Physiologically-inspired Neural Circuits for the Recognition of Dynamic Faces

Michael Stettler[1,2], Nick Taubert[1], Tahereh Azizpour[1], Ramona Siebert[1], Silvia Spadacenta[1], Peter Dicke[1], Peter Thier[1] and Martin A. Giese[1]

[1] Section for Computational Sensomotorics, Dept. of Cognitive Neurology, Centre for Integrative Neuroscience & Hertie Inst. for Clinical Brain Research, University Clinic Tübingen, Germany

[2] Internat. Max Planck Res. School for Intelligent Systems, Tübingen, Germany

michael.stettler@cin.uni-tuebingen.de, martin.giese@uni-tuebingen.de

Abstract. Dynamic faces are essential for the communication of humans and non-human primates. However, the exact neural circuits of their processing remain unclear. Based on previous models for cortical neural processes involved for social recognition (of static faces and dynamic bodies), we propose two alternative neural models for the recognition of dynamic faces: (i) an example-based mechanism that encodes dynamic facial expressions as sequences of learned keyframes using a recurrent neural network (RNN), and (ii) a norm-based mechanism, relying on neurons that represent differences between the actual facial shape and the neutral facial pose. We tested both models exploiting highly controlled facial monkey expressions, generated using a photo-realistic monkey avatar that was controlled by motion capture data from monkeys. We found that both models account for the recognition of normal and temporally reversed facial expressions from videos. However, if tested with expression morphs, and with expressions of reduced strength, both models made quite different prediction, the norm-based model showing an almost linear variation of the neuron activities with the expression strength and the morphing level for cross-expression morphs, while the example based model did not generalize well to such stimuli. These predictions can be tested easily in electrophysiological experiments, exploiting the developed stimulus set.

Keywords: Dynamic facial expressions · recognition · neural network model · norm-referenced encoding · visual cortex

1 Introduction

Dynamic facial expressions are central for the social communication of humans and non-human primates. In spite of this importance, the underlying detailed local neural circuits remain unclear. Single cells responding to dynamic facial expressions have been investigated so far only in very few studies in the superior temporal sulcus (STS) [1, 2], and the amygdala [3]. Advances in the technology

for stimulus generation, e.g the use of highly controlled dynamic avatars, in combination with the simultaneous recording of large numbers of neurons promise to clarify the underlying mechanisms. For the guidance of such experiments it seems useful to develop theoretical hypotheses about possible underlying neural computations and circuit structures for the recognition of dynamic faces.

We propose here two alternative models for the recognition of dynamic faces. Both models are derived from previous neural models that have provided good agreement with electrophysiological data from single-cell recordings in visual areas, either on the recognition of face identity in area IT [4, 5] or for the representation of head actions in area F5 of the macaque monkey [6, 7]. From these previous models we derive two alternative mechanisms for the recognition of dynamic facial expressions. We test these models with a highly controlled stimulus set of monkey facial expressions. We demonstrate that both models are feasible and derive predictions that can be used to distinguish the two models in electrophysiological experiments.

After a brief review of related work, we introduce in the following the model architectures and our simulation studies, followed by a discussion of implications for electrophysiological experiments.

2 Related work

Most physiologically-inspired models on the processing of faces investigate the recognition of static faces (e.g. [8]). Biologically-inspired models have been proposed for the recognition of dynamic bodies [9–11]. It seems likely that computational principles might shared between the processing of different social stimulus classes. Beyond this, a variety of conceptual models have been developed in psychology and the funcional imaging literature, e.g. the separate processing of static and dynamic aspects of faces [12], or the idea of a face space [13] and a norm-referenced ancoding of static faces relative to an average face (review see [14]). This work sets important general constraints for the modelling, but does not provide ideas about specific neural circuits. Dominant approaches on dynamic face processing in computer vision are based on deep recurrent neural networks [15], but typically not related to details of the brain.

3 Model architectures

Our model mimics the basic architecture of the visual pathway, from the retina up to higher-levels that contain neurons that are selective for dynamic facial expressions. Figure 1 A shows an overview of the model architecture. In the following, we first describe the mid-level feature extraction hierarchy, which is identical for both model versions. Then we discuss separately for the two models the circuits including the face-selective neurons that implement different computational principles.

