Neural Network model of Repetition Suppression using Dynamic Expectation Maximisation

R. Henson [1] & K. Friston [2]

1. MRC Cognition & Brain Sciences Unit, Cambridge, England, 2. Functional Imaging Laboratory, University College London, England

Introduction

Repetition Suppression (RS) refers to reduced haemodynamic responses to repeated versus initial stimulus presentations. Repetition also modulates evoked EEG/MEG responses but often only after the initial stimulus-specific transient[1]. Here we present a neural network model based on Dynamic Expectation Maximisation (DEM)[2] that simulates EEG/MEG and fMRI data.

Network

DEM deconvolves the inputs to a system to infer its hidden states (**D**-step). It also estimates the system's parameters (**E**-step) and hyperparameters (**M**-step), which encode uncertainty about the parameters. The parameters (*i.e.*, neuronal connection strengths) control the dynamics of the states (*i.e.*, neuronal activity). All three steps rest on minimising the system's free energy.

A simple network was constructed with a lower (sensory) layer and a higher layer. Each layer comprised neurons encoding 1) output to lower areas, $v^{(i)}$ (e.g, perceptual causes), 2) hidden states $dx^{(i)}/dt = f(x^{(i)},v^{(i+1)})$ and 3) prediction error, $e^{(i)} = v^{(i)} - g(x^{(i)},v^{(i+1)})$, where $g(x^{(i)},v^{(i+1)})$ is the prediction from top-down causes. The prediction error is passed "forward" to adjust the causes at higher levels. Our lower layer had 4 output neurons, representing the percept, and 2 hidden states; the higher layer had 1 output neuron, representing the cause. Stimuli were created by running the network in generative mode, using random values for the parameters of *f*. Random noise was then added to both *f* and *g*, and noised versions of the stimuli presented to the network multiple times for recognition.

Simulated ERPs reflect the mean local field potentials (LFPs) of error neurons (assumed to be superficial pyramidal neurons), firing-rates (peristimulus histograms, PSTH) are a non-negative function of the LFPs, and simulated BOLD is proportional to the integrated PSTH.

Results

Figure 1 (left) shows exemplar spatiotemporal responses (4 units by 24 timepoints) corresponding to the stimulus, and the network's prediction after 1 or 8 presentations. Figure 1 (right) shows the corresponding ERPs from the lLower and higher layers. Figures 2-4 show mean values across 100 simulations, for less (left) and more (right) degraded versions of the stimuli. Figure 2 (top) shows that free energy decreases across 8 presentations; the middle and lower panels show a more rapid decrease in total activity (summed firing-rates) in the lower layer, and greater total activity for more degraded stimuli. Figure 3 shows the PSTH for 1st and 8th presentations, with RS emerging sooner in the higher layer for less degraded stimuli. Figure 4 shows the predicted BOLD impulse responses (assuming PSTH time units of order 10ms), demonstrating RS together with a shorter peak latency in the lower layer.

Conclusion

Our network model simulates evoked responses for M/EEG and fMRI. RS reflects faster reduction in prediction error. Interesting predictions of the model include: RS asymptotes as a function of repetition, sooner in lower than higher layers, and repetition effects can onset earlier in higher than lower regions (with little degradation of stimuli). If correct, the model questions the use of RS to localise neuronal representations[3].

References

1. Henson, 2003, Prog. Neurobiol.
2. Friston, 2005, Philos Trans R Soc Lond B Biol Sci.
3. Grill-Spector, Henson & Martin, 2006, Trends Cog Sci.

Fig 1. Left: example stimulus (top) and 1st and 8th prediction (middle/bottom). Right: ERPs in lower (top) and higher (bottom) layer for 1st (dark) and 8th (light) presentation

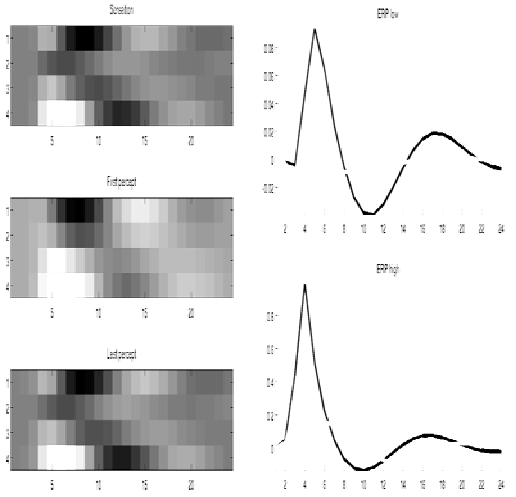

Fig 2. Mean Free Energy (top panels) and Total Activity in lower (middle panels) and higher (bottom panel) layers for less (left) and more (right) degraded stimuli

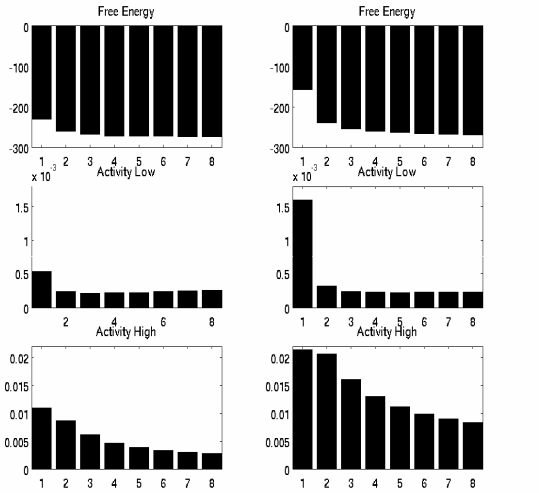

Fig 3. Mean Peristimulus Histograms (PSTH) of firing-rates in lower (top panels) and higher (bottom panels) layers for less (left) and more (right) degraded stimuli

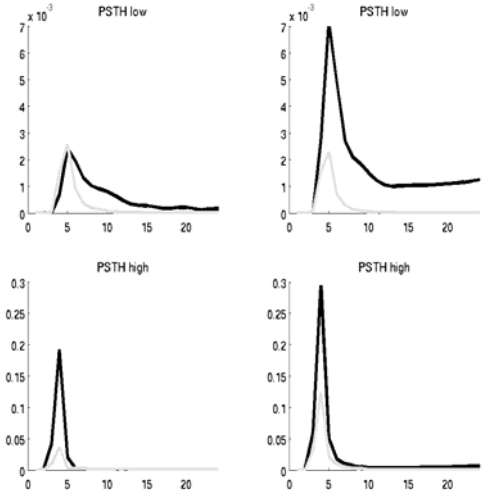

Fig 4. Predicted BOLD impulse response in lower (middle panels) and higher (bottom panels) layers for less (left) and more (right) degraded stimuli