Neurophysiologically-inspired computational model of the visual recognition of social behavior and intent

Albert Mukovskiy[1], Mohammad Hovaidi-Ardestani[1], Alessandro Salatiello[1,2], Michael Stettler[1,2], Rufin Vogels[3], Martin A. Giese[1]

[1]Section for Computational Sensomotorics, HIH&CIN, Department of Cognitive Neurology, University Clinic Tübingen, [2]IMPRS for Intelligent Systems, Tübingen, Germany, [3]Research Group Neurophysiology, Dept. Neurowetenschappen, KU Leuven, 3000 Leuven, Belgium

Introduction

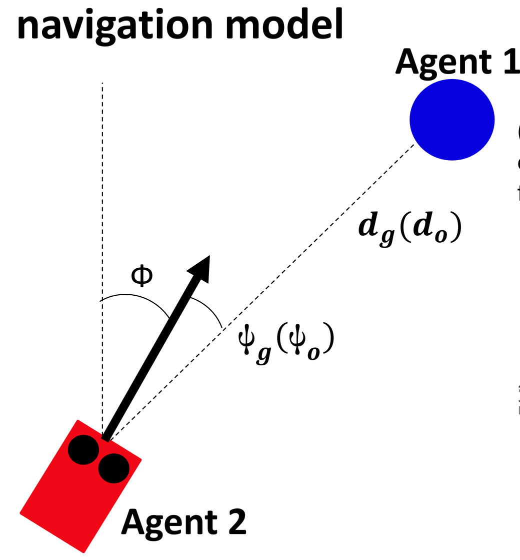
- Humans reliably attribute social interpretations to highly impoverished stimuli, such as interacting geometrical shapes, as shown in the classical experiments (Heider & Simmel, 1944).
- Perception of interaction has been explained by high-level cognitive processes, such as probabilistic reasoning (Baker et al., 2009)
- Perception of animacy from simple figures is dependent on a number of critical stimulus parameters (Tremoulet, Feldman, 2000, 2006; Henrik et al., 2014).
- The perception of basic interactive actions (e.g. 'chasing' or 'fighting') has been addressed in several studies (Gao & Scholl, 2013; Scholl & Tremoulet, 2000; McAleer & Pollick, 2000; Blythe et al. 1999); six types of interactive movements has been used repeatedly in these studies.
- Building on classical biologically-inspired models for action perception (Giese & Poggio, 2003), and a deep learning architecture (Simonyan & Zisserman, 2015) we propose a learning-based hierarchical NN model that analyses such stimuli directly from video sequences of the abstract and of the natural captured scenes.
- The model includes only simple physiologically plausible operations. The shape-recognition feed-forward pathway, modeled by a DeepNN (VGG16), followed by discriminative feature selection, an RBF NN and Neural Fields recognizing and tracking shape, orientation and position of moving agents.

Goal of the research

- Investigation if and how basic aspects of social and animacy perception can be accomplished by simple and physiologically plausible neural mechanisms, exploiting a hierarchical (deep) model of the visual pathway.

Generation of Stimuli

Modelling social interaction by a modified human navigation model



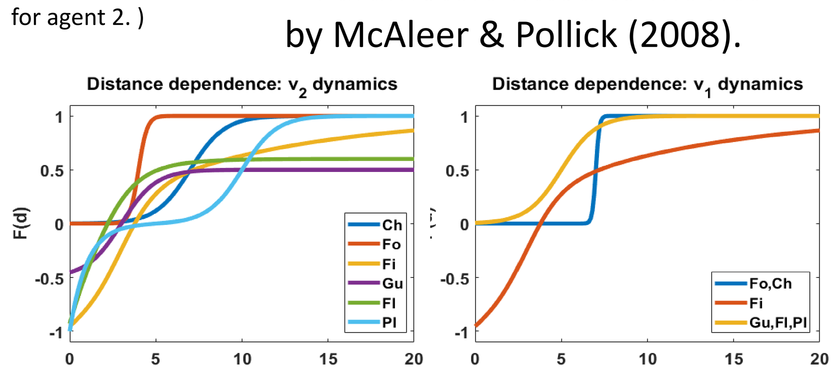
- Dynamics of heading direction (Fajen and Warren 2003):

