Toy Model Report

24 May 2019

Hypothesis

Structural plasticity offers significant advantages towards functional goals. Here, we try to build a proof of concept model, by using the features of the song bird vocal learning circuitry.

Plan

Song birds have a simple motor pathway (SMP) and a parallel pathway (AFP) that help with vocal learning. Based on evidences from literature, we model the SMP using Hebbian learning, and the AFP using reinforcement learning. The influence of AFP is eventually consolidated within the SMP, rendering it capable of producing song independently. Also, there is evidence that the AFP drives motor learning prior to the formation of SMP (in a crude sense). We attempt to provide a proof of concept that 1. there is a use of having 2 parallel pathways in learning, and 2. there is an advantage in having a head start for the AFP, by using the paradigm of arm exploration.

1 Models tested

1.1 Initial models

Initially, we model the two parallel pathways as shown in Fig 1. I provide here a brief overview without going into the details for the initial models. - HVC receives a syllable encoding as the input. (For now, the syllable encodings do not overlap).
- We design the HVC and RA as layers with less than 20 neurons each.
- The HVC and RA are fully connected, with modifiable weights.
- The SMP weights are modified via Hebbian learning and the AFP weights via RL.
- Reward is computed using a simple exponential function of the euclidean distance between the target and output.
- RL uses a convariance learning rule to update the weights.

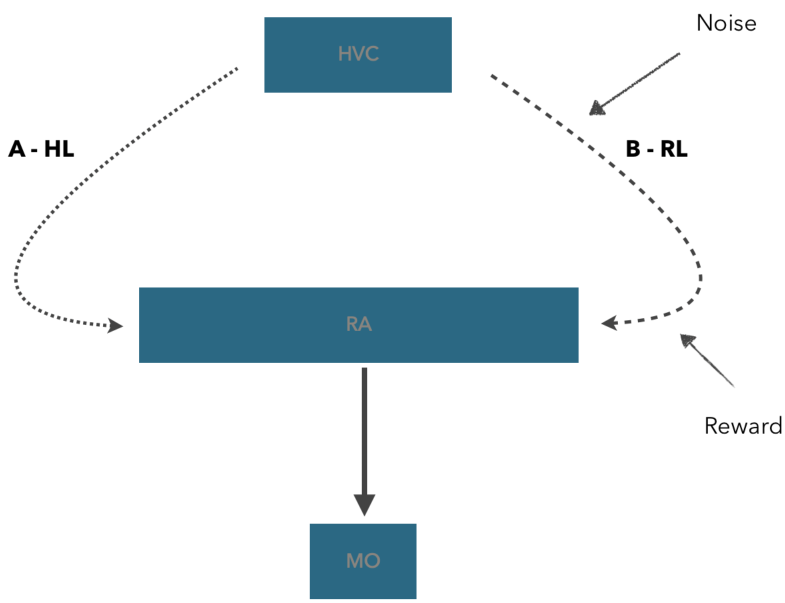


Figure 1: Schematic of model with simple motor pathway (A) and anterior frontal pathway (B). The input to HVC is the syllable encoding. The output of MO is a scalar assigned to the syllable. Pathway A works under HL and pathway B under RL.

- MO mainly takes a weighted sum of the RA output and provides a single output.
- The RA and MO have linear functions to transform the inputs to outputs.

Outcomes:-

With this model,

1. Mimicking AFP was causing the SMP to learn song, and become independent eventually.
2. There was an advantage in RL having a headstart.[Fig 2]

Cons:-

AFP was able to produce song independently, which is biologically inaccurate.

