Using the soft actor-critic algorithm to generate rapid eye movements with an unconstrained 6DOF biomimetic robotic eye

Miguel E. Teixeira*, A. John van Opstal†, and Alexandre Bernardino*,+, *Senior Member, IEEE*

***Abstract*— Understanding how the brain controls rapid eye movements, known as saccades, remains an open challenge. Biomimetic models have provided valuable insight into these control mechanisms. In this work, we develop an accurate, six-degree-of-freedom computational model of a biomimetic eye to explore the neural control signals underlying these movements. We hypothesise that saccade generation is governed by an optimisation of multiple costs, including accuracy, duration, energy, and tendon tension. To test this hypothesis, we trained a model-free reinforcement learning algorithm using biologically inspired input signals, in an open-loop manner. Our results show that the emerging control strategies successfully replicate human-like saccadic characteristics, including nonlinear *main sequence* relationships, compliance with Listing's law, straight trajectories, normometric pulse-step-like controls, and the antagonistic pairing of extraocular muscles, without explicitly enforcing these behaviours. Additionally, we analyse the impact of different reward costs and noise components on the resulting saccadic control strategies and resulting motions.**

I. Introduction

The neural control of eye movements is an intriguing topic of research. Unlike other skeletomotor systems, the eye is not subjected to external loads and is confined to a socket that restricts its movements to rotations.

Biorobotics aims to develop biologically inspired control strategies and designs for mechanical systems [1]. It can also offer insights into neurobiology, allowing researchers to study biological behaviours without invasive manipulations [2].

Here, we focus on the emergence of rapid, goal-directed eye movements and demonstrate that their properties closely resemble the stereotyped saccade trajectories of primates [3][4]. Prior research suggests that saccades result from a neural strategy that optimises a speed-accuracy trade-off [5].

Although existing approaches rely on model-based control theory [6][7][8], here we employ a model-free Reinforcement Learning (RL) approach that learns which input controls lead to optimal trajectories of a physics-derived nonlinear computational model of a 6 degrees-of-freedom (DOF) biomimetic eye system. The optimisation maximised a reward scalar, evaluated from costs such as end-point accuracy, duration, total energy use, and final tendon tension.

The emerging control strategies exhibit key biological properties, providing insight into how neural circuits may optimise saccadic movements. Furthermore, in this work, we refine our earlier proposed computational model [7][8], analyse the effects of different types of noise on the control signals, and evaluate how the different cost contributions affect the resulting control strategies.

II. Background

Saccades are fast and accurate ballistic eye movements that reach velocities of up to 700 deg/s in humans [9], and 1300 deg/s in monkeys [4]. Due to their short duration, ranging from 25 to 100 ms (depending on the saccadic amplitude) [10], they are assumed to be controlled in an open-loop fashion, without real-time visual feedback [3][11].

To elicit a saccade, the brain programmes a pulse-like signal, believed to originate in the midbrain superior colliculus [12], which is subsequently distributed to the six extra-ocular muscles (EOMs), grouped in antagonistically acting pairs: Medial and Lateral rectus (MR-LR), Superior and Inferior Oblique (SO-IO), and Superior and Inferior Rectus (SR-IR).

A. Main Sequence

The *main sequence* refers to the experimentally observed relationships between saccadic movement variables: amplitude, A, duration, D, and peak eye-velocity, ω_{pk}. Duration and amplitude obey an affine relation, Eq. (1a), indicating that the saccadic system is nonlinear. Further, the product of the peak velocity and duration follows a linear relationship with saccadic amplitude, Eq. (1b). As a result, the peak eye-velocity saturates at large amplitudes [13], Eq. (1c):

$$D = c_1 \cdot A + c_2 \tag{1a}$$

$$\omega_{pk} \cdot D = c_3 \cdot A \tag{1b}$$

$$\omega_{pk} = \frac{c_3 \cdot A}{c_1 \cdot A + c_2} \tag{1c}$$

with $c_1, c_2,$ and c_3 subject- and saccade-direction dependent fitting parameters.

As the peak velocity of a saccade is reached about $t_{pk} \approx$ 20-25 ms after the start, regardless of the saccade amplitude, velocity profiles are positively skewed and their skewness increases with saccade amplitude [14].

B. Straight Trajectories & Cross-Coupling

Oblique saccadic trajectories are approximately straight, indicating that eye movements are generated about a fixed axis of rotation. Thus, the three components of the vectorial angular velocity, $\boldsymbol{\omega}(t) = [\omega_x(t), \omega_y(t), \omega_z(t)]$, are scaled versions of each other, obeying:

$$\omega_k(t) = \alpha_k \|\boldsymbol{\omega}(t)\| \quad \text{with} \sum_{k=x,y,z} \alpha_k^2 = 1 \tag{2}$$

[*] Institute for Systems and Robotics, Instituto Superior Técnico, ISR, Lisbon, Portugal, [†] Section Neurophysics, Donders Centre for Neuroscience, Radboud University, Nijmegen, The Netherlands

[+] Correspondence: `alex@isr.tecnico.ulisboa.pt`

This relation reflects a second nonlinear property of the controller: component cross-coupling. For a linear controller, the components are independent — for instance, the profile of the horizontal velocity component is the same for all oblique saccades with the same azimuth angle. However, to satisfy Eq. (2), the three components must have matching durations and shapes. As a result, the smaller components are stretched to match the duration of the vector [6].

