Skeleton Model for the Neurodynamics of Visual Action Representations

Martin A. Giese

Section Computational Sensomotorics, Dept. of Cognitive Neurology
CIN & HIH, University Clinic Tübingen
Otfried-Müller-Str. 25, 72076 Tübingen, Germany
Martin.giese@uni-tuebingen.de

Abstract. The visual recognition of body motion in the primate brain requires the temporal integration of information over complex patterns, potentially exploiting recurrent neural networks consisting of shape- and optic-flow-selective neurons. The paper presents a mathematically simple neurodynamical model that approximates the mean-field dynamics of such networks. It is based on a two-dimensional neural field with appropriate lateral interaction kernel and an adaptation process for the individual neurons. The model accounts for a number of, so far not modeled, observations in the recognition of body motion, including perceptual multi-stability and the weakness of repetition suppression, as observed in single-cell recordings for the repeated presentation of action stimuli. In addition, the model predicts novel effects in the perceptual organization of action stimuli.

Keywords: Action recognition, biological motion, neural field, adaptation, superior temporal sulcus, premotor cortex.

1 Introduction

Body motion recognition is a central visual function with high importance for social communication and the learning of movements by imitation [1]. The cortical core circuit of visual action recognition might be based on a competitive network of neurons that are selective for motion and optic flow patterns, and which detect such patterns in a sequence-selective manner [2]. Consistent with this hypothesis is the observation of neurons in the superior temporal sulcus (STS) that respond selectively to snapshots of action movies [3-5], and which often show temporal sequence selectivity, i.e. they respond differently for action movies shown in normal and inverted temporal order [3]. Neurons with very similar properties were found in higher action-selective areas, such as area F5 in monkey premotor cortex [Pomper et al., SFN, 2011, abstract 914.02/QQ7]. Another interesting property of such visual action-selective neurons is that they often show view-dependence, i.e. they respond preferentially to one particular view, but much less to other views of the same action [3, 5, 6]. These observations constrain a simple neurodynamical model that accounts for the joint neural encoding of the view and the time structure of action stimuli.

S. Wermter et al. (Eds.): ICANN 2014, LNCS 8681, pp. 707–714, 2014.

On the behavioral side, body motion perception has interesting dynamic properties which so far have not been studied very much by theoreticians. Firstly, body motion perception can show multi-stability. This has been first demonstrated by Vanrie and collaborators [7], who showed that the same two-dimensional point-light body motion stimuli can be interpreted as locomoting in two different directions, e.g. towards or away from the observer. This ambiguous percept shows spontaneous perceptual switching between the two possible perceptual interpretations, in a similar manner as this is known for other multi-stable displays, such as the Necker cube or binocular rivalry [8]. This observation suggests the existence of an underlying multi-stable neural dynamics that gives rise to these perceptual switching, and to decisions between the two alternative perceptual interpretations.

Secondly, many perceptual processes, including object recognition, are characterized by adaptation when the same stimulus is presented repeatedly. This fact is fundamental for repetition suppression paradigms in fMRI experiments, which have been extensively applied in the field of visual object recognition [9]. For action stimuli, however, the results on fMRI repetition suppression have been ambiguous, and electrophysiological experiments have either failed to show substantial adaptation effects in action-selective areas, such as area F5 in single units [10], or they have reported only very week adaptation effects after a large number of stimulus repetitions [11]. This raises the question how adaptation interacts with the perceptual organization of body motion stimuli, and why adaptation effects are so much weaker in action recognition areas than in object recognition in the inferotemporal cortex (IT) [12].

We present in the following a relatively simple mathematical neurodynamical model that provides an account for these phenomena, and which offers a possible explanationation why adaptation effects for action stimuli might be much weaker than the ones found in experiments with static shape stimuli. In addition, the model provides a possibility to coarsely estimate the importance of noise and internal fluctuations (or top-down effects) in the causation of perceptual switches for ambiguous body motion stimuli.

The paper is structured as follows: We first review some related theoretical approaches. Then the model will be briefly described. In the subsequent section discusses the simulation results and relates them to the experimental literature, followed by some conclusions.