$$\ddot{\Phi}_i = b\dot{\Phi}_i - k_g\left(\Phi_i - \psi_{g,i}\right)\left(e^{-c_1 d_{g,i}} + c_2\right) + k_o \sum_{n=1}^{N_{obst}}(\Phi_i - \psi_{o,ni})\left(e^{-c_3|\Phi_i - \psi_{o,ni}|}\right)\left(e^{-c_4 d_{o,ni}}\right)$$

- Dynamics of forward speed:

$$\tau\dot{v}_i = -v_i + F_i(d) + c_i\,\varepsilon_i(t)$$

- Parameters fitted to movies by McAleer & Pollick (2008).

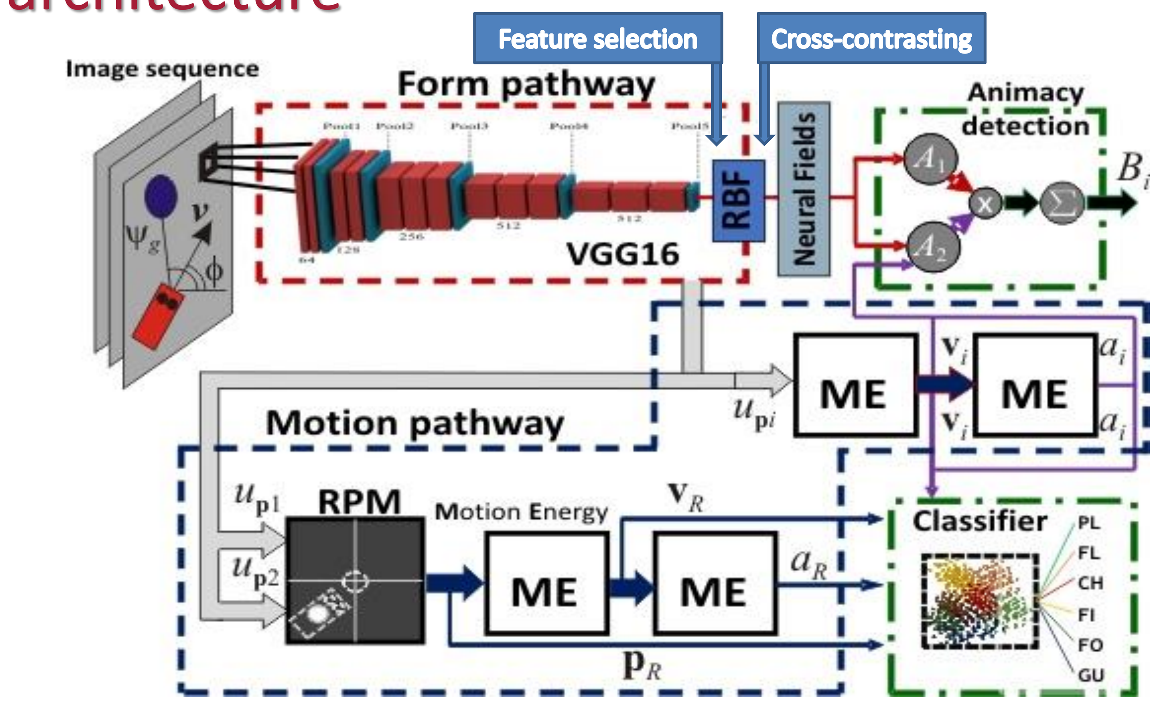


12 different interaction categories (8 best recognized classes):
Avoiding, Fithing, Chasing, Pushing, Dodging, Flirting, Walking (together), Tug of War