C. Donders' & Listing's Law

Any gaze direction can be fully specified by two DOFs: azimuth and elevation. However, the third DOF, known as cyclotorsion, cannot be left unchecked, as saccade sequences could otherwise lead to an accumulation of ocular torsion, resulting in significant localisation errors [15][16].

Donders empirically observed that the eye's cyclotorsion is fully determined by the other two DOFs. Thus, each gaze direction has a unique value of cyclotorsion, independent of the trajectory that brought it there [17].

A special case of Donders' Law, known as Listing's law, states that when the head is held upright and still, with the eyes looking at infinity, all eye orientations, when expressed as 3D Euler-Rodrigues rotation-vectors [6][18], $\boldsymbol{r} = (r_x, r_y, r_z)$, are constrained to Listing's plane (LP): $r_x = p \cdot r_y + q \cdot r_z$, defined by the normal vector $\boldsymbol{n} = (1, -p, -q)$.

D. Controller

Despite the complexity of the oculomotor system, Goldstein proposed that the horizontal oculomotor plant could be approximated by an overdamped second-order linear model, consisting of two viscoelastic Voigt elements in series. Each element i is characterised by both a spring constant κ_i and a damping coefficient γ_i, which can be expressed as a time constant $T_i = \gamma_i / \kappa_i$ [19]. When subject to a force, generated by a change in firing of oculomotor neurons, $\Delta R(t)$, the Laplace transfer characteristic for this system is written as:

$$\frac{\Theta(s)}{\Delta R(s)} = G \frac{sT_z + 1}{(sT_1 + 1)(sT_2 + 1)} \approx G \cdot \frac{1}{(sT_2 + 1)} \tag{3}$$

with s the complex Laplace variable, G a normalization factor, $T_1 \sim 20\,\text{ms}$, $T_2 \sim 200\,\text{ms}$ and $T_z = (\gamma_1 + \gamma_2)/(\kappa_1 + \kappa_2) \sim 70\,\text{ms}$ [10]. Here, T_1 quantifies the contribution of the globe's (small) moment of inertia, while T_2 describes the strong damping caused by fatty tissue and the drag from the optic nerve. Since $T_2 \gg T_1, T_z$, the dynamics of the plant are dominated by T_2; transients decay quickly and can be neglected.

Under the first-order approximation of Eq. (3), the neural input that would produce a pure step output (the 'ideal saccade'), $\Delta R^*(s)$, can be readily calculated, by letting $\Theta(s) = \mathcal{L}[AH(t)] = A/s$, with A the amplitude of the Heaviside step function, $H(t)$:

$$\Delta R^*(s) = \frac{A}{G}\left(T_2 + \frac{1}{s}\right) \cdot 1 \tag{4a}$$

$$\Delta R^*(t) = \frac{A}{G}\left(T_2 \delta(t) + H(t)\right), \quad t \geq 0 \tag{4b}$$

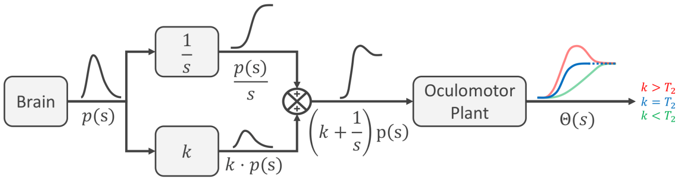


Fig. 1: *Simplifying scheme for the saccadic pulse-step generator of the LR muscle, depicting the pulse and its subsequent transformations (in Laplace format). The pulse, $p(s)$, is integrated, $\frac{p(s)}{s}$, and summed with a scaled version of itself, $k \cdot p(s)$. The resulting pulse-step signal drives the 1D oculomotor plant, the output of which, $\Theta(s)$, is depicted for three different values of k. Note that when $k = T_2$, the resulting movement has an optimal speed-accuracy trade-off. For simplicity, the antagonist action of MR is not included.*

According to Eq. (4b), an ideal saccadic eye movement is driven by a pulse-step control: a summation of a scaled Dirac pulse, $\delta(t)$, and a step, $H(t)$.

Robinson [20] hypothesised that the two components of the pulse-step signal originate from a common pulse, $p(s)$, which subsequently undergoes temporal integration, $\frac{p(s)}{s}$, and summation with a scaled version of itself, $k \cdot p(s)$. This hypothesis was later confirmed by experiments [21]. A scheme of this process is represented in Fig. 1, where the three outputs represent three values for k, the gain of the direct path to the LR muscle. The red ($k > T_2$) and green ($k < T_2$) outputs represent overshooting and undershooting trajectories, respectively, while the blue ($k = T_2$) output corresponds to the trajectory that optimises speed and accuracy of the movement.

Note that using the full expression described in Eq. (3), would introduce an additional 'slide' term, caused by the joint characteristic time T_z, characterised by an exponential decay of the pulse into the step of the oculomotor neuron signal [20].

E. Optimisation strategy

The characteristic properties of saccades have led researchers to hypothesise that they emerge from an optimisation strategy that aims to trade off speed, accuracy, and effort constraints. Here, we apply a model-free control based on RL to address this optimisation problem. The controller generates six pulse-like commands to the tendons, that drive a realistic nonlinear computational simulator of a biomimetic eye, to produce eye movements.

Since the algorithm does not rely on direct information about the model's dynamics, the commands are optimised using a trial-and-error strategy with feedback based on a numeric reward, as explained below.

