The Emergence of Movement Units Through Learning with Noisy Efferent Signals and Delayed Sensory Feedback

Michael Kositsky[a*], Andrew G. Barto[a]

[a]Department of Computer Science, University of Massachusetts,
Amherst, MA 01003–4610

Rapid human arm movements often have velocity profiles consisting of several bell-shaped acceleration-deceleration phases, sometimes overlapping in time and sometimes appearing separately. We show how such sub-movement sequences can emerge naturally as an optimal control policy is approximated by a reinforcement learning system in the face of uncertainty and feedback delay. The system learns to generate sequences of pulse-step commands, producing fast initial sub-movements followed by several slow corrective sub-movements that often begin before the initial sub-movement has completed. These results suggest how the nervous system might efficiently control a stochastic motor plant under uncertainty and feedback delay.

1. Introduction

It has been consistently observed that rapid human arm movements in both infants and adults often consist of several sub-movements, sometimes called “movement units” [22]. The tangential velocity profiles of such movements show sequences of several bell-shaped acceleration-deceleration phases, sometimes overlapping in the time domain and sometimes completely separate [10–12,15,18]. These data provide clues about how the nervous system efficiently produces fast and accurate movements in the presence of noise and significant feedback delay. Most of modeling efforts concerned with movement units have addressed only the kinematic aspects of movement [4,6].

We present a model and simulation results that suggest how movement units might emerge as the result of a learning process that successively approximates an optimal control policy in the face of uncertainty and feedback delay. We implemented a reinforcement learning model that learns to produce accurate rapid movements in the presence of feedback delay for a stochastic dynamic system with nonlinear damping. The dynamics of the system simulate physiological properties of muscles and spinal reflexes.

2. Fractional-power damping dynamics

To demonstrate the proposed learning scheme we used the fractional-power damping model of arm dynamics [23]. The simplest model that captures the most critical dynamical features is

*This work was supported by NIH Grant MH 48185–09.

a spring-mass system with the nonlinear damping term:

$$m\ddot{x} + b\dot{x}^{\frac{1}{5}} + k(x - u) = 0. \tag{1}$$

Here x is the position of the mass attached to the spring, $\dot{x}$ and $\ddot{x}$ are respectively the velocity and the acceleration of the object, m is the mass of the object (the mass of the spring is assumed to be equal to zero), b is the damping coefficient, k is the stiffness coefficient, u is the control signal which determines the resting, or equilibrium, position.

This model arises from simplifying assumptions about muscle mechanical properties and spinal reflex mechanisms. The fraction-power damping in the model represents the result of interaction between muscle properties and the friction-like property of the stretch reflex. In our simulations we used a spring-mass system with fractional-power nonlinear damping, which is the the simplest model that captures the above mentioned critical features [2]. Another essential characteristics of the neural signal transmission is accounted for by using a cascade of low-pass filters on motor commands [16].

3. The learning algorithm

The learning mechanism of the model is a reinforcement learning algorithm [21]. In contrast to supervised learning, reinforcement learning systems learn from very simple training feedback. In our simulations, the system is rewarded only upon successful completion of a movement, where use of a discount factor effectively caused the reward to increase as the time to completion decreased. The learning architecture of the model was motivated by the correspondence between reinforcement learning methods that use temporal difference algorithms and the activity of dopamine cells in the basal ganglia [1,9]. In particular, the model makes use of the actor-critic algorithm as described in [21].

To simulate delayed feedback, the learning model observes the state of the motor plant with a 200 ms delay. The learning system also makes control decisions at 200 ms intervals, maintaining a newly selected activation level until the next control decision. To simulate variability inherent in the motor control process, each activation level is modified by additive Gaussian noise. This intermittent action selection scheme is supported by biological evidence (e.g., [13]) as well as by computational considerations. It allows the use a compact internal representation and preserves the Markovian property of the underlying process.

We have implemented the actor-critic algorithm for a continuous state space and a finite set of actions, i.e., activation level magnitudes u evenly spaced every 1 cm between 0 cm and 10 cm. To represent functions defined over the continuous state space we have used a CMAC representation with 10 tilings, each tiling spans all three dimensions of the state space and has 10 tiles per dimension. The tilings have random offsets drawn from the uniform distribution. Learning is done in episodes. At the beginning of each episode the plant is at a fixed initial state, and the episode is complete when the plant reaches the target region of the state space. Table 1 shows the parameter values used in the simulations.

4. Results

The proposed reinforcement learning architecture successfully learned to move the mass quickly and accurately to the target in approximately 1,000 episodes. Figure 1 shows the corresponding learning curve.

