Layer-specific cortical mechanisms underlying visual perceptual learning

Monika Jozsa* **(mj555@cam.ac.uk)**
Engineering Department, University of Cambridge, Trumpington Street
Cambridge, UK

Clara Pecci Terroba* **(cp661@cam.ac.uk)**
Psychology Department, University of Cambridge, Downing Place
Cambridge, UK

Ke Jia (kjia@zju.edu.cn)
Liangzhu Laboratory, MOE Frontier Science Center for Brain Science and Brain-machine Integration
Zhejiang University, Hangzhou, China

Mengxin Wang (mengxin.wang@psy.ox.ac.uk)
Department of Experimental Psychology, University of Oxford, OX2 6GG
Oxford, UK

Zoe Kourtzi† **(zk240@cam.ac.uk)**
Psychology Department, University of Cambridge, Downing Place
Cambridge, UK

Yashar Ahmadian† **(ya311@cam.ac.uk)**
Engineering Department, University of Cambridge, Trumpington Street
Cambridge, UK

Abstract

Training in perceptual tasks can enhance the brain's representations of task-relevant features for improved decision making. Recent ultra-high-field neuroimaging studies have found that training in fine discrimination tasks increases signal-to-noise ratio of activity patterns in the superficial layers of primary visual cortex (V1), accompanied by an increase in GABAergic inhibition. However, the causal circuit mechanisms behind the layer-specific changes of representation and cortical excitation-inhibition remain unknown. Here, we theoretically study these questions by training a biologically-constrained model of V1 in a fine orientation discrimination task near a fixed orientation. Training led to (1) strengthening (weakening) of cortical inhibition (excitation), which was larger and more robust in the model's superficial layer, and (2) sharpening of the superficial-layer tuning curves at the trained orientation, as previously reported in neurophysiology experiments. Further, these changes correlated with improved network task performance. Finally, the mechanistic nature of our model allows making testable predictions about the causal pathways linking layer-specific changes in excitation and inhibition with representational or behavioral improvements.

Keywords: visual perceptual learning; cortical excitation and inhibition; recurrent network dynamics; deep learning

*These authors contributed equally.
†These authors contributed equally.

Jia *et al.* (2024, under review) used ultra-high-field (UHF-7T) brain imaging in combination with magnetic resonance spectroscopy (MRS) to investigate functional and neurochemical plasticity for perceptual decisions. In their experiment, participants performed a fine orientation discrimination task focused at a given trained orientation. The UHF-7T analysis showed that training increased the signal-to-noise ratio of activity patterns in the superficial layers of V1, consistent with (Schoups, Vogels, Qian, & Orban, 2001; Jia et al., 2020). Further, MRS analysis showed a significant increase in the ratio of GABA to glutamate concentration averaged across multiple visual areas, which was related to performance improvement (in alignment with, e.g., Ip and Bridge, 2022; Stagg et al., 2011) and to layer-specific changes in V1 representations. The low spatial resolution of MRS prevents the localization of these measurements to specific visual areas or cortical layers. To address this question, we optimized the structural parameters of a biologically-constrained model of V1 for the same task as was used in the experiment. Our study shows significant contribution to performance improvement of layer-specific changes in inhibition and excitation.

Model architecture

We modelled V1 as a recurrent network with two layers, representing the thalamorecipient middle (M) and superficial (S) layers of V1 (Fig. 1A); for reasons of parsimony, we do not model the feedback from V1' S layer to its M or deep (not modelled) layers. The layers are Stabilized Supralinear Networks (SSN) —a previously validated model of cortical circuitry (Rubin, Van Hooser, & Miller, 2015; Ahmadian

& Miller, 2021). The SSN is a network of excitatory (E) and inhibitory (I) neurons that have rectified supralinear power-law input/output functions: $\max(0,x)^n$ with $n > 1$. The layers share a retinotopic map that covers the area of the stimulus, which is a grating image. This map is discretized into a grid of "mini-columns", each with its own orientation preference according to an orientation map. The M layer cells[1] represent simple cells with four possible phase preferences, while the S layer cells represent complex cells without phase preference.

Visual information first enters the M layer which feeds forward to the S layer. The stimulus input to each M layer cell is given by the rectified inner product of the cell's Gabor filter (specified by the cell's phase and orientation preference) with the grating image. The feedforward projections from the M to the S layer are local and arise solely from E cells at the same retinotopic location, with all phase preferences contributing equally. Cells also received a stimulus-independent baseline input, which was identical across E or I cells.

