Movement Duration as an Emergent Property of Reward Directed Motor Control

Rigoux L., Sigaud O., Terekhov A., Guigon E.

Institut des Systèmes Intelligents et de Robotique,UPMC-Paris 6

CNRS UMR 7222, 4 place Jussieu, 75005 Paris, France

lionel.rigoux@isir.fr

The central nervous system elaborates well-coordinated movements in spite of unexpected and unpredictable events that interfere with their goal (e.g. intrinsic noise, external forces, target displacements, ...). Comparisons between unperturbed and perturbed motor acts reveal that the movement duration is dynamically adjusted in response to perturbations [1, 2, 3]. Models have been proposed to explain the determination of movement duration in a principled way [4, 5, 6]. Yet, as these models compute movement duration based on open-loop planning, they fail to explain actual movement duration that results from an online control policy, as for perturbed movements.

Here we propose to consider a movement as a reward-driven behavior like in reinforcement learning theory; more precisely, we assume that motor control is governed by an optimal feedback policy computed at each visited state with respect to a stationary cost-to-go function J involving a trade-off between effort and reward over an infinite horizon:

$$J(x(t)) = \int_t^\infty e^{-\gamma_0(s-t)} \left[\alpha u(s)^2 - r(x(s))\right] \, ds$$

where x is the state of the controlled object, u is the control signal, α is a weight on the effort term, r is a reward function which is null everywhere except for the target state ($r = r_0$), and the exponential term parametrized by γ_0 represents the uncertainty about the future and favors immediate actions over procrastination. The feedback controller, obtained by coupling this optimal policy with an optimal state estimator, continuously drives the system towards the rewarded state.

We tested the proposed controller with a two-segment planar model of the human arm actuated by six muscles (modeled as second-order low-pass filter force generators). Movements, simulated for different amplitudes and reward values (r_0), have classical properties such as slightly curved trajectories and bell-shaped velocity profiles (Fig. 1a). More importantly, the experimentally observed law of amplitude/duration scaling is replicated by the model and is modulated by the reward value (Fig. 1b).

The model was first tested on the target jump protocol described in [2]. This experiment shows that the total duration of a movement depends on the timing of an unexpected displacement of the target during the reach. As the perturbation is unpredictable, no global duration could be explicitly determined in advance. We estimated parameters of our model by fitting unperturbed trajectories and, then, applied sudden change in the reward location to simulate target jumps. Our model quantitatively reproduces lengthening of movement duration according to the perturbation timing (Fig. 2). The model was then tested on a force-field experiment [3]. Fig. 3 illustrates how a unique set of parameters, which produces rapid reaching movements in the null-field, explains the complex trajectories performed under viscous perturbations: movements do not stop before they achieve their goals, therefore their duration is highly dependent on the events (like external forces) which impede the action.

To summarize, we propose a computationnal view of motor control where movements emerge from a continuous actuation of the controlled system towards a rewarded state by an optimal feedback controller. This approach can explain how the movement duration is regulated according to target distance and encountered perturbations.

References

[1] M. A. Goodale, D. Pelisson, C. Prablanc, *Nature* **320**, 748 (1986).

[2] D. Liu, E. Todorov, *J Neurosci* **27**, 9354 (2007).

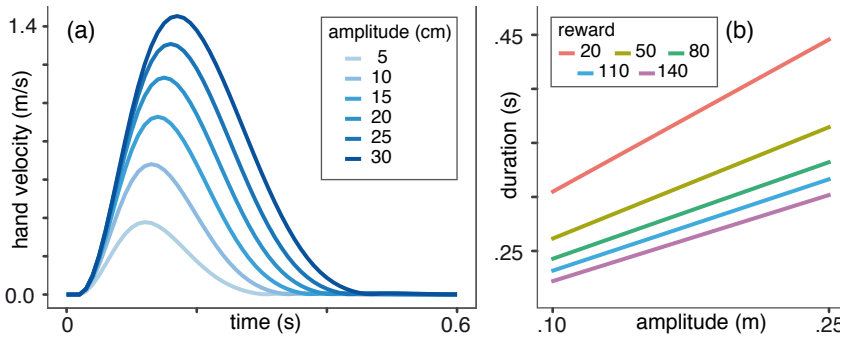
[3] R. Shadmehr, F. A. Mussa-Ivaldi, *J Neurosci* **14**, 3208 (1994).

[4] B. Hoff, *Biol Cybern* **71**, 481 (1994).

[5] C. M. Harris, D. M. Wolpert, *Biol Cybern* **95**, 21 (2006).

