Neural Model for the Recognition of Agency and Interaction from Motion

Mohammad Hovaidi Ardestani[1,2], Nitin Saini[1,2], and Martin Giese[1]

[1] Section for Computational Sensomotorics, CIN&HIH, Department of Cognitive Neurology, Univ. Clinic Tübingen, Germany.
[2]IMPRS for Cognitive and Systems Neuroscience, Univ. Clinic Tübingen, Germany

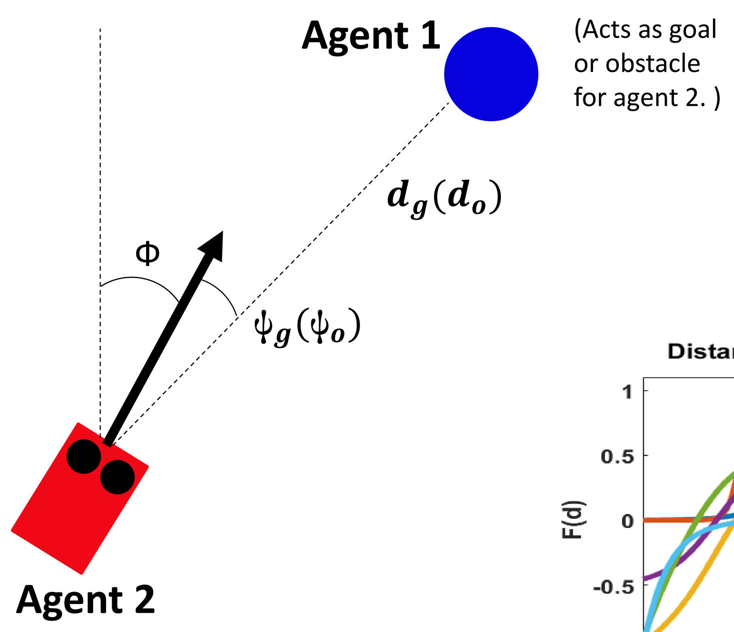
Introduction

- Humans reliably attribute social interpretations to highly impoverished stimuli, such as interacting geometrical shapes (Heider and Simmel, 1944).
- Perception of animacy from such 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, such as 'chasing' or 'fighting' has been addressed in several studies (Gao and Scholl 2013; Scholl and Tremoulet 2000; McAleer and Pollick 2000, Blythe et al. ,1999); a set of six types of interactive movements has been repeatedly used in these studies.
- This perception of interaction has been explained by high-level cognitive processes, such as probabilistic reasoning and inference.(Baker et al., 2009)
- Building on classical biologically-inspired models for object and action perception (Riesenhuber and Poggio, 1999; Giese and Poggio, 2003), we propose a learning-based hierarchical neural network model that analyses such stimuli based shape and motion features directly from video sequences.
- The model has a simple feed-forward architecture and comprises two processing streams for form and object motion in the retinal frame of reference.
- The model contains only simple physiologically plausible operations.

Goal of our work

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

Generation of Stimuli



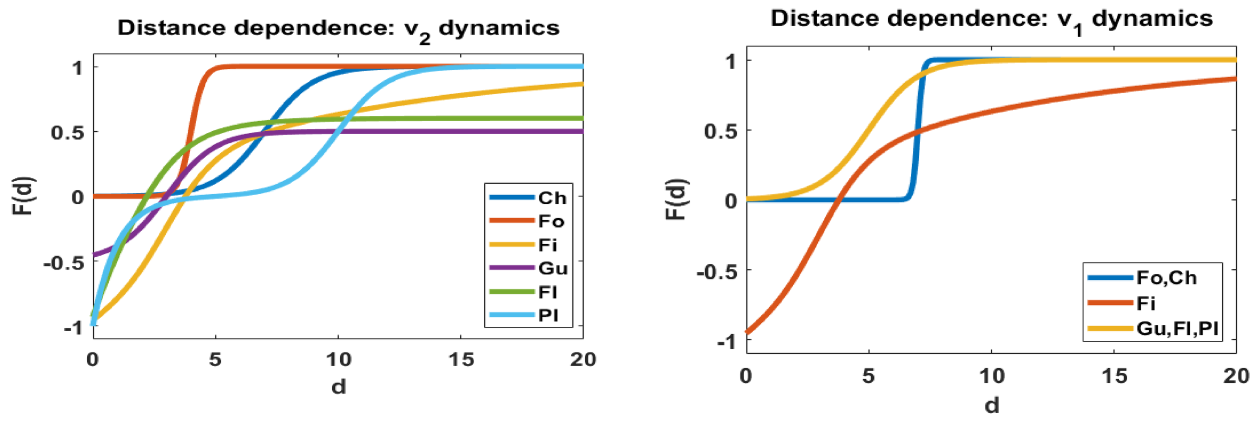
- Dynamics of heading direction (Fajen and Warren 2003):

