



Modeling and Control Human Arm With Fuzzy-Genetic Muscle Model Based on Reinforcement Learning: The Muscle Activation Method

Fereidoun Nowshiravan Rahatabad^{1*}

¹Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Abstract

Background: The central nervous system (CNS) is optimizing arm movements to reduce some kind of cost function. Simulating parts of the nervous system is one way of obtaining accurate information about the neurological and treatment of neuromuscular diseases. The primary purpose of this paper is to model and control the human arm in a reaching movement based on reinforcement learning (RL) theory.

Methods: First, Zajac's muscle model has improved by a fuzzy system. Second, the proposed muscle model applied to the 6 muscles, which are responsible for a two-link arm that moves in the horizontal plane. Third, the model parameters are approximated based on the genetic algorithm (GA). Experimental data recorded from healthy subjects for assessing the approach. At last, the RL algorithm has utilized to guide the arm for reaching tasks.

Results: The results show that: (1) The proposed system is temporally similar to a real arm movement. (2) The RL algorithm can generate the motor commands obtained from electromyographies (EMGs). (3) The similarity of obtained activation function from the system has compared with the real data activation function, which may prove the possibility of RL in the CNS (basal ganglia). Finally, in order to have a graphical and effective representation of the arm model, the virtual reality environment of MATLAB has been used.

Conclusion: Since the RL method is a representative of the brain's control function, it has some features, such as better settling time, not having any peak overshoot, and robustness.

Keywords: Musculoskeletal model; Reinforcement learning; Upper limb; EMG; Hill-type muscle model; Virtual reality.

*Correspondence to

Fereidoun Nowshiravan
Rahatabad, Department of
Biomedical Engineering,
Science and Research Branch,
Islamic Azad University, Tehran,
Iran,
Email: nooshiravan@gmail.com

Published online June 21, 2020



Citation: Nowshiravan Rahatabad F. Modeling and Control Human Arm With Fuzzy-Genetic Muscle Model Based on Reinforcement Learning: The Muscle Activation Method. Int Clin Neurosci J. 2020;7(3):138-146. doi:10.34172/icnj.2020.15.

Introduction

When a person performs a reaching movement, he/she chooses a unique trajectory among an infinite number of possible ones. It has suggested that the central nervous system (CNS) optimizes arm movements to reduce some kind of cost functions.¹⁻³ Cost functions usually contain movement-related variables that should optimize during the movement.⁴ Some computational control models have shown that CNS generates a set of motor commands to optimize cost functions.⁵⁻¹⁰ These models predict the arm trajectories the same as the experimental data, but they do not describe the process of learning.

The present paper discusses an optimal learning control method based on reinforcement learning (RL). RL acquired an optimal control strategy through a trial-and-error process with no explicit "teacher."¹¹ RL algorithms are considered as a model for Dopamine-based learning

in the brain,¹² where the dopaminergic projections from the substantia nigra to the basal ganglia functions as the prediction error.

The innovation of this study is to use the FGIHM model as an adaptive muscle model¹³ and control it by an RL method that is a model of brain function. Modeling of the CNS and the muscle-skeletal system is a step forward to understanding the complexity of the system. This knowledge can help clinicians soon to treat neurodegenerative diseases.

In a study,¹⁴ this classical RL algorithm applied to implement human arm movement. Major drawbacks of Classical RL in the biological system (human arm) are their requirements for many trials during learning procedure to achieve a simple point-to-point motion. In other words, the computational cost is so high, and the simulation process takes a long time. To overcome

the computational cost, a method of RL, as Q-learning¹⁵ was proposed. The Q-learning technique is one of the RL methods that utilize the action-value function. The action-value function gives the expected utility of taking a given action in a given state and following a fixed policy after that.¹⁶ Recently, a method of Q-learning, multi-agent state action reward state action (MA-SARSA),¹⁶ was introduced to increase the convergence of system states and avoid the local optima traps. This approach takes advantage of continuous reward functions to improve the learning speed in which a classical RL with discrete rewards does not.¹⁶ Another advantage of MA-SARSA is overcoming combinatorial explosion through multi-agents that are learning separately. The combinatorial explosion happens when multiple actuators are needed to control a complex agent in a dynamical environment optimally.

In other research combining an artificial bee colony algorithm called the BCR algorithm is used to control the two-link arm with 6 muscles.¹⁷ The chaos phenomena may occur in two-link arm control based on the RL algorithm by changing parameters of muscles.¹⁸ Another method to determine the muscle activity is muscle synergy.¹⁹ Some research to understanding the function of the neuromuscular system has used in anthropomorphic systems.²⁰

In this paper, a reaching movement by a two-link arm with 6 fuzzy-genetic muscles has investigated. The authors previously showed that Zajac's muscle model is unable to take account of planar movement details during the experiment because of few tuning parameters.¹³ In the proposed muscle model, 3 fuzzy parameters are added to Zajac's model to overcome the mentioned drawbacks. In order to validate the approach, an experiment is conducted on three right-hand male subjects (mean age 27), and tuning the model with a genetic algorithm (GA). Then MA-SARSA algorithm is used for finding each muscle activation function. The results show that: (1) MA-SARSA can control this model with the redundancy of responses for reaching movement with settling time, rise time, and peak overshoot the same as a human arm reaching movement. (2) RL can yield the activation level of muscle the same as the one extracted from real electromyography (EMG).

