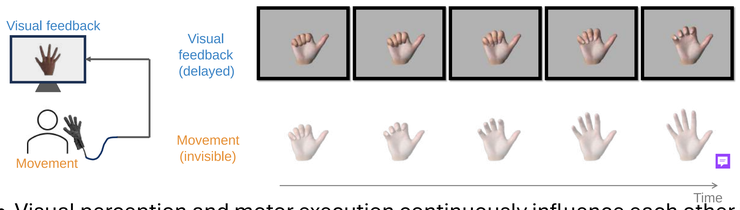
Modeling Action-Perception Coupling with Reciprocally Connected Neural Fields

Xinrui Jiang, Martin A. Giese
CIN & HIH, Department N3, University Clinic Tübingen, Germany

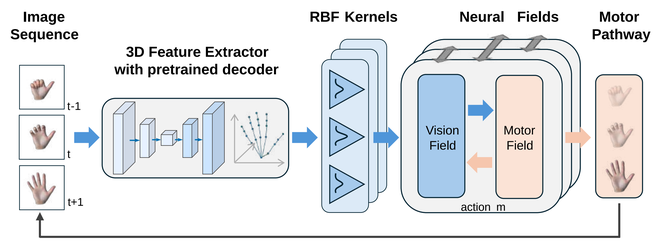
Introduction



- Visual perception and motor execution continuously influence each other in a dynamic, reciprocal manner.
- We proposed a neural dynamic model based on coupled recurrent neural networks, adapted from previous work [1].
- Using real video sequences as input, we reproduced multiple findings from psychophysical experiments that probe the interaction between perception and execution of actions.

Model Overview

A closed-loop system was simulated which takes image sequences of a real hand or avatar [2] as input and generates complete kinematics through a motor pathway to control a hand avatar.



Visual feature extraction: A deep neural network model [3] is used to estimates 3D joint angels and corresponding angular velocities of the hand.

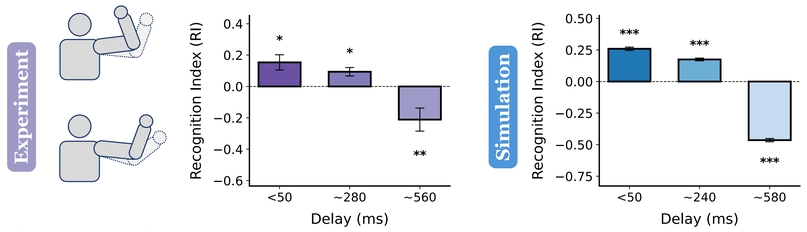
Neural input of posetures: Features are pass through Radial Basis Function (RBF) network, which is trained with key poses of different hand actions.

Dynamic neural representations (neural fields): Two sets of neural fields represent recognized movements (vision field) and executed movements (motor field). Evolving movement corresponds to a travelling activation pulse in these fields. Different movements are represented by different fields.

Coupling dynamics: Fields representing different actions inhibit each other. Vision and motor fields representing the same action are coupled reciprocally through special coupling kernels (see box).

Motor readout: The location of peak activity in the motor field is mapped onto the kinematics of an output pose using a nonlinear regression. Rendering the pose with avatar can produce input movies for the model, closing the loop.

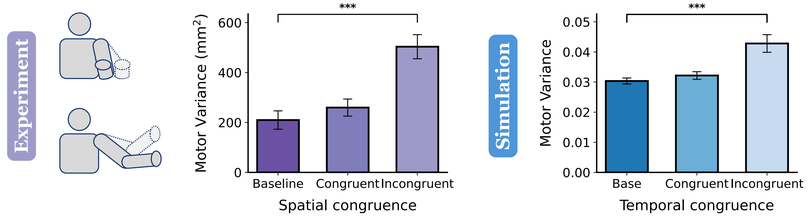
Variation of Temporal Congruence



Experiment: Participants observed noisy point-light displays of a waving arm that was either synchronous or delayed relative to their own arm movement [4].

Simulation: Noise was added to the visual neural field input. Recognition performance was assessed by the signal-to-noise ratio in the visual fields.

