Recurrent top-down synaptic connections at different spatial frequencies help disambiguate between dynamic emotions

Erwan David (david@psych.uni-frankfurt.de)
Department of Cognitive Psychology, Theodor-W-Adorno Platz, 6
60323 Frankfurt am Main, Germany

Yannick Bourrier (yannick.bourrier@univ-grenoble-alpes.fr)
LPNC & CNRS, LPNC UMR 5105
621 Avenue Centrale, 38400 Saint-Martin-d'Hères, France

Roman Vuillaume (roman.vuillaume@u-bourgogne.fr)
Laboratory ImViA EA 7535, avenue Alain Savary, 9
21078 Dijon, France

Martial Mermillod (martial.mermillod@univ-grenoble-alpes.fr)
Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC
38000 Grenoble, France

Abstract

The coarse-to-fine hypothesis posits that, in the Human visual system, a coarse representation of visual information is propagated quickly through the retina to the cortex, whereas a finer, more detailed representation is propagated more slowly. In a previous study we showed that recurrent synaptic connections help predict low intensity EFEs. Furthermore, a feedback loop coming from coarser information processing is postulated to influence later processing of finer features. In this paper, we intend to examine the value of coarser information and recurrence in the processing of dynamic Emotional Facial Expressions (EFE). In a step forward in studying the importance of recurrent connectivity in the coarse-to-fine model, we tested its advantage for discriminating emotions for different spatial frequencies and facial expression intensities. Using Artificial Neural Networks, we modeled recurrent synaptic connections with a recurrent feedback loop. Using a Gabor filter bank, we computed different levels of spatial frequency features. Our results replicate the advantage of recurrence at first facial expression intensities. Our main finding is that the recurrent model is also better when predicting high spatial frequencies features. Additionally, mid-to-low spatial frequencies are more useful to the prediction of EFEs. We conclude that feature processing feedback has a significant effect in disambiguating facial expressions when information is particularly complex, i.e., at high spatial frequencies and low EFE intensities.

Keywords: Proactive Brain, Neural Network modeling, Emotional Facial Expressions, Spatial Frequencies.

Introduction

The proactive brain hypothesis (Bar, 2007; Trapp & Bar, 2015) models the use of fast but coarse information streams, which are necessary to react quickly to stimuli and events by making predictions that are refined as finer information is propagated through the brain (Bar et al., 2006). It has been postulated that the existence of top-down cortical information feedback loops improves this proactive system and this is one explanation for the high density of the top-down recurrent connections found in the brain (Bullier, 2001; Sherman & Guillery, 2002). At the level of the lateral geniculate nucleus, the magnocellular pathway represents a fast but coarse route in the brain, whereas the parvocellular pathway represents a slower but fine route (Peyrin et al., 2010; Kauffmann, Ramanoël, & Peyrin, 2014; Kauffmann, Chauvin, Guyader, & Peyrin, 2015). With the help of artificial neural networks we intend to study the importance of recurrent connections for the prediction of dynamic Emotional Facial Expressions (EFE).

EFEs are complex stimuli which play an important role in our society. Multiple studies have shown that face identification (Vida & Maurer, 2015) and EFE detection (Mermillod, Bonin, Mondillon, Alleysson, & Vermeulen, 2010; Bayle, Schoendorff, Hénaff, & Krolak-Salmon, 2011) can be achieved without high spatial frequencies. The anticipation of dynamic stimuli such as EFEs is assumed to rely on neurological systems involving bottom-up and top-down processes related to visual processing (Kveraga, Ghuman, & Bar, 2007; Beffara et al., 2015). In this study our intention is to demonstrate the usefulness of plausible recurrence strategies in the processing and prediction of dynamic emotions.

Mermillod et al. (2019) implemented feedforward and recurrent artificial neural network models that took EFE features as inputs and the next EFE intensity in an emotional sequence (neutral to apex) as target. Recurrent models are characterized by the re-injection (along with input vector) of the hidden layer state provided by a previous intensity step. The authors demonstrated that such a feedback loop of processed data (i.e., hidden layer state reinjected) enhanced the predicion of facial emotions, in particular at initial intensities. In the present work, the inputs were Gabor magnitude responses obtained with a Gabor filter bank of different spatial and orientation frequencies. We segregated spatial frequency processing in order to measure which frequencies, high or low, are the most useful for predicting early EFE intensities.