1) Reward Function: Following [7], the reward function, $R(\boldsymbol{\vartheta}_F, \hat{\boldsymbol{\vartheta}}_G, \boldsymbol{u}, D, \boldsymbol{m})$, is defined as a linear combination of four costs: the accuracy of the final gaze direction $\boldsymbol{\vartheta}_F$ relative to the intended goal $\hat{\boldsymbol{\vartheta}}_G$, the duration of the movement D, the total energy expenditure associated with the control inputs $\boldsymbol{u}$, and the total magnitude of the final tendon tensions exerted on the globe, $\boldsymbol{m}$:

$$R(\boldsymbol{\vartheta}_F, \hat{\boldsymbol{\vartheta}}_G, \boldsymbol{u}, D, \boldsymbol{m}) = -\lambda_X \varphi_X(\boldsymbol{\vartheta}_F, \hat{\boldsymbol{\vartheta}}_G) \\ - \lambda_D \varphi_D(D) - \lambda_E \varphi_E(\boldsymbol{u}) - \lambda_F \varphi_F(\boldsymbol{m}) \quad (5)$$

with $\varphi_i(v)$ the costs functions used, and $\lambda_i > 0$ their weights.

The accuracy cost aims to penalise the deviation of the final gaze direction, $\boldsymbol{\vartheta}_F$, relative to the intended goal, $\hat{\boldsymbol{\vartheta}}_G$.

$$\varphi_X(\boldsymbol{\vartheta}_F, \hat{\boldsymbol{\vartheta}}_G) = \|\boldsymbol{\vartheta}_F - \hat{\boldsymbol{\vartheta}}_G\|^2 \quad (6)$$

Note that the gaze direction, $\boldsymbol{\vartheta}$, is fully determined by the elevation, θ_y, and azimuth, θ_z, angles, thus $(\boldsymbol{\vartheta}_F)^T = [\theta_y^F, \theta_z^F]$. Importantly, the cyclotorsion, θ_x, is neither specified nor constrained, during the learning process, unlike in [7][8].

Since eye motion should be as fast as possible, longer-duration movements are penalised. Studies on humans and monkeys have shown that the subjective value of a reward decreases over time according to a hyperbolic function [22]. As such, the duration cost is described as:

$$\varphi_D(D) = 1 - \frac{1}{1 + \beta D} \quad (7)$$

where D is the estimate for saccade duration, and $\beta = 0.6\ \mathrm{s}^{-1}$ an appropriate scaling parameter [23].

Since energy is a finite resource, it is assumed that the brain aims to limit its consumption during saccade execution.

The associated energy cost is quadratic in the velocity of the motor commands:

$$\varphi_E(\boldsymbol{u}) = \sum_{i=1}^{6} \int_0^D \left(\frac{du_i}{dt}(t') \right)^2 dt' = \sum_{i=1}^{6} K_i \quad (8)$$

where $\boldsymbol{u}^T(t) = [u_1, u_2, u_3, u_4, u_5, u_6](t)$ is a six-dimensional vector containing the independent instantaneous inputs applied to the environment (i.e., the elastic tendons of the computational biomimetic eye model). This term thus accounts for the expenditure of the total kinetic energy, K.

Finally, it was assumed that the eye should avoid static fixation states in which the tendons are unnecessarily strained. The tension cost aims to penalise tendon co-contraction, which amounts to an excess of energy stored in these elastic elements as increased stiffness:

$$\varphi_F(\boldsymbol{m}) = \sum_{i=1}^{6} m_i \quad (9)$$

where $\boldsymbol{m}^T = [m_1, m_2, m_3, m_4, m_5, m_6]$ is a positive-only six-dimensional vector containing the magnitudes of the tendon forces, computed in the final eye orientation.

III. Methods

A diagram of the methodology used in this work is presented in Fig. 2. Briefly, the system is divided into an agent and an environment. The agent uses the SAC algorithm [24][25] to generate six modifiable action signals that are subsequently applied to the environment, where they are transformed into pulse-step commands (Fig. 1) that drive the computational simulator, thus generating an eye movement. The eye trajectories resulting from each action are processed to calculate a reward signal, which is fed back to the agent. This feedback loop enables the agent to adapt its actions such that, in future iterations, it can achieve higher rewards.

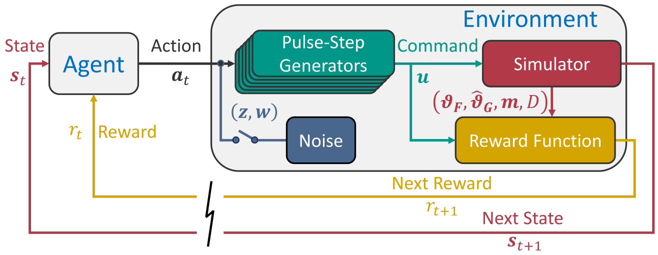


Fig. 2: *Schematic of the agent-environment interactions. The agent provides an action signal, $\boldsymbol{a}_t$, which generates six pulse-step signals that drive the simulator, $\boldsymbol{u}$. Each of the six pulses has 4 adjustable parameters, Eq. (10). The simulator generates the eye trajectory, which is used to compute a reward signal, r_{t+1}. The reward is fed back to the agent.*

In this work, an *episode* is defined as a complete agent-environment interaction: The agent is tasked with reaching a randomly defined target from the current eye orientation, prompting it to generate an action (i.e., six pulses with the current parameters). The episode ends when the agent receives the corresponding reward signal, leading to an update of the $6 \cdot 4$ pulse-step parameters [25], [24]. During training, the agent performs a user-defined number of continuous saccades, i.e. each episode's starting position corresponds to the settling position of the previous movement.

