

Neurodynamical Model for the Coupling of Action Perception and Execution

Mohammad Hovaidi-Ardestani^{1,2(✉)}, Vittorio Caggiano³, and Martin Giese¹

¹ Section of Computational Sensomotrics, Department of Cognitive Neurology, CIN and HIH, University Clinic Tübingen, Ottfried-Müller-Str. 25, 72076 Tübingen, Germany

Mohammad.Hovaidi-Ardestani@uni-tuebingen.de

² IMPRS for Cognitive and Systems Neuroscience, Tübingen, Germany

³ Computational Biology Center, IBM T.J. Watson Research Center, 1101 Kitchawan Road, Route 134, Room 30-048, 10598 Yorktown Heights, USA

Abstract. In cortical representations action perception and action execution are closely linked, as indicated by the presence of mirror neurons. Experiments show that concurrent action execution and action perception influence each other. We have developed a physiologically-inspired neural model that accounts for the neural encoding of perceived actions and motor plans, and their interactions. The core of the model is a set of coupled neural fields that represent either perceived actions or motor programs. We demonstrate that this model reproduces the results of a variety of quite different experiments investigating the interaction between action perception and execution. It also predicts the emergence and stability of synchronized coordinated behavior of two individuals that observe each other during action execution.

Keywords: Action perception · Motor program · Neural field · Recurrent neural network · Mirror neurons

1 Introduction

Perceptual and motor representations of actions are tightly coupled (e.g. [1]). This is supported by many results from behavioral and functional imaging studies, and physiologically by the existence of mirror neurons, e.g. in premotor and parietal cortex [2,3]. Behavioral and functional imaging studies show influences of motor execution on simultaneous action perception as well as influences in the opposite direction (e.g. [4–6]). Physiological data provides insights in the basis of the encoding of actions at the single-cell level [2,7,8]. This has motivated the development of neural models that account for action perception (e.g. [9,10]) as well as for the neural encoding of motor programs (e.g. [11]). Multiple conceptual models have been proposed that discuss the interaction between action perception and execution (e.g. [12–14]). Some implemented models have been proposed for these interactions in the context of robot systems (e.g. [15]).

We describe here a model that is based on electrophysiologically plausible mechanisms. It combines mechanisms from previous models that accounted separately for electrophysiological results from action recognition and the neural encoding of motor programs [9, 16, 17]. We demonstrate that our model provides a unifying account for multiple experiments on the interaction between action execution and action perception. The model might thus provide a starting point for the detailed quantitative investigation how motor plans interact with perceptual action representations at the level of single-cell mechanisms.

2 Model Architecture

The architecture of our model is illustrated in Fig. 1. The core of the model is a set of dynamically coupled neural fields that encode visually perceived actions and motor programs (Fig. 1B). Each encoded action is represented by a pair of neural fields, a *motor field* encoding the associated motor program, and a *vision field* that represents the visually perceived action. Within these fields the evolving action is represented by a stable traveling pulse solution that runs along the field. The different fields are dynamically coupled in a way that enforces a synchronization of the traveling peaks between the vision and motor field that encode the same action. Fields encoding different actions inhibit each other. The vision fields receive a feed-forward input from a visual pathway that recognizes shapes from gray-level images (Fig. 1A). The motor fields are read out by a neural network that models the motor pathway and produces joint angle trajectories that correspond to the evolving action. These angles are used to animate an avatar, which is rendered to produce an image sequence or movie that shows the action (C). The architecture thus models motor execution as well as action recognition. The following sections describe the individual components of the model in further detail.

2.1 Neural Vision and Motor Fields

The model assumes that individual actions can be encoded as visual patterns, or as motor program. Neurally, the patterns are encoded as stable traveling pulse solutions in dynamic neural fields. For the simulations in this paper these fields are defined over periodic spaces ($x, y \in [-\pi, \pi]$). We assume the encoding of M different actions (where M was 2 for the simulations). The vision field that encodes the precept of action m (assuming $1 \leq m \leq M$) is driven by an input signal distribution $s^m(x, t)$, which is produced by the output neurons of the visual pathway that are tuned for body postures of the action pattern m . The temporal evolution of the activation $u^m(x, t)$ of this visual field is determined by the neural field equation [18]:

$$\frac{\tau \partial u^m(x, t)}{\partial t} = -u^m(x, t) - h + w_u(x) * F(u^m(x, t)) + s^m(x, t) + c_u^m(x, t) \quad (1)$$

with the nonlinear saturating threshold function $F(u) = d_0 (1 - \exp(-u^2/2d_1))$ for $u > 0$, and $F(u) = 0$ otherwise, and $h > 0$ determining the resting level

activity. As interaction kernel we chose the asymmetric function: $w_u(x) = -a_0 + a_1 \left(\frac{1 + \cos(x - a_3)}{2} \right)^\gamma$ with $\gamma > 0$. The convolution operator is defined by $f(x) * g(x) = \int_{-\pi}^{\pi} f(x')g(x - x')dx'$. With this kernel for appropriate choice of the parameters, a traveling-pulse input signal $s^m(x, t)$ induces a traveling pulse equilibrium solution that moves synchronously with the input. This solution breaks down if the frames of the input movies appear in inverse or random temporal order [9]. The term $c_u^m(x, t)$ summarizes the inputs from the other fields and is further specified below.

The corresponding motor program is encoded by another neural field without feed forward input. It is defined by the equation:

$$\frac{\tau \partial v^m(y, t)}{\partial t} = -v^m(y, t) - h + w_v(y) * F(v^m(y, t)) + c_v^m(y, t). \quad (2)$$

The form of the interaction kernel w_v is identical to the one of w_u with slightly different parameters, resulting in stronger recurrent feedback. As consequence, once a local activation is established by a ‘go signal’ a self-stabilizing traveling peak solution emerges that propagates with constant speed along the y -dimension [19]. We associate the values of y with the body poses (joint angles) that emerge during the action, so that the traveling pulse encodes the temporal evolution of a motor program. The term $c_v^m(x, t)$ again specifies inputs from the other fields.

2.2 Coupling Structure

The cross connections between the vision and motor fields encoding the same actions were defined by the kernel function:

$$w_{uv}(x, y) = -b_0 + b_1 \left(\frac{1 + \cos(x - y)}{2} \right)^\gamma = w_{vu}(y, x). \quad (3)$$

This kernel results in a tendency of the activation peaks in both fields to propagate synchronously. The fields encoding different actions are coupled by the cross-inhibition kernel $w_I(x, y) = -c_0$ with $c_0 > 0$. As consequence the different encoded actions compete in the neural representation. Summarizing, the corresponding interaction terms in Eqs. 1 and 2 are given by the relationships

$$\begin{aligned} c_u^m(x, t) &= w_{uv}(x, y) *_y F(v^m(y, t)) + \sum_{m' \neq m} w_I(x, y) *_y (F(u^{m'}(y, t)) + F(v^{m'}(y, t))) \\ c_v^m(x, t) &= w_{vu}(x, y) *_y F(u^m(y, t)) + \sum_{m' \neq m} w_I(x, y) *_y (F(u^{m'}(y, t)) + F(v^{m'}(y, t))) \end{aligned}$$

where the operator $*_y$ indicates the convolution with respect to the variable y .

2.3 Vision and Motor Pathway

The input module of our model is given by a vision pathway that recognizes shapes from image sequences (Fig. 1A). This module is taken over from a previous model without motor pathway (see [9] for details). In brief, the vision

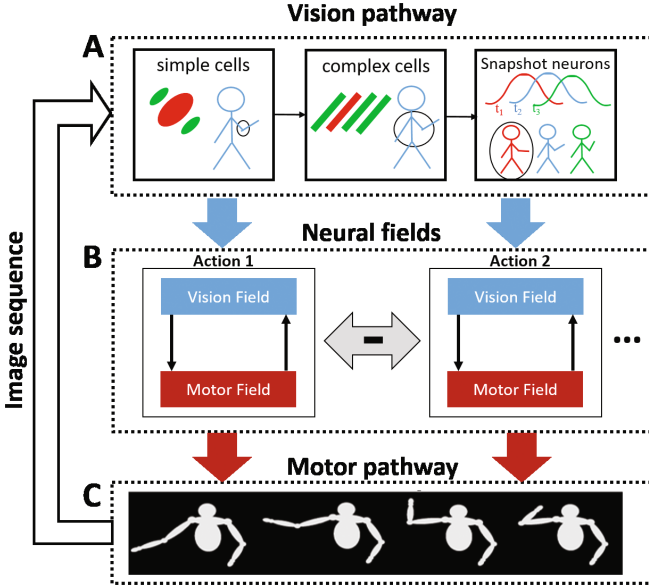


Fig. 1. Overview of the model architecture. **A** The form pathway taken over from a previous neural model [9] drives the input signals for the vision fields from image sequences. **B** The core of the model consists of coupled pairs of vision and motor fields that encode the same action. **C** Motor pathway that reads out the motor fields and generates joint angle trajectories, which are used to animate an avatar, which then can be rendered to produce visual input movies.