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

- Dynamics of forward speed:

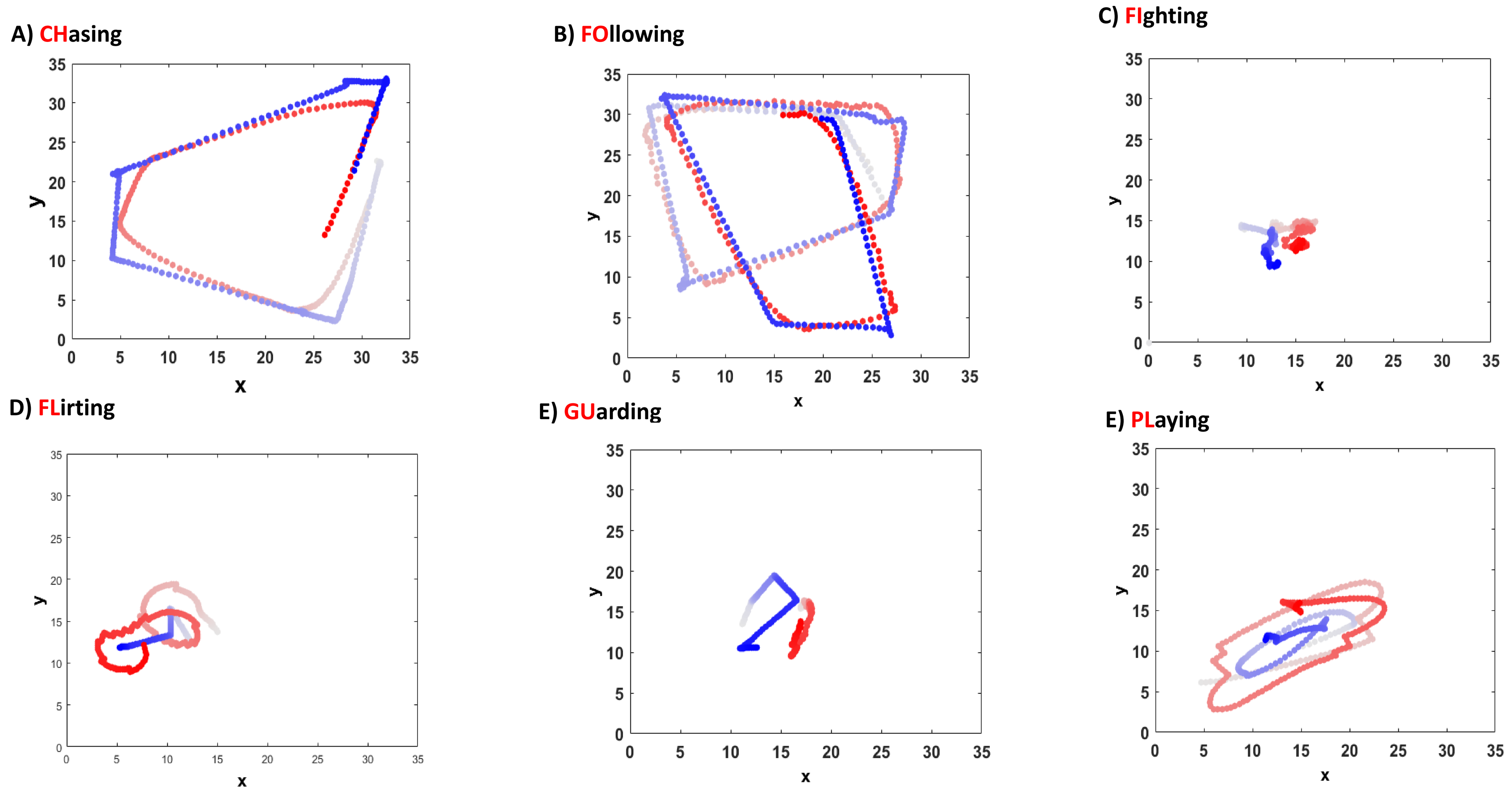
$$\tau\dot{v}_i = -v_i + F_i(d) + c_i\,\varepsilon_i(t)$$