A. Actions: Pulses

As suggested by Eq. (4a), an ideal step-like saccade requires a Dirac delta pulse, $\delta(t)$ ($p^*(s) = 1$), for the pulse-step generator. This would require the eye to achieve infinite accelerations, which is impossible. To relax this idealisation, we utilise a continuous pulse shape [26], $P(t)$, with a clear onset (at $t = 0$) and offset (at $t = t_f$) with zero first and second derivatives, reaching a single extremum of magnitude $2V$, $P(t_p) = 2V$. According to Fig. 1, the pulse is integrated and summed with a scaled version of itself, gain k, to yield the pulse-step input, $PS(t, V, t_p, t_f, k)$, given by:

$$PS(t, V, t_p, t_f, k) = \int_0^t P(t', V, t_p, t_f) dt' + k \cdot P(t, V, t_p, t_f) \quad (10)$$

The total energy expenditure, Eq. (8), associated with a single pulse-step, is given by:

$$K(V, t_p, t_f, k) = \int_0^D \left(\frac{dPS}{dt}(t, V, t_p, t_f, k) \right)^2 dt \quad (11)$$

To study the impact of noise on the inputs, two different components were included in the base pulse: additive noise, independent of the signal, and multiplicative noise, whose variance increases with the strength of the signal. The latter is commonly found in biological systems. The noise terms had standard deviations σ_{add} and σ_{mul}, respectively:

$$P^N(t, V, t_p, t_f) = sig(\sigma_{mul} \cdot z(t)) P(t, V, t_p, t_f) + \sigma_{add} \cdot w(t) \quad (12)$$

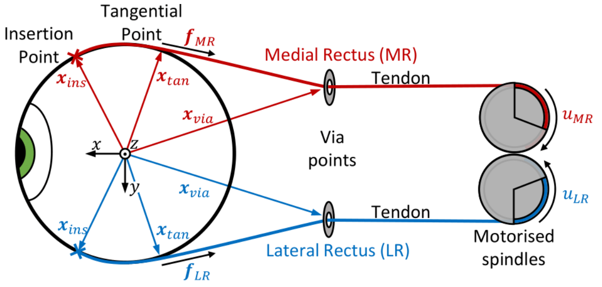


Fig. 3: *The computational simulator. The eyeball is driven by 6 elastic tendons connected to 6 'head'-fixed motorised spindles, driven by* $\boldsymbol{u}$*. Only the LR and MR tendons are shown, for simplicity. The tendons follow a wrap-around geometry on the eyeball, where the tendon follows the globe from its insertion point to a tangential point where it leaves the eyeball. Tendons are passed through a via point and directed to the spindles. The directions of the force vectors,* $\boldsymbol{f}_m$*, are determined by the difference between* $\boldsymbol{x}_{via}$ *and* $\boldsymbol{x}_{tan}$*.* $[\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}]$ *represents the Cartesian reference frame, centered in the eye.*

where $z(t), w(t) \sim \mathcal{N}(0,1)$ are independently sampled at each time instant, t. The multiplicative noise is squashed by a sigmoid function: $sig(x) = 1/(1+\exp(-2x))$, to prevent negative values. Note that when the pulse is stochastic, the costs become stochastic variables too.

B. The eye simulator

The computational simulator follows the physical equations derived from the Newton-Euler formulation for rigid bodies. It was implemented using the `Cython` programming language [27] and built upon previous works [7][8].

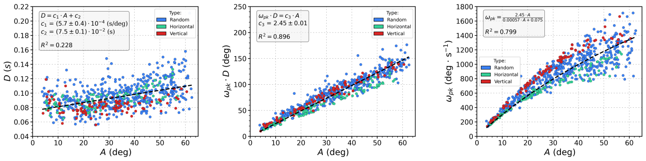
In this implementation, the eyeball is considered to be a spherical rigid body, and the EOMs are modelled as elastic cables, or tendons, that connect to both the eyeball and independent, freely-rotating spindles. As each spindle rotates, it winds its respective cable, changing its length and thus exerting a force on the eyeball. The agent-defined pulse-step commands set the rotation angles of these spindles. A simplified graphical representation for the LR and MR tendons of the simulator is shown in Fig. 3.

To enhance the anatomical accuracy of the model, a wrap-around geometry was introduced for the six tendons. By following this geometry, the elastic elements are forced to follow the surface of the eyeball (Fig. 3). This refinement prevents the unphysical scenario where tendons can pass through the interior of the eyeball as in [7]. More detailed information on this implementation can be found in [26].

C. Reward calculation.

The reward is a scalar quantity that numerically encodes the consequences of the actions; the goal of the agent is to maximise the reward. Each action performed by the agent results in an eye trajectory computed by the simulator, which is subsequently evaluated to obtain the relevant costs needed to compute the reward, Eq. (5). Note that the costs are computed at the end of the motion (duration, D). The movement duration was determined by applying a threshold criterion to the eye's absolute angular velocity, normalized with respect to its maximum value. The movement offset was found whenever this quantity dropped below $\|\hat{\boldsymbol{\omega}}_{thr}\| = 0.01$ and remained below the threshold for at least $w = 10$ ms.