The model is further constrained by known neuroanatomy: (1) only the S layer has long-range horizontal connections between the mini-columns; these connections fall off with distance and, at equal distance, neurons with more similar preferred orientations have stronger connections. E horizontal connections have longer range than the I ones, consistent with neuroanatomy (Lund, Yoshioka, & Levitt, 1993), but E and I connections have the same degree of orientation tuning. (2) only the S layer's responses are read out by the decision making output (simulating the findings that the primary origin of inter-area feedforward pathways are the superficial layers).

Denoting the full connectivity matrix of the whole network by W, the vector of neural time constants by τ, the external input vector by $\mathbf{h}$, and the vector of neural firing rates by $\mathbf{r}$, the latter obeys the differential equation $\tau\dot{\mathbf{r}} = -\mathbf{r} + (W\mathbf{r} + \mathbf{h})_+^n$.

Trained parameters consisted of: the overall strength of recurrent connections of different types in each layer, denoted by $J^{\text{mid}}_{a\to b}$ and $J^{\text{sup}}_{a\to b}$, with $a,b \in \{E,I\}$; f_E and f_I, the strength of feedforward E connections from the M layer E cells to the S layer E or I cells, respectively; c_E and c_I, the baseline inputs to E and I cells, respectively. We denote the set of trained SSN parameters by $\Theta_{\text{SSN}} = \{f_E, f_I, c_E, c_I, J^{\text{mid}}_{a\to b}, J^{\text{sup}}_{a\to b}\}$.

Training the model for orientation discrimination

The training protocol has two phases: pre-training on a general, coarse discrimination task, followed by training in the fine orientation discrimination task. The role of pre-training is to gain initial parameters for training that yield orientation discriminabilities comparable to naive experimental participants. In each trial, the model sequentially receives a pair of noisy gratings: a "reference" grating, with orientation α, followed by a "target" grating with orientation $\alpha + \beta_{\text{offset}}$. The network's task in both phases is to decide whether the target grating is clockwise ($\beta_{\text{offset}} > 0$) or counterclockwise ($\beta_{\text{offset}} < 0$) relative to the reference. In pre-training, α and the offset were sampled from wide ranges: $\alpha \in [15^\circ, 165^\circ]$, $\beta_{\text{offset}} \in [10^\circ, 20^\circ]$. For the training task, α was randomly jittered up to $\pm 5^\circ$ around the trained orientation (55°). β_{offset} was changed over time in steps of $\pm 0.1^\circ$ following a 3-up-1-down staircase algorithm.

Responses of S layer E cells to the reference and target gratings were subtracted and fed to a sigmoid perceptron (with weight vector $\mathbf{w}_{\text{sig}}$ and bias b_{sig}) outputting the predicted probability of the clockwise choice. The loss was the perceptron output's cross-entropy with regularization terms encouraging the maintenance (homeostasis) of the network's stability and mean activity level. The model was initialized by randomly generating an orientation map, and perturbing Θ_{SSN} by $\pm 10\%$ from baseline values (Holt, Miller, & Ahmadian, 2023). During pre-training, both Θ_{SSN} and $(\mathbf{w}_{\text{sig}}, b_{\text{sig}})$ were updated using SGD. The fine discrimination task inherited these from pre-training, but only trained Θ_{SSN}. The stopping criteria for pre-training was to achieve 75% accuracy for $\beta_{\text{offset}} < 6$ for the training task. The training run for fixed number (1000) of steps.

Results

We trained the model for 50 different initialization (runs). Training for the fine discrimination task led to a reduction in the discrimination threshold across runs (Fig. 1E), an average increase in inhibitory recurrent weights (both $J_{I\to E}$ and $J_{I\to I}$), and an average decrease in excitatory weights (both $J_{E\to E}$ and $J_{E\to I}$). These changes are consistent with the empirical finding by Jia *et al.* of an increase in the ratio of GABA to glutamate concentration averaged across visual areas. Further, we found that in the model these changes were stronger in the S layer, and only there the direction of change was consistent across all runs (Fig. 1B top vs. middle). The baseline inputs to E and I cells (c_E and c_I) decreased and increased, respectively (Fig. 1B bottom). Physiologically, an increase in extrasynaptic GABA (glutamate) may reduce (increase) both inputs; thus, this outcome is hard to reconcile with global changes in extrasynaptic neurotransmitters. The feedforward excitatory drive, f_I, from the M layer onto the I cells of the S layer decreased (Fig. 1B bottom — but f_E showed no consistent change across runs), indicating an overall suppression of inhibitory influence from the M layer on the S layer's output.

Jia et al. also found a positive correlation (r=0.456, p=0.029) between GABA/Glu ratio and task improvement. Our findings align with this result and further delineate this relationship across layers and specific types of connectivity (Fig. 1C). In particular, we found significant positive (negative) correlations between the changes in recurrent I (E) and the improvement in discrimination threshold, that were stronger in the S layer.