References

1. Heider, F. , Simmel, M.: **An Experimental Study of Apparent Behavior.** *The American Journal of Psychology* (1944).
2. Tremoulet, P.D., Feldman, J.: **Perception of animacy from the motion of a single object**. *Perception 29*, 943–951 (2000).
3. Tremoulet, P.D., Feldman, J.: **The influence of spatial context and the role of intentionality in the interpretation of animacy from motion**. *Perception and psychophysics* (2006).
4. Hernik, M., Fearon, P., & Csibra, G.: **Action anticipation in human infants reveals assumptions about anteroposterior body structure and action**. *Proc. Biological Sciences* (2014).
5. Gao, T. , Scholl, B. J.: **Perceiving animacy and intentionality**. In Rutherford, M.D. and Kuhlmeier, V.A., editors, *Social Perception*. The MIT Press (2013).
6. Scholl, B.J., Tremoulet, P.D.: **Perceptual causality and animacy.** *Trends in Cognitive Sciences*, 4(8):299–309 (2000).
7. McAleer, P., Pollick, F.E.: **Understanding intention from minimal displays of human activity**. *Behavior Research Methods, 40*, 830–839 (2008).
8. Fajen, B.R., Warren, W.H.: **Behavioral dynamics of steering, obstacle avoidance, and route selection**. *Journal of Experimental Psycholology: HPP* (2003).
9. Giese, M.A., Poggio, T.: **Neural mechanisms for the recognition of biological movements.** *Nat. Rev .Neurosci. 4*, 179–192 (2003).
10. Blythe, P.W., Todd, P.M., & Miller, G.F.: **How motion reveals intention: Categorizing social interactions.** *Simple heuristics that make us smart*, 257-285 (1999).
11. Baker, C.L., Saxe, R., Tenenbaum, J.B.: **Action understanding as inverse planning**. *Cognition, R.L. and Higher Cognition*, 113, 329–349 (2009).
12. Simonyan, K., Zisserman, A.: **Very Deep Convolutional Networks for Large-Scale Image Recognition**. *ICLR.* (2015).
13. Hovaidi-Ardestani, M., Saini, N., Martinez, A. & Giese, M.A.: **Neural model for the visual recognition of animacy and social interaction**. *ICANN,* Greece (2018).
14. Salatiello, A., Hovaidi-Ardestani, M. & Giese, M.A.: **A Dynamical Generative Model of Social Interactions**. *Frontiers in Neurorobotics*, *15*, 62 (2021).

Model architecture



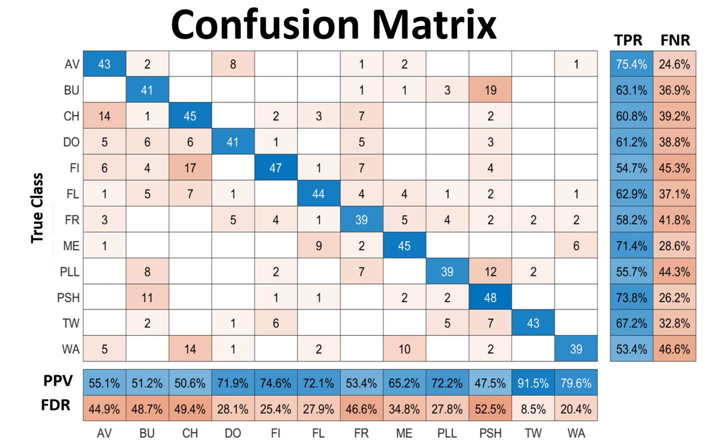
Neural model architecture extends standard model of the visual pathway by neural circuit that analyze perceived agency and classifies social interactions. Mid-level features recognized by first ***5 layers of* VGG16** (Simonyan & Zisserman, 2015). This module is trained for the ImageNet visual features detection. ***LDA-based weighted PCA*** is used for the stimulus-vs-background feature selection of the outputs of VGG16. ***RBF network*** recognizes position and orientation of agents for specific keyframes. ***Positive ICA based cross-contrasting*** of two agents channels enhances the position discriminative estimation. ***Neural Field/RNN*** used for the stabilization of agent tracking in the video sequence, by suppression of wrong detections.