In the simulations with noise, we applied this criterion to a low-pass filtered version of the absolute eye velocity [26].



(a) (b) (c)

Fig. 4: *Main sequence relations for model eye movements in all directions: (a) relationship between duration, D, and amplitude, A; (b) the relationship between* $\omega_{pk} \cdot D$ *and A; (c) saturation of* ω_{pk} *for large A. Fits were performed on all* 700 *data points (see text); results are provided in the legends. Note the direction dependence on eye-movement dynamics: vertical movements (red) are fastest.*

IV. Results

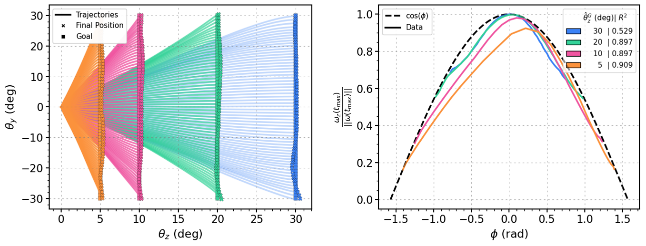
This section summarizes the main results from this work, after the agent had been trained for $2 \cdot 10^6$ episodes. In Secs. IV-A to IV-D, we compare the optimal behaviours generated by the agent with known biological behaviours. In Sec. IV-E we examine the relative contribution of the different terms of the reward function, and in Sec. IV-F we analyse the impact of noise, both in terms of the resulting eye movements.

The default weights used in the reward function were empirically determined. For this study, these were taken as: $\lambda_A = 100$, $\lambda_D = 3$, $\lambda_E = 1 \cdot 10^{-7}$ and $\lambda_F = 1 \cdot 10^{-4}$.

A. Main Sequence

Fig. 4 depicts the *main sequence* relations obtained from 700 random continuous saccades: 500 performed in random directions (blue), and 100 horizontal (green), and vertical (red) movements. The 700 responses were fitted to the relations of Eq. (1) (parameters in Fig. 4 legends), with the resulting fits displayed as dashed black lines.

Despite the smaller R^2 value in Fig. 4(a), compared to Figs. 4(b)-(c), it is noted that the duration of the motions increases with the amplitude. Furthermore, note that even though the affine relation between $\omega_{pk} \cdot D$ and A, Fig. 4(b), seems unaffected by the dynamics of the movement, the nonlinear relationship between ω_{pk} and A differs depending on the direction of motion, with vertical movements achieving higher peak velocities than horizontal motions. This difference can be explained by noting that while elasticity κ is taken constant for all tendons, vertical saccades require four tendons, while horizontal movements result from the activation of mainly two tendons.



(a)

(b)

Fig. 5: *Results from the cross-coupling analysis. (a) Trajectories (coloured lines) generated by the agent to target locations (squares). Crosses indicate final eye positions. (b) Quotient between the horizontal component and norm of the angular velocity vector, computed at the instant of maximum norm (coloured lines), and the ideal cosine from Eq. (2) (dashed line), as function of the movement direction, ϕ.*

B. Cross-Coupling

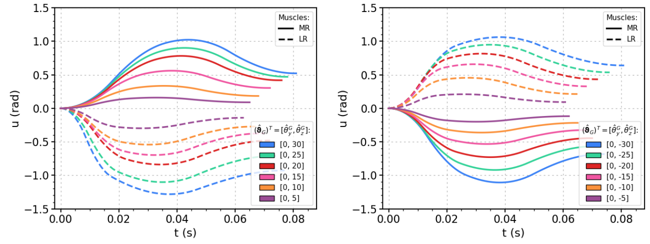
To test whether the emerging control strategies satisfy the cross-coupling relations of Eq. (2), oblique eye movements were performed from the origin to the endpoints $\hat{\vartheta}_G = [\hat{\theta}_z^G, \hat{\theta}_y^G]$ with $\hat{\theta}_z^G = \{5, 10, 20, 30\}$ deg, $\hat{\theta}_y^G \in [-30 : 1 : 30]$ deg (square markers in Fig. 5(a)). The quantities $\omega_z(t_{max})$ and $\|\boldsymbol{\omega}(t_{max})\|$ were evaluated at the time when the latter reached a maximum $t_{max} = \mathrm{argmax}_t \|\boldsymbol{\omega}(t)\|$. The data were then compared with the ideal cosine prediction for perfectly straight saccades (dashed line) using the coefficient of determination, R^2 (legend in Fig. 5(b)). The results indicate a strong amount of component cross-coupling that is very close to the prediction for straight oblique trajectories.

C. Antagonism

The antagonistic behaviour between muscle pairs, observed in real biological systems, can be inferred from the negative correlation between activation signals: when one muscle in a pair contracts, its antagonist will relax. This behaviour can be readily observed in the LR and MR commands generated by the agent in Fig. 6. It can also be seen that the signals have a pulse-step-like shape, in which the pulse amplitude and duration have characteristics similar to the *main sequence* for the eye-movement dynamics.

To further test this antagonistic nature for all tendons, the correlation coefficients between each antagonistic pair were calculated for 700 continuous saccadic motions, obtained in a similar way as in Sec. IV-A. The results are presented as histograms in Fig. 7 for all saccade directions.