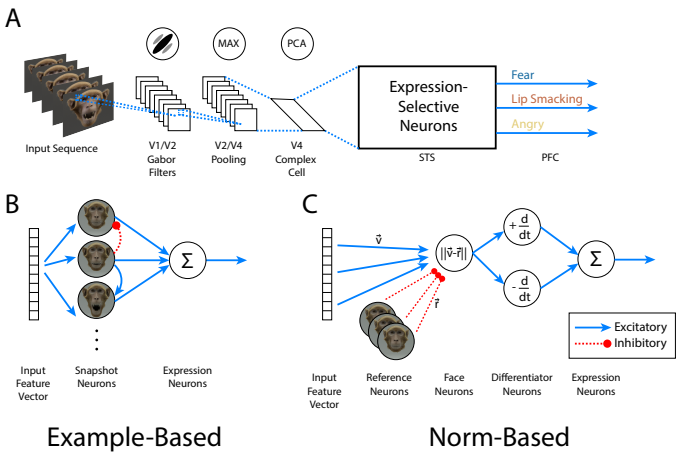

Fig. 1: Model architectures. A) Mid-level feature extraction hierarchy (common to both models). B) Example-based circuit, and C) Norm-based circuit for the recognition of one dynamic expression.

3.1 Shape feature extraction hierarchy

The first levels of our model hierarchy extract mid-level form features, similar to neurons area V4 in the macaque visual cortex. While more sophisticated models have been proposed for this part of the visual pathway (e.g. [16]), we applied here a highly simplified implementation that was sufficient for our stimuli since we wanted to focus on higher-level face-selective circuits, whose behavior is to some degree invariant against the chosen mid-level feature dictionary that is used as input. The feature extraction hierarchy of our model can be easily exchanged against more elaborate models from the literature. Our feature extraction hierarchy consists of 3 layers. The face region in our stimulus movies was down-sampled 200 x 200 pixels and converted to gray level for further processing.

First layer. The first layer consists of even and uneven Gabor filters with 8 different orientations, and 3 different spatial scales that differed by $\sqrt{2}$. A constant was subtracted from these filter functions to make them mean-free. Filter responses for the three spatial scales were computed on rectangular grids with 49, 69, and 97 points spaced equally along the sides of the image region. This layer models orientation-selective neurons such as V1 simple cells [17].

Second layer. The second layer models V1 complex cells and makes the responses of the first layer partially position- and spatial phase-invariant. For this purpose, the thresholded responses of the even and the uneven Gabor filters with the same orientation preference and spatial scale were pooled within a spatial region (receptive field) using a maximum operation. The receptive fields of these pooling neurons comprised 3 neurons in the precvious layer (respectively one

for the largest spatial scale). The receptive fields centers of the pooling neurons were again positioned in quadratic grids with 15, 10 and 14 grid points for the three spatial scales.

Third layer. The third layer of our model extracts informative mid-level features, combining a simple heuristic feature selecting algorithm with a Principal Components Analysis (PCA). These two steps can be implemented by a sparsely connected simple linear feed-forward neural network. For feature selection, we simply computed the standard deviation over all input signals from the previous layer (after thresholding) over our training set. Only those features were retained for which this variability measure exceeded a certain threshold. This eliminated uninformative features that are zero or constant over the training set. In total we retained 17 % of the original features.

The vector of the selected features was then subject to a PCA analysis. The activity of the selected features was projected to a subspace that corresponds to 97% of the total variance for the training set. The resulting (thresholded) PCA features provide the input to the expression-selective neurons that form the next layer of our model.

3.2 Expression-selective neurons

The next layers of the model were implemented in two different ways, implemennting two different computational principles. The first mechanism is based on the recognition of temporal sequences of learned snapshots from face movies by a recurrent neural network (RNN), referred to as *example-based* mechanism. The second mechanism is based on the concept of norm-referenced encoding, as discovered in the context of the representation of static images of faces [14]. The dynamic face is encoded by neurons that represent the difference between the actual face picture and a norm picture, in this case the shape of a neutral face. The details of the two implementation are described in the following.