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

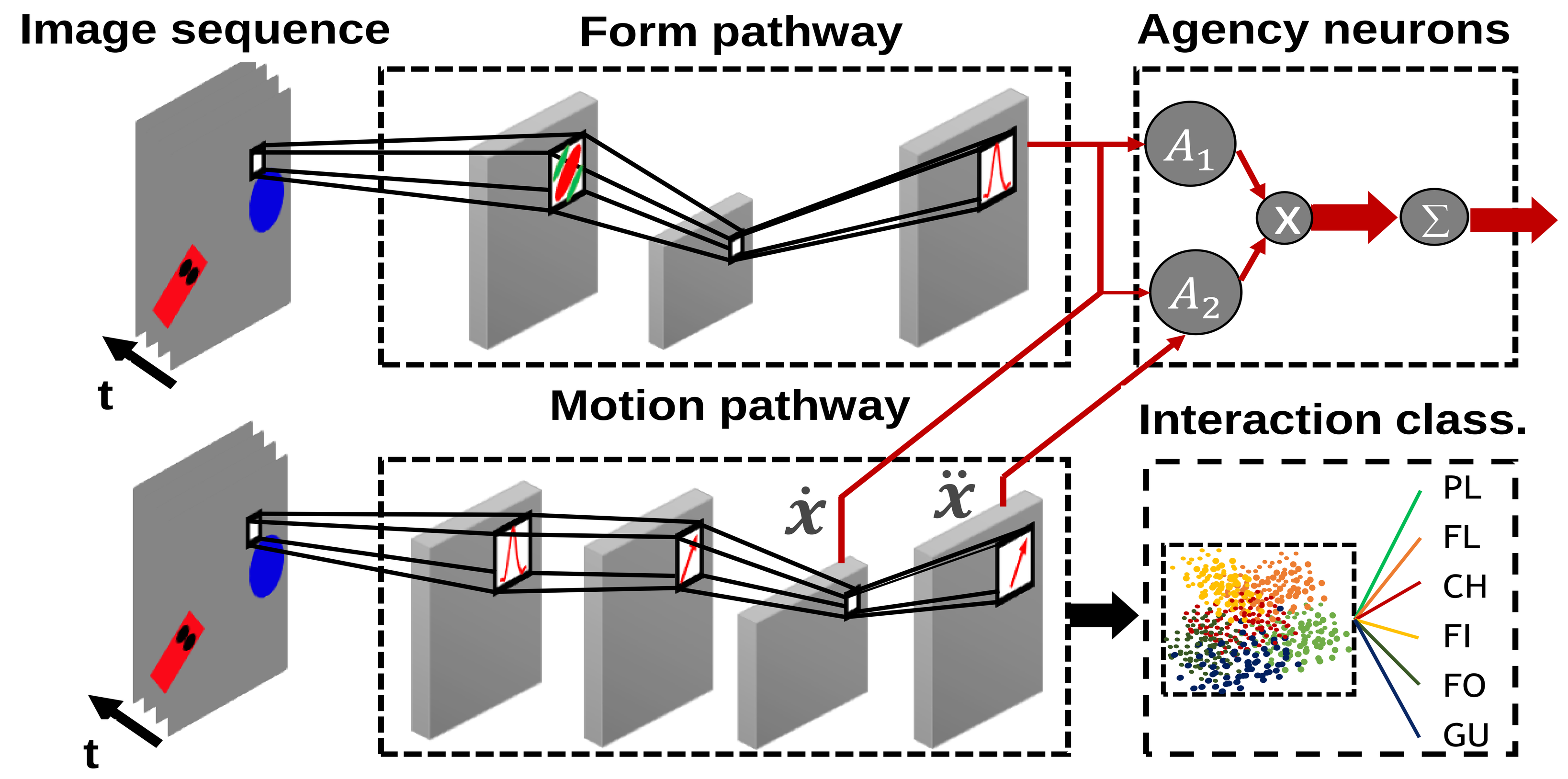


Sample trajectories from different intention categories

(Agent 1: blue, agent 2: red. Color saturation indicates time, the color fading with time.)



