Physiologically-inspired neural model for the processing of dynamic facial expressions

M.A. Giese, G. Ravishankar, S. Safavi, D. Endres

Facial expressions are essentially dynamic. However, most existing research has focused on static pictures of faces. The computational neural functions that underlie the processing of dynamic faces are largely unknown. Combining multiple physiologically relevant neural encoding principles, we propose a neural model that accomplishes the recognition of facial expressions robustly over different facial identities.

Our model is based on a physiologically plausible hierarchical model of the ventral stream for the extraction of form features, building on a previous model for the processing of identity from static pictures of faces [Giese & Leopold, 2005, Neurocomputing]. It combines norm-referenced as well as example based coding of patterns, and different physiologically-inspired mechanisms for the encoding of temporal sequences.

In example-based coding, 'snapshot neurons' that are selective for frames (snapshots) form the dynamic face sequence, they are modeled by radial basis function units (see figure). These neurons are laterally coupled, resulting in a network which is a dynamic neural field with an asymmetric interaction kernel. This makes the snapshot neurons sequence selective: we find only a weak response if frames occur in an incorrect temporal order. Facial expression neurons at highest level sum activity over the neural field that encodes one facial expression (e.g. ‘happy’ or ‘sad’).

In norm-referenced encoding, face-selective neurons encode distance and direction of the stimulus relative to a norm stimulus, here neutral expressions. This computational function can be implemented by a simple feed-forward neural network [Giese & Leopold, 2005, Neurocomputing]. For static face processing this norm-referenced mechanism accounts better for the neurophysiological data than an example-based mechanism. In the dynamic case, the evolution of facial expression corresponds to a vector with increasing length in the direction of the extreme expression; face neurons show monotonic increases (or decreases) of activity during the time-course of the expression. Their output is fed into ‘differentiator neurons’ which are detecting raising flanks in their input, thus becoming selective to dynamic facial expressions in the correct temporal order, while they fail to respond to static expressions and ones with inverse temporal order. This proposed mechanism is more efficient in terms of neural hardware, since it encodes only neutral faces and the extreme expressions.

The model is tested with movies showing real monkey expressions (‘threat’ and ‘coo-call’) and a standard data basis containing a large number of human expressions of different individuals. The performance of different physiologically plausible circuits for the recognition of dynamic facial expressions is evaluated, and specific predictions for the behavior of different classes of dynamic face-selective neurons are discussed, which might e.g. be suitable to distinguish different computational mechanisms based on single-cell recordings from dynamic face-selective neurons.


Acknowledgements: DFG, FP7-ICT-215866 SEARISE, FP7-249858-TP3 TANGO, FP7-ICT-248311 AMARSi.
Mechanisms for encoding of dynamic facial expressions. A: example-based, B: norm-reference encoding