Neural models for the (cross-species) recognition of dynamic facial expressions

Michael Stettler, Nick Taubert, Ramona Siebert, Silvia Spadacenta, Peter Dicke, Peter Thier, Martin A. Giese

Dynamic facial expression recognition is an essential skill of primate communication. While the neural mechanisms of the recognition of static pictures of faces has been extensively investigated, the neural circuits of the recognition of dynamic facial expressions remain largely unclear. We studied possible neural encoding mechanisms by neural modelling, exploiting highly controlled and realistic stimulus sets that were generated by computer graphics, and which are also used in electrophysiological experiments.

METHODS: Combining previous physiologically plausible neural models for the recognition of dynamic bodies (Giese & Poggio, 2003) and of static faces (Giese & Leopold, 2005) with architectures from computer vision (Simonyan, 2014), we devised two models for the recognition of dynamic facial expressions. The first model is *example-based* and encodes dynamic faces as temporal sequences of snapshots, which are recognized by a sequence-selective recurrent neural network. The second architecture is based on a *norm-referenced encoding* mechanism. Individual face pictures are neutrally encoded by face-space neurons that are tuned to the differences between the actual stimulus frame and a reference face showing a neutral facial expression. Dynamic expressions can be recognized by a simple circuit that differentiates the responses of these face-space neurons. Both models were tested using movies of highly realistic human and monkey face avatars that were animated using motion capture data from humans and monkeys. Expression strength and style was precisely controlled using motion morphing techniques (Taubert et al. 2020).

RESULTS: Both models recognize reliably dynamic facial expressions of humans and monkeys from movies. They make quite different predictions for the behaviour of face-selective neurons, especially for stimuli that interpolate between different expressions. The norm-referenced model shows a highly gradual, almost linear dependence of the neuron activity with the expressivity of the stimuli, which is not the case for the example-based model. In addition, we also explored in how far the models account for the experimental observation (Taubert et al. 2020) that humans recognize human expressions on monkey faces spontaneously without any prior training on such stimuli.

CONCLUSIONS: Both models are physiologically plausible and accomplish the recognition of dynamic faces from movies. The models make very different predictions about the behaviour of face-tuned single cells. Norm-referenced encoding might support the generalization of expressions across different basic face shapes, and even across faces from different species.

Acknowledgements: This work was supported by HFSP RGP0036/2016, EC CogIMon H2020 ICT-23-2014/644727 and ERC 2019-SyG-RELEVANCE-856495. MG was also supported by BMBF FKZ 01GQ1704 and BW-Stiftung NEU007/1 KONSENS-NHE. RS, SS, PD, and PT were supported by a grant from the DFG (TH 425/12-2), NVIDIA Corp.