- Hierarchical neural network with two pathways analyzing form and motion features.
- Mid-level features extracted by first 5 layers of VGG16, followed by discriminative feature selection, RBF mapping and 2-channels cross-contrasting, followed by the robust 2D tracking of position by Neural Field.
- Two top levels compute perceived animacy and classify perceived interaction.
- The choice of features for agency judgements was driven by results in the psychophysical literature: *absolute velocity and acceleration of agents, relative distance, velocity, and acceleration* (cf. McAleer & Pollick, 2008).
- Testing multiple types of classifiers at the top level.

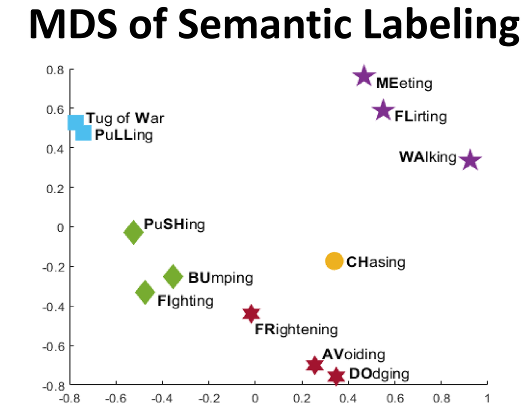
Psychophysical Experiment

- Fee labelling task: participants assigned descriptions to each test video freely.
- Classification task: (new) subjects classified using the most frequently chosen labels from free labelling task.
- Semantic similarity task: (New) participants rated (Likert scale) the pairwise similarity of the category labels.

Confusion Matrix



MDS of Semantic Labeling

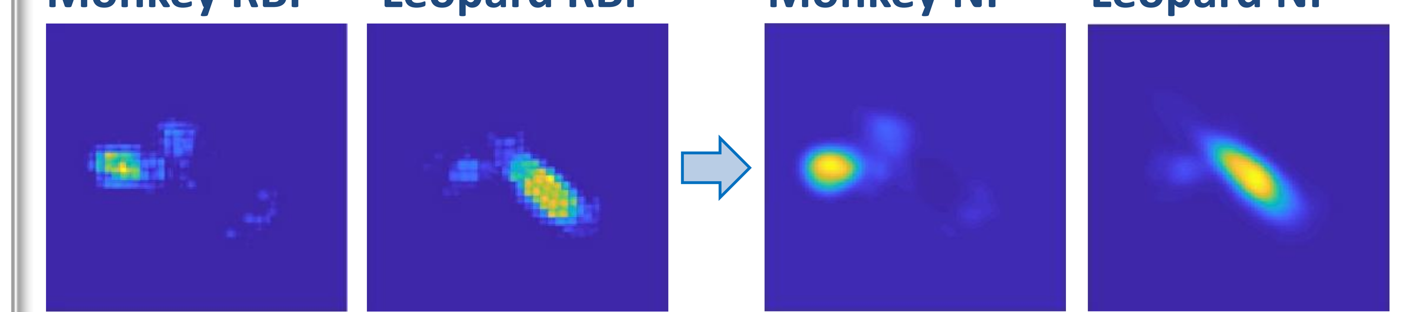


- Reliable classification, way above chance level.
- MDS results indicate that misclassified labels are semantically similar.
- Classes of semantically similar actions can be distinguished from videos.

Tracking of realistic stimuli

Realistic movies with articulating animals: Monkey follows leopard. The sequences generated by our realistic behaviors simulator are animated in Autodesk Maya. Sequence of snapshots sampled every 0.33 seconds.

Monkey RBF **Leopard RBF** **Monkey NF** **Leopard NF**



Activity of the neurons in the RBF net-works that detect the two agents (without enhanced feature selection).

Activity of the corresponding neurons in the neural field (without preceding cross-contrasting, *see next section*).