Neural tuning curves in the S layer, but not M layer, showed changes in shape consistent with increase in signal-to-noise ratio. In particular, their tuning width decreased, and their slope at the trained orientation increased, especially for neurons that had higher slopes at this orientation (Fig. 1D), consistent with previous studies (Schoups et al., 2001).

In the future, using various model ablations, we can test the causal role of the changes in the M and S layers' E/ I connectivity to provide predictions for experimental studies.

[1] Cell's firing rate represents the mean rate of a neural population.

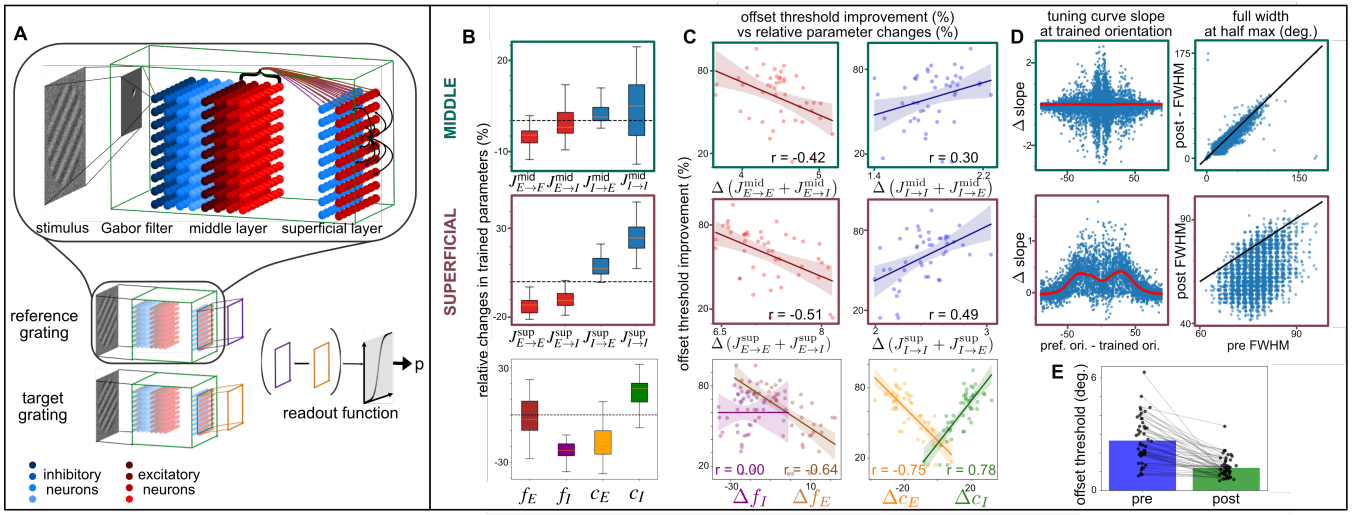

Figure 1: A: Model architecture and the task. B: Training-induced changes in network parameters (boxes/whiskers represent distributions over runs with different initial parameters). Top and middle rows in B-D correspond to M and S layers, respectively. C: Correlations between relative changes in parameters and decrease in discrimination threshold across runs. D: Changes in tuning curve width and slope at the trained orientation. E: Pre- and post-training discrimination thresholds for different runs.

Acknowledgments

This work was supported by grants to 1) YA & ZK from UKRI Biotechnology and Biological Sciences Research Council (BB/X013235/1) to 2) ZK from the Wellcome Trust (205067/Z/16/Z; 223131/Z/21/Z) .

References

Ahmadian, Y., & Miller, K. D. (2021). What is the dynamical regime of cerebral cortex? *Neuron*, *109*(21), 3373–3391. doi: 10.1016/j.neuron.2021.07.031

Holt, C. J., Miller, K. D., & Ahmadian, Y. (2023). The stabilized supralinear network accounts for the contrast dependence of visual cortical gamma oscillations. *bioRxiv*. doi: 10.1101/2023.05.11.540442

Jia, K., Zamboni, E., Kemper, V., Rua, C., Goncalves, N. R., Ng, A. K. T., ... Kourtzi, Z. (2020). Recurrent Processing Drives Perceptual Plasticity. *Current Biology*, *30*(21), 4177-4187.

Lund, J. S., Yoshioka, T., & Levitt, J. B. (1993). Comparison of Intrinsic Connectivity in Different Areas of Macaque Monkey Cerebral Cortex. *Cerebral Cortex*, *3*(2), 148–162.

Rubin, D. B., Van Hooser, S. D., & Miller, K. D. (2015). The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, *85*(2), 402–417. doi: 10.1016/j.neuron.2014.12.026

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in v1 neurons. *Nature*, *412*, 549–553.