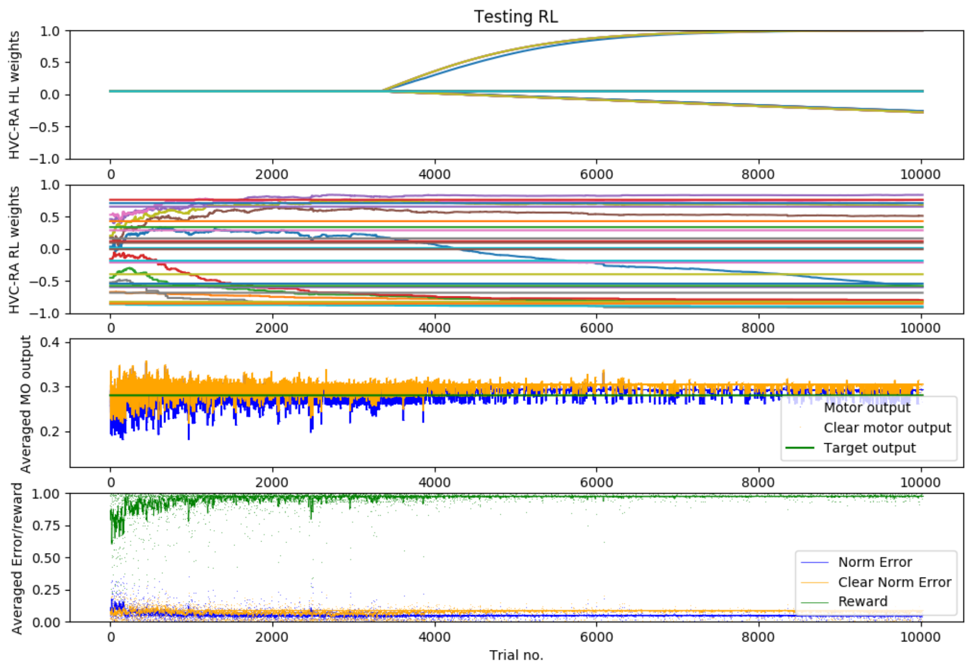


Figure 2: An initial model with SMP (HL) mimicking AFP (RL). SMP successfully learns the song and advantage is seen to the AFP having a headstart. However, the AFP is able to produce song independently, too, which is biologically inaccurate. In the first panel, we can see the SMP weights have a delayed growth. In panel 2, the development of the AFP weights are shown. The third panel shows the output converging to the target. The fourth panel shows the reward increasing over time, as the error decreases.

1.1.1 Modification: Reduced RL influence

To address the aforementioned problem, we tried to limit the RL pathway weights, to allow only local exploration. This led to

Outcomes:-

1. SMP's presence being necessary for song production.
2. AFP was not producing song independently.

Cons:-

No advantage was observed when RL was given a headstart. [Fig 3]

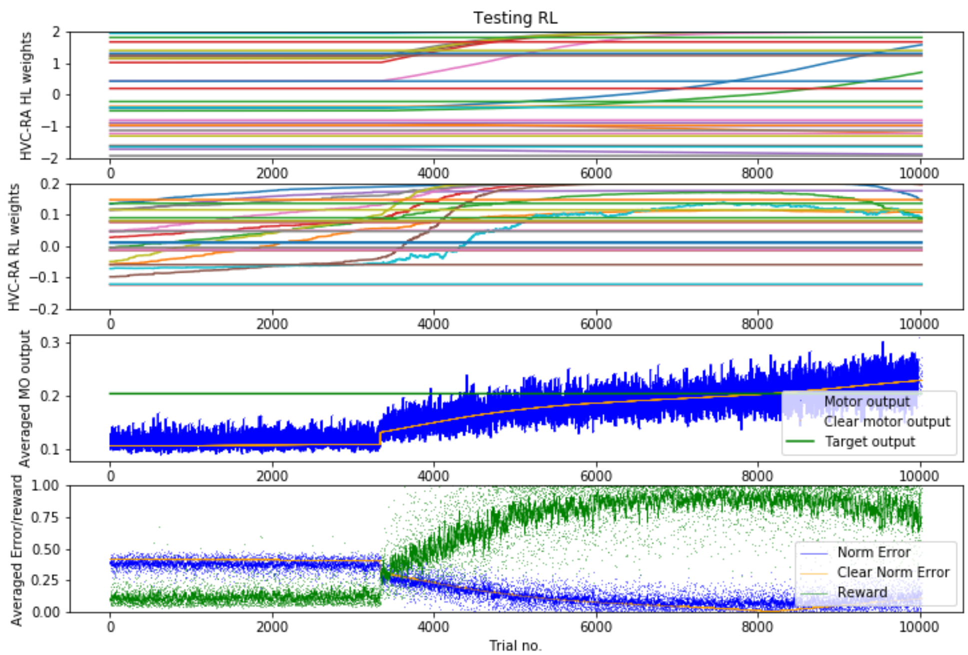


Figure 3: Model with limited AFP exploration i.e. controlled RL influence. While the song was learnt successfully, no advantage was observed when RL pathway had an early onset. In the first panel, we can see the SMP weights have a delayed growth. In panel 2, the development of the AFP weights are shown. The third panel shows the output converging to the target, only when pathway A starts to develop. The fourth panel shows the reward not increasing much when only AFP has influence. Once the SMP starts growing, the reward increases over time, as the error decreases.