2 Related Theoretical Work

The experimental literature on body motion perception and perceptual multi-stability is vast, and space allows here only to review a few related theoretical models. While initial models for body motion perception have been purely computational (e.g. [13]), more recently a number of neural models have been developed. Our model is based on the dynamical core circuit of a physiologically-inspired hierarchical recognition model that integrates form and motion features [2]. More recently it has been shown that architectures of this type can be made computationally sufficiently powerful to compete with state-of-the-art algorithms for action detection (e.g. [14]). Many neural models exist for the perceptual multistability of static stimuli, e.g. the Necker cube, or ambiguous coherent or apparent motion displays (e.g. [15-17]). Typically, such models are based on competitive dynamic neural networks. More recently, perceptual multi-stability for such phenomena has been analyzed using probabilistic approaches for the analysis of the activity of competing neural ensembles [18]. While we acknowledge that some phenomena, such as synchronized oscillations, might necessitate the use of spiking neuron models, we reside to a mean-field approximation for this paper because it results in a model that is in principle mathematically tractable, and permits a qualitative understanding of the underlying dynamical phenomena.

3 Model Architecture

The model is based on a two-dimensional neural field [19], that represents the view and the keyframe (time point) of body shapes within an action sequence. The model represents a two-dimensional extension of the dynamic layer of a model in [2], which did not represent views in a continuous manner. The input of this neural field is given by the responses of shape-selective neurons that are selective for particular body postures and views arising during action stimuli. The selectivity of such neurons can be established by learning [2]. For the simulations presented in this paper we assumed an idealized input signal, where we replaced the real input by moving peaks in the input distribution. However, the same model has also been tested also while embedded in the hierarchical visual recognition architecture from [2], using real stimuli as input. In the following we focus on the neural encoding of periodic body motions, such as walking. In this case, the neural field is periodic in the view as well as in the direction of the stimulus frame.

The proposed model is defined by two dynamic equations. The first defines an activation dynamics, which is modeled by a two-dimensional neural field, where the first dimension ϕ specifies the stimulus view, and where the second dimension θ defines the frame or snapshot (e.g. within the gait cycle) that is represented by corresponding neuron (or point within the neural field). The second equation specifies a 'point-wise' simple linear adaptation dynamics that is associated with each neuron (point) in the field. More specifically, the model is defined by the equations:

$$\begin{aligned}\tau_u \dot{u}(\phi,\theta,t) = -u(\phi,\theta,t) + w(\phi,\theta) * 1(u(\phi,\theta,t)) + s(\phi,\theta,t) - h \\ - \alpha a(\phi,\theta,t) + \xi(\phi,\theta,t)\end{aligned} \tag{1}$$

$$\tau_a \dot{a}(\phi,\theta,t) = -a(\phi,\theta,t) + 1(u(\phi,\theta,t)) \tag{2}$$

In equation (1) the variable u specifies the (average) membrane potential for a neuron ensemble representing view ϕ and snapshot (body configuration) θ. The constant h specifies the resting potential. The recurrent interaction kernel w specifies the interaction between different points in the field. It is symmetric with respect to the origin in the ϕ-direction (view), and asymmetric in the θ-direction (snapshot), with an additional strong inhibitory component. Its functional form is given by equations $w(\phi,\theta) = w_\phi(\phi) w_\theta(\theta) - w_0$, with the functions $w_\phi(\phi) = \exp((\cos\phi - 1)/\sigma_\phi)$ and $w_\theta(\theta) = \exp((\cos(\theta - \eta) - 1)/\sigma_\theta)$ and the global inhibition $w_0 > 0$. The parameters σ_ϕ and σ_θ specify the tuning width, and the parameter η specify the asymmetry of the kernel in θ-direction. It has been shown elsewhere that such asymmetric kernels, if designed appropriately, result in temporal sequence selectivity and a well-defined speed tuning curve of the neurons in the field with respect to this direction [20]. The symbol * signifies a spatial convolution, which is periodic since the field is periodic in both directions. The step threshold function $1(u)$ takes the value 1 for $u >$ 0, and zero otherwise.