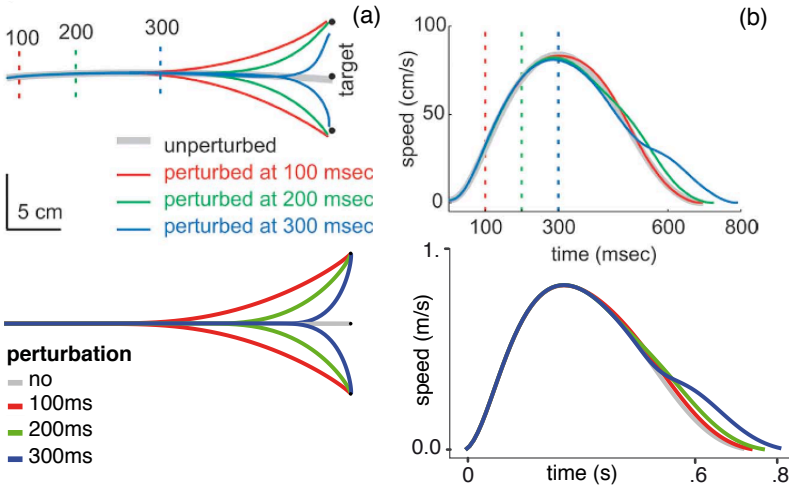
[6] R. Shadmehr, J. J. O. de Xivry, M. Xu-Wilson, T.-Y. Shih, *J Neurosci* **30**, 10507 (2010).

Figure 1 :



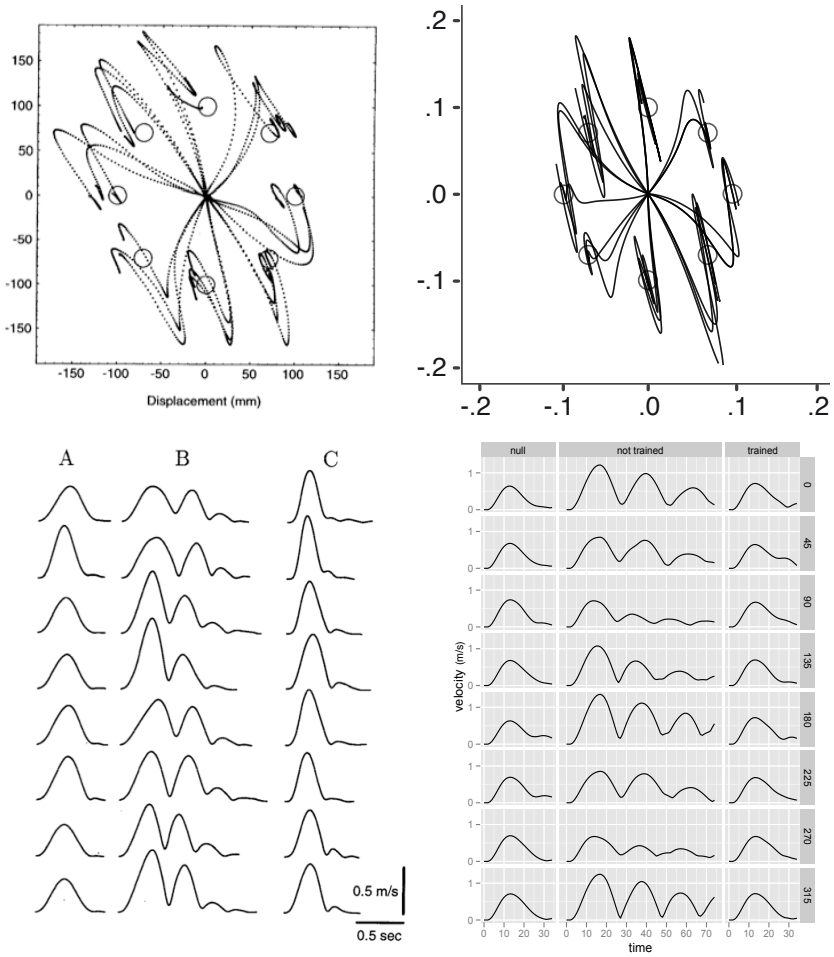
(a) Velocity profiles for different movement amplitudes. Note that the simulation continues after the target is reached.
(b) Changes in the movement duration as a function of the distance to the target for different reward values. Amplitude/duration scaling law emerges form the model. Reward devaluation produces bradykinesia as in [6].

Figure 2 :



Top: Experimental trajectories (a) and velocity profiles (b) for reaching movements toward a target which jumps unexpectedly left or right 100ms (red), 200ms (green) or 300ms (blue) after movement onset. Note that the arrival time is influenced by the timing of perturbation [2].
Bottom: Simulation results. Target jumps were modeled by a change of the reward location (not reward value). Difference in the corrective movement duration is also an emergent property of the model.

Figure 3 :



Left: Experimental results of a representative subject reaching to targets in eight directions [3]: hand trajectories during first exposure to the viscous force-field (top); corresponding velocity profiles (bottom) measured without force-field (A), with force-field before adaptation (B) and after adaptation (C).
Right: Simulation results. The most important property is the dynamic adaptation of the movement duration to the perturbation: the same set of parameters generates bell-shaped velocity profile hand movements in null-field and adapted force-field but longer movements with multi-peaked velocity profiles during initial exposure to the force-field. Trajectories were simulated with multiplicative motor noise and additive sensory noise, with a controller naive to the effect of the force-field on the dynamic of the arm.