1.2 Arm model

To address this issue, we further look into this local motor exploration aiding SMP, by using the arm exploration paradigm.

Task: A multi-segmented arm has the range to move in a circular area (with its total arm length as radius), and is given a target to reach.

Method: We start with random exploration in the motor space. [Fig 4].

- There is no network.
- The angles of each joint (between the arm segments) is shown w.r.t to the horizontal axis.
- A point in the circular area is chosen as the target.
- The two pathways SMP and AFP (only symbolic in this model) produce an

output each. The angles of the arm is configured according to the sum of the outputs of the SMP and AFP.
- The AFP perturbs its output slightly at each trial.
- The new end point of the arm is compared with the target to compute the reward. The AFP updates its output according to a learning rule.
- The SMP updates its output to mimic the angle produced (sum of AFP and SMP outputs).

$$\text{angles}_{\text{t}} = (\text{angles}_{hl} + \text{angles}_{rl} + \xi) \tag{1}$$

$$\text{pos} = \mathit{cumulativeSum}(\text{lengths} * \mathit{cos}(\text{angles}_t), \text{lengths} * \mathit{sin}(\text{angles}_t)) \tag{2}$$

$$E = ||\text{pos} - \text{Target}||/\text{Output\_range} \tag{3}$$

$$R = e^{f(-E)} \tag{4}$$

$$\Delta angle_{hl} = \eta_{hl} * (angles_t - angles_{hl}) \tag{5}$$

$$\Delta angle_{rl} = \eta_{rl} * (R_{curr} - \bar{R}_{prev}) * \xi \tag{6}$$

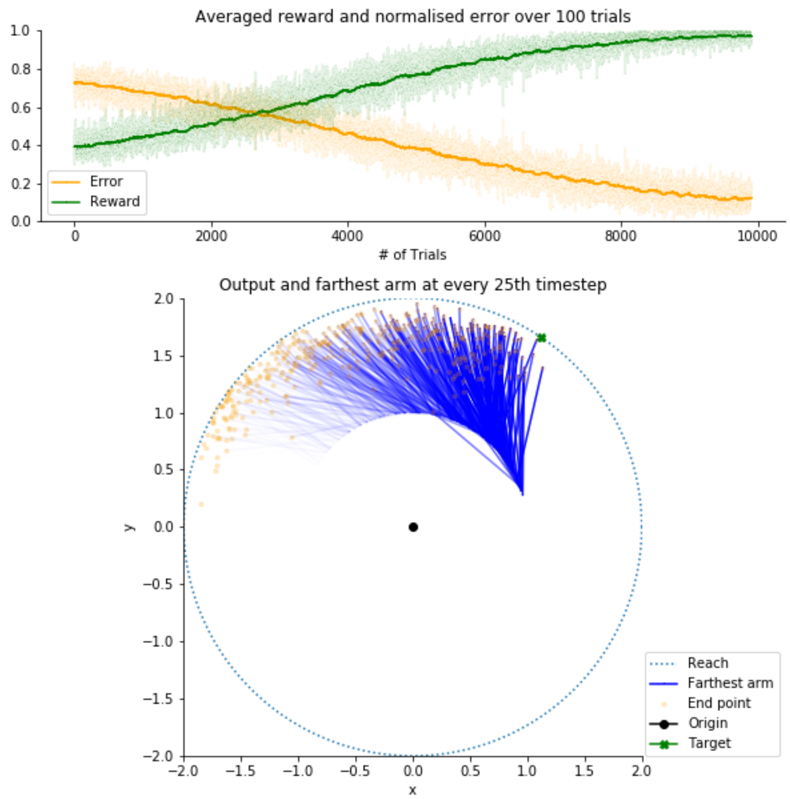
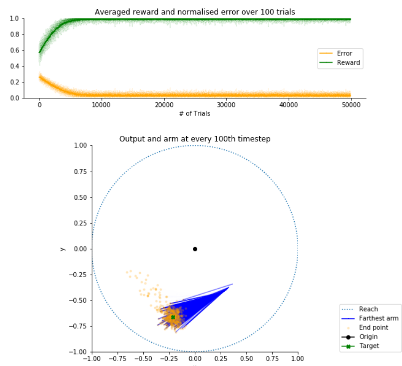
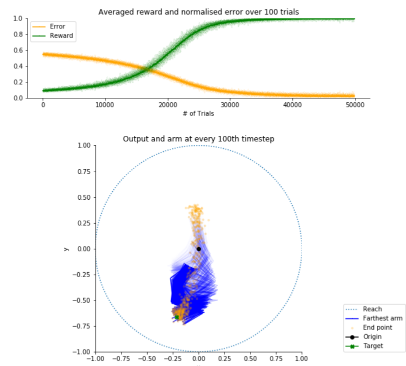


Figure 4: Arm model with random exploration in the motor space. The first panel shows reward increasing over trials. The second panel shows the final segment of the arm (blue) and the endpoint (orange). Over the trials, the arm reaches the target.