Previous studies have shown that low spatial frequencies are more useful for object, scene and face categorization (e.g., Mermillod, Guyader, & Chauvin, 2005; Mermillod, Vuilleumier, Peyrin, Alleysson, & Marendaz, 2009). As such, we expect models to perform better when the inputs are low spatial

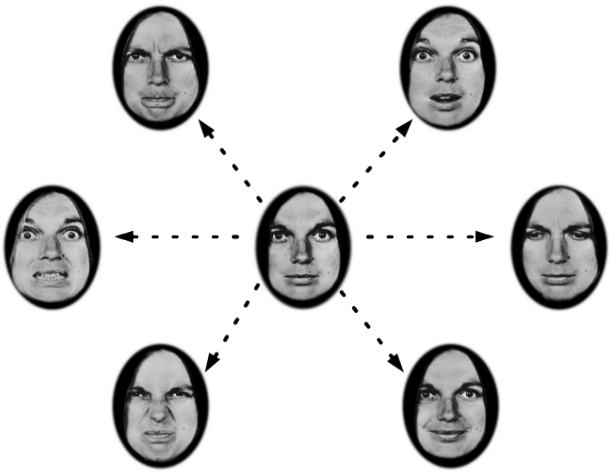

Figure 1: EFEs were organized in a star configuration with a neutral expression in the center. Stimuli increased in intensity until reaching apexes by means of a morphing transformation.

frequencies. We expect a recurrent model to be more efficient at predicting EFE features in complicated situations such as early emotion intensities (Mermillod et al., 2019).

Mermillod et al. (2019) have previously shown the importance of recurrence for recognizing dynamic emotions. In this article, we extend these investigations in an attempt to validate the coarse-to-fine feedback loop hypothesis with computational models.

Method

Stimuli

In this study, we used the Pictures of Facial Affect database (POFA, Ekman & Friesen, 1993). It consists of photographs of 10 actors (six women and four men) displaying six emotions (anger, disgust, sadness, joy, fear, and surprise) as well as a neutral state. This facial emotion database has been extensively used in the scientific literature, and more particularly in previous computational studies on facial expression processing.

We used the POFA database to create a continuum from neutral to an emotional expression apex (10 intensities) As depicted in Figure 1, EFEs of any given actor share the same neutral state. After preprocessing (described below), images in the database were presented as vectors of average energy responses obtained with a Gabor filter bank.

Processing

We obtained EFE intensities via *morphing*, by performing a linear interpolation between neutral and emotional apex pictures. In a second step, we extracted significant information related to spatial frequency and orientation similar to the first stages of human visual processing (DeValois & DeValois, 1990; Mermillod, Guyader, & Chauvin, 2004).

The POFA has the advantage of including neutral emotional faces which we can use as starting points in an emotional continuum. To obtain this continuum we used an in-house *C++* tool tasked with automatically detecting 68 facial fiducial landmarks (face, eye, nose, mouth contour). Thanks to these points, we obtained a set of triangles mapping facial features. We interpolated between the positions of the vertices of these triangles and the pixel content within triangles via affine transformations and alpha blending, respectively. The interpolation was performed between triangles corresponding to the same fiducial points in a neutral and an apex picture. We obtained a ten-step interpolation from neutral to emotion apex.

We constructed a Gabor filter bank and projected pictures and filters in the spectral domain with a fast Fourier transform. Pictures and filters, were then multiplied together, as a fast alternative to convolution. This operation ensured that we could control the information provided at the level of the perceptual layer, which would not have been possible using convolutional layers. For each stimulus and Gabor filter, we obtained a measure of average energy response. The filter bank consisted of seven spatial frequencies and eight spatial orientations. With this method we transformed each stimulus into a vector of 56 average energy responses. The seven spatial frequencies were chosen so they spanned low to high frequencies (increasing by one octave on each step) as perceived by the human visual system:

- Low: 0.35 cycles per degree of field of view, 0.72 cy./deg.
- Medium: 1.4 cy./deg., 2.81 cy./deg., 5.62 cy./deg.
- High: 11.25 cy./deg., 22.5 cy./deg.