The stimulus input signal s models an (idealized) activity distribution over shape-selective neurons. The value $s(\theta,\phi,t)$ defines the average input activity of shape-selective neurons (neuron ensembles) that respond maximally to the body configuration appearing at (normalized) time θ of the gait / action cycle, and with the view angle ϕ. The stimulus input was modeled in an idealized manner, assuming peaks of activity with amplitude s_0 that propagate in the θ-direction with speed v. More specifically, these peaks were specified by the equation: $s(\theta,\phi,t) = s_0 \exp((\cos(\theta - \theta_c(t)) - 1)/\sigma_S) \exp((\cos(\phi - \phi_c) - 1)/\sigma_S)$, where the peak center θ_c in θ-direction was moving with speed v, and where the ϕ_c corresponds to the view angle of the body in the corresponding frame. The parameters σ_s defines the width of the idealized input peak. For ambiguous action stimuli that simultaneously activate two different competing views [7], we added a second peak with view angle the $-\phi_c$ to the input signal distribution. The noise distribution ξ is defined by a Gaussian process whose kernel function is the product of a spatial kernel function that was fitted in order to reproduce coarsely the correlation statistics, dependent on the tuning similarity of the neurons [21], and a delta function with respect to time.

The dynamic neural field define by equation (1) stabilizes a stimulus-locked travelling peak solution in θ-direction, i.e. an activation peak that follows the stimulus peak, if the frames of an action stimulus appear in the correct temporal order, and if the speed of the input peak falls in the range of preferred speeds that is determined by the parameters of the neural field. If the frames of an action movie are shown in the wrong temporal order, or if the speed of the presentation of the movie deviates very strongly from the natural speed of the action (that matches the preferred speed of the field) the activation in the neural field remains relatively small [2]. In addition, the lateral interactions in (view) ϕ-direction result in a winner-takes-all competition along this dimension, resulting in a decision for one stimulus view for stimuli that are ambiguous and activate simultaneously interpretations corresponding to multiple views.

Equation (2) specifies a simple adaptation process, independently for each point in the neural field. The adaptation variable $a(\phi,\theta)$ feeds back negatively in the activation field with a strength that is determined by the positive parameter α. It is driven by the output activity of the corresponding point in the neural field. The time scales of the activation and the adaptation dynamics were given by positive parameters τ_u = 120 ms and τ_a = 2.4 s.

4 Simulation Results

I) Multi-stability: The proposed neural model defines a multi-stable dynamics. For ambiguous stimuli activating the views $\pm\phi_c$ that deviate sufficiently from the side view ($\phi_c = 0$) of a walker, the field has two alternative stable travelling pulse solutions, moving together with the stimulus peaks in the θ-direction with speed v. One of these solutions is depicted in the first three panels of Fig. 1A, where the selected view is the one corresponding to the view angle (ϕ_c = -72 deg). Strong noise can induce a spontaneous switch to the other stable travelling pulse solution, corresponding to (ϕ_c = 72 deg). Such a spontaneous switch is taking place in panels 3 and 4 of Fig. 1A. After the switch the activity peak propagates with speed v along a horizontal line that correspond to the view angle 72 deg.

This behavior is confirmed by an analysis of the sum of the activity in the field. Fig 1B shows the sums of the thresholded activity over all values of θ, and the regions with positive respectively negative values of the variable ϕ. Due to the noise, these sum activity show strong fluctuations. A spontaneous transition to the other stable solution occurs within time interval close to t = 8 s, corresponding to a perceptual switch between the two alternative views that are compatible with the stimulus.,

In addition, the model predicts an interesting bifurcation (see Fig. 1C): For view angle differences below ±21 deg the bistability disappears, and only a single stable traveling pulse solutions exists that follows the average of the compatible stimulus views $\pm\phi_c$. Initial psychophysical observations seem to confirm this bifurcation.