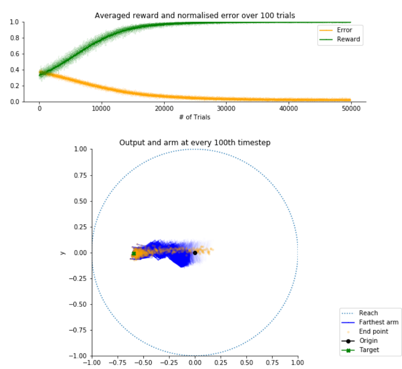
Outcome:- While, this leads to the target being reached eventually, in the case of a few no. of segments (n = 2), this is not the case as the no. of segments increases. This random exploration leads to higher exploration of certain regions of the available range (sensory space), more than others (typically centred around the origin). This is due to the fact that uniform exploration in the motor space, doesn't translate into uniform exploration in the sensory space, as the transformation between them is non-linear. Multiple arm configurations can result in the same endpoint. [Fig 5 and Fig 6.]



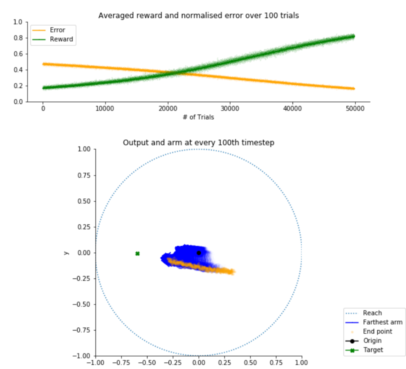
(a) 2 segments



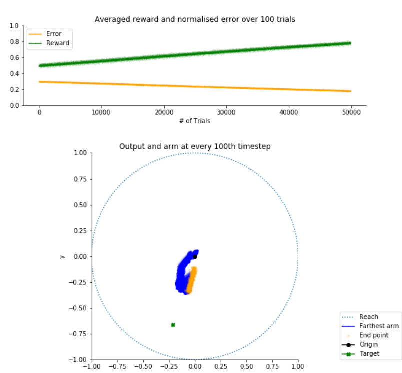
(b) 5 segments



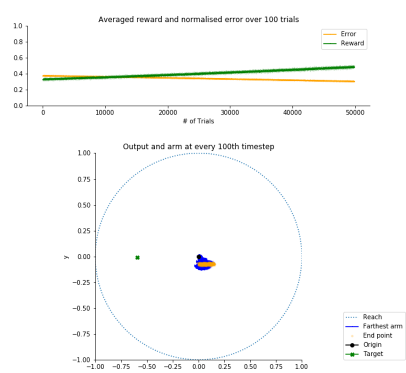
(c) 10 segments



(d) 25 segments



(e) 50 segments



(f) 100 segments

Figure 5: The arm is being trained to reach a target (shown in green). Exploration area reduces with an increase in no. of arm segments.