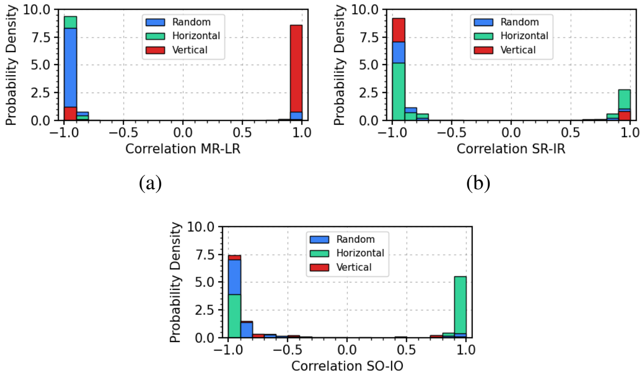
The data comply with the antagonist-agonist behaviour observed in the real system, as most correlation coefficients are negative, with the highest peak corresponding to correlations within the range $[-1, -0.9]$. Note that positive correlations were only obtained for movements that did not belong to a tendon's pulling direction. Thus, the LR-MR pair could show a positive correlation for vertical saccades, to which they hardly contributed. Similarly, the SR-IR and SO-IO pairs only had positive correlations for horizontal saccades.



(a)

(b)

Fig. 6: *Control signals of the LR (dashed) and MR (solid) tendons as generated by the agent, for horizontal saccades that start from the origin to six different goals (legend). Note the pulse-step shapes of the controls and their clear antagonistic behaviours.*



(a)

(b)

(c)

Fig. 7: *Normalized histograms of correlations between pulse-step controls for MR-LR, SR-IR, and SO-IO antagonists, computed for 500 random-direction (blue), and 100 random horizontal (green) and vertical (red) continuous saccades.*

D. Donders' & Listing's Law

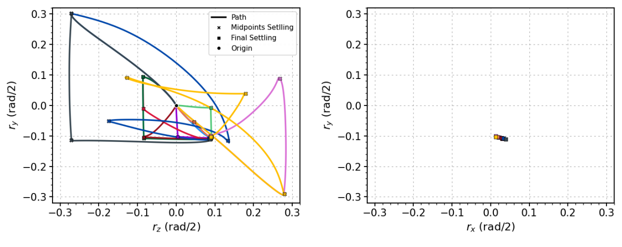
Donders' law states that the cyclotorsion, r_x, for any gaze orientation, is fully specified by the remaining two DOFs: $r_x = f(r_y, r_z)$, regardless of the trajectory taken by the eye to attain that orientation.

To analyse whether this behaviour is present in the biomimetic system, the agent had to reach the same target gaze direction, starting from the origin but taking different paths, represented by different colours in Fig. 8(a).

Note that the final values of cyclotorsion, r_x, were all very similar (Fig.8(b)), with an average $\overline{r_x} = 0.023$ rad/2, and a standard deviation $\sigma_{r_x} = 0.008$ rad/2.

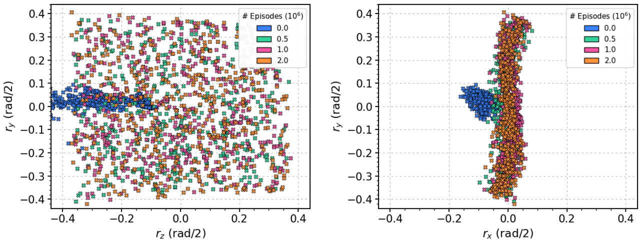
As noted in Sec. II-C, Listing's Law reduces $f(r_y, r_z)$ to a plane. Fig. 9 shows the final eye orientations of 500 continuous saccades as rotation vectors in the $[r_z, r_y]$ and $[r_x, r_y]$ planes, generated by the agent at four different periods during the training (colours). Note that at the start of the training (blue) resulting eye movements had no relation to the actual targets. Yet, after about 500,000 episodes (green) the orientations converged towards the final results.

Note also that although Listing behaviour was not explic-



(a) (b)

Fig. 8: *(a) Instantaneous eye orientations (expressed as rotation vectors) for several trajectories generated by the agent, towards a common target $(\hat{\boldsymbol{\vartheta}}_G)^T = [10, -10]$ deg $\simeq [8.77, -8.77] \cdot 10^{-2}$ rad/2. (b) The eye's final orientations in the $[r_x, r_y]$ plane. Note that cyclotorsion, r_x, varied little for the different trajectories.*



(a) (b)

Fig. 9: *Rotation-vector end points of eye orientations after 500 random continuous saccades during agent training. (a) Front view, $[r_y, r_z]$; (b) Side view $[r_x, r_y]$ of Listing's plane. Results were saved at $\{0, 0.5, 1, 2\} \cdot 10^6$ episodes.*

