Learning sequence disambiguation with synaptic traces in associative neural networks

Ramón Martínez[1], Anders Lansner[1, 2], Pawel Herman[1]

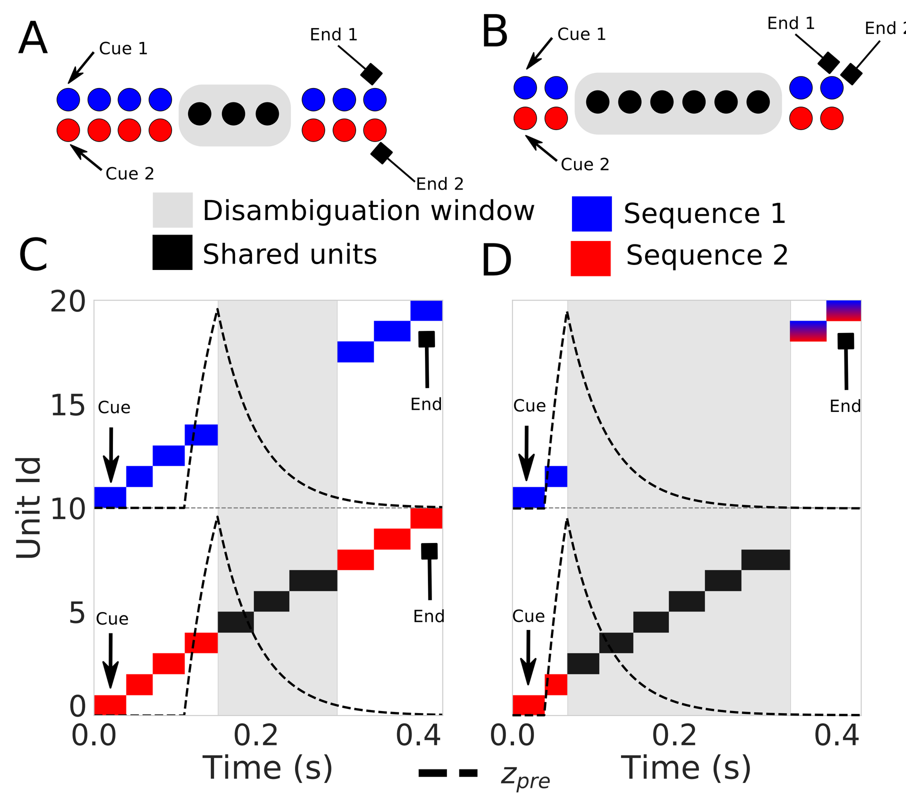
1) KTH, Royal Institute of Technology 2) Stockholm University, Mathematics and KTH Royal Institute of Technology

Sequence Learning

Sequence disambiguation or using past context to determine the trajectory of a sequence has been deemed one of themost important problems that a sequence prediction network should solve [1]. There have been a few attempts at the problem of sequence disambiguation in the attractor network framework but most of them rely on non-local learning rules or require an unfeasible large number of parameter. We present here a sequence learning system that works with probabilistic associative learning and is able to accomplish sequence disambiguation by using dynamical information in the form of synaptic traces.

Sequence disambiguation

For testing the system disambiguation capabilities we used a framework were we varied the length of a disambiguation window that the system has to overcome.



(A, B) Schema of a short and long disambiguation windows respectively. (C, D) The synaptic trace preserves the information for long enough to allow disambiguation.

References

[1] Levy, W. B 579 - 590. Hippocampus (1996)

[2] Tully, Philip J., Henrik Lindén, Matthias H. Hennig, and Anders Lansner. e1004954. *PLoS Comput Biol 12, no. 5 (2016)*

[3] Lansner, Lanserng. and Ekeberg Örjan 1(01) 77 - 87. *International journal of neural systems (1989)*

[4] Martinez, Ramon Heberto, Pawel Herman, and Anders Lansner 545871 *bioRxiv (2019)*

The Model

Based on previous work [2, 4] we present network whose dynamical evolution is controlled by the equations below. The model contains a current s for each unit that evolves according to the their interaction mediated by weights $\mathbf{W}$ and a bias term $\vec{\beta}$. Furthermore, to induce an structured sequential transition the model is subjected to intrinsic adaptation $\vec{\mathbf{a}}$ and a mechanism of winner-takes-all given by a strict max in $\vec{\mathbf{o}}$.