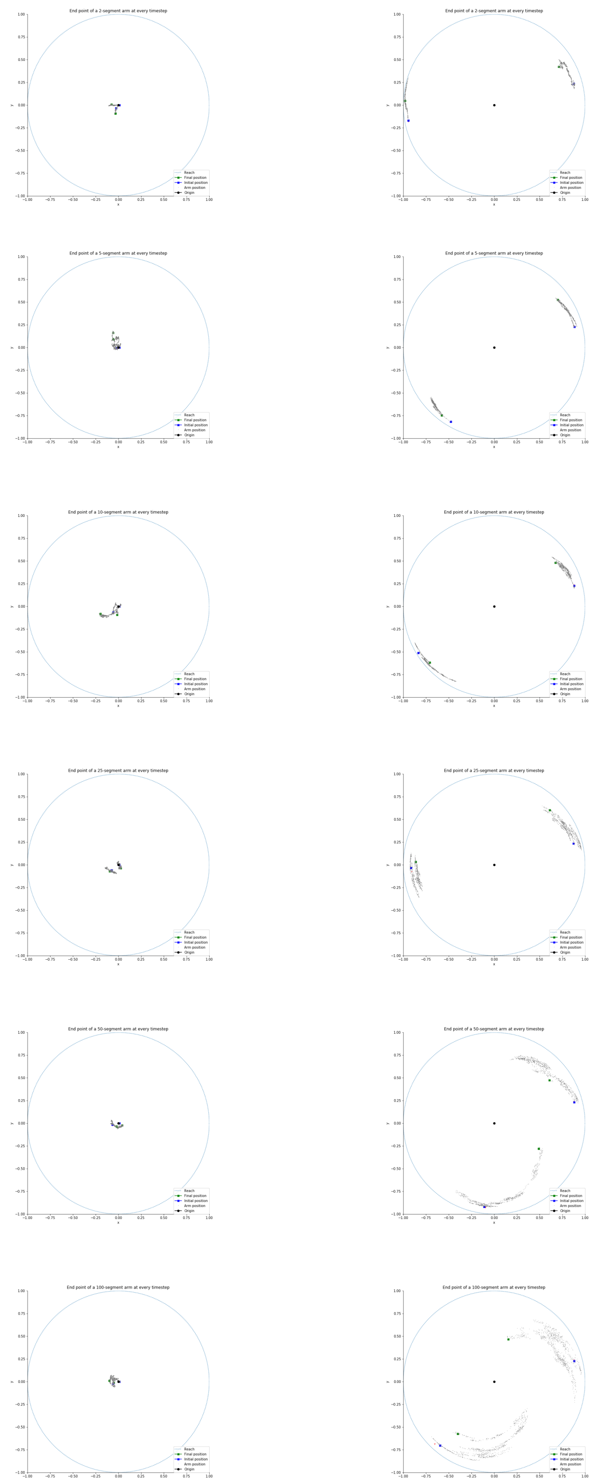


Figure 6: The arm is allowed to explore randomly. With the same uniform noise, exploration area tends to get concentrated towards the centre when no. of arm segments increases. Contrast between an internal point (left) and 2 relatively peripheral points (right) is shown.

2 Model to test now

The constraints introduced by random motor exploration led us to explore alternatives. We plan to test two approaches now.

2.1 Approach II: Activity model

Now, what I get from Arthur's suggestion of having motor exploration based on activity profiles, is written here. We use 2 Gaussians of different amplitudes to represent the output of the AFP and SMP. The weighted sum of these two gaussians would be the angle configuration. The mean for the 2 gaussians can be updated differently, according to RL or HL.

Doubts: I do not know if this approach will work on a high no. of segments.

2.1.1 Input

HVC - syllable encodings; governs which syllable will be produced.

2.1.2 Assumptions

For an abridged version for preliminary trial: 1 syllable (target), and 1 segment in arm. ($\implies$ we won't need an explicit input for the 1 syllable case).

2.1.3 Workings

There is an activity function (Gaussian) for both the AFP and SMP, with an attenuating factor.

$$Act_{RL} = \frac{1}{\rho_{RL}} N(\mu_{RL}, \sigma_{RL}) + noise \tag{7}$$

$$Act_{HL} = \frac{1}{\rho_{HL}} N(\mu_{HL}, \sigma_{HL}) \tag{8}$$

Initialise the standard deviation σ_{RL} high, indicating the higher variability in AFP.
Initialise σ_{HL} low, indicating the specificity of SMP driven motor activity.
The attenuating factor ρ_{HL} decreases over time, indicating the increasing strength of SMP connections.
ρ_{RL} increases over time, indicating the decreasing influence of LMAN.

$$Act_{total} = Act_{RL} + Act_{HL} \tag{9}$$

Take weighted average,

$$\theta_t = \frac{\Sigma_0^{2\pi} \theta . Act_{total}}{\Sigma_0^{2\pi} Act_{total}} \tag{10}$$

The angle (motor command) produced is a weighted average of the sum of the 2 activity functions. The effect is the polar coordinate reached on the plane of the arm.

$$Output = [r \cos \theta_t, r \sin \theta_t]$$

$$R = f(Output)$$

A reward is computed as an exponential function of the normalised error, where the error is the Euclidean distance between the target and the output coordinates. The mean of the activity functions are updated as follows.

$$\Delta\mu_{HL} = \eta_1(\theta_t - \mu_{HL}) \tag{11}$$

Hebbian learning tries to mimic the output of the whole system.

$$\Delta\mu_{RL} = \eta_2(\theta_t - \mu_{RL})(R - R_{prev}) * noise \tag{12}$$

Reinforcement tries to move towards higher reward.