II) Adaptation: The adaptation dynamics was fitted using data from single cell recordings in inferotemporal (IT) cortex [12]. To account for the recognition of static stimuli, the interaction kernel was made symmetrical (choosing $\eta = 0$), and a static stimulus distribution s was used. The time course and the maximum rate of adaptation in these experiments were coarsely matched (Fig. 2A). The adaptation results in a flattening of the tuning curve, not just in a multiplicative rescaling (Fig. 2B), consistent with the data [12]. Applying the same adaptation mechanisms to the original model in absence of internal noise ($\xi \equiv 0$) was insufficient to account for spontaneous switches, suggesting that these perceptual switches are not adaptation-induced.

The repetition of a single action stimulus (one gait or action cycle), following the procedure in fMRI studies and in [10], results in a very small adaptation effect that is difficult to detect in presence of noise (red curve, Fig 2C). (The noise level here was far below the one required for inducing perceptual switches.) Using a special stimulus that repeats a fragment from an action movie with a duration of about 200 ms very quickly, but keeping the total stimulus time (3 s) constant, results in a much stronger adaptation effect. The model thus predicts that such stimuli might be more efficient adaptors than the repetition of whole actions.

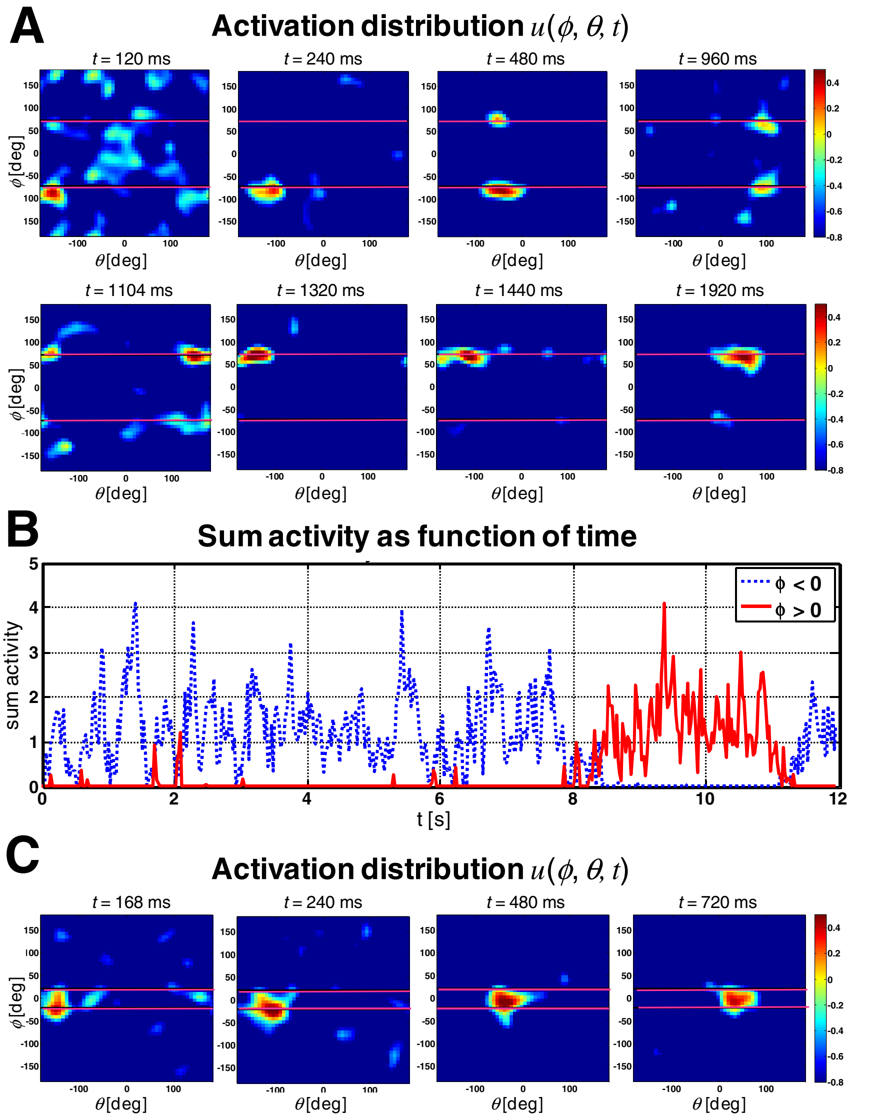

Fig. 1. Multi-stability **of the neural field dynamics. A** For an ambiguous body motion stimulus that equally activates two view directions (ϕ_c = ±72 deg relative to the side view; indicated by the pink lines) the solution peak first propagates along with the stimulus peak at ϕ = -72 deg. A spontaneous transition to the other stable travelling pulse solution, centered at the view angle ϕ = 72 deg occurs after some time (panel 4)). **B** The sum activity over frames and over positive, respectively negative view angles shows strong random fluctuations, resulting in a spontaneous