Example-based model. For this model the recognition is accomplished by using a RNN that detects temporal sequences of learned keyframes from facial expressions (Fig. 1 B). This type of mechanism has been shown to reproduce data from cortical neurons during the recognition of body actions (e.g. [9, 7]. It has been show to reproduce activity data from action selective neurons in the STS and in premotor cortex.

Expressions were represented by a total of 50 keyframes. The structure of the example-based encoding circuit for one expression is shown in Fig. 1 B. The output from the previous layer, signified as vector $\mathbf{z}$, is providing input to Radial Basis Function units (RBFs) that were trained by setting their centers to the vectors $\mathbf{z}_n^p$ for the individual expression frames of expression p. The actual outputs of these neurons are then given by: $f_k^p = \exp((|\mathbf{z} - \mathbf{z}_k^p|^2/(2\sigma^2)$. (For sufficent selectivity to distinguish temporally distant keyframes we chose $\sigma = 0.1$.). The outputs of the RBFs were thresholded with a linear threshold function.

The output signals of the RBFs were used as input of a recurrent neural network, or discretely approximated neural field [18], that obeys the dynamics:

$$\tau \dot{u}_n^p(t) = -u_n^p(t) + \sum_m w(n-m)[u_m^p(t)]_+ + s_n^p(t) - h - w_c I_c^p(t) \qquad (1)$$

The activity $u_n^p(t)$ is the activity of the neuron in the RNN that encodes keyframe n of the facial expression type p. The resting level constant was $h = 1$, and the time constant $\tau = 5$ (using an Euler approximation). The lateral interaction was asymmetric, inducing a network dynamics that is sequence selective. It was given by the function $w(n) = A\exp(-(n-C)^2/\ (2\sigma_{\text{ker}}^2)) - B$ with the parameters $A = 1$, $B = 0.5$, and $C = 3.5$. A cross inhibition term leads to competition between the neural subnetworks that encode different expressions. It was defined by the equation: $I_c^p(t) = \sum_{'\neq n}[\mathbf{u}_m^{p'}(t)]_+$ with the cross-inhibition weight $w_c = 0.5$.

The input signals $s_n^p(t)$ was computed from the output signals of the RBF units that recognize the key frames of expression p. These can be described by the vector $\mathbf{b}^p = [b_1^p, ..., b_{50}^p]^T$ for the actual time step. The components of this vector were smoothed along the neuron axis using a Gaussian filter with a width (standard deviation) of 2 neurons.

The neurons in this RNN are called *snapshot neurons* in the following. They are expression-selective and fire phasically during the evolution of the expressions. In addition, they are sequence-selective, i.e. they fire less strongly if the same keyframe is presented as part of temporally inverted sequence.

The thresholded output signals of the snapshot neurons belonging to the same expression are integrated by an *expression neuron* that computes the maximum over all snapshot neuron outputs (cf. Fig. 1 B). These neurons have a time constant $\tau_v = 4$, so that their outputs can be describe by the dynamics:

$$\tau_v \dot{v}_p(t) = -v_p(t) + \sum_n [u_n^p(t)]_+ \qquad (2)$$

The expressions neuron thus fire continuously during the evolution of the corresponding expression p, but only weakly during the other ones.

Norm-based model. The second mechanism for the recognition of dynamic expressions is inspired by results on the norm-referenced encoding of face identity by cortical neurons in area IT [4]. We have proposed before a neural model that accounts for these electrophysiological results on the norm-referenced encoding [5]. We demonstrate here that the same principles can be extended to account for the recognition of dynamic facial expressions.

The circuit for norm-based encoding is shown in Fig.1 C. The idea of norm-referenced encoding is to represent the shape of faces in terms of differences relative to a reference face, where we assume that this is the shape of a neutral expression. In our model postulates a special class of Reference Neurons that represent the output feature vector $\mathbf{z}_0^p$ from the mid-level feature hierarchy for a neutral face picture (e.g. at the beginning of the movies). Representations of this pattern could be established by learning or robust averaging, for example

exploiting the fact that the neutral expression is much more frequent than the intermediate keyframes of dynamic expressions. The prediction from this mechanism is thus the existence of a sub-population of neurons that represent the neutral expression in terms of its multi-dimensional feature vector. The output activities of these neurons is signified by the reference vector $\mathbf{r} = \mathbf{z}_0^p$.
A physiologically plausible way for the encoding of the vectorial difference $\mathbf{d}(t) = \mathbf{z}(t) - \mathbf{r}$ between the actual feature input $\mathbf{z}(t)$ and the reference vector $\mathbf{r}$ can be based on neurons with the tuning function:

$$f_p = \|\mathbf{d}\| \left(\frac{\frac{\mathbf{d}^T}{\|\mathbf{d}\|}\mathbf{n}_p + 1}{2} \right)^{\nu} \tag{3}$$