pathway consists of a hierarchy of neural shape detectors. The complexity of the extracted features and the position and scaling invariance increase along the hierarchy. The highest level of this pathway is composed from radial basis function (RBF) units that have been trained with snapshots of the learned action movies. These neurons thus detect instantaneous body shapes in image sequences, where the underlying neural network is trained in a supervised manner. Dropping for a moment the index m , assume that the vector $\mathbf{z}(t)$ is formed by the activations of the shape-selective RBF units that encode one particular action pattern at time t , and that the vector $\mathbf{s}(t)$ signifies input signal $s(x, t)$, sampled at a sufficient number of discrete points along the variable x . We learned a linear mapping of the form $\mathbf{s}(t) = \mathbf{Rz}(t)$ between these vectors using sparse regression. Training data pairs consisted of vectors $\mathbf{z}(t)$ of the RBF outputs for equidistantly sampled key frames from the training action movies. Vectors $\mathbf{s}(t)$ were derived from appropriately positioned idealized Gaussian input signals. For learned training patterns the outputs of this linear network define a moving positive input peak, while the input signal $s(x, t)$ remains very small for actions that deviate from the training action. In total, we learned M separate linear mappings from the RBF outputs of the units encoding the keyframes of action m to the corresponding input signal distributions $s^m(x, t)$.

The motor pathway computes joint angles from the position of the activation peak in the motor field along the variable y . This variable parameterizes the temporal evolution of the action. Dropping again the index m , we learned by Support Vector Regression a mapping of the position of the activation peaks $y_{\max}(t) = \arg \max_y v(y, t)$ onto the joint angles of the corresponding body postures. The motor fields encoding different actions compete in a winner-takes-all fashion, and we used only the output of the most activated motor field for the computation of the joint angles. In order to close the loop between action control and perception we used the joint angles to animate an avatar, which then was rendered to produce input movies for the visual pathway.

3 Simulations in Comparison with Experimental Data

We simulated the results of four experiments that studied the interaction between action perception and execution. In the following, simulation results from the model are presented side-by-side with the original data, always using the same model parameters.

(i) Influence of action execution on action perception: In the underlying experiment arm actions were presented as point-light stimuli in noise while the observers performed the same action in a virtual reality setup. The spatio-temporal coherence between the executed and the visually observed action was systematically varied, either by delaying the observed action in time or by rotating it in the image plane relative to the executed action. (See [6] for further details.) Fig. 2A shows a recognition index (RI) that measures the facilitation ($RI > 0$) or inhibition ($RI < 0$) of the visual detection by concurrent motor execution in comparison with a baseline without motor execution. For increasing spatial (Fig. 2A) as well as temporal (Fig. 2B) incoherence between the executed and observed actions the facilitation by concurrent motion execution goes over into an inhibitory interaction. The same behavior is reproduced by our model, simulating the masked point-light stimulus by a noisy traveling input peak (Fig. 2 C, D).

(ii) Influence of action perception on action execution: The underlying experiment measured the variability of motor execution when participants moved their arms periodically in one direction while they saw another person performing a periodic arm movement in the same or in orthogonal direction [4]. As illustrated in Fig. 3A, compared to a baseline without concurrent visual stimulation, the variability of the motor pattern increases when the visually observed arm movement is inconsistent (orthogonal) to the executed pattern. The same increase in variability is obtained from the model (Fig. 3B) (quantified as variability of the timing of the corresponding activation peak in the motor field).