Artificial Neural Networks

In order to study whether information feedback loops are plausible in the human visual system we modeled three Artificial Neural Networks (ANNs, PyTorch Paszke et al., 2017). To do this, we used a simple Multi-Layer Perceptron similar to that used in Mermillod et al. (2019) in order to ensure a fair comparisons of the different spatial frequencies:

1. MLP: a Multi-Layer Perceptron.
2. SRN: a Simple Recurrent Network.
3. MLP$_{aug}$: an MLP augmented with the same number of parameters as the SRN, taking three time-steps' worth of emotion intensity at once as input.

The MLP is a feedforward ANN with an input layer, a hidden layer, and an output layer. It represents a baseline performance, in order to predict emotion vectors on a single information step. The SRN is based on the Elman network (Elman, 1990). It is similar to the MLP, with the exception of the hidden layer, which is considered as a context layer, whose content is used alongside the input layer content as input to the next time-step. The MLP$_{aug}$ has three times as many inputs as the MLP. It therefore has the same number of parameters as the SRN and does not rely on feedback loops, since it receives three time-steps' worth of intensity at once.

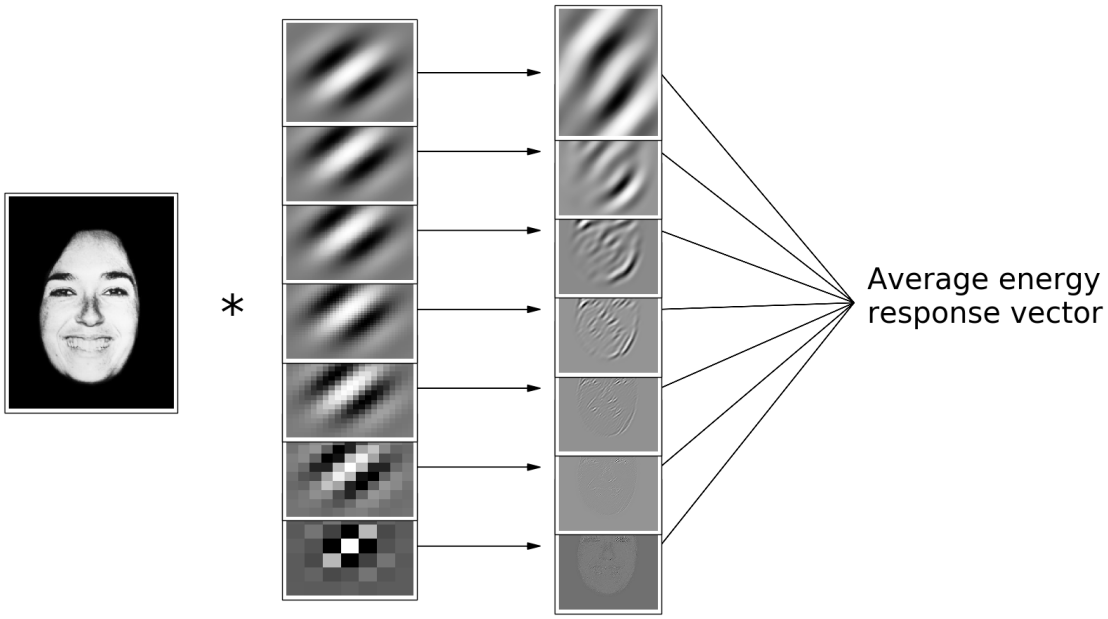

Figure 2: EFE stimuli were multiplied in the Fourier domain with a Gabor filter banks. This figure shows one of the eight orientations used in our experiment. This operation was repeated for all stimuli with the entire Gabor filter bank to obtain a vector of 56 average energy responses.