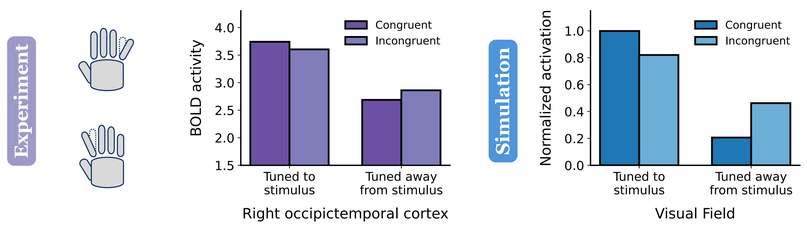
Variation of Spatial Congruence



Experiment: Observing arm movements that were *spatially incongruent* with own movement increased the variability of participants' movements [5].

Simulation: *Posture incongruence was simulated approximately.* Motor variance was measured by the variability in peak activation locations in the motor neural field.

Variation of Action Congruence



Experiment: Participants observed congruent or incongruent finger movements (i.e., index vs. little finger), while brain activity was measured with 3T fMRI in right occipitotemporal cortex [6].

Simulation: Model activity was analyzed by comparing peak activations in visual fields tuned to the observed action versus those tuned to different actions, mimicking BOLD responses for visual representations.

Model Details

Neural fields are defined over one-dimensional periodic spaces $x, y \in (-\pi, \pi]$. Activity in vision field $u_m(x,t)$ and motor field $v_m(y,t)$ follow the dynamics:

$$\tau \frac{\partial u^m(x,t)}{\partial t} = -u^m(x,t) - h + w_{uu}(x) *_x F(u^m(x,t)) + s_u^m(x,t) + w_{uv}(x,y) *_y F(v^m(y,t)) + I_m$$

$$\tau \frac{\partial v^m(y,t)}{\partial t} = -v^m(y,t) - h + w_{vv}(y) *_y F(v^m(y,t)) + s_v^m(y,t) + w_{vu}(x,y) *_x F(u^m(x,t)) + I_m$$

Where:

- m : index of encoded action
- $s(x,t)$: Action-dependent intrinsic inputs
- I_m: Inhibition from neural fields encoding other actions
- $*$: Convolution operation $f(x) *_x g(x) = \int_{-\pi}^{\pi} f(x')g(x-x')dx'$
- $F(\cdot)$: Activation function

The connection kernels are defined as (for $i, j = u, v$):

$$w_{ii}(x,x') = -A_{ii} + B_{ii} \exp\left(-\frac{(x-x')^2}{2\sigma_{ii}^2}\right) - \gamma_{ii} \frac{B_{ii}(x-x')}{\sigma_{ii}^2} \exp\left(-\frac{(x-x')^2}{2\sigma_{ii}^2}\right)$$

$$w_{ij}(x,y) = -C_{ij} + D_{ij} \exp\left(-\frac{(x-y)^2}{2\sigma_{ij}^2}\right)$$

Asymmetric lateral connectivity supports traveling pulse solution. Reciprocal interactions support synchronously traveling peaks in both fields.

Discussions

- Model accounts for a range of distinct effects within a unified framework.
- This model has a relatively simple mathematical structure, which enables a clearer understanding of the underlying dynamics—unlike end-to-end trained recurrent neural networks.
- The strict separation of visual and motor representations assumed by the model may be artificial; in the cortex, a more gradual transition along the visuomotor hierarchy likely exists.

References

1. Hovaidi-Ardestani, M. et al. (2017). Neurodynamical model for the coupling of action perception and execution. ICANN, 19–26.
2. Romero, J. et al. (2017). Embodied hands: Modeling and capturing hands and bodies together. SIGGRAPH Asia, 36(6), 245:1–245:17.
3. Chen, X. et al. (2022). MobRecon: Mobile-friendly hand mesh reconstruction from monocular image. CVPR, 20783–20792.
4. Christensen, A. et al. (2011). Spatiotemporal tuning of the facilitation of biological motion perception by concurrent motor execution. Journal of Neuroscience, 31, 3493-3499.
5. Kilner, J. M. et al. (2003). An interference effect of observed biological movement on action. Current Biology, 13, 522-525.
6. Yon, D. et al. (2018). Action sharpens sensory representations of expected outcomes. Nature Communications, 9, 4288.

Acknowledgements: The work was funded by ERC 2019-SyG-RELEVANCE-856495. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Xinrui Jiang.