Neural model for multi-stability in visual action recognition

Martin Giese*, Leonid Fedorov* and Rufin Vogels**

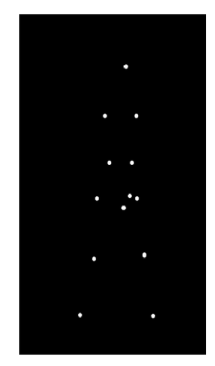
* Section for Computational Sensomotorics, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research & Centre for Integrative Neuroscience, University Clinic Tübingen, Germany.
**Laboratorium voor Neuro- en Psychofysiologie, KU Leuven, Belgium

Hertie-Institut für klinische Hirnforschung
Universitätsklinikum Tübingen
CIN Werner Reichardt Centre for Integrative Neuroscience
Eberhard Karls Universität Tübingen

Introduction

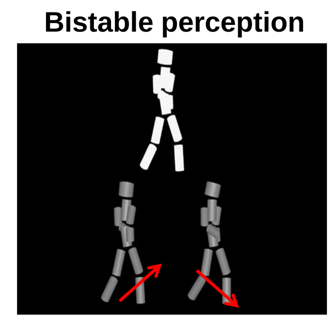
- Action perception often treated under the viewpoint of pattern recognition: classification of spatio-temporal visual patterns.
- Example: perception of 'biological motion'; Johansson was originally interested in dynamic pattern formation ('Gestalt'), not in information encoding in point-light stimuli (Johansson, 1973;Jansson et al.1994; Poljac et al. 2011).
- Body motion perception shows interesting dynamical properties:
  I) **Multi-stability**: Switching between multiple percepts of the same stimulus (e.g. in terms of walking direction) (Vanrie et al. 2004; Schouten et al. 2011); disambiguation by shading cues.
  II) **Adaptation**: repetition of the same action results in high-level after-effects and reduction of neural activity (in the BOLD signal) (repetition suppression) (Jordan et al. 2006; Troje et al. 2006; Jastorff et al. 2009; Grossman et al., 2010).
- No or very weak adaptation effects observed in single-cell studies on mirror neurons in area F5 (premotor cortex) (Caggiano et al.2013; Kilner et al. 2014).
- Ambiguous fMRI adaptation results for repetition suppression in human mirror neuron system (e.g. Dinstein et al. 2008; Lingnau & Caramazza, 2009).
  ⇔ Strong adaptation (decay of activation by 10-20%) for shape-selective neurons in area IT (de Baene & Vogels, 2011).

Biological motion



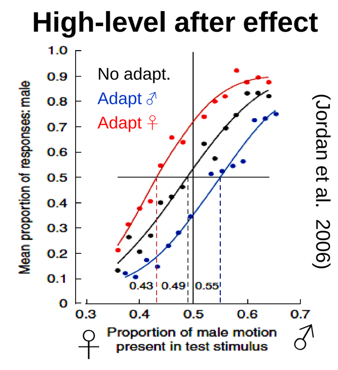
(Johansson, 1973)

Bistable perception



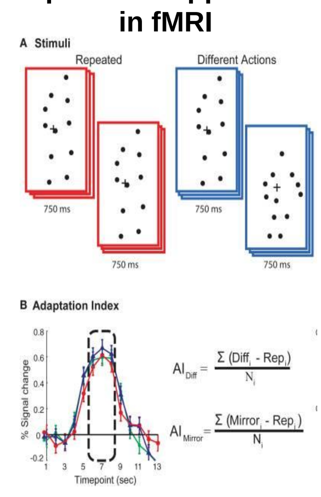
(Vangeneugden et al. 2011)

High-level after effect



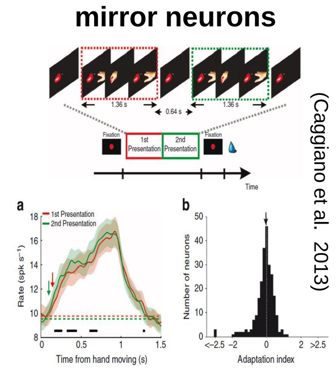
(Jordan et al. 2006)

Repetition suppression in fMRI



(Grossman et al. 2010)

Adaptation in F5 mirror neurons



(Caggiano et al. 2013)