The input data of the network is an energy response vector at an intensity step t, while the target is the energy response vector at time-step $t+1$ To measure performances with different spatial frequencies, the input vector consists of eight spatial orientations extracted from the 56-element energy response vector described above. Because the input layer of MLP_{aug} receives three intensities, the first intensity vectors are filled with zeros for non-existent intensities (intensities -2 and -1 at time step 0 and intensity -1 at time step 1). In all three models, the size of the hidden/context layer is twice that of the input layer as implemented by Elman (1990).

The purpose of the models is to predict the energy response vectors at the next time-step (emotional intensity). The loss is the root-mean square error (RMSE) between the expected energy response vector and that predicted by the network. Therefore, the models predict the evolution of the emotion on a continuum, rather than classifying the emotion itself. A classification accuracy measure is obtained by comparing the RMSE of a model's prediction with all six emotion vectors at the predicted intensity (for the same actor). If the predicted value is closest to the emotion vector of the expected class, then the prediction is labeled positive, otherwise the prediction is negative.

Training methodology

A set of ANN models was trained for each of the seven spatial frequencies separately. The database for a training instance was thus made up of 600 eight-element vectors (ten actors by six emotions by ten intensities). This data subset was normalized between 0 and 1 (min-max range normalization) for each spatial frequency independently in order to fit the range of the activation function (sigmoid) at the output of the network.

We used the Adam optimizer to update the model's parameters with a learning rate of 0.0005 (other parameters: *betas*=(0.9, 0.999), *epsilon*=1e-8, *weight decay*=0). This number is lower than the default value suggested by the authors of the optimizer in order to avoid overfitting.

The ANNs were trained for 10,000 epochs (one epoch is one pass over the training dataset).At each iteration, one emotion intensity was passed through a model starting from neutral data up to intensity $n-1$ (predicting the tenth one). Since the first prediction was based on the neutral face, which was common to all the emotions, we expected the first time-step prediction to be equal to chance level (1/6).

In a leave-one-out strategy, one actor's data was set aside for testing. It was therefore kept independent from the learning protocol and served to measure the generalization performances of the networks.

Analyses

In this study, we analyzed prediction success rates as a function of ANN types and emotion intensity across the information for the seven spatial frequencies extracted as described above. For each spatial frequency, our statistical models answered two questions: 1) is there a main effect of ANN types, 2) is there an interaction effect between ANN types and emotion intensity.

The results of this experiment were analyzed with Generalized Linear Mixed Models (GLMMs Jaeger, 2008) in order to model the binomial nature of the binary data (success

rates) and the random effects inherent to our set-up. As well as determining significance levels, the results of the analysis of main and interaction effects were not trivial. We therefore adopted the method used by Nuthmann and Malcolm (2016) by comparing different multi-level statistical models using likelihood ratio tests.

Three GLMMs were specified: an *intensity* model including solely the emotion intensity as a fixed effect, a *net-intensity* model including the ANN type and emotion intensity as fixed effects and a *full* model including these two variables and their interaction as fixed effects. If the *full* model has a better fit than the *intensity* model this would lead us to believe that network type had an effect (main or interaction); if the *full* model has a better fit than the *net-intensity* model then there would be an interaction effect. The complexity of the models by inclusion of random effects (actor, emotion) was reduced until they converged. Unless specified otherwise in the following section, the main effect of ANN type and the interaction effect are significant.

We planned post-hoc comparisons with the least mean squares method (*R*, R Core Team, 2018; package *lsmeans*, Lenth, 2016) to study interaction effects. Planned contrasts compare emotions of the same intensity step across ANN types. Post-hoc tests have been corrected with the Tukey method.