As a learning rule we use the Bayesian Confidence Propagator Neural Netowkr (BCPNN) [3]. The nature of the BCPNN learning rule is such that it connects in an excitatory fashion patterns that more often that not appear together (in a probabilistic sense) and connects in an inhibitory fashion patterns that do not. More importantly for sequence disambiguation the BCPNN automatically accounts for balanced connections in sequential forks.

$$\tau_s \frac{d\vec{s}}{dt} = \vec{\beta} + \mathbf{W} \cdot \vec{\mathbf{z}}_{pre} - g_a \vec{\mathbf{a}} - \vec{\mathbf{s}} + \sigma d\vec{\xi}(t)$$

$$\tau_a \frac{d\vec{\mathbf{a}}}{dt} = \vec{\mathbf{o}} - \vec{\mathbf{a}}$$

$$o_i = \begin{cases} 1, & s_i = \max\limits_{hypercolumn}(\vec{\mathbf{s}}), \\ 0, & \text{otherwise} \end{cases}$$

$$\tau_{z_{pre}} \frac{d\vec{\mathbf{z}}_{pre}}{dt} = \vec{\mathbf{o}} - \vec{\mathbf{z}}_{pre}$$

$$\tau_{z_{post}} \frac{d\vec{\mathbf{z}}_{post}}{dt} = \vec{\mathbf{o}} - \vec{\mathbf{z}}_{post}$$

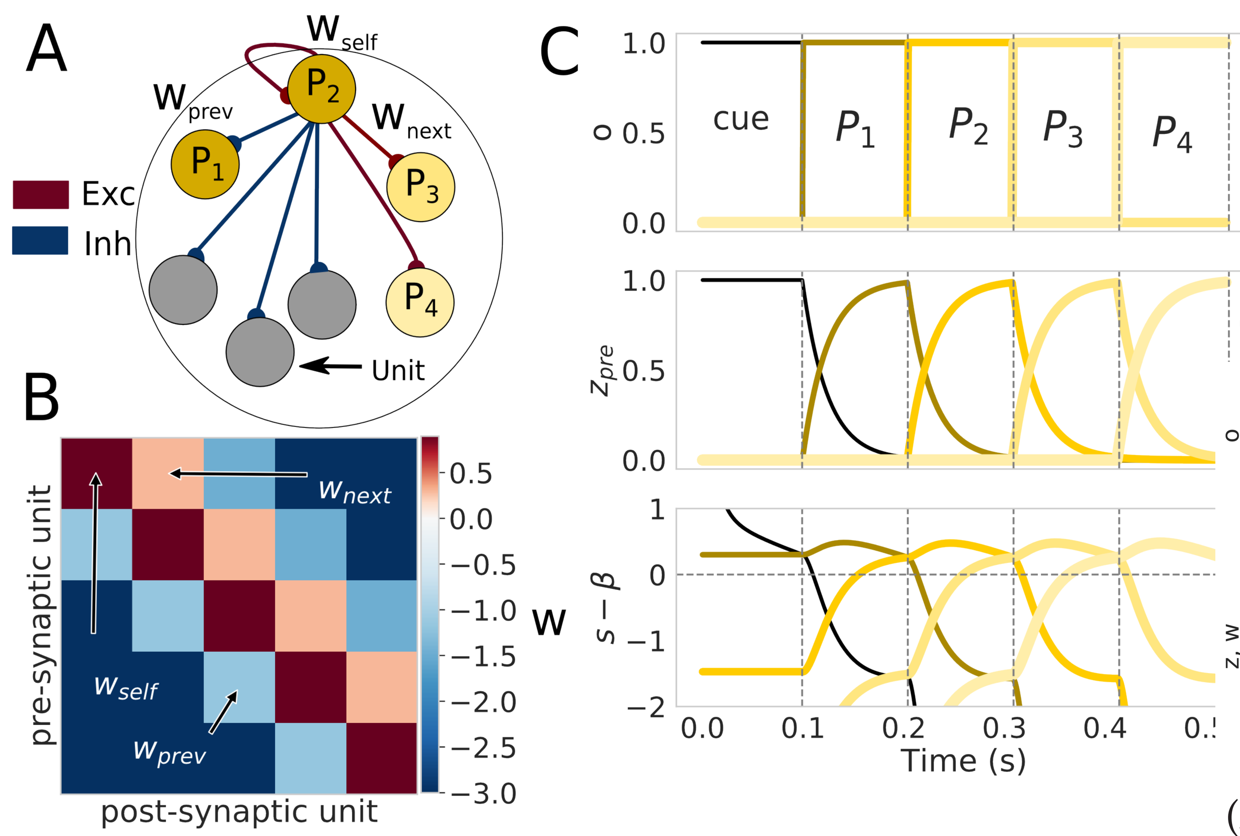
$$t \frac{d\vec{\mathbf{p}}_{pre}}{dt} = \vec{\mathbf{z}}_{pre} - \vec{\mathbf{p}}_{pre}$$

$$t \frac{d\mathbf{P}}{dt} = \vec{\mathbf{p}}_{pre} \otimes \vec{\mathbf{p}}_{post} - \mathbf{P}$$