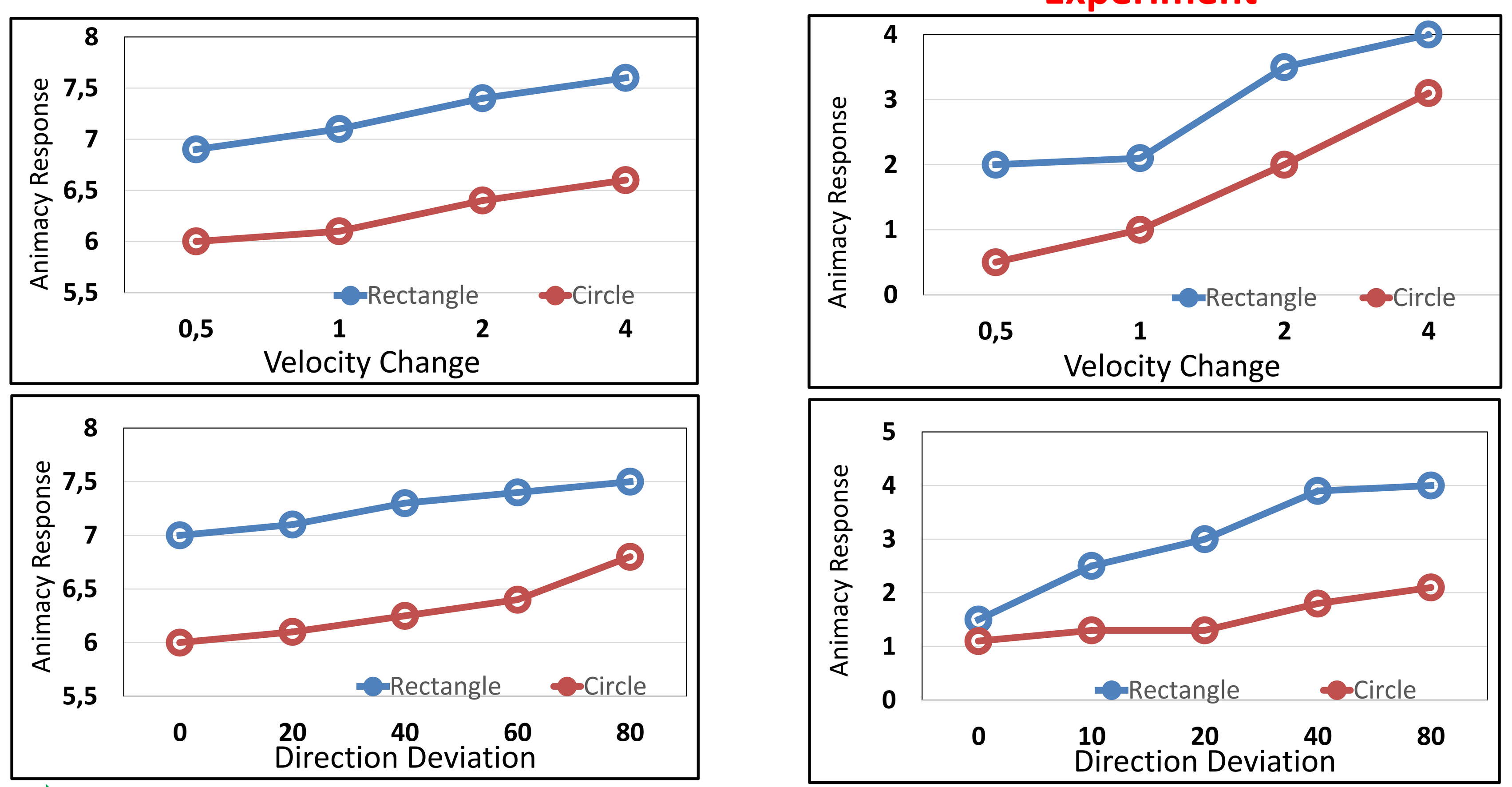
Model architecture



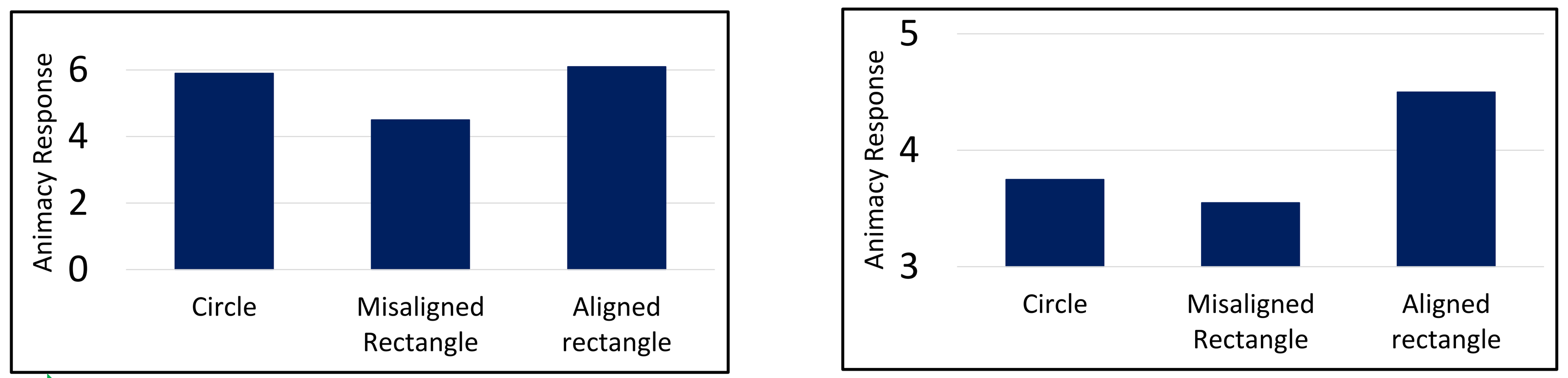
- Hierarchical neural network with two pathways that analyze form and motion features.
- Two top levels that compute perceived animacy and classify perceived interaction.
- The choice of features for the computation of agency judgements was driven by results in the psychophysical literature.
- Critical features: absolute velocity and acceleration of agents, relative distance, velocity, and acceleration (McAleer and Pollick 2008).
- Testing multiple types of classifiers at the top level.

Results

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



Consistent with the psychophysical results, the activity of the output 'agency neuron' increases with size of velocity and direction changes (testing trajectories where the agent followed a line and then suddenly ' changed direction or speed).