switching between the two stable solutions for *t* around 8 s. **C** For view angles ϕ_c that deviate less form the side view than ±21 deg the bistability disappears (bifurcation). Only a single stable travelling peak solution exists that follows the midpoint of the two peaks of the input signal distribution in ϕ direction (which are indicated by the pink lines).

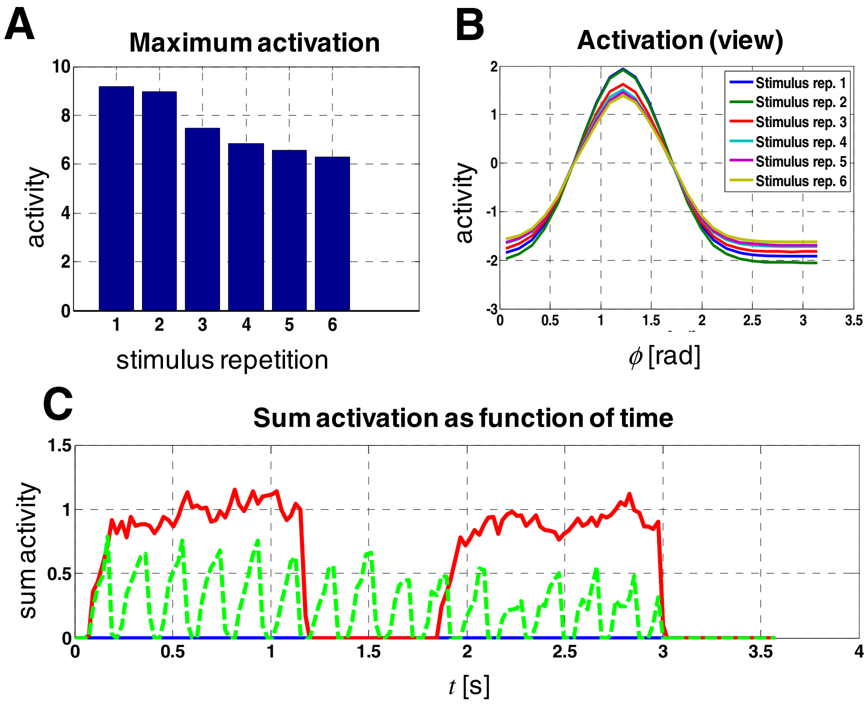

Fig. 2. Activation-dependent **adaptation. A** Simulating adaptation effects in area IT [12], using static stimuli and a field with symmetric interaction kernel ($v = 0$), the repeated presentation of static patterns results in a maximum adaptation of the neuron activity of about 30 %. **B** The adaptation results in a widening of the tuning curve, consistent with the experimental data from area IT. **C** Using the same adaptation dynamics, only very weak adaptation is found for repetition of action stimuli (red curve: sum activity for one stimulus repetition). A different stimulus with fast repetition of the same short action fragment (duration of about 200 ms) results in much stronger adaptation for the same total stimulus duration.

5 Conclusions

This paper proposed a neurodynamical model that captures several aspects of the perceptual dynamics in body motion perception. It provides a unifying explanation for several phenomena: i) multi-stability in the perception of views of body motion; ii) a possible cause for the difficulty to demonstrate adaptation effects for the stimulus repetition of action stimuli. In addition, the model makes a number of predictions: a) It predicts a bifurcation of the dynamics in dependence of the deviation of the stimulus views from the side view. b) It suggests a new action stimulus that might result in stronger adaptation effects than simple stimulus repetition. Furthermore, the proposed model is mathematically quite simple and thus accessible for mathematical analysis. Experimental testing of these predictions and such a mathematical analysis are the topics of ongoing work.

