Oscillatory Timing Models in RL-Automata

Michael Tarlton
Oslo Metropolitan University
michaelt@oslomet.no
https://Poster.Tarlton.info

Summary

Time mechanisms in neuronal assemblies exhibit complexity-rich and multi-dimensional dynamics, on which all neuronal communication is based. By modeling the simplest properties of neuronal spiking communication, we may be able to create emergent properties of timing and learning in deep and entangled assemblies.

***I.e.* small, simplistic time-mechanisms such as Spike Timing Dependent Plasticity (STDP) may compose more complicated modes of timing in an assembly at scale. Assuming biological models are composed of these self-adjusting mechanisms as a means of adapting in changing environments, learning the properties of time may offer new methods of credit-assignment for online and continuous learning models.**

The *Striatal Beat Frequency (SBF) model* [1], is a neuroscientific model for encoding time-separated events in a distributed architecture of sub-circuits. This well-supported model provides an explanation for flexible and distributed encodings of time information at multiple scales. The spike-based communication of this model allows for implementation in SNNs and learning with STDP. Here, we abstract the model's features into an automata framework for continuous and discrete periodicity finding problems.

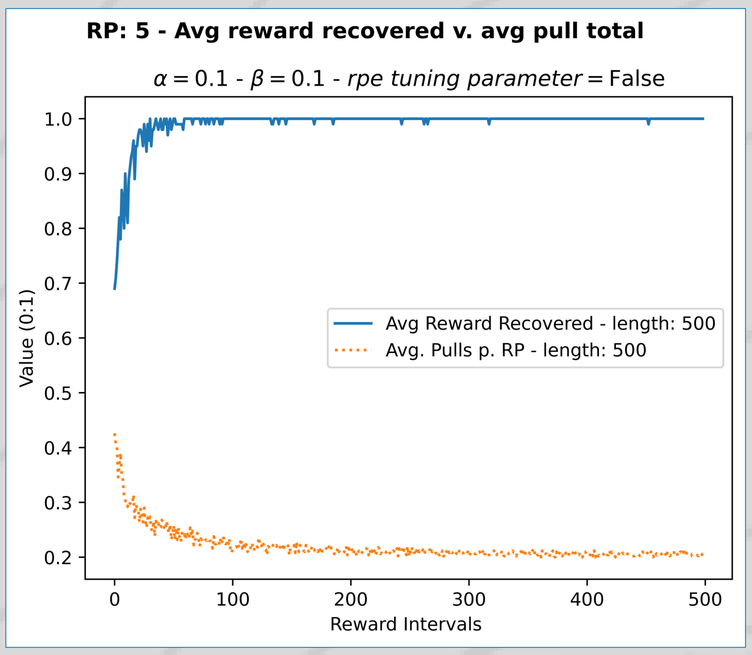
*We adapt the SBF model into a naïve reinforcement learning automata: the **SBF-Automata (SBF-A)**, and study the automata's ability to learn and reproduce time intervals and for static and changing environments.*

Why?

A key goal of the SBF model is to build an understanding of how neural circuits are able to encode timing information in the suprasecond to minutes range from the microsecond activity of neurons. Many previous neurological time models rely on some dedicated "clock" mechanism as well as "cold storage" memory to hold the relevant time information. This schema reflects the Von Neumann style architecture found in computing, and poorly reflect the atomically distributed and multiplexed nature of information in the brain.

In contrast, the SBF model uses the intrinsic firing patterns of neural assemblies as the mechanism for time interval detection, as well as a opening for the resultant time information to be encoded within the state of the network [2]. This proves usable in an STDP model [3], and in the plastic regime could allow for compact, robust, and scalable properties consistent with those found in time cells [4].

Does it Work?



Here we show proof of concept, in which an oscillator period equal to the reward period is present. Because SBF-A acts as an analog fourier transform, the accuracy increases with number of oscillators used [7]. With discrete tonic oscillators, accurate performance requires thousands of oscillators, but with phasic activity oscillators we can obtain adequate performance with less than a hundred.

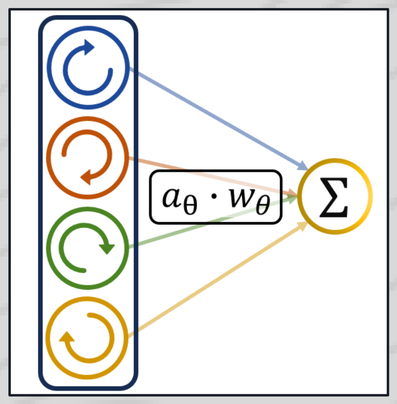
The SBF-A Model

For an environment in which a reward is made available with some periodicity t_{RP}. The automata consists of a block oscillatory units, each of which activate with a unique periodic cycle. These oscillatory nodes inform the executive node, which checks the environment for reward with a probability based on the weighted periodic input from the oscillatory nodes.

When the automata acts on the environment, the weights are updated depending on if a reward was discovered. Where the nodes that were active at the time of the decision, are increased or decreased in weight, and the opposite is done on the inactive nodes.

Effectively, this acts as an analog Fourier Transform, where the frequency of target time period is encoded in the distributed weights of compnent frequencies.

