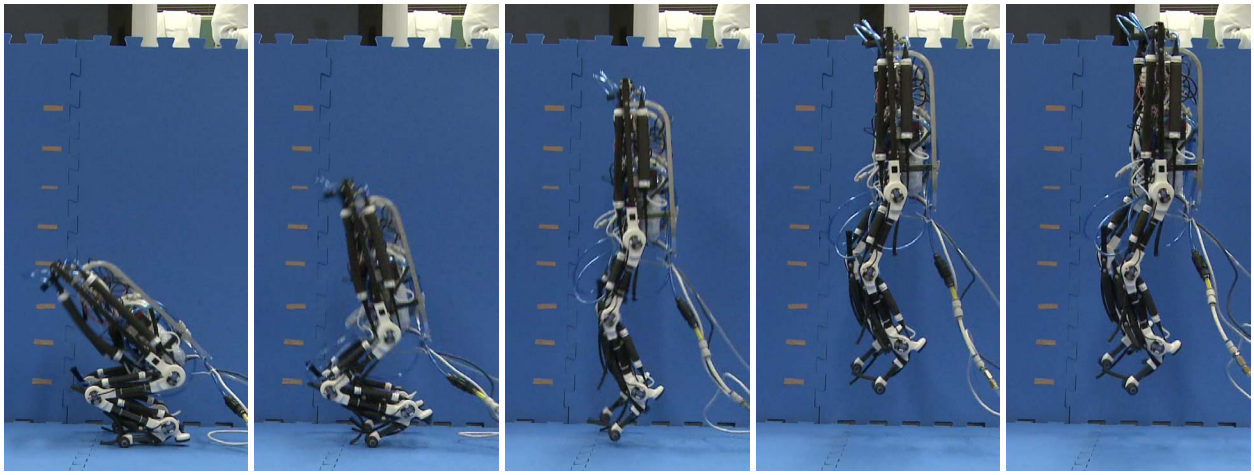
$$(evaluation\ value) = \max_{t\in[0,1.5]} (z(t) - z(0)) \qquad (1)$$

とし，姿勢センサの中心位置における鉛直方向の変位の最大値とした．

4·3 結果と考察 3 回行った探索の結果，すべてにおいて跳躍運動が獲得された．学習曲線を図 5 に示す．それぞれの探索における最高跳躍高さは，434mm，430mm，440mm であった．学習に要した時間は 1 探索あたりおよそ 25 分であった．

1 探索目に獲得された最も評価値の高い跳躍運動（97 試行目）のスナップショットを図 6 に示す．また，1 探索目と 2 探索目のランダム探索 100 回における評価値（跳躍高さ）のヒストグラムを図 7 に示す．学習曲線で次第に評価値が高くなっていること，スナップショットにおいて鉛直方向への跳躍が実現されていることや，ヒストグラムが示すように 430mm 付近の跳躍が偶然起こる可能性は低いことから，跳躍運動が獲得されていることがわかる．

図 8，9 は，この運動における筋指令，空気圧人工筋の圧力，各関節角の時系列データである．図 8 より，こ



t = 0.00 [s]　　t = 0.10 [s]　　t = 0.20 [s]　　t = 0.30 [s]　　t = 0.40 [s]

Fig. 6　Snapshot of acquired jumping motion (1st exploration, 97th trial).

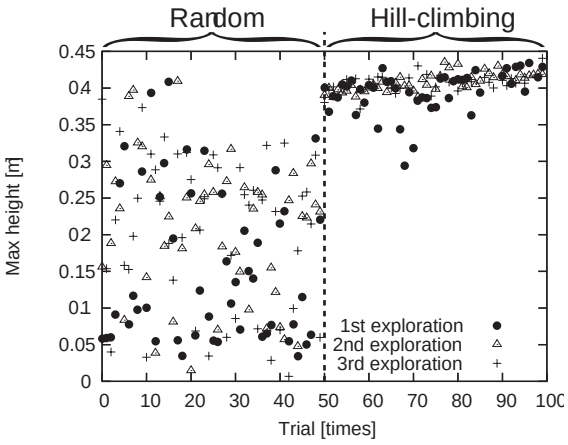


Fig. 5　Learning curve of each exploration.