Questions

- **Unifying neural model for these neuro-dynamic effects in action perception?** (No account by existing models (e.g. Giese & Poggio, 2003; Lange & Lappe, 2006; Jhuang et al. 2007).
- **Mathematical framework for treatment of the underlying multi-stability?**
- **Why is adaptation in action-selective neurons so small compared to shape-selective neurons in area IT ?**

Stability analysis for 2D neural field

- Start from a simplified model without adaptation / noise; step threshold $1(u)$:

$$\boxed{\tau_u \dot{u}(\phi,\theta,t) = -u(\phi,\theta,t) + w(\phi,\theta) * 1(u(\phi,\theta,t)) + s(\phi,\theta,t)}$$

$$s(\phi,\theta,t) = S(\phi,\theta - vt)$$
$$u(\phi,\theta,t) = U(\phi,\theta - vt,t)$$
$$w(\phi,\theta) = w_s(\phi,\theta) - \tau_u v \nabla_\theta w_s(\phi,\theta)$$

- In moving coordinate system: equivalent kernel w_s:

$$\tau_u \dot{U}(\phi,\theta,t) = -U(\phi,\theta,t) + w_s(\phi,\theta) * 1(U(\phi,\theta,t)) + S(\phi,\theta)$$

- Reformulate dynamics using a level set approach using **divergence theorem** (Coombes et al. 2011) :

$$w_s(\phi,\theta) * 1(u(\phi,\theta,t)) = \int_B w_s(\mathbf{r}-\mathbf{r}')\,d\mathbf{r}' = \oint_{\partial B} \mathbf{F}(\mathbf{r}(\sigma)-\mathbf{r}(\sigma'))\,\mathbf{n}(\sigma')\,d\sigma'$$

$$\mathbf{z}(\mathbf{r},t) = \nabla_{\mathbf{r}} u(\mathbf{r}(\sigma),t)$$
$$\mathbf{n} = \mathbf{z}(\mathbf{r},t) / |\mathbf{z}(\mathbf{r},t)|$$

$$\operatorname{div} \mathbf{F}(\mathbf{r}) = w_s(\mathbf{r}) \qquad \mathbf{r} = (\phi,\theta)$$

- Dynamics of boundary of excited region (level set):

$$\tau_u \dot{u}(\mathbf{r}(\sigma),t) = -u(\mathbf{r}(\sigma),t) + \oint_{\partial B} \mathbf{F}(\mathbf{r}(\sigma)-\mathbf{r}(\sigma'))\,\mathbf{n}(\sigma')\,d\sigma' + S(\mathbf{r}(\sigma))$$
$$\tau_u \dot{\mathbf{z}}(\mathbf{r}(\sigma),t) = -\mathbf{z}(\mathbf{r}(\sigma),t) + \oint_{\partial B} w(\mathbf{r}(\sigma)-\mathbf{r}(\sigma'))\,\mathbf{n}(\sigma')\,d\sigma' + \nabla_{\mathbf{r}} S(\mathbf{r}(\sigma))$$

Stability condition:

$$\operatorname{Re}\{\lambda(\mathbf{M})\} < 0 \text{ with } \mathbf{M} = \begin{pmatrix} \operatorname{Re}\{\tilde{\mathbf{A}}\} & -\operatorname{Im}\{\tilde{\mathbf{A}}\} \\ \operatorname{Im}\{\tilde{\mathbf{A}}\} & \operatorname{Re}\{\tilde{\mathbf{A}}\} \end{pmatrix} \text{ and } \tilde{\mathbf{A}} = \mathrm{DFT}(\mathbf{A}(\sigma,\sigma'))$$

$$\mathbf{A}(\sigma,\sigma') = -(\tau_u / |\mathbf{z}^*(\mathbf{r}^*(\sigma),t)|)\,\mathbf{n}^*(\sigma) \cdot \left[\left(\int_{\partial B} \mathbf{H}(\sigma,\sigma')\cdot\mathbf{n}^*(\sigma')\,d\sigma'\right)^T \delta(\sigma-\sigma') - \mathbf{n}^*(\sigma')^T \cdot \mathbf{H}^T(\sigma,\sigma')\right]$$

and where $\mathbf{H}(\sigma,\sigma') = \left.\frac{\partial \mathbf{F}}{\partial \mathbf{r}}\right|_{\mathbf{r}^*(\sigma)-\mathbf{r}^*(\sigma')}$