$$\eta 1 << \eta 2$$

$$\mu_{HL} = (\mu_{HL} + \Delta\mu_{HL})\% angle_{max} \tag{13}$$

$$\mu_{RL} = (\mu_{RL} + \Delta\mu_{RL})\% angle_{max} \tag{14}$$

$$\rho_{RL} + +$$

$$\rho_{HL} - -$$

2.1.4 Choices

- For the activity function, we can either have a change in the attenuating factor, or a change in the standard deviation. I have chosen to change the attenuating factor, as (from what I understood in the discussion) want to depict a reduction in the influence of the LMAN.

2.1.5 Issues

- Initially, when there is no SMP influence, there is no difference in the weighted average from one trial to another, as the change in reward is zero, and the produced angle and the RL activity mean are the same. Solution: Use noise appropriately.
- A low ammount of noise could translate into a huge difference in the arm endpoint or vice versa, depending on the no. of segments.

- Learning rule for RL. — Tested, works with external noise on limited trials.

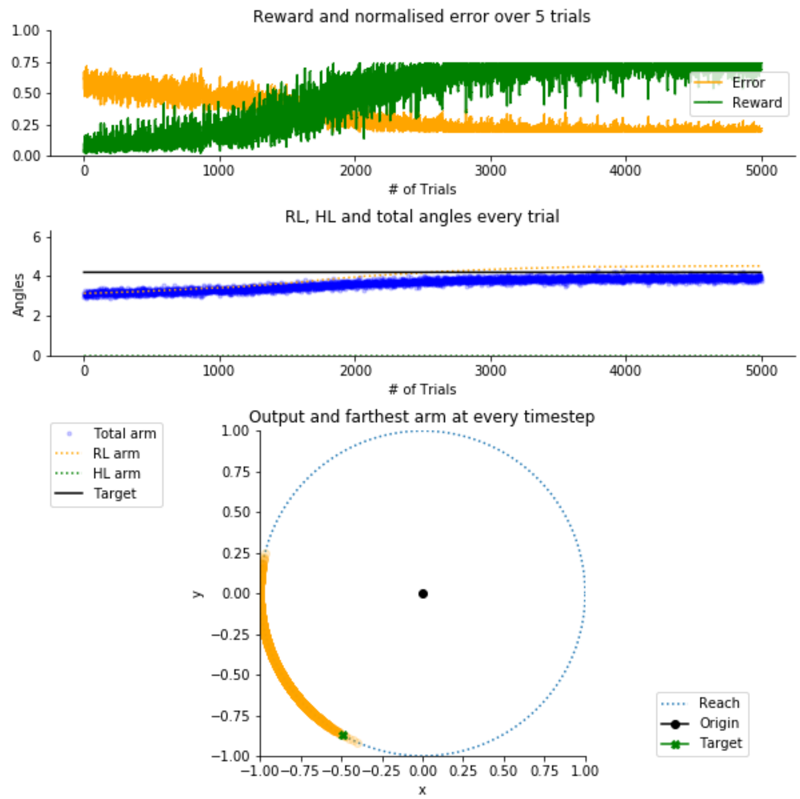


Figure 7: A starting attempt at the activity profile idea with external noise.

2.2 Approach I: Goal babbling

We plan to explore sensory exploration to solve the scalability problem.

Vague idea:
- Explore in a limited local area (RL) using an inverse model. Choose a random position. Use an inverse kinematics tool to find the angle configuration for the intial few positions. Keeping a history, explore in the local region.
- At a slower rate, HL mimics the changes in RL, and shifts the local area of

exploration.
- Attempt using Jacobian to find the direction of change for a given angle configuration, for a given point. Is it possible to make this computation less expensive OR compute approximations for the adjacent areas?

References to check out:
- Tutorial on CMA-ES: –https://arxiv.org/abs/1604.00772
–https://medium.com/@kkstrack/cmaes-algorithm-in-c-using-mlpack-1a233af7a1f7
–https://www.youtube.com/watch?v=aP31Q7o2UGU
- Topology-Conserving Maps for Learning Visuo-Motor-Coordination [ritter1989.pdf]
- Visual servoing of redundant manipulator with Jacobian matrix estimation using self-organizing map [premkumar2010.pdf]
- Inverse kinematic and Jacobian [Jacobian.pdf]
- An inverse kinematics tool: FABRIK https://www.sciencedirect.com/science/article/pii/S1524070311000178F