Materials and Methods

The Kinematics of Muscle-Joint Space

The two-link arm model with 6 muscles has shown in Figure 1 (This model is a simplified model of the human arm in horizontal plane movements(x)). To simplify the model: (1) the movements of the arm are bounded to the horizontal plane. (2) The effect of gravity has omitted. (3) The frictions of joints are linear. (4) The mass effect of each muscle has not considered.

The geometrical relationship between the length of

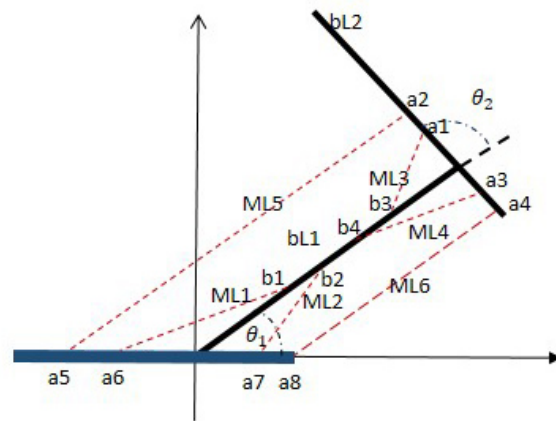


Figure 1. The Two-Link Arm With 6 Muscles. The following muscles used in this model: a single-joint flexor and extensor at the shoulder (ML1 & ML2: Pectoralis and deltoid) and the elbow (ML3 & ML4: biceps long head and Triceps Brachii medial-head) and also two double-joint muscles spanning both the shoulder and elbow joints (ML5 & ML6: biceps short head and triceps long head). The letters (ai, bi) mark the coordinates of the origin and insertion points for each muscle group.²¹

muscle groups and their connection coordinates is given as follows²¹:

$$\begin{aligned}
 ML1 &= (a6^2 + b1^2 + 2a6 \times b1 \cos \theta_1)^{\frac{1}{2}} \\
 ML2 &= (a7^2 + b2^2 - 2a7 \times b2 \cos \theta_1)^{\frac{1}{2}} \\
 ML3 &= (a1^2 + b3^2 + 2a1 \times b3 \cos \theta_2)^{\frac{1}{2}} \\
 ML4 &= (a3^2 + b4^2 - 2a3 \times b4 \cos \theta_2)^{\frac{1}{2}} \\
 ML5 &= (a5^2 + a2^2 + b1^2 + 2a5 \times b1 \cos \theta_1 + \\
 & 2a2 \times b1 \cos \theta_2 + 2a5 \times a2 \cos(\theta_1 + \theta_2))^{\frac{1}{2}} \\
 ML6 &= (a8^2 + a4^2 + b1^2 - 2a8 \times b1 \cos \theta_1 - \\
 & 2a4 \times b1 \cos \theta_2 + 2a8 \times a4 \cos(\theta_1 + \theta_2))^{\frac{1}{2}}
 \end{aligned} \tag{1}$$

where ML_i denotes the length of each muscle, (θ_1, θ_2) are the joint angles. $a1 \sim 8, b1 \sim 4$ represent each position at which the muscles connect to the bones. The time derivative of equation (1) can be expressed as follows:

$$\dot{l} = Q(\theta) \cdot \dot{\theta} \tag{2}$$

where $\dot{l} \in \mathbb{R}^{6 \times 1}$ is the muscles contractile velocity vector,

$\dot{\theta} \in \mathbb{R}^{2 \times 1}$ is the joints angular velocity vector and $Q(\theta) \in \mathbb{R}^{6 \times 2}$ represents the jacobian matrix from the joint space to the muscle space which is as follows²¹:

$$Q^T(\theta) = \begin{pmatrix} r_{11} & r_{12} & 0 & 0 & r_{15} & r_{16} \\ 0 & 0 & r_{23} & r_{24} & r_{25} & r_{26} \end{pmatrix}$$

where:

$$\begin{aligned} r_{11} &= -\frac{a6 \times b1 \sin \theta_1}{ML1} \\ r_{12} &= \frac{a7 \times b2 \sin \theta_1}{ML2} \\ sr_{23} &= -\frac{a1 \times b3 \sin \theta_2}{ML3} \\ r_{24} &= \frac{a3 \times b4 \sin \theta_2}{ML4} \\ r_{15} &= \frac{(-a5 \times bL_1 \sin \theta_1 - a5 \times a2 \sin(\theta_1 + \theta_2))}{ML5} \\ r_{25} &= \frac{(-a2 \times bL_1 \sin \theta_2 - a5 \times a2 \sin(\theta_1 + \theta_2))}{ML5} \\ r_{16} &= \frac{(a8 \times bL_1 \sin \theta_1 - a8 \times a4 \sin(\theta_1 + \theta_2))}{ML6} \\ r_{26} &= \frac{(a4 \times bL_1 \sin \theta_2 - a8 \times a4 \sin(\theta_1 + \theta_2))}{ML6} \end{aligned} \quad (3)$$

The relation between muscle forces and joint torques can be expressed as follows:

$$\hat{o} = WF \quad (4)$$

where $F \in \mathbb{R}^{6 \times 1}$ is the vector of muscle forces, $\tau \in \mathbb{R}^{2 \times 1}$ is the vector of joint torques and $W = Q(\theta)^T$ from Eq. (3).