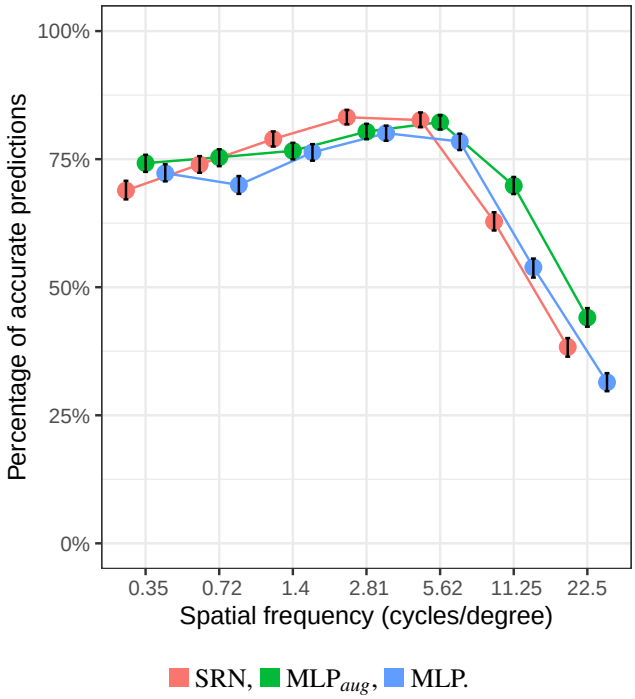

Figure 3: Average model performances as a function of spatial frequencies (from lowest to highest). Error bars denote 95% Confidence intervals (95%CI)

Results

As shown in Figure 3, prediction accuracy was lowest for high spatial frequencies but plateaued with medium to low frequencies. Th main effect of ANN types was significant across all spatial frequencies ($p < 0.0001$) and an interaction effect between ANN types and spatial frequencies also emerged ($p < 0.0001$). The average SRN performances were better than that of the MLP in the case of spatial frequencies 2.81 cy./deg. and above ($ps < 0.0001$), but below that of the MLP$_{aug}$ for the three highest spatial frequencies ($ps < 0.001$).

Our results show that a feedback loop significantly improves performances. The feedback loop gave the SRN the opportunity to carry information over from previous time-steps. As was expected, the accuracy of the MLP$_{aug}$ shows that processing multiple intensity steps at once was also beneficial.

Results of Figure 3 are averaged over emotion intensities. To learn more about the difference between models at emotion onset particularly, we present in Figure 4 the performances of the models as a function of emotional intensities and spatial frequencies.

Learning from the lowest spatial frequencies considered in this study (0.35 cy./deg) the SRN converged to a non-optimal solution and showed a small but significant accuracy decrease when predicting intensities six to nine ($p < 0.01$). Overall the performances of the models are quite similar at this low frequency.

The second low spatial frequency set of data (0.7 cy./deg) had the SRN in a better position compared to the alternative models when predicting the third emotion intensity ($p < 0.0001$). The intensity prediction of the MLP$_{aug}$ showed a significant increase in accuracy compared to the MLP and SRN ($p < 0.01$); at intensity five the MLP$_{aug}$ and SRN performed similarly and better than the MLP ($p < 0.01$). Beyond this point, the models had similar performances. Considering the medium to low spatial frequency (1.4 cy./deg), the SRN had an edge over the MLP and MLP$_{aug}$ when predicting the third intensity ($ps < 0.005$). The models performed at a similar level after that.

Results with the next medium spatial frequency step (2.81 cy./deg.) still showed an advantage for the SRN over the other networks in predicting the third intensity ($ps < 0.005$), though the performances of the three networks did not differ significantly after that: they increased with the intensities and reached a plateau before the last five intensities. The interaction effect between intensity and network type failed to reach a significant level here.

Predicting from the next medium spatial frequency data (5.62 cy./deg.), the SRN was significantly more accurate in predicting the third intensity ($ps < 0.05$) and its performances were similar to those of the MLP$_{aug}$ for the rest of the intensities, whereas the MLP scored lower in the cases of the third and the last three intensities ($ps < 0.05$).

