Neural model for the visual recognition of agency and social interaction

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

1) Section Computational Sensomotorics, Department of Cognitive Neurology, CIN&HIH, University Clinic Tübingen, Germany

2) IMPRS for Cognitive and Systems Neuroscience, Univ. of Tübingen, Germany

INTRODUCTION: Humans derive spontaneously judgements about agency and social interactions from strongly impoverished stimuli, as impressively demonstrated by the seminal work by Heider & Simmel (1944). The neural circuits that derive such judgements from image sequences are entirely unknown. It has been hypothesized that this visual function is based on high-level cognitive processes, such as probabilistic reasoning. Taking an alternative approach, we show that such functions can be accomplished by relatively simple computations that can be implemented by physiologically plausible neural circuits, exploiting an appropriate hierarchical (deep) neural model of the visual pathway.

METHODS: Using a deep neural network for the construction of dictionaries of low and mid-level feature detectors, we built a hierarchical neural model that reproduces elementary psychophysical results on animacy and social perception from abstract stimuli. The lower hierarchy levels of the model consist of position-specific neural feature detectors that re selective for oriented contours and intermediately complex shape features. The next-higher level is formed by position-variant neurons that are selective for shapes and their orientation in the image plain. The output of these neurons is processed by a form and a motion pathway that computes the body axis, the relative positions, speeds and accelerations of moving agents, exploiting established neural circuits (gain fields, motion energy detectors). The top level of the model combines these extracted features using simple feed-forward neural circuits and elementary classifiers.

RESULTS: Based on input video sequences, the model successfully reproduces results of Tremoulet and Feldman (2000) on the dependence of perceived animacy on motion parameters and the body axis. The animacy percept is stronger for agents that abruptly change their direction and speed, and if the movement is aligned with the body axis. In addition, the model classifies correctly six categories of social interactions that have been frequently tested in the psychophysical literature (following, fighting, chasing, playing, guarding, and flirting) (e.g. Scholl & McCarthy, 2012; McAleer et al., 2008). In addition, we show that the model can be extended for the processing of simple interactions in real-world movies.

CONCLUSION: Using simple physiologically plausible neural circuits, the model accounts simultaneously for a variety of effects related to animacy and social interaction perception. Even in its simple form the model proves that animacy and social interaction judgements partly might be derived by very elementary operations in hierarchical neural vision systems, without a need of sophisticated or accurate probabilistic inference. The model makes specific predictions about neurons involved in the visual processing of abstract social stimuli.

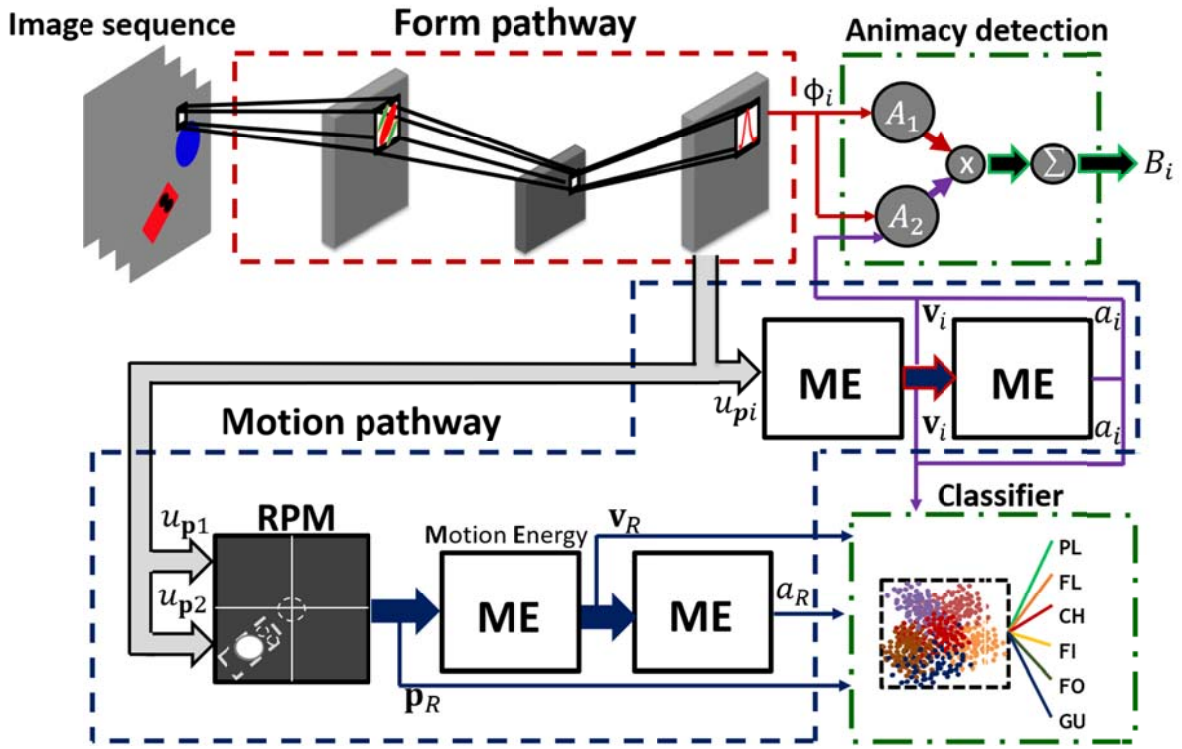

Fig.1. Model consisting of a form and a motion pathway. ME signifies a layer of motion energy detectors, and RPM the relative position map. The top level of the model is formed by neural detectors for the perceived animacy, and a network that classifies six different types of interactions.

Acknowledgements:

This work was supported by: HFSP RGP0036/2016; the European Commission HBP FP7-ICT2013-FET-F/ 604102 and COGIMON H2020-644727, and the DFG GZ: GI 305/4-1 and KA 1258/15-1.