The Modeling of the Two-Link Arm in the Horizontal Plane

The following relation for the two-link planar arm has existed if Lagrange's equation used:

$$Mq + v(q, \dot{q}) + G(q) = \tau \quad (5)$$

where M is the inertial matrix of the arm, V is the nonlinear term which includes the Coriolis and centrifugal force; τ is the input torque for arm joints, and G presents part of the moment encountering gravity

force.²² Substituting (4) in (5), yields:

$$Mq + v(q, \dot{q}) + G(q) = WF \quad (6)$$

where F is the force of the 6 muscles, and W is the Jacobian matrix from the joint space to the muscle space.

Experimental Setup and EMG Signal Processing to Obtain Muscle Activation

In order to set parameters of the muscle model for each person, EMGs, shoulder joint, and elbow joint angles have been recorded simultaneously by the BIOPAC system, Model MP150.²³

Each subject has asked to sit behind a special table. The height of this table is adjustable that the shoulder and the person's body meet the 90° angle (Figure 2).

Three healthy subjects have tested in this experiment (male subjects aged from 21 to.²⁷ All the cases are right-hand and without neural-muscular problems. Subjects have asked to perform the task without any concentration to minimize the co-activation of antagonist's muscles.

Ag-AgCl surface electrodes have used to obtain EMGs. For bipolar recording, the electrodes of 8 mm (8 mmAg-AgCl BIOPAC-EL208S) diameter were attached to the subject's skin.²³ The places of electrodes are used.^{24,25} EMGs has recorded from muscles of short biceps head (BSH), long biceps head (BLH), Triceps Brachii medial-head (TRIA), long triceps head (TRIO), pectoralis major (PMJ), and deltoid (DEL). Besides, the angles of the elbow joint and shoulder recorded by the electrogoniometer (Figure 3).

The EMG signals amplified by 5000 gain²⁰ with a sampling rate of 1 kHz. The recording duration was 15 seconds. There was 20 seconds rest between recordings.

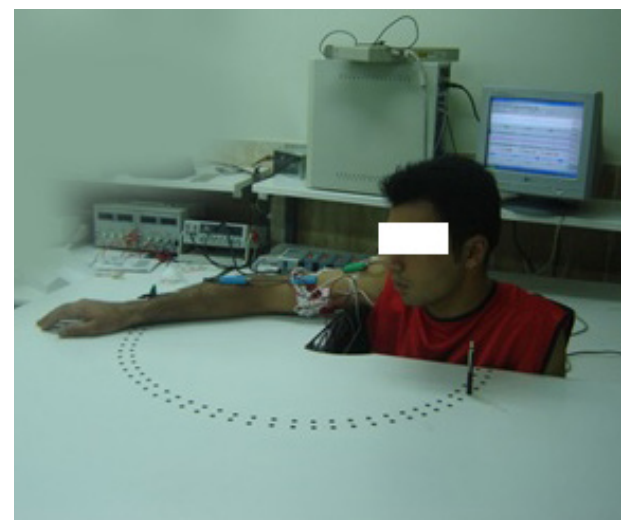


Figure 2. The Experimental Set up for the Simultaneous Recording of Electromyogram Signals, Elbow Joint, and Shoulder Joint Angles.

Subjects have asked to move their hands between 2 marked points. The distance between the start point and endpoints is 50 cm.

In order to obtain normalized muscle activation level (NAL) from raw EMG, the following steps have done²⁷: (1) A high-pass filter. (2) Full signal rectification (absolute value). (3) A low pass. (4) Signal normalization to the maximum EMG of each channel. Figure 4 shows the block diagram of the NAL method. The raw EMG signal with arbitrary amplitude has mapped into a normalized signal in the range of zero to one.

The above procedure has applied to a biceps signal, which has shown in Figure 5.

Enhancement of Zajac’s Muscle Model With Fuzzy System

A Zajac-type muscle model is used to generate muscle force.²⁸ According to this model, the muscle force (F_{muscle}) is the sum of active one (F_{α}) and passive one (F_p) as follows²⁶:

$$F_p = \left[\left(\frac{F_{max}}{\exp(K_{sh}) - 1} \right) \left(\exp\left(\frac{K_{sh}(l - l_o)}{0.5l_o} \right) - 1 \right) \right]^+$$