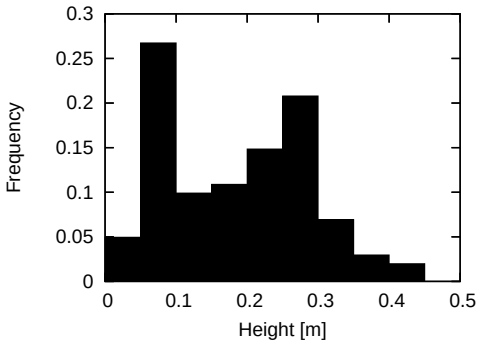


Fig. 7　Histogram of evaluation values (jump height) of random trials (random 100 trials of 1st and 2nd exploration).

れらのグラフにより，まず筋肉 1・3・4 を活動させ，筋肉 2 の活動を抑えることで脚を伸ばし，跳び上がっていることがわかる．また，図 8 より，筋指令（空気圧指令）に対し圧力の反応が 0.04s ほど遅れている．このよ

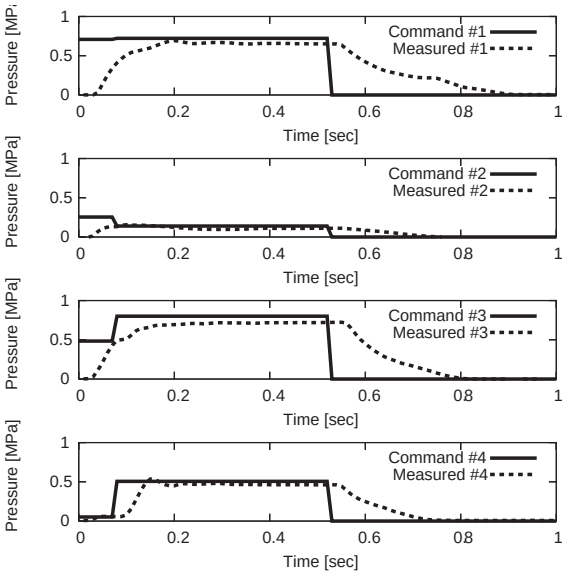


Fig. 8　Muscle command and pressure of the best motion (1st exploration, 97th trial).

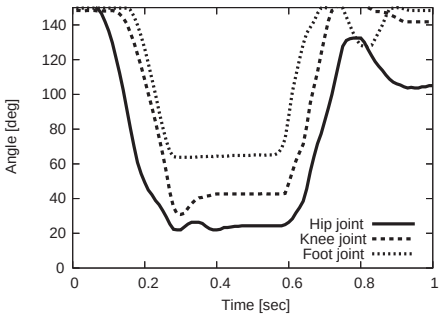


Fig. 9　Joint angle of the best motion (1st exploration, 97th trial).

うに非線形性を有するシステムであるため，解析的アプローチでは扱いにくいことが推測される．

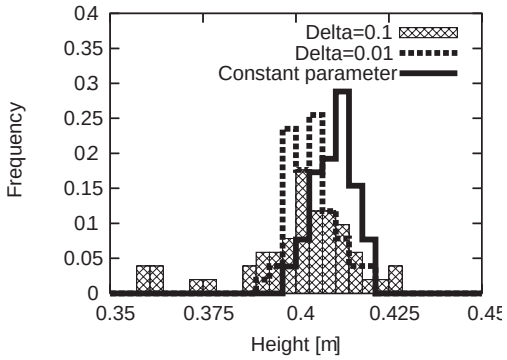


Fig. 10 Histogram of evaluated value with constant parameter, delta=0.1 or delta=0.01.

今回の学習においては前半でランダム探索，後半で山登り探索を行っている．これは前半のランダム探索により広域な全探索範囲から大まかな解の候補を抽出し，後半で抽出した解をさらに洗練させる戦略をとっている．図 5 の学習曲線からランダム探索と山登り探索についての評価関数の変化を見ると，確かにランダム探索により大まかな解を抽出，その後の山登り探索でその解の洗練化を行っていることが見て取れる．