(iii) Spontaneous coordination in multi-person interaction: A classical experiment in interactive sensorimotor control [20] shows that two people that observe each other during the execution of a periodic leg movement tend spontaneously to synchronize their movements. In addition, the variability of the

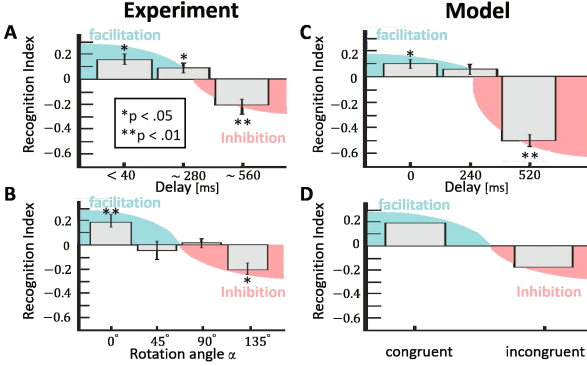


Fig. 2. Influence of concurrent motor execution on the visual detection of action patterns. The experimentally measured Recognition Index (RI) indicates transitions from facilitation to inhibition of visual detection by concurrent motor execution, when the temporal coherence (panel **A**) or the spatial congruence (panel **B**) of the visual pattern with the executed patterns are progressively reduced ([6]). Similar RI computed from the model output shows qualitatively the same behavior (panels **C** and **D**).

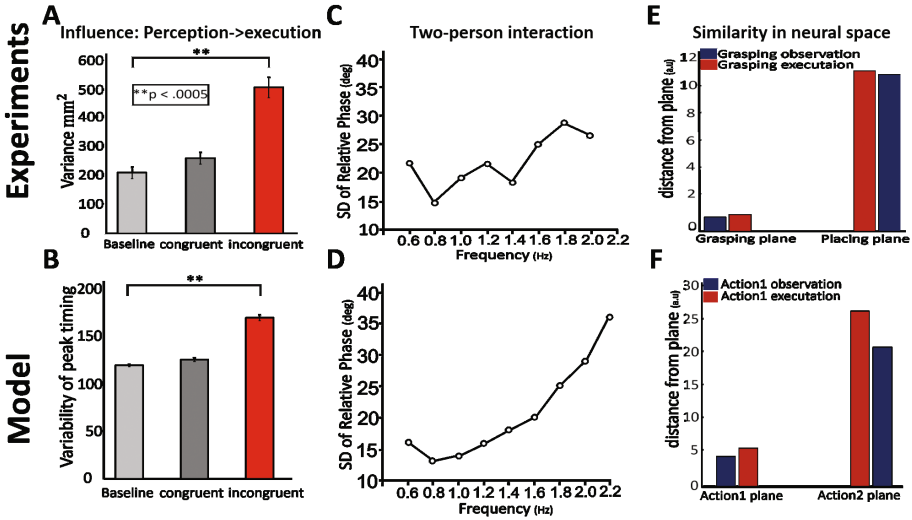


Fig. 3. Reproduction of experimental effects: **A** Motor variability of executed actions increases during observation of incongruent actions [4]. **B** Timing variability of motor peak in the model shows similar behavior. **C** Frequency dependence of standard deviation (SD) of relative phase for the spontaneous synchronization of two agents who observe each other [20]. **D** Corresponding model result derived from activity in motor fields. **E** Neural trajectories for grasping execution and observation are close to ‘grasping’ plane, but far away from ‘placing’ plane [8]. **F** Same behavior is observed for the neural trajectories computed from the model neurons. (Details see text.)

relative phase of the synchronized movements is frequency-dependent. Figure 3C shows the original data for the frequency dependence. In order to simulate this interactive behavior of two agents, we implemented two separate models and defined the visual input of either model by the movie that was generated by the motor output of the other. Like in the experiment, the two simulated agents spontaneously synchronize. Figure 3D shows that, in addition, the model predicts correctly frequency dependence of the variability of the relative phase (as consequence of the selectivity of the neural fields for the propagation speed of the moving peaks).

(iv) Reproduction of the population dynamics of F5 mirror neurons: Our last simulation reproduces electrophysiological data from action-selective (mirror) neurons in area F5 [8]. To generate this data, the responses of 489 mirror neurons, relative to the baseline activity, were combined into a population activity vector that varies over time. Using principle components analysis, the dimensionality of the ‘neural state space’ that is spanned up by these vectors was reduced to three. (Higher-dimensional approximations led to very similar results; see [8] for details.) In this neural state space the trajectories for the execution and observation of a first action (‘grasping’) were lying close to the same plane, while the trajectory for the observation of another action (‘placing’) evolved in an orthogonal pane. This is quantified in Fig. 3 E, which illustrates the average distances of the neural trajectories from the planes that fit best the trajectories for the observation of ‘grasping’ and ‘placing’. A very similar topology of the neural trajectories emerges for our model, if we concatenate the activities of all neural field neurons into a population vector and apply the same techniques for dimension reduction (Fig. 3F). Thus neural trajectories for the perception and the execution of the same action are close to the same plane, while neural trajectories for different actions evolve in orthogonal subspaces.