Table 1
Parameter values used in the simulations.

description	value	description	value
mass, m	1 kg	initial velocity	0 cm/s
damping coefficient, b	$3\ \mathrm{N}\,(\mathrm{s}/\mathrm{m})^{\frac{1}{5}}$	target position	5 cm
stiffness, k	$30\ \mathrm{N}/\mathrm{m}$	target velocity	5 cm
the basic simulation time step	1 ms	target position radius	0.5 cm
the feedback delay, Δ	200 ms	threshold velocity radius	0.1 cm/s
initial position	0 cm	standard deviation of the noise	1 cm

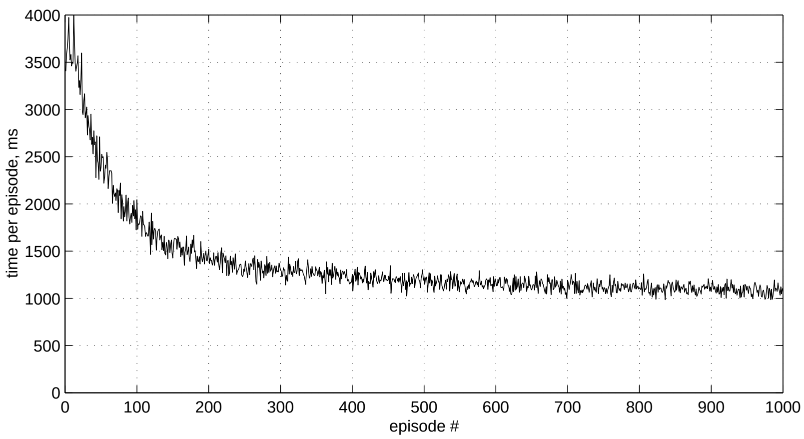

Figure 1. The learning curve averaged over 100 trials. The performance is measured in time-per-episode.

Figure 2 shows a typical movement accomplished by the controller after learning. The movement shown in Figure 2 has two acceleration-deceleration phases called movement units or sub-movements.

Corrective sub-movements may occur before the plant reaches zero velocity. The controller generates this corrective sub-movement “on the fly,” i.e., before the initial fast sub-movement is completed. Figure 3 shows a sample movement accomplished by the controller after learning where such overlapping sub-movements occur. This can be seen clearly in panel (b) of Figure 3 where the velocity profile of the movement is shown. Each of the sub-movements appears as a bell-shaped unit in the tangential velocity plot.

5. Discussion

The model learns to produce control sequences consisting of pairs of high activation steps followed by low activation steps. This feature stands in good agreement with pulse-step models of motor control [7,8,20]. Each of the pulse-step combinations produces a sub-movement characterized by a bell-shaped unit in the velocity profile.

The first sub-movement is always fast and covers most of the distance from the initial position to the target. All of the subsequent sub-movements are much slower and cover much shorter segments in the position space. This is in accord with the so-called dual control model [12,14,