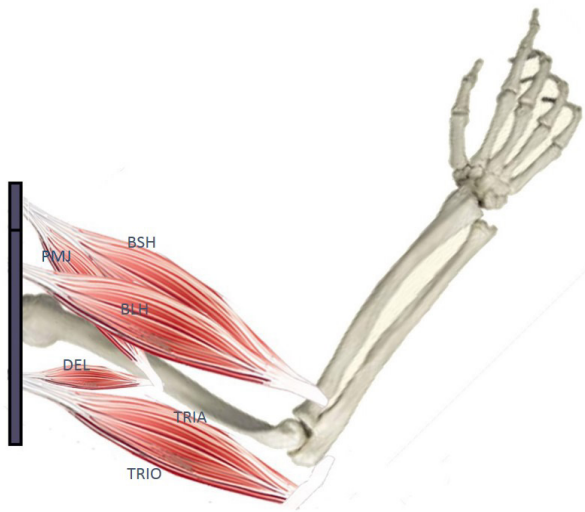


Figure 3. The Two-Link Arm With 6 Muscles. In this figure the following abbreviations are used; biceps short head (BSH), biceps long head (BLH), Triceps Brachii medial-head (TRIA), long triceps head (TRIO), pectoralis major (PMJ), and deltoid (DEL).^{25,26}

$$F_{\alpha} = \alpha F_{max} \left[1 - 4 \left(\frac{l - l_o}{l_o} \right)^2 \right]^+ \tag{7}$$

$$[u]^+ = \max[u, 0]$$

Where α is the muscle activation level of the muscle in the range of 0–1 (0 for passive muscle, 1 for fully activated muscle), F_{max} is the maximum force of muscle, l_o is the optimal fascicle length. The exponential shape determined by the variable K_{sh} , which has set $K_{sh}=3$. Finally, the output F_{muscle} depends on the velocity, F_{α} and F_p as following²⁶:

$$F_{muscle} = [a_1 + a_2 \tan(a_3 + a_4 \dot{l})](F_{\alpha} + F_p) \tag{8}$$

where $\alpha_1=0.80, \alpha_2=0.50, \alpha_3=0.43, \alpha_4=58$ s/m. The above relation investigated by Zajac.²⁸ In order to enhance the Zajac model, 3 fuzzy scaling coefficients added to F_p, F_{α} and F_{muscle} . The F_p and F_{α} coefficients are dependent on length, and the F_{muscle} coefficient is dependent on the velocity of muscle length. By applying these coefficients, the tuning parameter of the model is increased (by setting fuzzy scaling coefficients to one, the relation reduced to Zajac’s muscle model).

In order to calculate the coefficients, a fuzzy system designed with Mamdani inference system (Figure 6).²⁸ The fuzzification and defuzzification are performed by Gaussian membership functions as follows:

$$f(x; \sigma, c) = e^{-\frac{(x-c)^2}{2\sigma^2}} \tag{9}$$

This function has 2 parameters (c and σ), which are determined by the GA.

Tuning the Fuzzy Muscle Parameters by Genetic Algorithm

A GA method has used to modify scaling fuzzy coefficients of the model (section 2.4). These coefficients change based on input and output membership functions of the model. Membership functions depend on the length and muscular contraction velocity. GA adjusts muscle coefficients, which are muscle parameters for mapping EMG to elbow and shoulder joint angles.

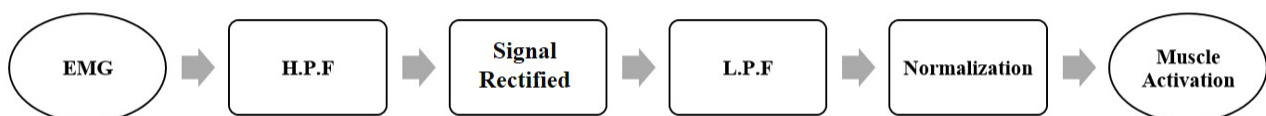


Figure 4. The muscle activation level is estimating from raw EMG.²⁷

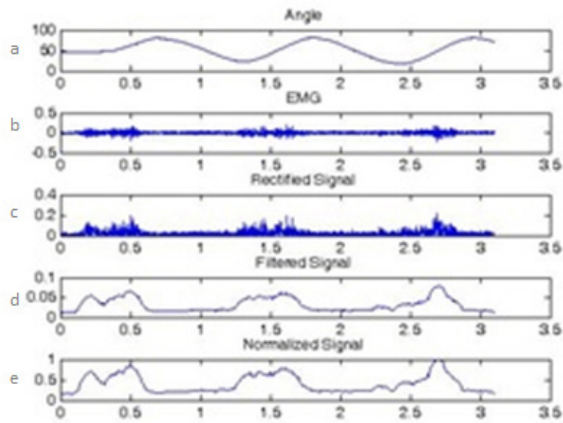


Figure 5 . (a) elbow joint angle, (b) biceps EMG, (c) rectified EMG signal, (d) applying the low pass filters with cut off frequency 15 Hz, and (e) the normalized form of the fourth diagram.