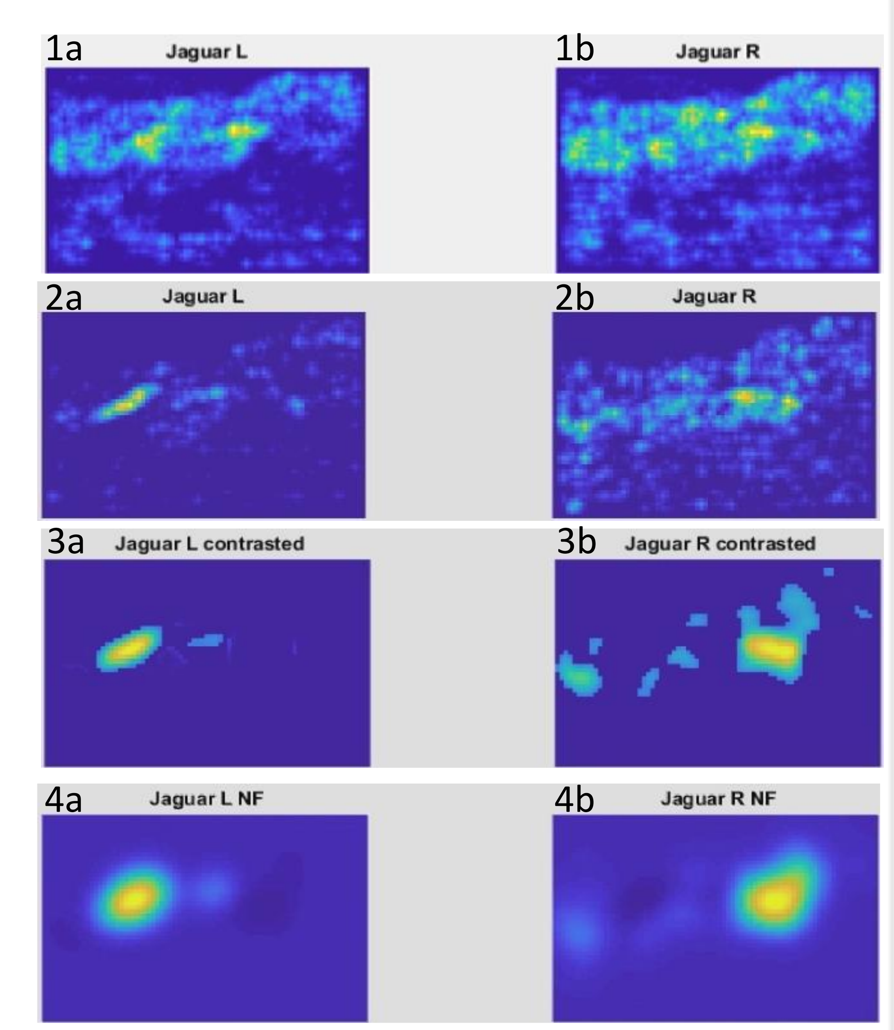
Tracking in the natural environments

Scenes used for the extension of the model to real videos including natural backgrounds.

The examples of feature selection and cross-contrasting are shown below for the first snapshot of the scene above.

__Discriminative feature selection:__ as input it takes the local populations of feature detectors of the last layer of VGG16. Feature selection is based on a combination of LDA (separating agent features from the background) and an appropriate rescaling of features in PCA space.

