Generating Slow 3D Reaching Movements by Minimizing Energy Consumption

Jhon P.F Charaja[1], Isabell Wochner[1], and Daniel F.B. Haufle[1]
[1]Hertie Institute for Clinical Brain Research, and Centre for Integrative Neuroscience, University of Tübingen, Germany
[1] Email: jhon.charaja-casas@uni-tuebingen.de

Summary

Generating 3D-reaching movements while accounting for muscle dynamics nonlinearities and model redundancy has been a significant challenge. While reinforcement learning combined with the minimum variability principle has tackled this issue for fast movements, it is not well-suited for rehabilitation applications involving impaired patients, who typically exhibit moderate-to-slow movements. To address this limitation, we introduce an energy consumption penalty into the reward function and employ a synergetic action representation to reduce the control dimensionality. Our findings indicate that incentivizing energy-efficient movements reduces peak velocity to three-fourths of the maximum velocity, and that a purely synergetic controller can generate precise and stable movements. We expect that these insights will contribute to more accurate models for predicting movement dynamics in impaired individuals.

Introduction

The biomechanics of the human arm allows for an infinite number of movement solutions to reach a given target position. Despite this redundancy, human movements exhibit stereotypical characteristics, such as bell-shaped velocity profiles. Most simulation studies investigating such goal-oriented movements typically resort to simplified conditions by approximating muscle dynamics to torque-driven models or restricting movements to a planar workspace [1, 2]. Ikkala et al. have employed a 3D musculoskeletal arm model in conjunction with the minimum variance principle [3]. The resulting control policy generates 3D reaching movements that replicate the stereotypical bell-shaped velocity profile [3]. However, despite the symmetric velocity profiles, the observed maximum velocities are significantly higher than those typically reached in daily life. This is primarily because the control policy was designed to minimize movement time given a position tolerance. To address this, our approach incorporates an energy consumption penalty to regulate maximum velocity. Additionally, we consider a purely synergetic action representation to simplify the control task while preserving the dynamics of the musculoskeletal system.

Methods

We use MuJoCo physics engine to simulate the muscle dynamics and the musculoskeletal human arm model developed by MyoSuite [4]. Wrist and fingers were fixed, resulting in a model with 24 muscle actuators and 16 joints. To identify muscle synergies, we generated movements across the entire kinematic workspace by a self-organization control method (DEP [5]), then computed 10 muscle synergies, representing 90% of variability in the recorded muscle activation data. We learned a control policy with reinforcement learning (SAC) and multiplied the output by the synergy matrix [6]. We add signal-dependent noise to the resulting control signals which are send to the muscles with a control frequency of 10 ms. The reaching task is considered successful if the hand remains within the position tolerance for 200 ms.

The reward function is $r = r_{\text{reaching}} + \alpha_{\text{penalty}} r_{\text{optimal}}$. The first term rewards fast reaching of the goal, the second term penalizes squared muscle activation to encourage optimal movement. α_{penalty} increases the priority of energy-efficient movements and is independent factor we investigate in this study.

Results and Discussion

Slower movements can be achieved by our approach. Progressively increasing α_{penalty} leads to longer movement times and lower peak velocities (See Figure 1). However, the effect of the penalty coefficient on peak velocity reduction saturates around $\alpha_{\text{penalty}} > 600$ ($\approx 0.25 V_{\max}$). Further increasing the penalty coefficient causes training instabilities. Also interesting to note is that the purely synergetic controller successfully generated precise and stable reaching movements in fast and slow movements, thereby benefiting from substantially reduced training time (3h vs 30h) compared to directly controlling all 24 muscles, confirming previous observations [6].

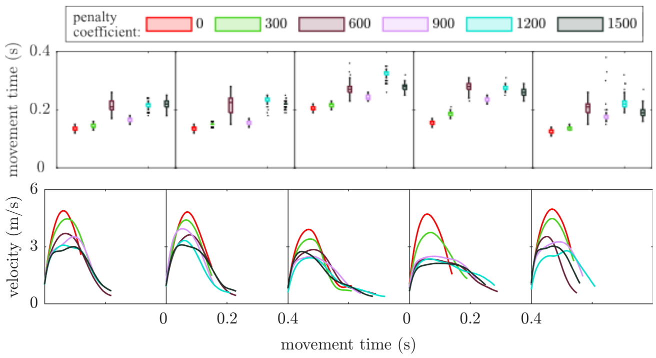

Figure 1: Effect of penalty coefficient (α_{penalty}) on velocity profiles and movement time. Five targets were used with 100 rollouts considered for each penalty coefficient.

Conclusions

Our reward function implements a trade-off between fast movements, incentivized by penalizing movement time, and slow movement, incentivized by energy efficiency. However, this strategy encountered a limitation in further reducing maximum velocity. This may be because, after increasing the penalty coefficient beyond a certain point, the two penalties become equivalent, preventing the agent from increasing movement time without expending additional energy to maintain posture

References

[1] Fisher, F. (2021). *Scientific Report*, **11**: 14445.
[2] Ueyama, Y. (2021). *Scientific Report*, **11**: 18564.
[3] Ikkala, A. et al (2022). *ACM*, **35**: No.90, pp. 1-14.
[4] Caggiano V. et al. (2022). *PMLR*, **168**: 492-507.
[5] Der R. et al. (2015). *PNAS*, **112(45)**: E6224-E6232.
[6] Berg C. et al. (2023). *Auton Robot*, **48**: 28.