Dissociable Dynamic Effects of Expectation During Statistical Learning Across Cortical Layers

Hannah H McDermott (h.mcdermott@fu-berlin.de)
Freie Universität Berlin, Berlin, Germany
Berlin School of Mind and Brain, Berlin, Germany
Maastricht University, Maastricht, the Netherlands

Mahdi Enan (m.enan@maastrichtuniversity.nl)
Maastricht University, Maastricht, the Netherlands

Federico De Martino (f.demartino@maastrichtuniversity.nl)
Maastricht University, Maastricht, the Netherlands

Ryszard Auksztulewicz (ryszard.auksztulewicz@maastrichtuniversity.nl)
Maastricht University, Maastricht, the Netherlands
Freie Universität Berlin, Berlin, Germany

Abstract

The brain seemingly generates internal predictions, to optimise behaviour. Predictive processing has been repeatedly demonstrated in non-invasive studies on human volunteers and in animal models. One commonly reported phenomenon is expectation suppression (ES) or the suppression of neural activity in response to expected stimuli. Our recent EEG study yielded direct evidence that ES is expressed as two opposing mechanisms, amounting to sharpening vs. dampening of neural activity by stimulus expectation, albeit at different time points both within and across trials. In this high-field neuroimaging study, we test if these dissociable dynamics of expectation effects can be explained in the context of hierarchical learning mechanisms. Healthy volunteers (N=15) completed an associative learning task with paired visual and auditory stimuli. Univariate analyses examined region- and layer-specific activity for expected vs. unexpected sounds. General Linear Model (GLM) analyses of activation across conditions revealed differences between valid and invalid trials across the auditory cortex and occipital cortex.

Keywords: expectation suppression; predictive processing; laminar fMRI;

Introduction

Predictive processing (PP) posits that the brain generates internal predictions to optimise behaviour, based on prior beliefs and incoming sensory information (Friston 2005). Within PP, expectation suppression is the attenuation of neural responses to expected stimuli (Todorovic & de Lange, 2012). In light of competing 'dampening' and 'sharpening' accounts of ES (which suggest that ES is underpinned by the reduction of neurons tuned towards, or away from the stimuli, respectively), the Opposing Process Theory (OPT) (Press et al., 2020) seeks to reconcile these accounts and theorises that both processes may occur at different time points in the predictive process. The (OPT) has suggested that initial sharpening is followed by later dampening of the neural representations of the expected stimulus as learning progresses. Our recent EEG study (McDermott et al., 2024) yielded direct evidence that ES is expressed as two opposing mechanisms, sharpening followed by dampening within trials, but dampening followed by sharpening across trials as learning takes place. We hypothesise that within-trial latency differences in valid vs. invalid trials might be mediated by relatively lower- vs. higher-level regions, while across-trial timing differences suggest that higher-order regions may learn statistics faster than lower-order regions. This is supported by earlier research which proposes a distinct mechanism, i.e., top-down pattern-separability, through which predictions exert their effects. (Nigam & Schwiedrzik, 2024). In this 7T fMRI study, we test if these dissociable dynamics of expectation effects can be explained in the context of hierarchical learning mechanisms.

Methods

Experimental Paradigm. Fifteen healthy participants were exposed to pairs of stimuli, comprising artificially generated natural scenes drawn from different scene categories, followed by artificially generated monosyllabic words split into male and female voices. In each trial, participants were presented with a Leading stimulus ie. an image from one category, followed by a Trailing stimulus, ie. a spoken monosyllabic word (Fig. 1). Categories were paired in a 2:1 manner, where two different 'Leading' categories could result in one 'Trailing' category with 75% validity. Participants did not undergo a training session, such that the neural signatures of early learning processes may be investigated using laminar fMRI. Each participant completed eight blocks consisting of 64 trials per block, totalling ~ 90 mins of testing.