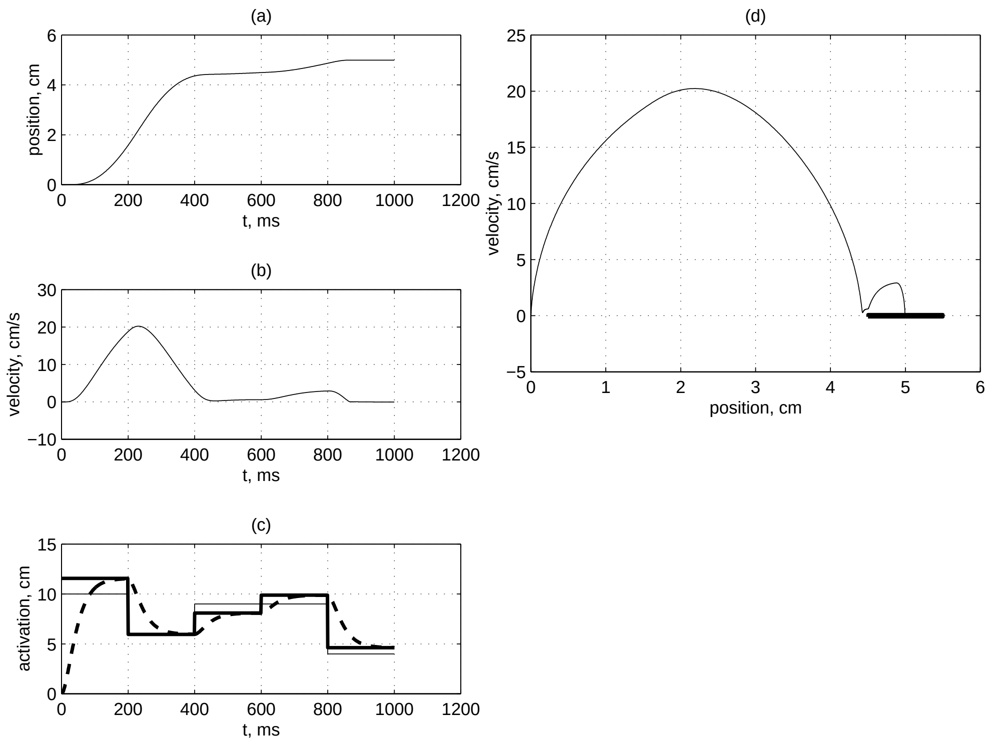

Figure 2. A sample movement accomplished by the controller after learning. Panels (a) and (b) respectively show the position and velocity time courses. Panel (c) shows the motor command time course, where the thin solid line shows the activation level selected by the learning model, the thick solid line shows the perturbed activation which is sent as the control signal to the motor plant, and the dashed line shows the activation after the temporal filtering is applied. Panel (d) shows the phase-plane trajectory of the movement. The thick bar at the lower-right corner shows the target region.

17], where the initial part of a movement is conducted in a ballistic manner and the final part is conducted under closed-loop control.

Additional evidence for this kind of dual control strategy comes from experiments in which subjects were given visual feedback only during the initial stage of movement, which did not provide much improvement compared to trials in which subjects were deprived of visual feedback during the entire movement [3,5]. In another set of experiments, proprioceptive feedback was altered by stimulations of muscle tendons. Movement accuracy decreased only when the stimulation was applied at the final stages of movement [19].

Our model does not follow the dual control strategy, but the movement patterns which emerge naturally from the existing constraints and conditions are similar to those produced while following the dual control strategy. The reinforcement learning component is encouraged by the reward structure (and the use of discounting) to accomplish each movement as quickly as possible. On the other hand, it faces high uncertainty in the system behavior. The uncertainty in the system behavior is high for states with high velocities and is low for states with low velocities. Thus, in a low velocity state the information available to the model determines the actual state of the plant very accurately. If the model were to adopt a policy in which it attempts to directly

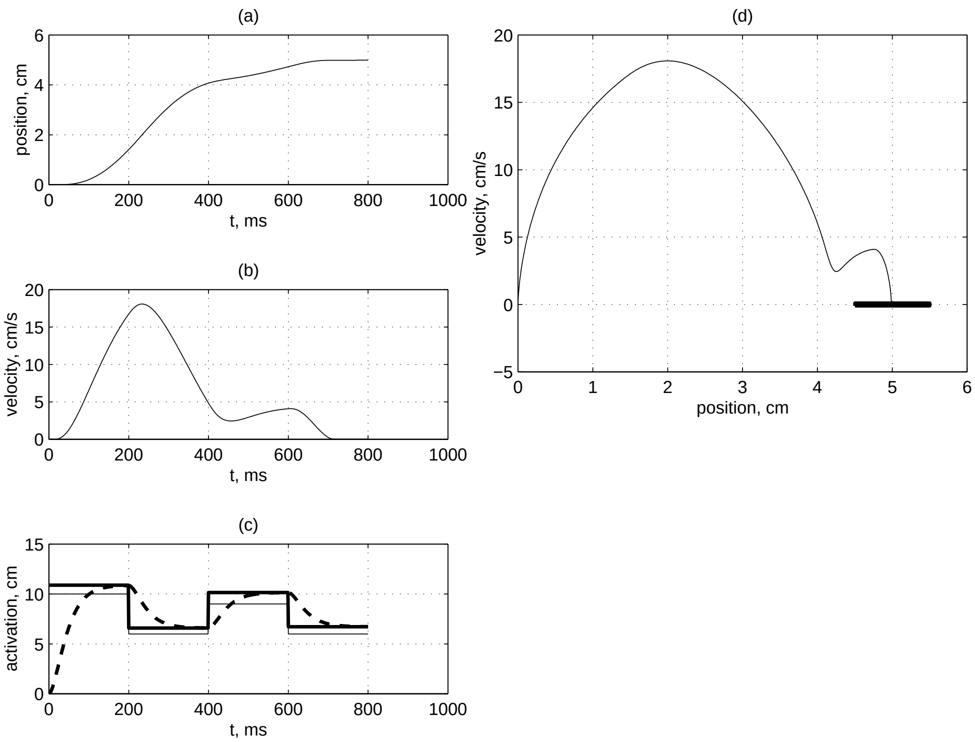

Figure 3. A sample movement accomplished by the controller after learning with a well expressed predictive correction.

hit the target in one fast sub-movement, then very often it would miss the target and spend long additional time to accomplish the task. The optimal policy in this situation is to drive the plant close to the target by one fast sub-movement and then apply a few slow sub-movements to accurately drive plant into the target region. This can be accomplished because the trajectories of the latter sub-movements do not leave the low velocity area of phase space.