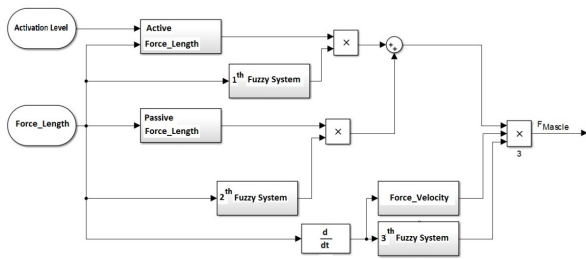


Figure 6. The Proposed Fuzzy Muscle Model. This model has based on the Zajac muscle model with three fuzzy coefficients.²⁹

Control of the Modeled Arm Based on Reinforcement Learning Method

In order to control the model, MA-SARSA is utilized.³¹ The block diagram of this algorithm has shown in Figure 7. In this method, agent updating is performed based on receiving rewards and punishments from system conditions and applying them to all of the agents and adjusting each one separately.

or using this algorithm, the reward functions have defined as follows:

$$Reward = \frac{\hat{a}}{1 + d^n} \tag{10}$$

where β is a gain that controls the magnitude of the reward, d is the distance from the goal position, and n is the shape factor parameter. MA-SARSA is using a continuous penalty function for every action that does not support the goal as follows:

$$Punishment = -\beta d^n - v / (1 + \beta d^n) - f \tag{11}$$

where ϵ is the bias term, and v is the velocity of the tipping arm.³²

Simulation Procedure

The two-link arm model with 6 fuzzy-genetic muscles (TLA6FGM) simulated in Simulink and SimMechanics toolbox of MATLAB software: MATLAB/Simulink/SimMechanic 2008 (ver 7.6.0.324). For solving differential equations, Runge-Kutta, with the constant step of 0.01 second, is used. The GA preceded 800 iterations, and in the 100th and 200th iteration, random noise added to its output, and the fitness function has been set based on mean square error. Physical parameters of arm and muscle connection points derived from.²¹ Three subjects used, and they have asked to do extension and flexion in the horizontal plane.

Before implementing the RL method, first, the TLA6FGM was tuned for each subject. Finally, the tuned TLA6FGM controlled with the RL algorithm. The RL algorithm, based on an MA-SARSA, has performed for 1500 episodes, which has 1500 steps.

Results

In Figure 8, the proposed model of the arm with 6 fuzzy muscles controlled by two controllers; state feedback (PID) and RL. Figure 8a shows the RL controller superiority over the state feedback controller in Figure 8b.

The angles and angular velocity of the model controlled by RL and state feedback have shown in Figure 9. The dash (--) lines came from the experiment, and the red line is RL controller responses, and the blue lines are for the state feedback.

Figure 10 compares the reaching distance of the tip in the 2 controllers, and the error estimated based on the Euclidean distance from the target.

In Figures 11 and 12, the activation level of the RL controller and state feedback compared for every 6 muscles.

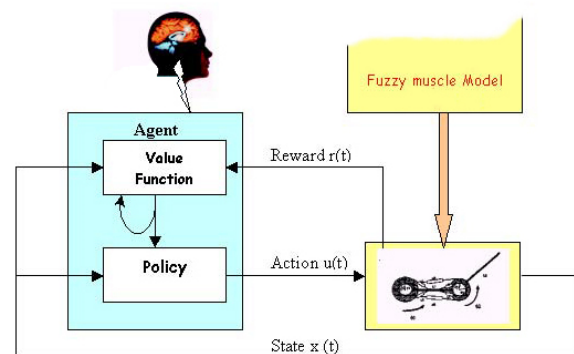


Figure 7. The Basic Model of the MA-SARSA.³¹

In Figure 13 the trace of the two-link arm with 6 models controlled by RL has shown. Figures 13a and 13b show arm motions with 500, and 1500 trails, and Figure 13c shows the tangential velocity profile of the hand from time 0-0.52 seconds.

Figure 14 shows the robustness of MA-SARSA and PID-GA under the disturbance field, which applied to the tip of the arm.

In order to have better insight into the arm movement, the virtual reality of MATLAB has used to demonstrate the performance of the model. Figure 15 shows the postures of the arm in different viewpoints.

Conclusion

In the present paper, a fuzzy-genetic muscle model based on Zajac’s model was proposed (FGMZM). The essential advantages of this model are: (1) FGMZM is more robust than Zajac to the input noise. (2) It can be customized depending on the number of adjustable parameters. FGMZM has been used to model the two-link arm with 6 muscles. Finally, the arm model has controlled with MA-SARSA algorithm, and the results have compared with the PID-GA state feedback control algorithm.

In a comparison between the MA-SARSA method and the PID-GA state feedback controller (Figure 8), the

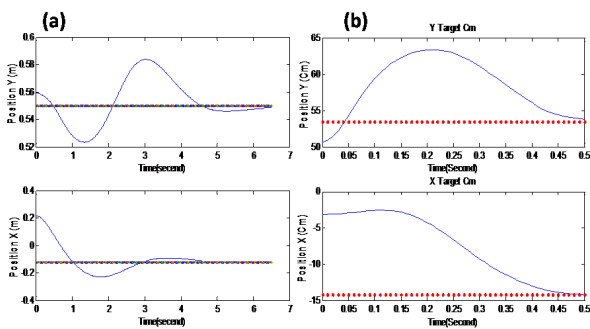


Figure 8 . End-Point Reaching Movement From (0.2192, 0.5592) to (-0.14, 0.55). The left two graphs belong to the X and Y position in MA-SARSA controller, and the right Column two graphs belong to the state feedback controller.