Acknowledgments. This work was supported by the EU projects ABC: PEOPLE-2011-ITN PITN-GA-011-290011; HBP: FP7-ICT-2013-FET-F/ 604102; Koroibot FP7-ICT-2013-10/ 611909, and by DFG GI 305/4-1, DFG GZ: KA 1258/15-1, and BMBF, FKZ: 01GQ1002A**.** I thank J. Vangeneugden, V. Caggiano, P Thier, and R. Vogels for very helpful discussions.

References

1. Blake, R., Shiffrar, M.: Perception of human motion. Annu. Rev. Psychol. 58, 47–73 (2007)
2. Giese, M.A., Poggio, T.: Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. 4(3), 179–192 (2003)
3. Barraclough, N.E., Keith, R.H., Xiao, D., Oram, M.W., Perrett, D.: Visual adaptation to goal-directed hand actions. J. Cogn. Neurosci. 21(9), 1806–1820 (2009)
4. Singer, J.M., Sheinberg, D.: Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. J. Neurosci. 30(8), 3133–3145 (2010)
5. Vangeneugden, J., De Mazière, P.A., Van Hulle, M.M., Jaeggli, T., Van Gool, L., Vogels, R.: Distinct mechanisms for coding of visual actions in macaque temporal cortex. J. Neurosci. 31(2), 385–401 (2011)
6. Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J.K., Thier, P., Giese, M.A., Casile, A.: View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. Curr. Biol. 21(2), 144–148 (2011)
7. Dekeyser, M., Verfaillie, K.: Bistability and biasing effects in the perception of ambiguous point-light walkers. Perception 33(5), 547–560 (2004)
8. Leopold, D.A., Logothetis, N.: Multistable phenomena: changing views in perception. Trends. Cogn. Sci. 3(7), 254–264 (1999)
9. Grill-Spector, K., Henson, R., Martin, A.: Repetition and the brain: Neural models of stimulus-specific effects. Trends. Cogn. Sci. 10(1), 14–23 (2006)
10. Caggiano, V., Pomper, J.K., Fleischer, F., Fogassi, L., Giese, M., Thier, P.: Mirror neurons in monkey area F5 do not adapt to the observation of repeated actions. Nat. Commun. 4, 1433 (2013)
11. Kilner, J.M., Kraskov, A., Lemon, R.: Do monkey F5 mirror neurons show changes in firing rate during repeated observation of natural actions? J. Neurophysiol. 111(6), 1214–1226 (2013)
12. De Baene, W., Vogels, R.: Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. Cereb. Cortex. 20(9), 2145–2165 (2010)
13. Marr, D., Vaina, L.: Representation and recognition of the movements of shapes. Proc. R. Soc. Lond. B Biol. Sci. 214(1197), 501–524 (1982)
14. Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., Steele, A.D., Serre, T.: Automated home-cage behavioural phenotyping of mice. Nat. Commun. 1, 68 (2010)
15. Kawamoto, A.H., Anderson, J.A.: A neural network model of multi-stable perception. Acta. Psy. 59, 35–65 (1985)
16. Kelso, J.A.S.: Dynamic Patterns. MIT Press, Cambridge (1995)
17. Giese, M.A.: Dynamic Neural Field Theory for Motion Perception. Kluwer Academic Publishers, Dordrecht (1998)
18. Gigante, G., Mattia, M., Braun, J., Del Giudice, P.: Bistable perception modeled as competing stochastic integrations at two levels. Comput. Biol. 5(7), e1000430 (2009)
19. Amari, S.: Dynamics of pattern formation in lateral-inhibition type neural fields. Biol. Cybern. 27(2), 77–87 (1977)
20. Xie, X., Giese, M.: Nonlinear dynamics of direction-selective recurrent neural media. Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys. 65(5 Pt 1), 51904 (2002)
21. Cohen, M.R., Kohn, A.: Measuring and interpreting neuronal correlations. Nat. Neurosci. 14(7), 811–819 (2011)