The spatial frequency data for 11.25 cy./deg. showed that, globally, the MLP$_{aug}$ had an advantage when predicting emotion intensities. It was more accurate than the MLP at all intensity steps ($ps < 0.02$) and was better than the SRN at

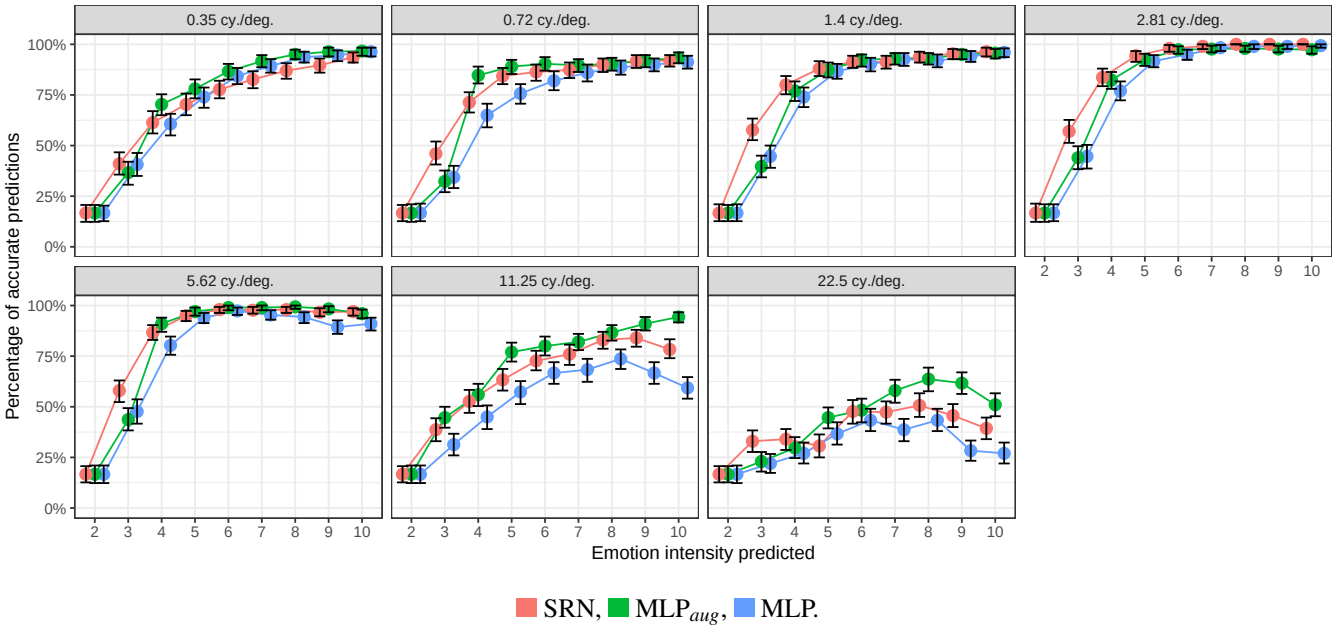

Figure 4: Average model performances as a function of spatial frequencies (from lowest to highest) and emotion intensity (neutral to apex). Emotion intensity starts at two because the first intensity is used as input to the model to predict the second. Error bars denote 95%CI

predicting intensities five, nine and ten ($ps < 0.05$). While the SRN's prediction accuracies appeared to be better than those of the MLP, they were only significantly so for the last three intensities ($ps < 0.02$).

When the models predicted the highest spatial frequency data (22.5 cy./deg.), the SRN displayed a significant advantage ($ps < 0.02$) compared to other the models when predicting the third emotion intensity. On the other hand, the MLP$_{aug}$ demonstrated the best performances over the last four emotion intensities ($ps < 0.02$) and the SRN was significantly better at the task than a simple MLP for the last two intensities ($ps < 0.005$).