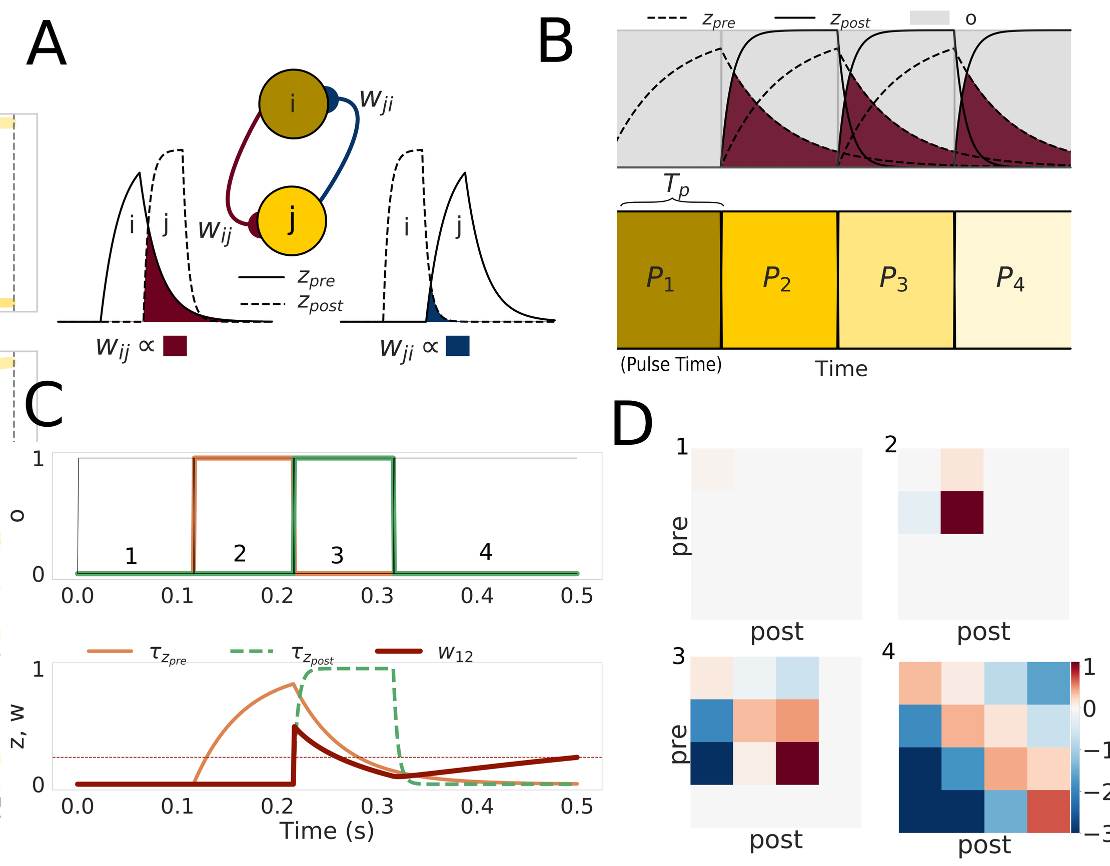
$$t \frac{d\vec{\mathbf{p}}_{post}}{dt} = \vec{\mathbf{z}}_{post} - \vec{\mathbf{p}}_{post}$$

$$\mathbf{W} = \log \left(\frac{\mathbf{P}}{\vec{\mathbf{p}}_{pre} \otimes \vec{\mathbf{p}}_{post}} \right)$$

$$\vec{\beta} = \log \left(\vec{\mathbf{p}}_{post} \right)$$



(A) The network. (B) Typical structure of the weight matrix. (C) Recall illustrated with the activity of o, z and the current respectively.



(A) Illustration of the learning rule. (B) The training protocol. (C) Weight evolution responds to coincidences in time. (D) Evolution of the weight matrix as patterns are presented to the network.