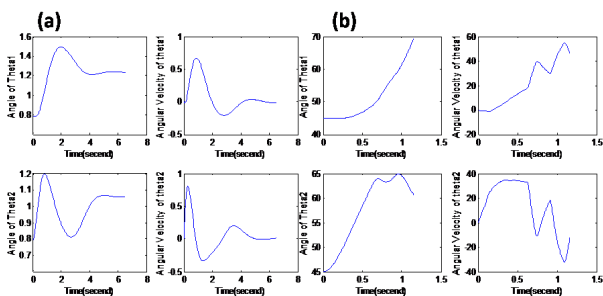


Figure 9. The Angles and Angular Velocities of the 2 Joints. a) The PID state feedback, b) The reinforcement learning controller.

MA-SARSA has better settling time (less than a second like human reaching arm movement) than the PID-GA controller (more than 5 seconds).

Moreover, MA-SARSA does not have any peak overshoot (Figure 9). This comes from a speed regulation process in MA-SARSA, which is similar to a bell-shaped curve (Figure 13c). The speed pattern at the beginning and the end of the process is low, but in the middle, the motion speed is high. In PID-GA state feedback, peak overshoot has always been seen because this controller utilized the error correction method for controlling procedure.

Disturbance effects to the tip of the arm and its drift from the no-load trajectory has shown in Figure 13. MA-SARSA has better robustness to the noise than the PID-GA state feedback controller because MA-SARSA changed the controller structure. Structural changes come from the learning process of trial-and-error, which allows the controller to study all the possible conditions.

The effect of the learning process on the reaching movement has illustrated in Figures 13a and 13b with sets of weight parameters after the 500th and 1500th trials. At the early phase of the learning (500th trial), the tip reached the target, but it passed away because the speed of tip at the reaching point is not acceptable. As the number of trials increases up to 1500, MA-SARSA can reach the goal with suitable cost function (Figure 13b). The velocity profile is almost bell-shaped, and if the number of trials increases, the velocity profile became smoother bell-shape typically (Figure 13c), and the hand trajectory tends to follow an approximately straight line.

The MA-SARSA controller drives out different parameter sets in which all of them achieved the reaching movement goals. Among them, there is a specific set that matches mostly to the biological controller. Hence, MA-SARSA has the capability of modeling a part of the brain named basal ganglia,³² which plays a vital role in the skilled movement.

Tahara et al suggested a task-space feedback for controlling two-link arm driven by 6 muscles.²¹ In comparison, when the arm reaches the target based on the task-space feedback control, it has to swing around the target, and sometimes overshoot has happened. This

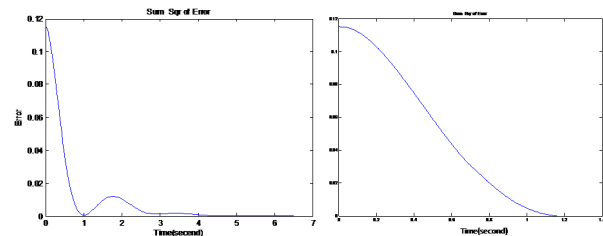


Figure 10. Endpoint Reaching Task Error . The right graph is related to the reinforcement learning controller, and the left one is related to the state feedback controller.

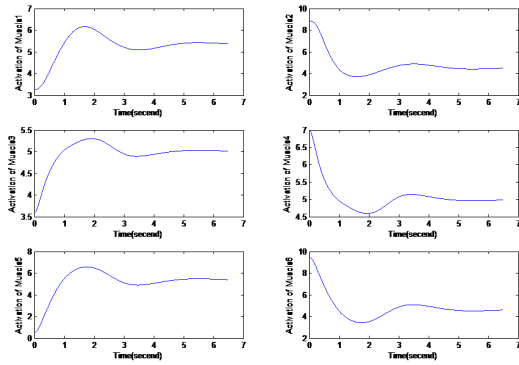


Figure 11. The Activation Level of the RL Controller for Every 6 Muscles.

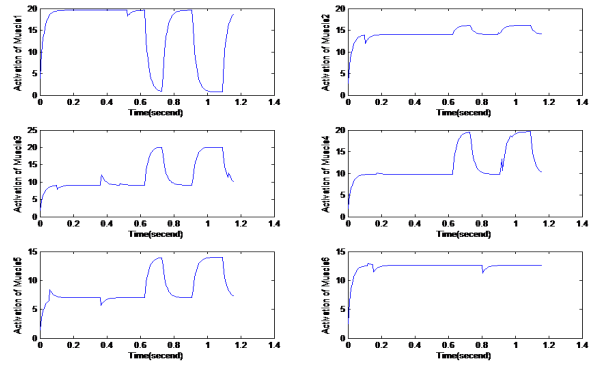


Figure 12. The Activation Level of (PID-GA) State Feedback for Every 6 Muscles

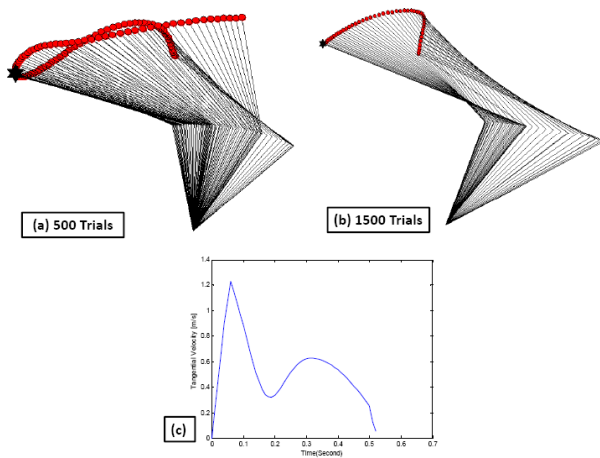


Figure 13. Reaching Motions During the Learning Process. Upper and lower plots in each of (a), (b) illustrate arm motions and tangential velocity profile of the hand from time 0-0.52 seconds. (c) Shows tangent velocity of hand movement after 1500 trials.