This expression defines the output activity of a directionally tuned neuron whose maximum firing rate depends linearly on the norm of the difference vector $\mathbf{d}$. The unit vector $\mathbf{n}_p$ defines the preferred direction of the neuron in a multi-dimensional space. The term in the parenthesis is a direction tuning function that is maximal if the difference vector is aligned with this preferred direction $\mathbf{n}_p$. The positive parameter ν determines the width of the direction tuning in multi-dimensional space. As shown by [5], using $\nu = 1$ results in reasonable fits of neural data from area IT. In this special case if $\|.\|$ signifies the 1-norm, this tuning function can be implemented by a simple two-layer network with linear rectifying units:

$$f_p = 0.5\ (\mathbf{1} + \mathbf{n}_p)^T [\mathbf{d}]_+ + 0.5(\mathbf{1} - \mathbf{n}_p)^T [-\mathbf{d}]_+ \tag{4}$$

It has been shown in [5] that the tuning fits the properties of face-selective neurons in area IT. We call this type of neuron *face neuron* in our model. The activity of these neurons is expression-selective and increases towards the frame with the maximum expression strength. However, these neurons also respond to static pictures of faces.

A very simple way to generate from the responses $f_p(t)$ responses that are selective for dynamically changing expressions is to add a simple output network that consists of *differentiator neurons* that respond phasically to increasing or decreasing changes of the activity $f_p(t)$, and to sum the output signals from these neurons to obtain the *expression neuron* output: $v_p(t) = \frac{\mathrm{d}f_p}{\mathrm{d}t} + \frac{\mathrm{d}(-f_p)}{\mathrm{d}t}$ (cf. Fig. 1 C). In fact, such differentiating neurons have been observed and the dependence of this behavior on channel dynamics has been analyzed in detail (e.g. [19]).

4 Results

In the following, we describe briefly the applied stimulus set, and then show a number of simulation results that show that both models are suitable to classify dynamic facial expressions, while they result in fundamentally different prediction of the behavior of single neurons, especially for morphed stimuli.

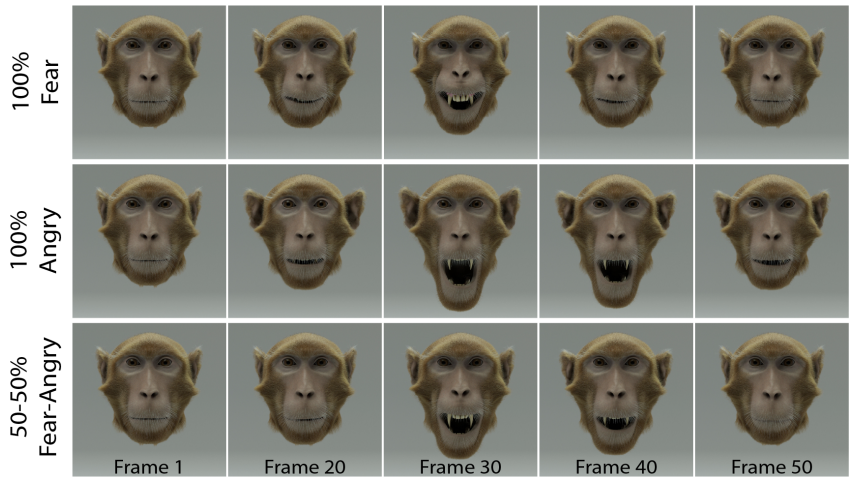

Fig. 2: Monkey Avatar Head. Movies of the three training expressions Fear, Lip Smacking and Angry, and morph (50 % - 50 %) between Fear and Angry.