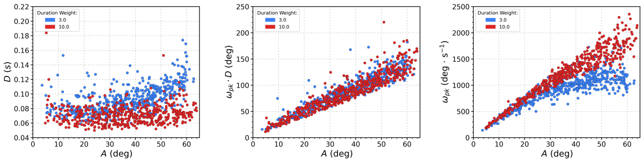
itly included as a reward in the cost-evaluation (in contrast to [7][8]), the final eye orientations quickly converged to a plane (green-to-orange markers in Fig. 9): the side view shows clusters of data around $r_x \sim 0$ rad/2, despite that movements were made across the oculomotor range $[r_y, r_z] \in [-0.4, 0.4]$ rad/2 (Fig. 9(a)). If $\overline{\sigma_{r_x}}(y)$ represents the mean value, over 20 different runs, of the standard deviation of r_x, after the model was trained for $y \cdot 10^6$ episodes, then: $\overline{\sigma_{r_x}}(0) = (3.3 \pm 0.1) \cdot 10^{-2}$ rad/2, $\overline{\sigma_{r_x}}(0.5) = (2.4 \pm 0.1) \cdot 10^{-2}$ rad/2, $\overline{\sigma_{r_x}}(1) = (2.41 \pm 0.08) \cdot 10^{-2}$ rad/2, $\overline{\sigma_{r_x}}(2) = (2.25 \pm 0.08) \cdot 10^{-2}$ rad/2.

E. Cost weights

We selectively adjusted the weights associated with duration, λ_D, energy, λ_E, and tendon tension, λ_F, to study their effects on the agent's behaviour. Each weight was modified separately, with the other three remaining at their default values specified in Sec. IV.

1) Duration: We tested two different weights for the duration cost, $\lambda_D = \{3, 10\}$. Figure 10 shows the *main sequence* relationships of 500 random continuous saccades, generated by the agents trained with these different weights.

Increasing λ_D (red markers) forced the agent to reach targets in a shorter amount of time. Consequently, it generated eye movements with shorter durations (Fig. 10(a)), resulting in higher peak velocities (Fig. 10(c)). Yet, the relationship between $\omega_{pk} \cdot D$ and A remained virtually unaffected (Fig. 10(b)).



(a) (b) (c)

Fig. 10: *Main sequence for 500 random continuous saccades for duration weights, $\lambda_D = \{3, 10\}$ (blue and red markers, respectively). (a) A - D; (b) A - $\omega_{pk} \cdot D$; (c) A - ω_{pk}.*

2) Energy: Altering λ_E produced the opposite effect as for λ_D: an increase in λ_E caused the emerging movements to attain lower peak velocities, and typically longer durations than those created with a lower λ_E (data not shown). It is also worth noting that with $\lambda_E = 0$, all eye movements had the same duration, and the peak velocities did not saturate, regardless the motion's amplitude. In other words, the controller became linear [26].

3) Tension: For the tension weight, λ_F, we verified that changes did not significantly alter the *main sequence* relations. However, we noted that an increase in λ_F led to a thinner Listing's plane (data not shown) [26].

F. Influence of noise

Previous works [26][22] have analysed the influence of additive and multiplicative noise sources in a simple linear 1D system. In general, multiplicative noise has an effect similar to that of the energy cost, while additive noise mainly affects movement duration.

Inspired by this analysis, we examined the influence of these two qualitatively different noise sources on the movements produced by the fully nonlinear, 6DOF biomimetic eye system.

1) Multiplicative Noise: Since the standard deviation of an input endowed with multiplicative noise increases with the magnitude of the signal, stronger commands, which result in faster motions, will be associated with higher end-point uncertainty. Therefore, it is expected that increasing σ_{mul} in Eq. 12 should force the agent to generate smaller commands (i.e., slower saccades).

The results shown in Fig. 11 corroborate this hypothesis since eye movements in the presence of multiplicative noise indeed had longer durations and achieved lower peak velocities (orange markers Figs. 11(a)-(b)). Moreover, the velocity profiles associated with multiplicative noise exhibit an increased skewness compared to the noiseless paradigm (Fig. 11(c)).

2) Additive Noise: When considering the influence of the duration term on the reward function, no significant differences were observed when additive noise was introduced. However, when the duration cost was removed ($\lambda_D = 0$),

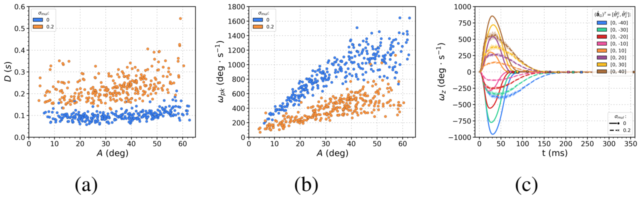


Fig. 11: *Main sequence relations for 300 random continuous saccades with (orange) and without (blue) multiplicative noise. (a) A - D; (b) A - ω_{pk}. (c) Velocity profiles for different targets $\hat{\vartheta}_G$. Saccades with $\sigma_{mul} > 0$ are markedly slower.*

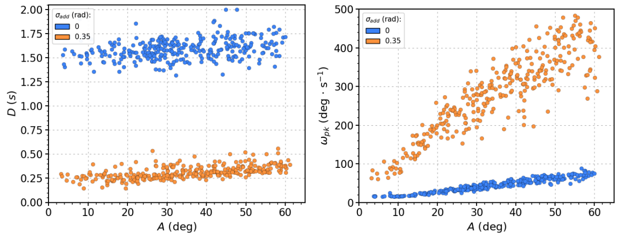


Fig. 12: *Main sequence relations with (orange) and without (blue) additive noise for* 300 *random continuous saccades and $\lambda_D = 0$. (a) A - D; (b) A - ω_{pk}.*

the introduction of additive noise resulted in faster motions with higher peak velocities, as can be observed in Fig. 12.