REFERENCES

1. A. G. Barto. Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, and D. G. Beiser, editors, *Models of Information Processing in the Basal Ganglia*, pages 215–232. MIT Press, Cambridge, MA, 1995.
2. A. G. Barto, A. H. Fagg, N. Sitkoff, and J. C. Houk. A cerebellar model of timing and prediction in the control of reaching. *Neural Computation*, 11:565–594, 1999.
3. D. Beaubaton and L. Hay. Contribution of visual information to feedforward and feedback processes in rapid pointing movements. *Human Movement Science*, 5:19–34, 1986.
4. N. E. Berthier. Learning to reach: a mathematical model. *Developmental Psychology*, 32:811–832, 1996.
5. L. G. Carlton. Processing of visual feedback information for movement control. *Journal of Experimental Psychology: Human Perception and Performance*, 7:1019–1030, 1981.
6. E. R. F. W. Crossman and P. J. Goodeve. Feedback control of hand-movements and Fitts' law. *Quarterly Journal of Experimental Psychology*, 35A:251–278, 1983. Paper presented at the meeting of the Experimental Psychology Society, Oxford, England, July 1963.

7. C. Ghez. Contributions of central programs to rapid limb movement in the cat. In H. Asanuma and V. J. Wilson, editors, *Integration in the Nervous System*, pages 305–320. Igaku-Shoin, Tokyo, 1979.
8. G. L. Gottlieb. A computational model of the simplest motor program. *Journal of Motor Behavior*, 25:153–161, 1993.
9. J. C. Houk, J. L. Adams, and A. G. Barto. A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, and D. G. Beiser, editors, *Models of Information Processing in the Basal Ganglia*, pages 249–270. MIT Press, Cambridge, MA, 1995.
10. S. W. Keele. Movement control in skilled motor performance. *Psychological Bulletin*, 70:387–403, 1968.
11. D. Lee, N. L. Port, and A. P. Georgopoulos. Manual interception of moving targets. II. on-line control of overlapping submovements. *Experimental Brain Research*, 116:421–433, 1997.
12. D. E. Meyer, S. Kornblum, R. A. Abrams, C. E. Wright, and J. E. K. Smith. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review*, 95(3):340–370, 1988.
13. R. C. Miall, D. J. Weir, and J. F. Stein. Intermittency in human manual tracking tasks. *Journal of Motor Behavior*, 25:53–63, 1993.
14. T. E. Milner. A model for the generation of movements requiring endpoint precision. *Neoruscience*, 49:487–496, 1992.
15. K. E. Novak, L. E. Miller, and J. C. Houk. Kinematic properties of rapid hand movements in a knob turning task. *Experimental Brain Research*, 132:419–433, 2000.
16. L. D. Partridge. Integration in the central nervous system. In J. H. U. Brown and S. S. Gann, editors, *Engineering Principles in physiology*, pages 47–98. Academic Press, New York, 1973.
17. R. Plamondon and A. M. Alimi. Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Science*, 20:279–349, 1997.
18. J. Pratt, A. L. Chasteen, and R. A. Abrams. Rapid aimed limb movements: age differences and practice effects in component submovements. *Psychology and Aging*, 9:325–334, 1994.
19. C. Redon, L. Hay, and J.-L. Velay. Proprioceptive control of goal directed movements in man studied by means of vibratory muscle tendon stimulation. *Journal of Motor Behavior*, 23:101–108, 1991.
20. D. A. Robinson. Oculomotor control signals. In G. Lennerstrand and P. B. y Rita, editors, *Basic Mechanisms of Ocular Mobility and Their Clinical Implications*, pages 337–374. Pergamon Press, Oxford, 1975.
21. R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.
22. C. von Hofsten. Development of visually directed reaching: The approach phase. *Journal of Human Movement Studies*, 5:160–168, 1979.
23. C. H. Wu, J. C. Houk, K. Y. Young, and L. E. Miller. Nonlinear damping of limb motion. In J. M. Winters and S. L.-Y. Woo, editors, *Multiple Muscle Systems: Biomechanics and Movement Organization*, pages 214–235. Springer-Verlag, New York, 1990.