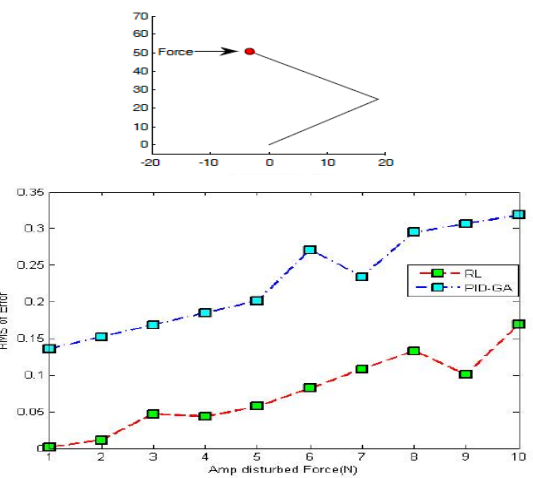


Figure 14. Robustness Analysis. Two-link arm model with 6 muscles at the 1500 trail in the learning under disturbed conditions to the tip is shown. The horizontal axis shows the amplitude of the disturbance, and the vertical axis shows the MSE error of the trajectory from the normal condition trajectory.

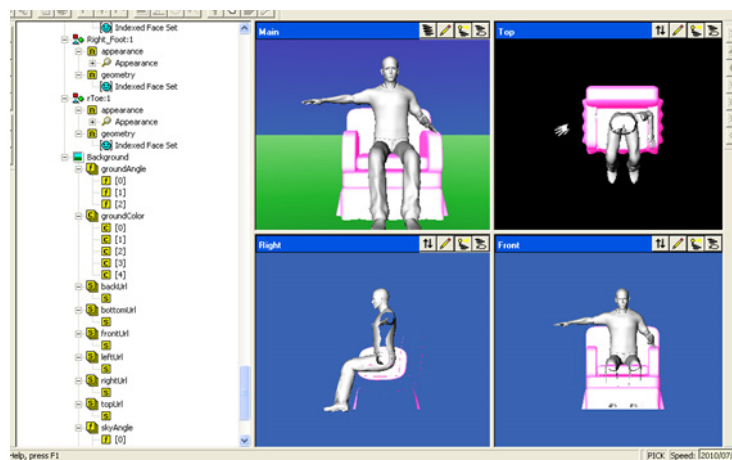


Figure 15. Graphical Representation of arm in Virtual Reality of MATLAB Software.

phenomenon has not seen in the human arm movements. In this sense, the RL control method is more reliable with the reality of human arm behavior than Task-space feedback control.

For further investment, it has suggested using the biological controller in a hierarchical structure to study the controlling procedure of the human brain more precisely. Besides, an experimental setup should design to implement and validate such a structure.

The virtual reality of MATLAB has a few block sets which are not proper for human movement demography. Therefore, the authors define some new block sets for animating human movements. In the future, this toolbox will be completed and presented for other research colleagues to invest the human body movements.

Conflict of Interest

The authors declare that they have no conflict of interests.

Ethical Statement

The Ethics Committee approved the study of Islamic Azad University- Science and Research Branch with ethical code IR.IAU.SRB.REC.1395.23.