Basic model

- Model architecture extends physiologically-inspired model for the recognition of body motion stimuli (Giese & Poggio, 2003).
- Hierarchical model with two pathways processing form and motion information.
- Snapshot neurons recognize body shapes, which are activated in sequence.
- Integration of information over time by **dynamic neural field** with asymmetric lateral interaction kernel ⇒ stable of stimulus-locked **travelling pulse solution** if input frames appear in correct temporal order, otherwise very small irregular or lurching activity.
- Outputs summed by 'motion pattern neurons'.
- Neurons with such properties found in the STS and area F5 (Barraclough et al., 2009; Vangeneugden et al. 2009; Singer & Sheinberg, 2010; Caggiano et al., subm.).
- Mathematical formulation as Amari field with travelling input peak (Amari, 1977; Zhang, 1996; Giese & Poggio, 2003):

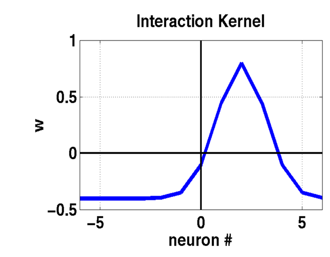
Dynamic neural field: (Amari, 1977)

Moving stimulus

$$\tau \dot{u}(\theta,t) = -u(\theta,t) + s(\theta - vt) + \int w(\theta-\theta')\,1(u(\theta',t))\,dx'$$

Lateral interaction kernel (asymmetric !!)

Interaction Kernel



Stability condition:

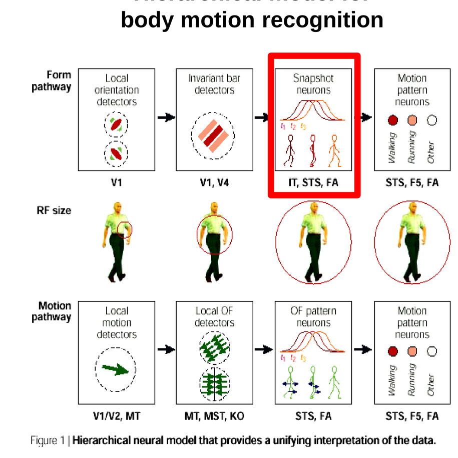
$$(G(0,s)-c_1^*)(G(0,s)-c_2^*) \neq G(a^*,s)G(-a^*,s)$$

with $\operatorname{Re}\{s\} > 0$ $\quad a^* = x_2^* - x_1^*$

$$G(y,s) = -\frac{1}{\tau v}\int_{\pm\infty}^{y} w(y')\,e^{\frac{1+\tau s}{\tau v}(y-y')}\,dy'$$

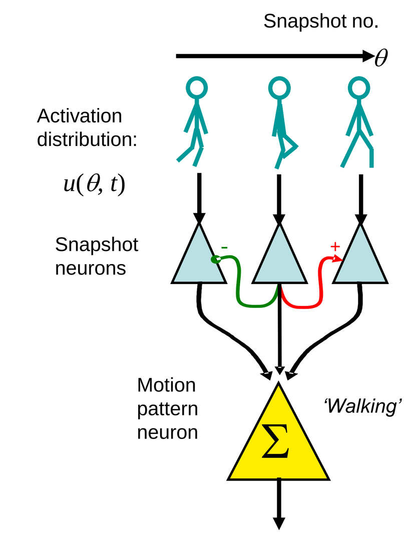
(Xie & Giese, 2002)

Hierarchical model for body motion recognition

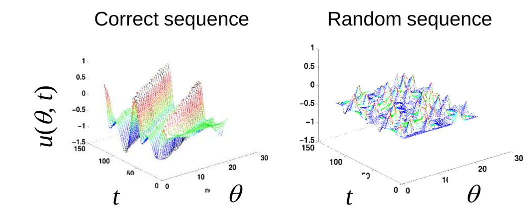


(Giese & Poggio, 2003)

Circuit for temporal integration



Sequence-selective neural field



Model extension

- Extension of highest hierarchy layer (neural field); assumption of idealized input (so far).
- Two dimensional field: θ: snapshot number, ϕ: view angle
- Noise / fluctuations ξ modelled by Gaussian process.
- Adaptation process I based on **firing rate fatigue** (neuron thresholds increase after firing) (de Baene & Vogels, 2011).
- Adaptation process II based on **input fatigue** (synapses become less efficient after use) (de Baene & Vogels, 2011)
- Spike rate adaptation needed to model detailed shape of the activity profiles.