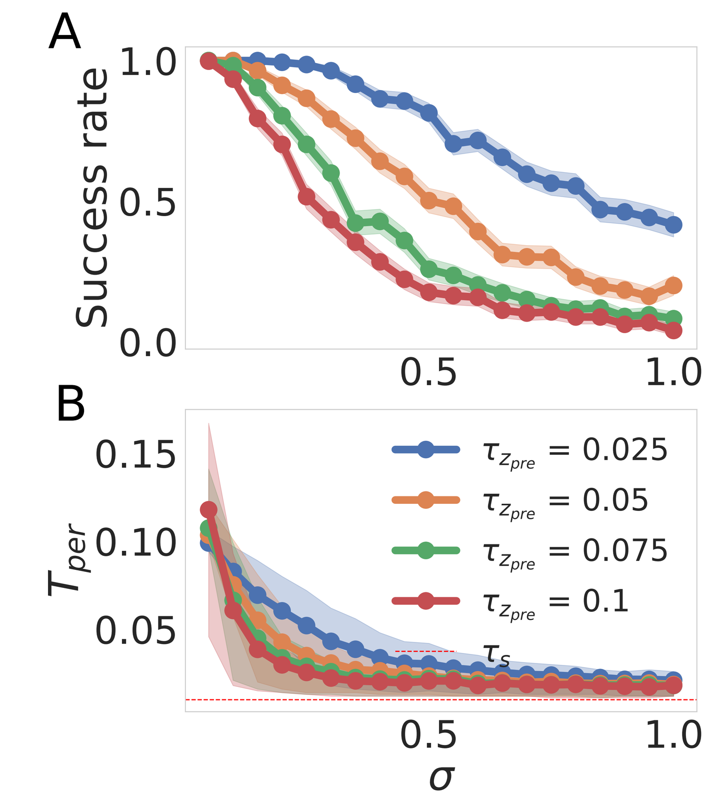
Funding

This work was supported by the Erasmus Mundus Joint Doctoral Program Eurospin.

Education and Culture DG

ERASMUS MUNDUS

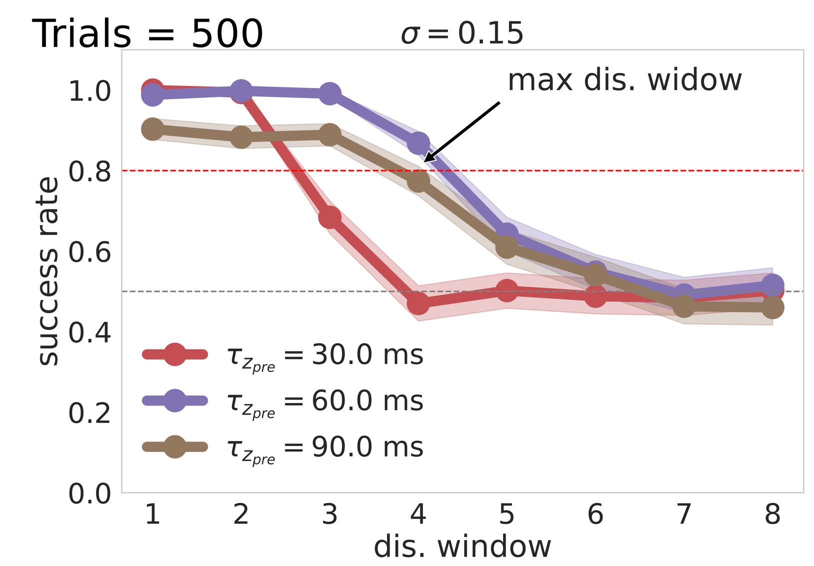
Effects of noise



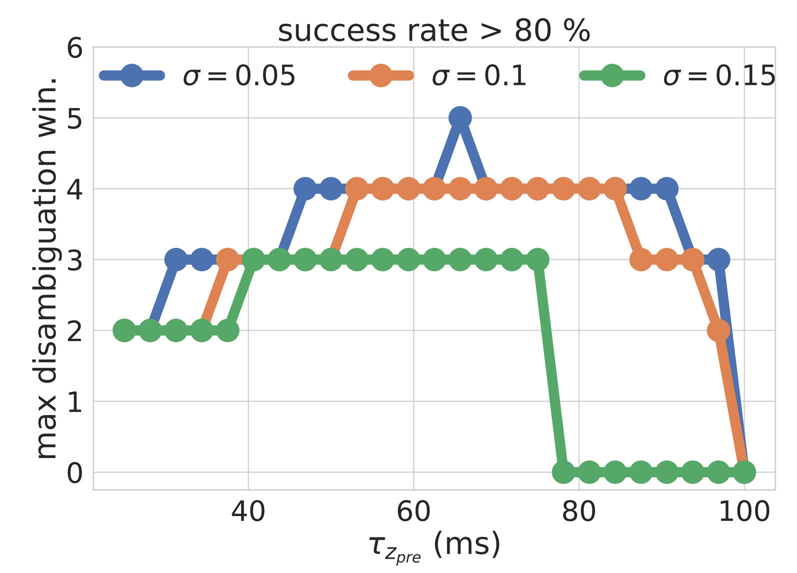
(A) Larger values of the time constant of the synaptic trace ($\tau_{z_{pre}}$) make the system more brittle. (B) The time an attractor remains activated decays in a systematic way in a noisier system.

Results

We tested the disambiguation power of the system for different arrangements of the synaptic trace time constant ($\tau_{z_{pre}}$), disambiguation windows and three noise regimes: small ($\sigma = 0.05$), medium ($\sigma = 0.1$) and large ($\sigma = 0.15$).



Success rate as a function of the disambiguation window length for different values of $\tau_{z_{pre}}$.



Max disambiguation window depending as a function of $\tau_{z_{pre}}$ in different noise regimes.