また，山登り探索における探索範囲を表す δ の変化が探索に与える影響を調べた．図 10 は，先ほどの 1 探索 16 試行目のパラメータで 50 回実行した場合と，このパラメータを中心として $\delta = 0.1, 0.01$ で局所ランダム探索を 50 回行った場合の評価値の分布を示している．パラメータ一定の場合の分布は，このパラメータにおいてノイズがどの程度評価値にのっているかを表しており，ノイズの分布ととらえられる．$\delta = 0.01$ の場合は，パラメータ一定の場合の分布と比較すると両者の分布が似ているために，評価値の変化が探索による微小変位によるものなのか，それともノイズによるものなのかの区別がつかなくなっている．よって学習が進まないと考えられる．$\delta = 0.1$ の場合は，パラメータ一定の場合の分布の外にも分布が広がっていることから，大きな変化が評価値変化があればノイズではなく探索によるものであると区別でき，学習が進む可能性が高いことがわかる．

以上より，探索においてパラメータの変化に対応する評価値の変化が実世界のノイズにかき消されるほど小さい場合，学習が進まなくなる可能性が高く，評価値に混入するノイズの大きさによって山登り探索の探索範囲を変化させる必要があると考えられる．

5. 結論と今後の展望

本研究は，運動表現として「仮想目標切替パタン」を筋骨格系へ拡張した「筋指令切替パタン」を提案し，従来の解析的アプローチでは制御が困難な筋骨格ロボット Mowgli2 による跳躍運動の学習を行った．学習の結果，100 回（約 25 分）の試行により跳躍運動が獲得され，本手法がこのような実機のロボットでの学習に対して有効であることが示された．また，ランダム探索によって大域的に解を探索し，山登り探索によって解の洗練化をおこなうという戦略が実際に行われていることを示した．さらに，探索範囲を表すパラメータを変化させて分布をとり，実世界におけるノイズと探索範囲との関係について論じた．

本研究では鉛直方向の跳躍運動の学習を行ったが，通常は跳躍と併せて着地を行う必要があり，今後は着地運動の学習についてもできるようにする予定である．また，実用的な移動手段とするため，様々な方向や距離へ跳躍する方法を学習により獲得することが考えられる．また，さらなる試行回数の削減や，より多くのノイズが評価値に含まれる場合においても学習できるようにするなどの発展が考えられる．

文　献

(1) Kei Kakitani, Koji Terada and Yasuo Kuniyoshi, Learning Multi-DOF Motion by Exploration of Virtual Goal Switching Pattern (in japanese), *Proceedings of JSME Robotics and Mechatronics Conference 2007*, 2A1-L04.

(2) Ryuma Niiyama, Akihiko Nagakubo and Yasuo Kuniyoshi, Mowgli: A Bipedal Jumping and Landing Robot with an Artificial Musculoskeletal System, *Proc. of the 2007 IEEE Int. Conf. on Robotics and Automation (ICRA 2007)*, pp. 2546–2551.

(3) Jun Morimoto and Kenji Doya, Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning, *Robotics and Autonomous Systems*, Vol. 36 (2001), pp. 37–51.

(4) Hajime Kimura, Toru Yamashita, Shgenobu Kobayashi, Reinforcement Learning of Walking Behavior for a Four-Legged Robot, *Proceedings of the 40th IEEE Conference on Decision and Control*, (2005), TuM01-3.

(5) Kenji Doya, Selforganization of Motional Patterns (in japanese), *Proceedings of the 16th Symposium of the Society of Instrument and Control Engineers*, (1987), pp. 961–964.

(6) Mitsuru Higashimori, Manabu Harada, Idaku Ishii and Makoto Kaneko, Jumping Pattern Gerenation for a Serial Link Robot, *Journal of the Robotics Society of Japan*, Vol. 23, No. 8 (2005), pp. 84–92.

(7) Kazuyuki Ito, Fumitoshi Matsuno and Akio Gofuku, A Study of Reinforcement Learning for Redundant Robots -New Framework of Reinforcement Learning that Utilize Body Image- (in japanese), *Journal of the Robotics Society of Japan*, Vol.22, No.5 (2004), pp. 672–689.

(8) Kei Kakitani, Koji Terada and Yasuo Kuniyoshi, Acquisition of Multiple Motion Strategies by A Real Robot Using Virtual Goal Switching Pattern (in japanese), *Proceedings of JSME Robotics and Mechatronics Conference 2008*, 2P2-G13.