Activation dynamics:
$$\tau_u \dot{u}(\phi,\theta,t) = -u(\phi,\theta,t) + w(\phi,\theta) * 1(u(\phi,\theta,t)) + c_v v(\phi,\theta,t) + s(\phi,\theta,t) + \xi(\phi,\theta,t) - \alpha a(\phi,\theta,t)$$

Firing-rate fatigue (FF) adaptation:
$$\tau_a \dot{a}(\phi,\theta,t) = H_a(-a(\phi,\theta,t) + [u(\phi,\theta,t)]_+)$$

Input-fatigue (IF) adaptation:
$$\tau_b \dot{b}(\phi,\theta,t) = H_b(-b(\phi,\theta,t) + [r(\phi,\theta,t)]_+)$$
$$s(\phi,\theta,t) = k(\phi,\theta) * (g(b(\phi,\theta,t))\,r(\phi,\theta,t))$$

Spike rate adaptation:
$$\tau_s \dot{v}(\phi,\theta,t) = -v(\phi,\theta,t) + [\dot{s}(\phi,\theta,t)]_+$$

Results

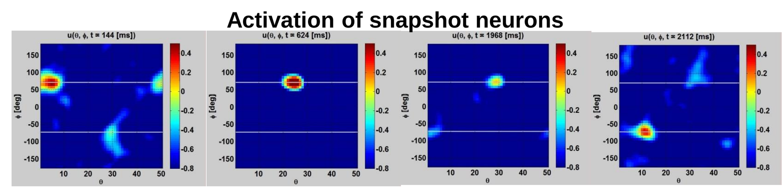
Multi-stability:

- Reproduction of perceptual switching.
- Two competing travelling pulse solutions (attractors).
- Switches mainly induced by noise, not by the adaptation mechanisms.

Adaptation results (static stimuli, area IT):

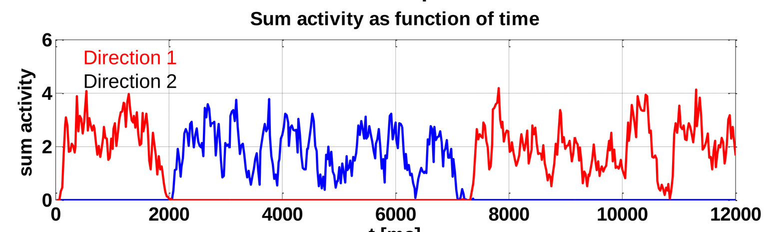
- Simulations used to fit the parameters of the adaptation dynamics.
- Reproduction of signal shape in IT.
- Match of short-term and long-term time courses of adaptation.
- Model with dominant input fatigue mechanisms reproduces interaction for adaptation strength dependent on effective / ineffective adaptors; effect not reproduced by model with firing rate (FR) fatigue

Activation of snapshot neurons



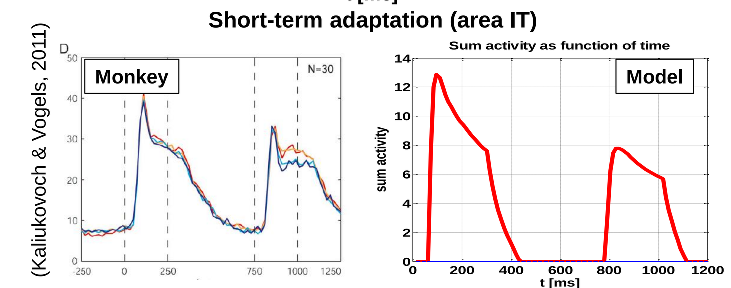
Activation of moton pattern neurons

Sum activity as function of time



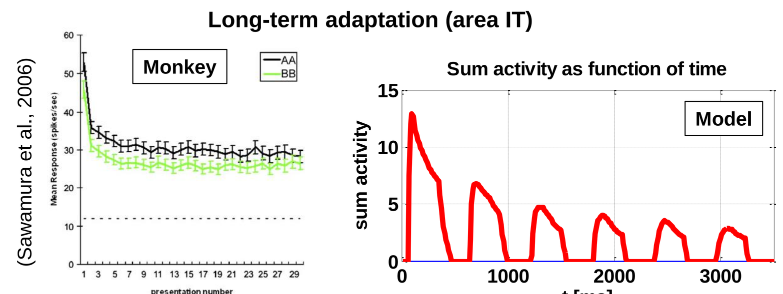
(Giese, ICANN, 2014)