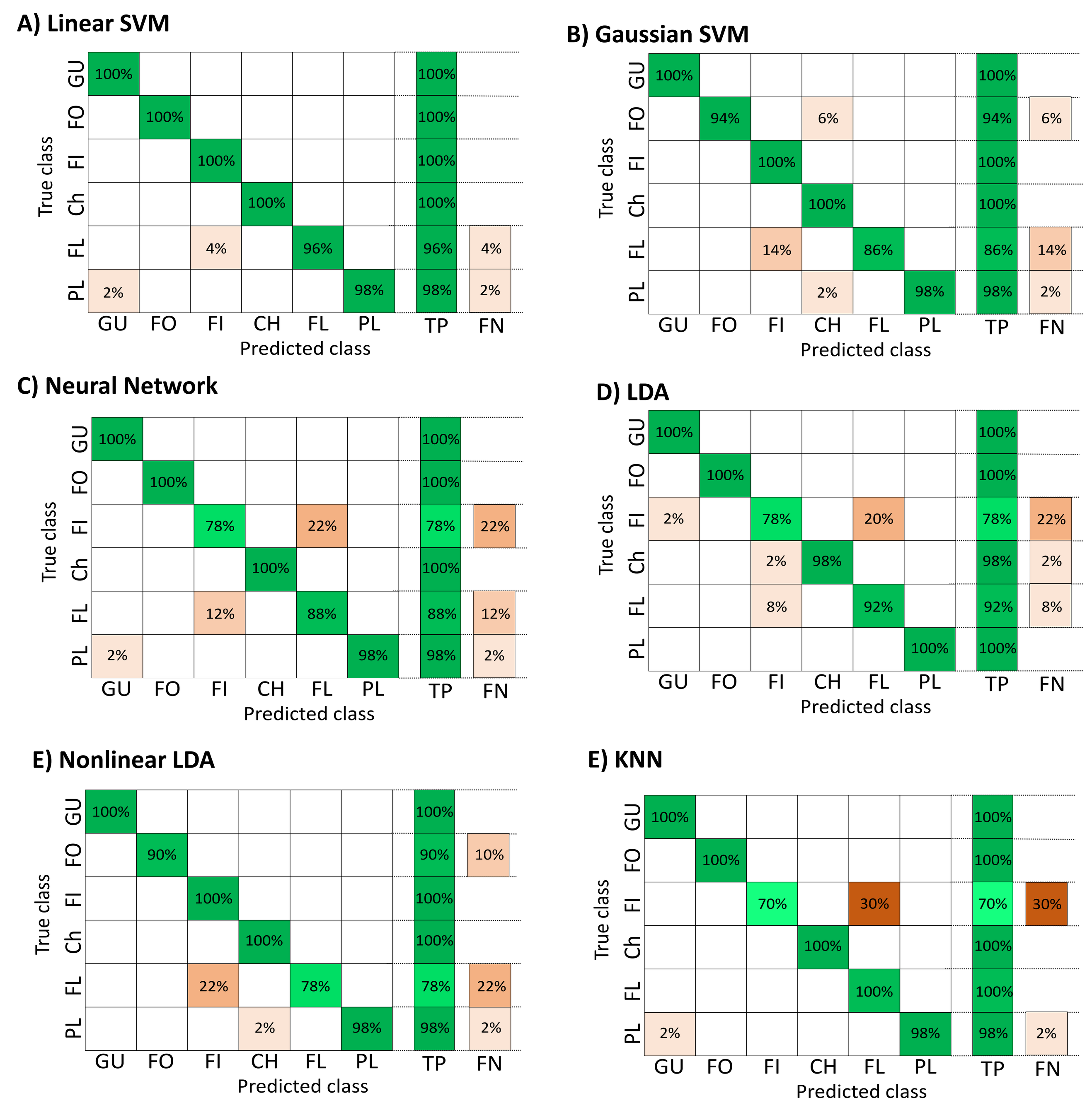


Reproduction of increased animacy perception, compared to a moving circle (that does not have a body axis), if object has a body axis that is aligned with its velocity vector.

Social interaction classification

Confusion matrices for the six tested classifiers

TP: true positive rate, FN: false negative rate. 50 videos per class.



- Highest confusion rates between 'flirting' and 'chasing'; sometimes also 'playing' and 'guarding'.
- For all classifier types accuracy is at least 94 %, best classification result is obtained with linear support vector machine, reaching an accuracy of 99 %.
- All original videos from McAleer and Pollick (2008) were classified correctly, though they were not part of the training set.

Accuracy of 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

✓ While our model is quite simple-minded and does not use any sophisticated computations it was able to reproduce several important characteristics of human perception of agency and of social interactions from strongly impoverished displays.
✓ The performance of form and motion recognition has to be improved by embedding deeper hierarchies in the two pathways that are trained with a richer set of image data.

References

1. Heider, F. and 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. and 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., and Csibra, G.: Action anticipation in human infants reveals assumptions about anteroposterior body structure and action. Proceedings. Biological sciences (2014)
5. Gao, T. and 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. and 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. Riesenhuber, M., Poggio, T.: Hierarchical models of object recognition in cortex. Nat. Neurosci. 2, 1019–1025 (1999)
11. Blythe, P. W., Todd, P. M., & Miller, G. F. : How motion reveals intention: Categorizing social interactions. Simple heuristics that make us smart ,pp. 257-285(1999)
12. Baker, C.L., Saxe, R., Tenenbaum, J.B.: Action understanding as inverse planning. Cognition, Reinforcement learning and higher cognition 113, 329–349 (2009)

Acknowledgements

This work was supported by: HFSP RGP0036/2016; the European Commission HBP FP7-ICT2013-FET-F/ 604102 and COGIMON H2020- 644727, and DFG KA 1258/15-1.