Figure 1 – Oscillator Block and Executive Unit



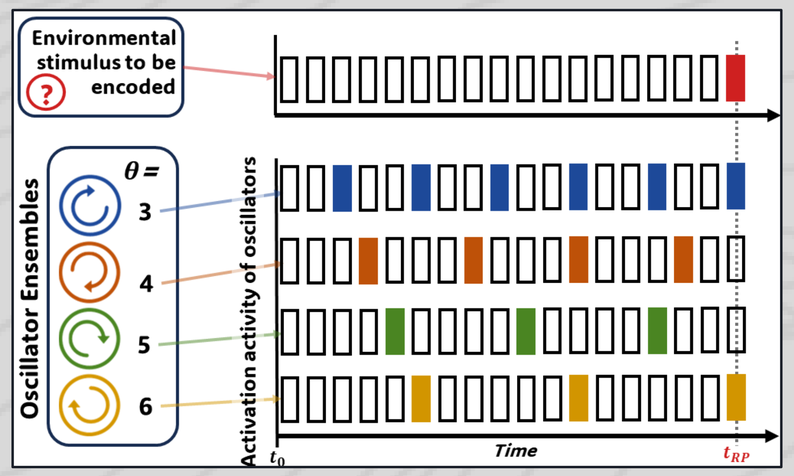


Figure 2 – Example for Discrete Activity

Algorithm 1 – Normalized Weighted Vote

(A) **No Punishment** α : Learning control parameter for rewarded action. **for** $t = 0, T$ **do**
  Check "active" agents with check cycles in phase with current time step, where $t \ mod \ tc = 0$ Take sum of weights of active agents, $w_t = \sum w_a$ Take uniformly random probability P of polling environment for reward
  if $P > w_t$ **then**
    Do nothing
  end
  if $P \leq w_t$ **then**
    if $RP \ mod \ t = 0$ **then timestep is on a reward interval**: **Decrease weights of inactive agents**:

$$w_i = w_i * (1 - \alpha)$$

    Increase weights of active agents:

$$w_a = w_a + \alpha \frac{\sum w_i}{n_a}$$

    ;
    else timestep is **not** on a reward interval;
    Do nothing
  end
  Move to next timestep

Our basic algorithm is a simple normalized weighted vote that occurs at every time time-step.

We extend this with update rules for punishment, when the automata acted but no reward was available, the weight is redistributed from active units to inactive units.

We add furhter extensions to our algorithm including Reward Prediction Error (RPE) based techniques to further extend the model's preformance in changing environments.

Receptive Fields & Temporal Rescaling

- The naïve discrete automata method is severely limited in ability as the collective node contribution at each timestep is dependent on initial phase distributions and their specific effectivity to identifying the reward interval.
- We expanded our model to reflect the phasic activity of time cells, giving each oscillator a phasic activation curve.
- Biological time cells are able to temporally rescale by tuning the mean peak and variance of their phase in response to dopaminergic feedback while training to a specific interval or towards a broader tuning in the face of complex environmental input and reaction [6].
- Precession of an oscillator's phase tuning allows entrainment to slower frequency oscillations which reflect changes at a high-frequency timescale with respect to other encoded time durations, allowing for multiplexed and multiscale time encodings [2].

Oscillatory Phase Distributions

- It is important to find a distribution of oscillator phases which can maximally recover reward, with respect to phase of the reward cycle.
- Our initial approach was to use prime numbered phases in order to avoid overlap in activity domains.
- $1/f$ noise (a power law distribution) can be found in oscillatory distributions throughout the brain [5] including in circuits with proposed involvement in the SBF model [1].
- Additionally, we found the need for an "inhibitory node" which votes for "no-action" at every timestep, thus absorbing surplus activity in the network.
  - The lossless weight redistribution of the algorithm may force the automata to perform unnecessary actions. The addition of the no-action oscillator effectivity acts as a broad inhibitory signal, optimizing the energy efficiency of the automata.

References

[1] Gu, Bon-Mi, Hedderik van Rijn, and Warren H. Meck. "Oscillatory Multiplexing of Neural Population Codes for Interval Timing and Working Memory." *Neuroscience and Biobehavioral Reviews* 48 (January 2015): 160–85.
[2] Yin, Bin, Zhuanghua Shi, Yaxin Wang, and Warren H. Meck. "Oscillation/Coincidence-Detection Models of Reward-Related Timing in Corticostriatal Circuits." *Timing & Time Perception* 1, no. aop (July 18, 2022): 1–43.
[3] Xu, Wei, and Stuart N. Baker. "Timing Intervals Using Population Synchrony and Spike Timing Dependent Plasticity." *Frontiers in Computational Neuroscience* 10 (2016).
[4] Mello, Gustavo B. M., Sofia Soares, and Joseph J. Paton. "A Scalable Population Code for Time in the Striatum." *Current Biology* 25, no. 9 (May 4, 2015): 1113–22.
[5] Buzsáki, György, and Mihály Vöröslakos. "Brain Rhythms Have Come of Age." *Neuron* 111, no. 7 (April 5, 2023): 922–26.
[6] Petter, Elijah A., Samuel J. Gershman, and Warren H. Meck. "Integrating Models of Interval Timing and Reinforcement Learning." *Trends in Cognitive Sciences*, Special Issue: Time in the Brain, 22, no. 10 (October 1, 2018): 911–22.
[7] Oprisan, Sorinel A., Dereck Novo, Mona Buhusi, and Catalin V. Buhusi. "Resource Allocation in the Noise-Free Striatal Beat Frequency Model of Interval Timing." Timing & Time Perception 11, no. 1–4 (July 21, 2022): 103–23.

Publications

Extended Abstract 447 - Neurological Based Timing Mechanism for Reinforcement Learning
Further Publication TBA.
Follow on twitter or message for future updates!