Short-term adaptation (area IT)



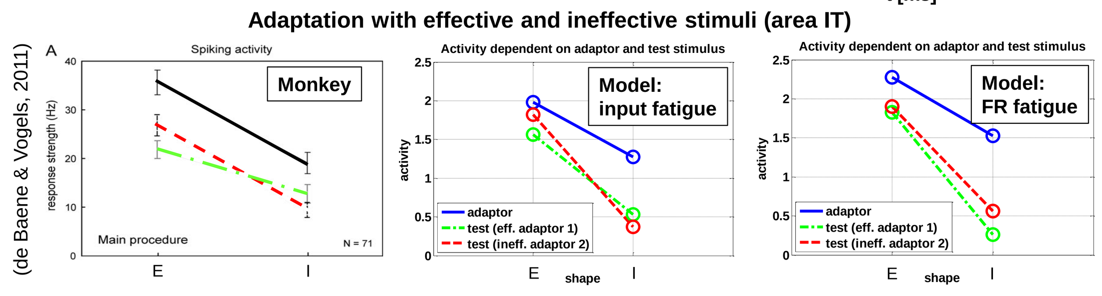
(Kaliukovoch & Vogels, 2011)

Long-term adaptation (area IT)



(Sawamura et al., 2006)

Adaptation with effective and ineffective stimuli (area IT)



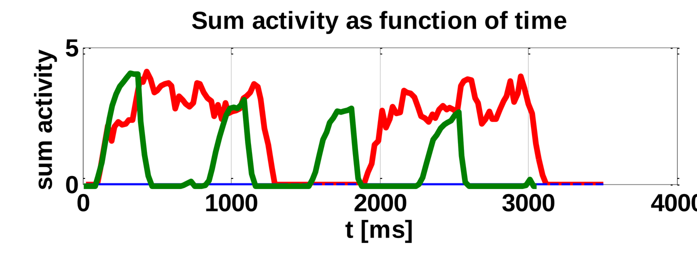
(de Baene & Vogels, 2011)

Adaptation results (action representation):

- We took over the parameters of the adaptation mechanisms in the model for area IT in the model for the representation of actions.
- Testing of models variants with dominant input and firing rate fatigue.
- Very small adaptation effects for both model variants if a single action is repeated (red curves).
- Much stronger adaptation in both model variants for **new adaptation stimulus** (green curves): action fragment (360 ms) repeated as fast as possible.

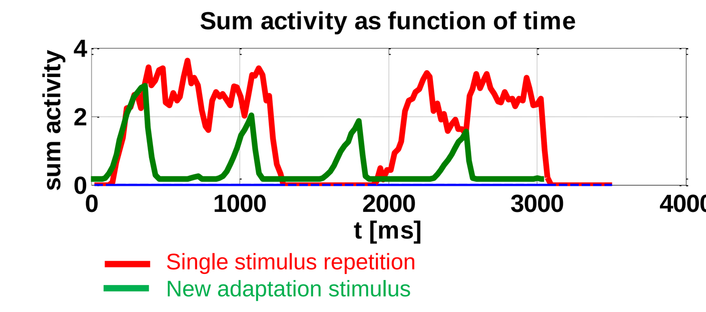
Activation for action stimulus (firing rate fatigue)

Sum activity as function of time



Activation for action stimulus (input fatigue)

Sum activity as function of time



Conclusions

- By appropriate extensions, our previous learning-based recognition model we can account for multi-stability + adaptation effects.
- Key concept: 2D neural field with dimensions snapshot number and view angle.
- Complex interaction between adaptation and dynamic pattern encoding.
- Weaker adaptation effects for action stimuli than for static shape stimuli, assuming the same neural mechanisms for adaptation.
- Reason: transient activation of snapshot neurons by action stimuli.
- Prediction of a new potentially more efficient adaptation stimulus for action-selective neurons.