4.1 Stimulus Generation

In order to test our model we developed a novel highly-realistic monkey head model that was animated using motion capture data from monkeys. The head model was based on an MRI scan of a monkey. The resulting surface mesh model was regularized and optimized for animation. A sophisticated multi-channel texture model for the skin and fur animation were added. The face motion was based on motion capture (VICON), exploiting 43 markers that were placed on the face of a rhesus monkey that executed different facial expressions. By interaction with an experimenter, three expressions (prototypes) were recorded: Fear, Lip Smacking and a Threat/Angry expression (Figure 2). Exploiting a muscle-like deformable ribbon structure, the motion capture data was transferred to the face mesh, resulting in highly realistic face motion. A recent study shows that the generated facial expressions are perceived by animals as almost as realistic as videos of real monkey expressions, placing the method on the 'good side' of the uncanny valley (Siebert et al., in press).

In addition to the original expression, we generated morphs between them exploiting a Bayesian motion morphing algorithm [20]. In addition, expressions with reduced strength were generated by morphing the original expression with a neutral facial expressions. This allowed us to study the behavior of the models for gradual variations between the expressions, and for expressions with reduced strength. Expressions always started with neutral, evolved to a maximally expressive frame, and went back to the neutral face. Natural durations of the expressions were between 2 and 3s. All expressions were time-normalized to 50 frames for our analysis of the model. For the experiments reported here, we used

the prototype expressions, and expressions with reduced strength that included 25-50-75-100 % of each prototype. In addition, morphs between the expressions Fear and Angry with contributions of 0-25-50-75-100 % of the Fear prototype were generated.

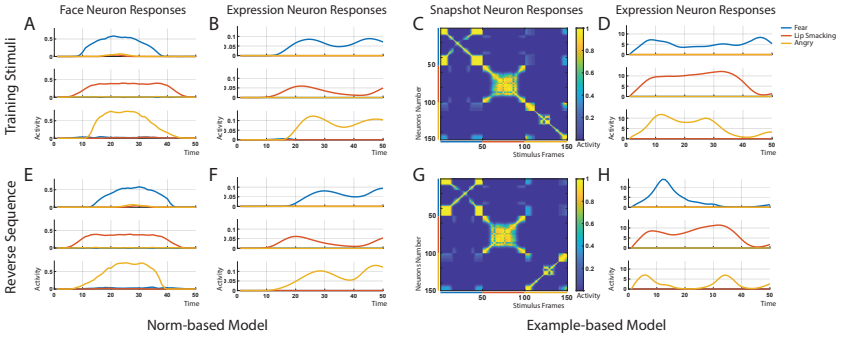

Fig. 3: Activity of model neurons. Presentation of the prototypical expressions: Fear, Lip Smacking and Angry expressions. Upper panels shows data for the original, and the lower panels the reversely played expression movies. Norm-based model: A, E) Face neurons, and B, F) Expression Neurons. Example-based model: C, G) Snapshot neurons, and D,H) Expression Neurons.

4.2 Simulation results

We first tested whether the models can classify the prototypical expressions that were used for training correctly. Fig. 3 A shows the responses of the face neurons that are selective for individual expressions. They show a bell-shaped increase and decrease of activity that is selective for the encoded expression. Panel B shows the response of the corresponding expression neurons, which is also selective for the individual expressions. Opposed to the face neurons, these neurons remain silent during the presentation of static pictures showing the extreme frames of the expressions. Panels C and D show the responses for the example-based model. The activity of the snapshot neurons for the three test expressions is shown in Fig 3 C. Only the frames of the learned expression cause a travelling pulse solution in the corresponding part of the RNN, while the neurons remain silent for the other test patterns. This induces a high selectivity of the responses of the corresponding expression neurons (panel D).

We tested the model also with temporally reversed face sequences in order to investigate the sequence selectivity. The results are shown in Fig. 3 E-H. Due to the high temporal symmetry of facial expressions (backwards-played expressions look very similar, but not identical to forward-played expressions), the responses of the face neurons in the norm-based model and also the one of the expression

neurons are very similar the responses in panels A-D. Most prominently, the responses of the snapshot neurons show now a travelling pulse that runs in opposite direction. The small differences between forward and backwards movie result in slightly lower amplitudes of the expression neuron responses, especially for the example-based model (panel H), but interestingly also for the norm-based model.