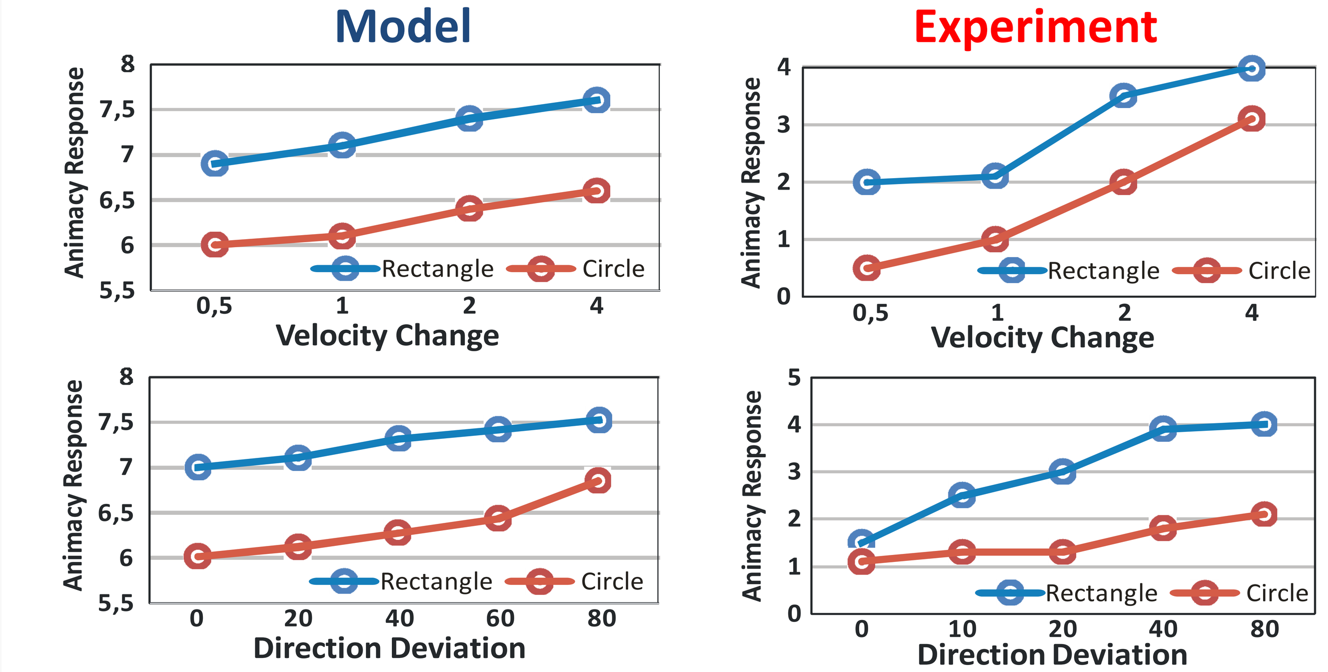
__Cross-contrasting:__ In order to separate the patterns generated by the two interacting agents we exploited a simplified form of positive ICA, resulting in clearly separated activity patterns for the individual agents.



1a & **1b**: the output of RBF, without discriminative feature selection.
2a & **2b**: the output of RBF, with LDA-based feature selection.
3a & **3b**: 2-channel cross-contrasting of the RBF network outputs (*2a*, *2b*).
4a & **4b**: NFs activation

Results on abstract stimuli

Perception of animacy from the motion of a single object (Tremoulet, Feldman 2000)



- Consistent with the psychophysical results, activity of the output 'agency neuron' increases with size of velocity and direction changes of the agent.
- Reproduction of increased animacy perception for stimuli that have a body axis, as opposed to a moving circle (which does not have a body axis), if motion is aligned with body axis.

Social interaction classification

- 6 social interactions regularly used in psychophysics.
- Highest confusion rates between 'flirting' and 'chasing'; sometimes also 'playing' and 'guarding'.
- Minimum achieved accuracy: 94 %; best classification result with linear support vector machine: 99 %.
- All original videos from McAleer and Pollick (2008) were classified correctly, even though they were not part of the training set.

Accuracy: different classifiers

Classifier	Accuracy
Linear SVM	99.0%
Gaussian kernel SVM	96.3%
LDA	94.7%
KNN	94.7%
Nonlinear LDA	94.3%
Neural Network	94.0%

Conclusions

- ✓New psychophysically validated simulator generates 12 reliably distinguishable categories of social interactions.
- ✓Simple physiologically plausible neural model reproduces several important characteristics of human agency perception and of social interaction recognition from abstract displays.
- ✓Model suitable also for the recognition of articulating bodies and the real animals in a rich natural backgrounds.
- ✓Model makes precise predictions about the behavior of neurons involved in interaction perception, which can be verified in electrophysiological experiments.

Acknowledgements

This work was supported by: ERC 2019-SyG-RELEVANCE-856495; HFSP RGP0036/2016; BMBF FKZ 01GQ1704; DFG GZ: KA 1258/15-1; SSTeP-KiZ BMG: ZMWI1-2520DAT700, and NVIDIA Corporation.