References

Amari S. (1977). "Dynamics of pattern formation in lateral inhibition type neural fields". *Biol Cyb*, 27:77–87.
Barraclough N.E., Keith R.H., Xiao D., Oram M.W., Perrett D.I..(2009) „Visual adaptation to goal-directed hand actions." *J Cogn Neurosci*. 21(9), 1806-1820.
de Baene W., Vogels R.(2010). "Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. " *Cer Cortex 20(9),* 2145-2165.
Caggiano V, Pomper JK, Fleischer F, Fogassi L, Giese M, Thier P. (2013) "Mirror neurons in monkey area F5 do not adapt to the observation of repeated actions." *Nat Commun.* 4:1433.
Coombes S., Schmidt H., Bojak I. (2012) "Interface dynamics in planar neural field models," *Journal of Mathematical Neuroscience*, 2:9.
Dinstein I. (2008) Human cortex: reflections of mirror neurons. *Curr Biol*.18(20):R956-9.
Giese M.A. (2014). "Skeleton model for the neurodynamics of visual action representations." *ICANN 2014.*
Giese M.A., Poggio T. (2003). "Neural mechanisms for the recognition of biological movements and action." *Nat Rev Neurosci*, *4*, 179-192.
Grossman E.D., Jardine N.L., Pyles J.A. (2010) "fMR-Adaptation Reveals Invariant Coding of Biological Motion on the Human STS." *Front Hum Neurosci.* 23;4:15.
Jansson, G. Bergström, S.S., Epstein, W. (1994) "Perceiving events and objects." Lawrence Erlbaum Associates,: Hillsdale, NJ.
Jastorff J., Kourtzi Z., Giese M.A. (2009) Visual learning shapes the processing of complex movement stimuli in the human brain. *J Neurosci.* 29(44):14026-38.
H Jhuang, T Serre, L Wolf, T Poggio (2007) "A biologically inspired system for action recognition." Proceedings of IEEE 11th International Conference on Computer Vision (ICCV 2007), 1-8.
Johansson, G. (1973). "Visual perception of biological motion and a model for its analysis". *Perc Psychophys.* 14: 201–211.
Jordan H., Fallah M., Stoner G.R. (2006) Adaptation of gender derived from biological motion. Nat Neurosci. 9(6):738-9.
.
Kaliukhovich D.A, Vogels R. (2011) "Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex." *Cereb Cortex.* 21(7):1547-58.
Kilner J.M., Kraskov A., Lemon R.N. (2014) Do monkey F5 mirror neurons show changes in firing rate during repeated observation of natural actions? *J Neurophysiol*. 111(6):1214-26.
Lange J., Lappe M. (2006). "A model of biological motion perception from configural form cues". *J. Neurosci.* 26: 2894-2906.
Lingnau A., Gesierich B., Caramazza A. (2009) "Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans." *Proc Natl Acad Sci U S A.* 106(24):9925-30.
Poljac, E., Verfaillie, K., Wagemans, J. (2011) Integrating biological motion: the role of grouping in the perception of point-light actions *PloS One* 6 (10), e25867-e25867.
Sawamura H., Orban G.A., Vogels R. (2006) "Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm." *Neuron* 49(2):307-18.
Schouten, B., Troje, N. F., Verfaille, K. (2011) "The facing bias in biological motion perception: Structure, kinematics, and body parts." *Atten, Perc & Psychophys. 73:130-14.*
Singer J.M., Sheinberg D.L.(2010) „Temporal cortex neurons encode articulated actions as slow.sequences of integrated poses." *J Neurosci*. 30(8),3133-3145.
Vangeneugden J. et al. (2012). "Activity in areas MT+ and EBA, but not pSTS, allow prediction of perceptual states during ambiguous biological motion." *SfN* 127.04
Vanrie J. et al. (2004). "Bistability and biasing effects in the perception of ambiguous point-light walkers". *Perc* 33: 547-560.
Xie X., Giese M. (2002). "Nonlinear dynamics of direction-selective recurrent neural media." *Phys Rev E Stat Nonlin Soft Matter Phys.* 65 (5 Pt 1):051904 (2002)., *65*(5, Pt. 1).
Zhang K. (1996). "Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory". *J Neurosci* 16: 2112-2126.

Support

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; CogIMon H2020 ICT-644727