V. Discussion

This paper presents a novel approach to the study of the neural control of eye movements of a 6DOF nonlinear biometic eye, using a combination of computational modelling and reinforcement learning with the soft actor-critic algorithm. The results obtained from the simulations show that the proposed method can generate realistic eye movements, with the agent learning to generate neurobiologically realistic pulse-step control signals that drive the eye to the desired target locations with appropriate kinematics and nonlinear dynamics. The model was able to reproduce a number of key features of human saccadic eye movements, including the *main sequence*, component cross-coupling for straight oblique trajectories, and antagonistic muscle pairing. The model was also able to reproduce Listing's law, which is a fundamental principle of 3D eye-movement control. The results suggest that the model is a promising tool for studying the neural control of eye movements, and could be used to investigate the effects of different neural control strategies on eye movement behaviour and the dynamic effects on saccades for different eye-plant dynamics or specific eye-muscle surgery. Future work will focus on further refining and extending the model (e.g., more realistic muscle properties, eye-head control, visual input) and exploring its potential applications in the study of eye-movement control.

Acknowledgments

Supported by EU Horizon 2020 program, ERC Adv. Grant 2016 (nr. 693400 'ORIENT'); the Donders Centre for Neuroscience, Radboud University (AJVO); the FCT HAVATAR project (DOI: 10.54499/PTDC/EEI-ROB/1155/2020); LARSyS funding (DOI: 10.54499/LA/P/0083/2020, 10.54499/UIDP/50009/2020 and 10.54499/UIDB/50009/2020); the Portuguese Recovery and Resilience Plan (RRP), through project number 62, Center for Responsible AI.

References

[1] H. Bao et al, "Adaptive attitude determination of bionic polarization integrated navigation system based on reinforcement learning strategy," *Mathematical Foundations of Computing*, 2023.
[2] B. Webb, "Can robots make good models of biological behaviour?" *Behavioral and Brain Sciences*, 2001.
[3] R. W. Baloh et al, "Quantitative measurement of saccade amplitude, duration, and velocity," *Neurology*, 1975.
[4] A. F. Fuchs, "Saccadic and smooth pursuit eye movements in the monkey," *Journal of Physiology*, 1967.
[5] C. M. Harris and D. M. Wolpert, "The main sequence of saccades optimizes speed-accuracy trade-off," *Biological Cybernetics*, 2006.
[6] A. John et al, "Modelling 3d saccade generation by feedforward optimal control," *PLOS Computational Biology*, 2021.
[7] R. J. Alitappeh et al, "Emergence of human oculomotor behavior in a cable-driven biomimetic robotic eye using optimal control," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.
[8] A. J. Van Opstal and el al, "Realistic 3d human saccades generated by a 6-dof biomimetic robotic eye under optimal control," *Frontiers in Robotics and AI*, 2024.
[9] A. K. Horn and R. J. Leigh, "The anatomy and physiology of the ocular motor system," 2011, ch. 2.
[10] A. J. Van Opstal, *The auditory system and human sound-localization behavior*. Elsevier Academic Press, 2016.
[11] J. Syka et al, "Responses of neurons in the superior colliculus of the cat to stationary and moving visual stimuli," *Vision Research*, 1979.
[12] H. H. L. M. Goossens and A. J. Van Opstal, "Optimal control of saccades by spatial-temporal activity patterns in the monkey superior colliculus," *PLoS Computational Biology*, 2012.
[13] A. Bahill et al, "The main sequence, a tool for studying human eye movements," *Mathematical Biosciences*, 1975.
[14] A. J. Van Opstal and J. A. M. Van Gisbergen, "Skewness of saccadic velocity profiles: A unifying parameter for normal and slow saccades," *Vision Research*, 1987.
[15] J. D. Crawford and D. Guitton, "Visual-motor transformations required for accurate and kinematically correct saccades," *Journal of Neurophysiology*, 1997.
[16] D. B. Tweed et al, "Non-commutativity in the brain," *Nature*, 1999.
[17] F. Donders and W. Moore, *On the Anomalies of Accommodation and Refraction of the Eye: With a Preliminary Essay on Physiological Dioptrics*. New Sydenham Society, 1864.
[18] J. S. Dai, "Euler–rodrigues formula variations, quaternion conjugation and intrinsic connections," *Mechanism and Machine Theory*, 2015.
[19] H. Goldstein, "The neural encoding of saccades in the rhesus monkey," Master's thesis, Johns Hopkins Univ. Baltimore, MD, USA., 1983.
[20] D. A. Robinson, "Models of the saccadic eye movement control system," *Kybernetik*, 1973.
[21] S. C. Cannon and D. A. Robinson, "Loss of the neural integrator of the oculomotor system from brain stem lesions in monkey," *Journal of Neurophysiology*, 1987.
[22] R. Shadmehr and S. Mussa-Ivaldi, *Biological Learning and Control: How the Brain Builds Representations, Predicts Events, and Makes Decisions*. The MIT Press, 2012.
[23] H. Granado et al, "Learning open-loop saccadic control of a 3d biomimetic eye using the actor-critic algorithm." IEEE, 2023.
[24] T. Haarnoja et al, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," 2018.
[25] ——, "Soft actor-critic algorithms and applications," 2018.
[26] M. E. Teixeira, "How does the brain control the eye movements? an analysis-by-synthesis approach." Master's thesis, IST Lisbon, 2024.
[27] S. Behnel et al, "Cython: The best of both worlds," *Computing in Science Engineering*, 2011.