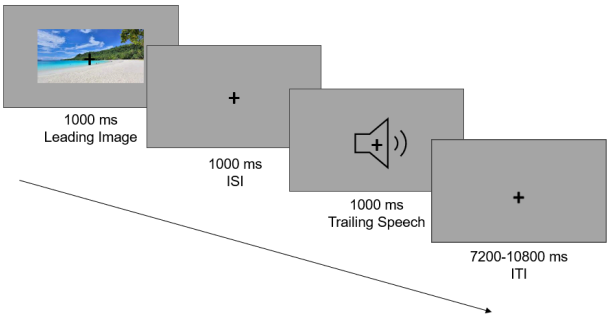

Figure 1: Associative learning experimental paradigm. Leading image category (e.g. beach)

predicted trailing speech category (male/female voice) with 75% validity.

Data analysis. Anatomical and functional data was analysed using BrainVoyager (BV, version 23.2, Brain Innovation, Maastricht, the Netherlands) and custom Python scripts. Preprocessing included slice scan time correction using sinc interpolation, 3D motion correction using trilinear interpolation and high-pass filtering (GLM Fourier) with a cut-off of 7 cycles per run. We conducted a whole-brain voxel-wise GLM contrast to compare valid versus invalid conditions averaged across all 8 runs. Thresholding was applied at $p < 0.001$ and results were visualised in coronal and transverse views (Fig. 2).

Results

Whole-brain voxel-wise analyses revealed significant activation differences between valid and invalid trials The resulting activation map identified clusters of significant activation primarily in the auditory cortex and occipital cortex. Significant clusters were observed with peak activation at $t = 5.57$ in Heschl's gyrus.

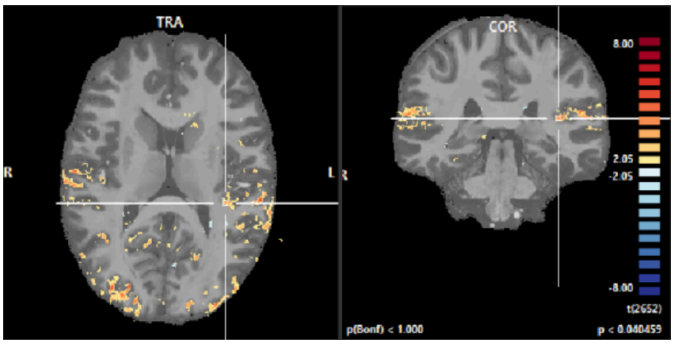

Figure 2: Contrast map of activation in valid vs invalid conditions.

Discussion

The present study aims to assess dissociable dynamics of expectation effects in the context of hierarchical learning mechanisms. In line with expectations, there are significant differences in activation between valid and invalid trials, suggesting expectation suppression effects. This leads the way for more extensive analyses, including decoding analyses to parse out the opposing effects of sharpening and dampening both across and within trials. Based on previous research (Press et al., 2020) and our previous EEG study, we expect that sharpening effects will lead to higher decoding accuracy for valid trailing trials earlier within trials (due to increased signal-to-noise ratio), and dampening effects will lead to higher decoding for invalid trials later within trials (due to overall reduced neural responses). We also predict, based on the findings of Nigam & Schwiedrzik (2024) that sharpening will be found in deep layers of the in higher-order regions (e.g. the Superior Temporal Gyrus) which likely signal predictions, while dampening will be found in superficial layers of lower-order regions (e.g. A1) which likely signal prediction errors.

References

Friston, K. A. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences, 360*(1456), 815-836.

McDermott, H. H., De Martino, F., Schwiedrzik, C. M., & Ausztulewicz, A. (2024). Dissociable dynamic effects of expectation during statistical learning. *eLife,* **13**:RP103689

Nigam, T. & Schwiedrzik, C. M. (2024). Predictions enable top-down pattern separation in the macaque face-processing hierarchy. *Nature Communications, 15*(1), 7196.

Press, C., Kok, P., & Yon, D. (2020). The perceptual prediction paradox. *Trends in Cognitive Sciences, 24*(1), 13-24.

Todorovic, A. & de Lamge, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience, 32*(39), 13389-13395.