Interesting differential predictions emerged when the models were tested with the stimuli of variable expression strength, which were generated by morphing between the prototypes and neutral facial expressions. Here the Face neurons as well as the Expression neurons in the norm-based model show a gradual, almost linear variation of their activations with the expression level (Fig. 4 A and B. Strongly deviating from this behavior, the snapshot neurons do not generate a travelling activity pulse for all stimuli with reduced expressivity levels. Only for the expression level 75 % some activity emerges for the snapshot neurons that represent frames that deviate from the neutral expression. As consequence, the expression neurons do not show significant activity for the conditions with reduced expression strength. This behavior could not be improved by making the snapshots less selective in order to support generalization to more dissimilar patterns. It was not possible with this model to obtain pattern- and sequence-selectivity together with generalization to patterns with reduced expression strength.

The norm-based model showed also very smooth an gradual generalization between different expressions. This is shown in Fig. 4 E-F that shows the responses of the Face and the Expression neurons for morphs between the Fear and the Angry expression. Both neuron types show a very smooth change of their activity with the morph level, which is antagonistic for the neurons with selectivity for the two expressions. Also in this case, the example-based model failed to show generalization to stimuli with intermediate morph levels (not shown).

5 Conclusions

Based on previous models that are grounded in electrophysiological data, we have proposed two alternative mechanisms for the processing of dynamic facial expressions. Both mechanisms are consistent with physiological data from other cortical structures that process social stimuli, static faces and dynamic bodies. Both models were able to recognize monkey expressions from movies. Also the recognition of reversed movies of facial expressions could be accounted for by both models. Testing the models with morphed expressions, and expressions with reduced expression strength, however, resulted in fundamentally different predictions. The norm-based model showed smooth and almost linear variation of the activity patterns with the expression strength and the morph level, while the example-based model had problems to generalize to such morphed patterns. In addition, the models make specific predictions about the activity dynamics of the different postulated neuron classes. For example, an example-based mechanism predicts a traveling pulse of activity, as observed e.g. in premotor cortex [6]. The

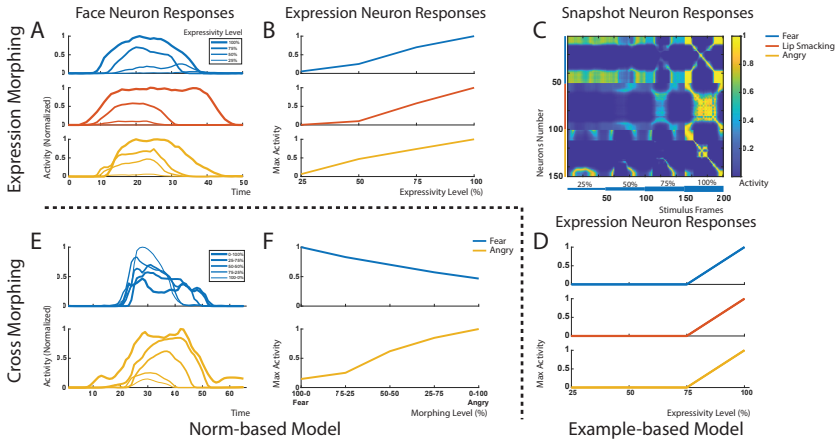

Fig. 4: Upper panels: Neuron activities for stimuli with different expressivity levels (25-50-75-100%). Lower panels: Neuron activities for cross-expression morphs of the neurons of the norm-based model. A) Normalised activity of Face neurons, and B) maximum activity of Expression neurons as function of the expressivity level for the norm-based model. C) Normalised activities of snapshot neurons, and D) maximum activity of Expression neurons in example-based model. E) Normalised activity of Face neurons, and F) maximum activity of Expression neurons as function of the morph level form morphs between Fear and Angry expressions.

norm-based mechanism predicts a linear tuning of the activity with the distance from the neutral reference pattern in morphing space for the Face neurons, etc.