References

- Flash T, Hogan N. The coordination of arm movements: an experimentally confirmed mathematical model. *J Neurosci.* 1985;5(7):1688-703. doi: 10.1523/JNEUROSCI.05-07-01688.1985.
- Harris CM, Wolpert DM. Signal-dependent noise determines motor planning. *Nature.* 1998;394(6695):780-4. doi: 10.1038/29528.
- Uno Y, Kawato M, Suzuki R. Formation and control of optimal trajectory in human multijoint arm movement. *Biol Cybern.* 1989;61(2):89-101. doi: 10.1007/BF00204593.
- Kambara H, Kim K, Shin D, Sato M, Koike Y. Learning and generation of goal-directed arm reaching from scratch. *Neural Netw.* 2009;22(4):348-61. doi: 10.1016/j.neunet.2008.11.004.
- Flash T. The control of hand equilibrium trajectories in multi-joint arm movements. *Biol Cybern.* 1987;57(4-5):257-74. doi: 10.1007/bf00338819.
- Gribble PL, Ostry DJ, Sanguineti V, Laboisière R. Are complex control signals required for human arm movement? *J Neurophysiol.* 1998;79(3):1409-24. doi: 10.1152/jn.1998.79.3.1409.
- Hogan N. An organizing principle for a class of voluntary movements. *J Neurosci.* 1984;4(11):2745-54. doi: 10.1523/JNEUROSCI.04-11-02745.1984.
- Miyamoto H, Nakano E, Wolpert DM, Kawato M. TOPS (Task Optimization in the Presence of Signal-Dependent Noise) model. *Syst Comp Jpn.* 2004;35(11):48-58. doi: 10.1002/scj.10377.
- Todorov E, Jordan MI. Optimal feedback control as a theory of motor coordination. *Nat Neurosci.* 2002;5(11):1226-35. doi: 10.1038/nn963.
- Wada Y, Kawato M. A neural network model for arm trajectory formation using forward and inverse dynamics models. *Neural Netw.* 1993;6(7):919-32. doi: 10.1016/S0893-6080(09)80003-8.
- Sutton RS, Barto AG. Reinforcement Learning: An Introduction. MIT Press; 1998.
- Stocco A, Lebiere C, Anderson JR. Dopamine, Learning, and Production Rules: The Basal Ganglia and the Flexible Control of Information Transfer in the Brain. Association for the Advancement of Artificial Intelligence; 2009.
- Nowshiravan Rahatabad F, Fallah A, Jafari AH. Human arm modeling with antagonistic muscle, 16th Iranian Conference on Biomedical Engineering; 2009; Tehran, Tehran University of Medical Science. [Persian].
- Izawa J, Kondo T, Ito K. Biological arm motion through reinforcement learning. *Biol Cybern.* 2004;91(1):10-22. doi: 10.1007/s00422-004-0485-3.
- Watkins CJCH. Learning from Delayed Rewards [thesis]. England: University of Cambridge; 1989.
- Strehl AL, Li L, Wiewiora E, Langford J, Littman ML. PAC Model-Free Reinforcement Learning. Proceedings of the 23rd International Conference on Machine Learning; 2006.
- Nowshiravan Rahatabad F, Rangraz P. Combination of reinforcement learning and bee algorithm for controlling two-link arm with six muscle: simplified human arm model in the horizontal plane. *Phys Eng Sci Med.* 2020;43(1):135-42. doi: 10.1007/s13246-019-00828-4.
- Nowshiravan Rahatabad F, Fallah A, Jafari AH. A study of chaotic phenomena in human-like reaching movements. *Int J Bifurcat Chaos.* 2011;21(11):3293-303. doi: 10.1142/s0218127411030532.
- Nowshiravan Rahatabad F, Maghooli K, Rahimi Balan T. Relationship between muscle synergies and skill of the Basketball players. *Iran J Med Phys.* 2020. doi: 10.22038/ijmp.2020.42349.1631.
- Wang T, Chen X, Qin W. A novel adaptive control for reaching movements of an anthropomorphic arm driven by pneumatic artificial muscles. *Appl Soft Comput.* 2019;83:105623. doi: 10.1016/j.asoc.2019.105623.
- Tahara K, Luo ZW, Arimoto S, Kino H. Task-Space Feedback Control for A Two-Link Arm Driven by Six Muscles with Variable Damping and Elastic Properties. *IEEE International Conference on Robotics and Automation*; 2005. p. 223-7. doi: 10.1109/ROBOT.2005.1570123.
- Taghirad HD. An Introduction to Modern Control, K. N. Toosi University Publication 2003. (In Persian with English abstract).
- Biopac Systems website. Available from: <https://www.biopac.com/>.
- Konrad P. The ABC of EMG: A Practical Introduction to Kinesiological Electromyography. USA: Noraxon Inc; 2005.
- Weiss L, Silver JK, Weiss J. Easy EMG. New York: Elsevier; 2004.
- Kim HK, Carmena JM, Biggs SJ, Hanson TL, Nicolelis MA, Srinivasan MA. The muscle activation method: an approach to impedance control of brain-machine interfaces through a musculoskeletal model of the arm. *IEEE Trans Biomed Eng.* 2007;54(8):1520-9. doi: 10.1109/tbme.2007.900818.
- Rosen J, Brand M, Fuchs MB, Arcan M. A myosignal-based powered exoskeleton system. *IEEE Trans Syst Man Cybern A Syst Hum.* 2001;31(3):210-22. doi: 10.1109/3468.925661.
- Zajac FE. Muscle and tendon: properties, models, scaling, and application to biomechanics and motor control. *Crit Rev Biomed Eng.* 1989;17(4):359-411.
- Jang JSR, Sun CT, Mizutani E. Neuro-Fuzzy Soft Computing. Englewood Cliffs, NJ: Prentice-Hall; 1997.
- Doya K, Kimura H, Kawato M. Neural mechanisms of

- learning and control. IEEE Control Syst. 2001;21(4):42-54. doi: 10.1109/37.939943.
31. Martin JA, de Lope J. A Distributed Reinforcement Learning Control Architecture for Multi-Link Robots Experimental Validation. 4th International Conference on Informatics in Control, Automation, and Robotics, ICINCO; 2007; Angers, France.
32. Nowshiravan Rahatabad F, Jafari AH, Fallah A, Razjouyan J. A fuzzy-genetic model for estimating forces from electromyographical activity of antagonistic muscles due to planar lower arm movements: the effect of nonlinear muscle properties. Biosystems. 2012;107(1):56-63. doi: 10.1016/j.biosystems.2011.09.004.