4 Conclusion

The proposed model is consistent with the behavior of action-selective neurons in the superior temporal sulcus and mirror neurons in area F5 of monkeys ([16, 17]). It provides a unifying account for a whole spectrum of experiments on the interaction between action perception and execution. Future work needs to give up the strict separation of visual and motor fields, potentially exploiting inhomogeneous neural field models.

Acknowledgments. We thank A. Christensen for helpful comments. Funded by EC, HBP FP7-ICT-2013-FET-F/ 604102, HFSP RGP0036/2016, German Federal Ministry of Education and Research: BMBF, FKZ: 01GQ1002A; Deutsche Forschungsgemeinschaft: DFG GI 305/4-1, DFG GZ: KA 1258/15-1.

References

1. Prinz, W.: Perception and action planning. *Eur. J. Cogn. Psychol.* **9**, 129–154 (1997)
2. Rizzolatti, G., Fogassi, L., Gallese, V.: Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* **2**, 661–670 (2001)
3. Giese, M.A., Rizzolatti, G.: Neural and computational mechanisms of action processing: Interaction between visual and motor representations. *Neuron* **88**, 167–180 (2015)
4. Kilner, J.M., Paulignan, Y., Blakemore, S.J.: An interference effect of observed biological movement on action. *Curr. Biol.* **13**, 522–525 (2003)
5. Calvo-Merino, B., Grèzes, J., Glaser, D.E., Passingham, R.E., Haggard, P.: Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.* **16**, 1905–1910 (2006)
6. Christensen, A., Ilg, W., Giese, M.A.: Spatiotemporal tuning of the facilitation of biological motion perception by concurrent motor execution. *J. Neurosci.* **31**, 3493–3499 (2011)
7. Barraclough, N.E., Keith, R.H., Xiao, D., Oram, M.W., Perrett, D.I.: Visual adaptation to Goal-directed hand actions. *J. Cogn. Neurosci.* **21**, 1805–1819 (2009)
8. Caggiano, V., Fleischer, F., Pomper, J.K., Giese, M.A., Thier, P.: Mirror neurons in monkey premotor area F5 show tuning for critical features of visual causality perception. *Curr. Biol.* **26**, 3077–3082 (2016)
9. Giese, M.A., Poggio, T.: Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* **4**, 179–192 (2003)
10. Jhuang, H., Serre, T., Wolf, L., Poggio, T.: A biologically inspired system for action recognition. In: *IEEE International Conference on Computer Vision*, vol. 1, pp. 1–8 (2007)
11. Chersi, F., Ferrari, P.F., Fogassi, L.: Neuronal chains for actions in the parietal lobe: a computational model. *PLoS ONE* **6**, e27652 (2011)
12. Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W.: Codes and their vicissitudes. *Behav. Brain Sci.* **24**, 910–926 (2001)
13. Wolpert, D.M., Doya, K., Kawato, M.: A unifying computational framework for motor control and social interaction. *Philos. Trans. Royal Soc. London B Biol. Sci.* **358**, 593–602 (2003)
14. Kilner, J.M., Friston, K.J., Frith, C.D.: The mirror-neuron system: a Bayesian perspective. *Neuroreport* **18**, 619–623 (2007)
15. Erlhagen, W., Bicho, E.: The dynamic neural field approach to cognitive robotics. *J. Neural Eng.* **3**, R36 (2006)
16. Cisek, P., Kalaska, J.F.: Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* **33**, 269–298 (2010)
17. Fleischer, F., Caggiano, V., Thier, P., Giese, M.A.: Physiologically inspired model for the Visual recognition of transitive hand actions. *J. Neurosci.* **33**, 6563–6580 (2013)
18. Amari, S.: Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* **27**, 77–87 (1977)
19. Zhang, K.: Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996)
20. Schmidt, R.C., Carello, C., Turvey, M.T.: Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *J. Exp. Psychol. Hum. Percept. Perform.* **16**, 227–247 (1990)