Obviously, the proposed models are only a very simple proof-of-concept demonstration of the discussed encoding principles that need a much more thorough investigation. First, the initial stages of the models have to be replaced by more powerful deep recognition networks that result in mid-level feature dictionaries that make recognition more robust against variations in lighting, texture, etc. Second, the proposed encoding principles have to be tested on much larger sets of dynamic face stimuli, including specifically human facial expressions, to test whether the proposed coding principles can be extended to more challenging recognition problems. Only this will allow to verify whether the proposed norm-based encoding has computational advantages to the more common example-based encoding that underlies many popular RNN-based technical solutions. In addition, such extended models need to be tested against optic flow-based recognition models [21]. Predictions from such models then can be tested against the measured behavior of recorded dynamic face-selective neurons. Even it its present simple form, however, our model makes some interesting predictions that can help to guide the search for the tuning properties of neurons in areas (patches) with neurons that are selective for dynamic facial expressions.

Likewise, the developed stimulus sets will likely be useful to characterize the computational properties of such neurons.

Acknowledgements: This work was supported by HFSP RGP0036/2016 and EC CogIMon H2020 ICT-23-2014/644727. It was also supported by BMBF FKZ 01GQ1704, BW-Stiftung NEU007/1 KONSENS-NHE and ERC 2019-SyG-RELEVANCE-856495. NVIDIA Corp.

References

1. Nick E Barraclough and David I Perrett. From single cells to social perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1571):1739–1752, 2011.
2. Asif A Ghazanfar, Chandramouli Chandrasekaran, and Ryan J Morrill. Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: Implications for the evolution of audiovisual speech. *European Journal of Neuroscience*, 31(10):1807–1817, 2010.
3. Clayton P Mosher, Prisca E Zimmerman, and Katalin M Gothard. Neurons in the monkey amygdala detect eye contact during naturalistic social interactions. *Current Biology*, 24(20):2459–2464, 2014.
4. David A Leopold, Igor V Bondar, and Martin A Giese. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, 442(7102):572–575, 2006.
5. Martin A Giese and David A Leopold. Physiologically inspired neural model for the encoding of face spaces. *Neurocomputing*, 65:93–101, 2005.
6. Vittorio Caggiano, Falk Fleischer, Joern K Pomper, Martin A Giese, and Peter Thier. Mirror neurons in monkey premotor area f5 show tuning for critical features of visual causality perception. *Current Biology*, 26(22):3077–3082, 2016.
7. Falk Fleischer, Vittorio Caggiano, Peter Thier, and Martin A Giese. Physiologically inspired model for the visual recognition of transitive hand actions. *Journal of Neuroscience*, 33(15):6563–6580, 2013.
8. Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber, and Tomaso Poggio. Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, 29(3):411–426, 2007.
9. Martin A Giese and Tomaso Poggio. Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, 4(3):179–192, 2003.
10. Joachim Lange and Markus Lappe. A model of biological motion perception from configural form cues. *Journal of Neuroscience*, 26(11):2894–2906, 2006.
11. Hueihan Jhuang, Thomas Serre, Lior Wolf, and Tomaso Poggio. A biologically inspired system for action recognition. In *2007 IEEE 11th international conference on computer vision*, pages 1–8. Ieee, 2007.
12. James V Haxby, Elizabeth A Hoffman, and M Ida Gobbini. The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6):223–233, 2000.
13. Tim Valentine, Michael B Lewis, and Peter J Hills. Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology*, 69(10):1996–2019, 2016.
14. David A Leopold and Gillian Rhodes. A comparative view of face perception. *Journal of Comparative Psychology*, 124(3):233, 2010.

15. Shan Li and Weihong Deng. Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 2020.
16. Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J Majaj, Rishi Rajalingham, Elias B Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Kailyn Schmidt, et al. Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, page 407007, 2018.
17. Judson P Jones and Larry A Palmer. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of neurophysiology*, 58(6):1187–1211, 1987.
18. Shun-ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics*, 27(2):77–87, 1977.
19. Stéphanie Ratté, Milad Lankarany, Young-Ah Rho, Adam Patterson, and Steven A Prescott. Subthreshold membrane currents confer distinct tuning properties that enable neurons to encode the integral or derivative of their input. *Frontiers in cellular neuroscience*, 8:452, 2015.
20. Nick Taubert, Andrea Christensen, Dominik Endres, and Martin A Giese. Online simulation of emotional interactive behaviors with hierarchical gaussian process dynamical models. In *Proceedings of the ACM Symposium on Applied Perception*, pages 25–32, 2012.
21. John L Barron, David J Fleet, and Steven S Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77, 1994.