Discussion

We replicated the results of Mermillod et al. (2019) showing that recurrent connections are particularly useful in the first stages of recognizing dynamic EFEs when signs of emotion are subtle. Our most important finding indicates that the positive effect of recurrent connections when processing ambiguous information also applies to specific spatial frequencies. Indeed, the SRN model performed better than the MLP models when presented with higher spatial frequencies (11.25 and 22.5 cy./deg.). As demonstrated by the high accuracy observed with all networks, peaking at first emotion intensities, we show that mid-to-low spatial frequencies are the most useful for predicting emotion, a finding which is consistent with humans performances (Bayle et al., 2011). The results validate our hypotheses that recurrent connections provide an advantage when processing stimuli that are ambiguous due to their poor perceptual features (high spatial frequencies) as well as at early stages of EFEs. The high performances of the MLP$_{aug}$, a non-recurrent ANN receiving three intensity-steps at once as input, indicates that the SRN probably does not consider information prior to $t-2$. Indeed, the augmented MLP performs comparably or better as soon as three emotion intensities are available for processing. The results for the MLP$_{aug}$ do not challenge the usefulness of recurrent connections, but instead help identify the nature of the process the recurrent model converges to.

Future works

To expand on this study we intend to implement an MLP with as many parameters as the SRN by increasing the number of neuron units in the hidden layer but without modifying the content of the input layer (contrary to the procedure used to obtain the MLP$_{aug}$). This will serve to test whether the superiority in performances of the SRN stems from an advantage in using recurrent connections or is simply due to in an increase in computational power.

While the POFA database guaranteed that the emotions evolved linearly from a neutral point to the apex, we plan in future works to replicate these results using other databases such as ADFES (van der Schalk, Hawk, Fischer, & Doosje, 2011). ADFES is a database of videos in which actors display EFEs starting from a resting state. It is interesting as it is free from morphing artifacts and because EFEs do not evolve linearly and do not reach apex at the same speed (Schmidt & Cohn, 2001; Cohn & Schmidt, 2004). As the continuum evolves naturally rather than linearly, the task should be more complex but should also shed further light the proactive role of feedback loops for fast EFE recognition. Additionally,

we will rely on morphing again to perform testing using the KDEF (Lundqvist, Flykt, & Öhman, 1998). This will allow us to verify if our findings persist when there is more facial diversity (images from 40 actors are used). In this context we will study the usefulness of recurrent connections on a per-emotion basis.

Moreover, in order to continue testing the coarse-to-fine hypothesis, we intend to replicate this study but vary the spatial frequencies along with emotion intensity. We will test if processing low emotion intensities with coarse data helps predict higher intensities presented with fine data.

Conclusion

We have demonstrated that a model with recurrent connections processing new visual information along with previous information performs significantly better than a non-recurrent model, in particular when it comes to ambiguous stimuli. Our results support interpretations of top-down recurrent connections in the brain as a way to refine predictions as cortical areas receive more recent or finer visual information. However, the important conclusion of this article is that the computational advantage of recurrent connections seems to be related to specific spatial frequency channels, as it has been previously assumed for humans (Bar, 2007; Beffara et al., 2012; Trapp & Bar, 2015).

Acknowledgments

Martial Mermillod received partial support for this work from MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003).

References

Bar, M. (2007, July). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280–289. Retrieved 2016-04-29, from http://linkinghub.elsevier.com/retrieve/pii/S1364661307001295 doi: 10.1016/j.tics.2007.05.005

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006, January). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. Retrieved 2016-04-25, from http://www.pnas.org/cgi/doi/10.1073/pnas.0507062103 doi: 10.1073/pnas.0507062103

Bayle, D. J., Schoendorff, B., Hénaff, M.-A., & Krolak-Salmon, P. (2011). Emotional facial expression detection in the peripheral visual field. *PloS one*, *6*(6).

Beffara, B., Ouellet, M., Vermeulen, N., Basu, A., Morisseau, T., & Mermillod, M. (2012). Enhanced embodied response following ambiguous emotional processing. *Cognitive processing*, *13*(1), 103–106.

Beffara, B., Wicker, B., Vermeulen, N., Ouellet, M., Bret, A., Molina, M. J. F., & Mermillod, M. (2015). Reduction of interference effect by low spatial frequency information priming in an emotional stroop task. *Journal of vision*, *15*(6), 16–16.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2), 96–107. Retrieved 2016-05-26, from http://www.sciencedirect.com/science/article/pii/S0165017301000856

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, *2*(02), 121–132.

DeValois, R. L., & DeValois, K. K. (1990). *Spatial vision* (2nd ed ed.) (No. 14). New York: Oxford Univ. Press.

Ekman, P., & Friesen, W. (1993). Pictures of facial affect. *Consulting Psychologists Press, Palo Alto: CA*.

Elman, J. L. (1990). Finding structure in time. *Cognitive science*, *14*(2), 179–211. Retrieved 2016-04-25, from http://www.sciencedirect.com/science/article/pii/036402139090002E

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language*, *59*(4), 434–446.

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, *107*, 49–57.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in integrative neuroscience*, *8*, 37.

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and cognition*, *65*(2), 145–168.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. doi: 10.18637/jss.v069.i01

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces (KDEF)*. Stockholm: Karolinska Institute and Hospital, Section of Psychology.

Mermillod, M., Bonin, P., Mondillon, L., Alleysson, D., & Vermeulen, N. (2010, August). Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation. *Neurocomputing*, *73*(13-15), 2522–2531. Retrieved 2016-02-03, from http://linkinghub.elsevier.com/retrieve/pii/S0925231210002559 doi: 10.1016/j.neucom.2010.06.002

Mermillod, M., Bourrier, Y., David, E., Kauffmann, L., Chauvin, A., Guyader, N., ... Peyrin, C. (2019). The importance of recurrent top-down synaptic connections for the anticipation of dynamic emotions. *Neural Networks*, *109*, 19–30.

Mermillod, M., Guyader, N., & Chauvin, A. (2004). Does the energy spectrum from gabor wavelet filtering represent sufficient information for neural network recognition and classification tasks? In *Connectionist models of cognition and perception ii* (pp. 148–156). World Scientific.

Mermillod, M., Guyader, N., & Chauvin, A. (2005, March). The coarse-to-fine hypothesis revisited: Evi-

dence from neuro-computational modeling. *Brain and Cognition*, *57*(2), 151–157. Retrieved 2016-06-15, from http://linkinghub.elsevier.com/retrieve/pii/S0278262604002325 doi: 10.1016/j.bandc.2004.08.035

Mermillod, M., Vuilleumier, P., Peyrin, C., Alleysson, D., & Marendaz, C. (2009). The importance of low spatial frequency information for recognising fearful facial expressions. *Connection Science*, *21*(1), 75–83.

Nuthmann, A., & Malcolm, G. L. (2016). Eye guidance during real-world scene search: The role color plays in central and peripheral vision. *Journal of Vision*, *16*(2), 3–3.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... Lerer, A. (2017). Automatic differentiation in pytorch.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... Vuilleumier, P. (2010, December). The Neural Substrates and Timing of Top–Down Processes during Coarse-to-Fine Categorization of Visual Scenes: A Combined fMRI and ERP Study. *Journal of Cognitive Neuroscience*, *22*(12), 2768–2780. Retrieved 2016-05-26, from http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2010.21424 doi: 10.1162/jocn.2010.21424

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Schmidt, K. L., & Cohn, J. F. (2001). Dynamics of facial expression: Normative characteristics and individual differences. In *null* (p. 140). IEEE. Retrieved 2016-06-10, from http://www.computer.org/csdl/proceedings/icme/2001/1198/00/11980140-abs.html

Sherman, S. M., & Guillery, R. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1428), 1695–1708.

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1), 190–198.

van der Schalk, J., Hawk, S. T., Fischer, A. H., & Doosje, B. (2011). Moving faces, looking places: Validation of the Amsterdam Dynamic Facial Expression Set (ADFES). *Emotion*, *11*(4), 907–920. Retrieved 2016-05-02, from http://doi.apa.org/getdoi.cfm?doi=10.1037/a0023853 doi: 10.1037/a0023853

Vida, M. D., & Maurer, D. (2015, July). A comparison of spatial frequency tuning for judgments of eye gaze and facial identity. *Vision Research*, *112*, 45–54. Retrieved 2016-05-04, from http://linkinghub.elsevier.com/retrieve/pii/S0042698915001728 doi: 10.1016/j.visres.2015.04.018