City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/19419/

Link to published version:

New Approaches for Studying Cortical Representations

Dimitris A. Pinotsis[1,2] and Earl K. Miller[1]

[1]The Picower Institute for Learning & Memory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

[2]The Wellcome Trust Centre for Neuroimaging, University College London, Queen Square, London WC1N 3BG

pinotsis@mit.edu

Abstract

We review two new approaches for studying cortical representations of sensory stimuli. These exploit optimization algorithms and auto-encoders from machine learning and high resolution electrophysiology data. We show how these approaches can shed new light into the information processing and maintenance taking place in neuronal populations. These approaches allow us to study: 1. Changes in the precision of error representations as a result of neuromodulation. 2. Differences in the cortical connectivity underlying memory representations for different stimuli.

Introduction

Following the advent of modern recording systems and multi-electrode arrays, it has become apparent that the brain processes information at both the single cell and neuronal population levels. We here review two approaches that shed new light on information processing at the population level. These exploit optimization algorithms and auto-encoders from machine learning and animal data obtained using: 1. Microelectrodes that penetrate different cortical layers. 2. Multi-electrode arrays that can sample from multiple locations within a cortical area.

Error Representations in Superficial Cortical Layers

Different cortical layers are thought to represent sensory stimuli in different ways. The exact role of these representations remains unknown. According to Hierarchical Bayesian Inference, activity of deep pyramidal cell (PC) populations might represent states associated with sensory stimuli while superficial PC activity might represent errors. A way to test this hypothesis is through a model that can distinguish between activities in deep and superficial cortical layers. This model should also make predictions about the role of different cortical representations in these layers. To test these predictions, one can then perform some experimental manipulation. We describe this below. We also review an approach for fitting mass models to data recorded using microelectrodes.

Our approach included the following steps. We first developed a biophysical neural mass model that could predict data from different cortical layers. We then established the validity of this model using a more detailed compartmental model. We showed that both models predicted the same activity. Compartmental models cannot be fit to real data from microelectrodes due to the high number of their parameters. They are thus limited and only used for simulations. Using a maximum likelihood algorithm, we then fitted the mass model to two different datasets obtained with microelectrodes. We showed that our model could correctly predict activity from supragranular and infragranular layers in both cases. We analyzed cross spectral density data from: 1. the prefrontal cortex (PFC) during the delay period of monkey memory guided saccade task (Pinotsis et al., 2016). 2. The primary visual cortex (V1) during a visual perception task with optogenetic activation of the basal forebrain in a mouse (Pinto et al., 2013).

A Test for Hierarchical Bayesian Inference

Our model predicted that superficial PC activity during the visual perception task should be enhanced in the presence of cholinergic neuromodulation. Neuromodulation is a candidate mechanism that can explain changes in the precision of error representations. Hierarchical Bayesian Inference suggests that these representations might be subserved by superficial PC activity. Our model described the precision of these representations by a parameter controlling the local inhibition in the superficial PC population. This follows from a correspondence between parameters describing learning in Hierarchical Bayesian Inference and the biophysics captured by the neural mass model. The model predicted that local inhibition the superficial PCs should decrease as a result of optogenetic activation of cholinergic neurons in the basal forebrain that project to V1. This prediction was confirmed by our analysis (Figure 1; see also Pinotsis et al., in press):
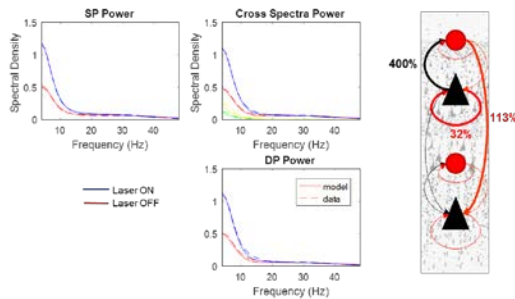

Figure 1. Left. Cross Spectral Density Data and Model Fits from Pairs of Superficial and Deep Contacts (Laser ON and OFF).Right. : Conditional parameter estimates and their trial specific changes.

Memory Representations of Different Cued Locations in Frontal Areas

Above, we considered error representations in the context of information processing performed by the brain. We focused on Hierarchical Bayesian Inference. In the remaining part of this paper, we consider memory representations in the context of information maintenance. We focus on combining ideas from biophysical modeling and variational auto-encoders. We also focus on spatial aspects of cortical representations and review a new approach for describing their structure.

Memories, thoughts, rules, etc. are thought to be represented by a group of co-activated neurons, called neural ensemble. Describing neural ensembles is a challenge: the complexity of the underlying cortical micro-circuitry is immense. Current approaches often describe ensembles in a piecemeal fashion. They focus on single neurons and employ local measures like pairwise correlations between neurons or recording sites (Buschman et al., 2012). We introduced an alternative approach that identifies and describes neural ensembles in a holistic fashion (Pinotsis et al., preprint).

Neural ensembles and cortical connectivity

We analyzed data from multiple-electrode recordings obtained during the delay period of a classic test of working memory: spatial delayed response. We used a neural field model and studied the spatial organization of neural activity induced by different cued locations within three cortical areas: PFC, frontal eye fields (FEF) and supplementary eye field (SEF). We identified neural ensembles associated with different cued locations. We described the connectivity patterns that underlie ensemble activity. We also obtained estimates of connectivity weights, that is, parameters that describe the strength of the connections between the recording sites within each cortical area.

A new explanation for the oblique effect in psychophysics

We characterized ensemble connectivity in terms of characteristic path lengths. We showed that these graph theoretic measures captured behaviorally relevant information. In FEF, cues on the horizontal axis had shorter characteristic path lengths than others. This connectivity could explain the oblique effect (psychophysics performance is better for stimuli on than off the horizontal axis). This in turn might be the result of a more parsimonious micro-circuitry. We concluded that maintaining horizontal cued locations might require less energy and result in faster information processing. Also, characteristic path length values appeared to partition the space of possible cued locations into three subsets: locations on the horizontal axis and in the left and right visual hemifields. They were

shortest for the horizontal axis, next shortest for locations in the right visual hemisphere and longest for locations in the left visual hemisphere. Characteristic path length values for each pair of locations on the horizontal axis in all three areas were similar in value (Figure 2).
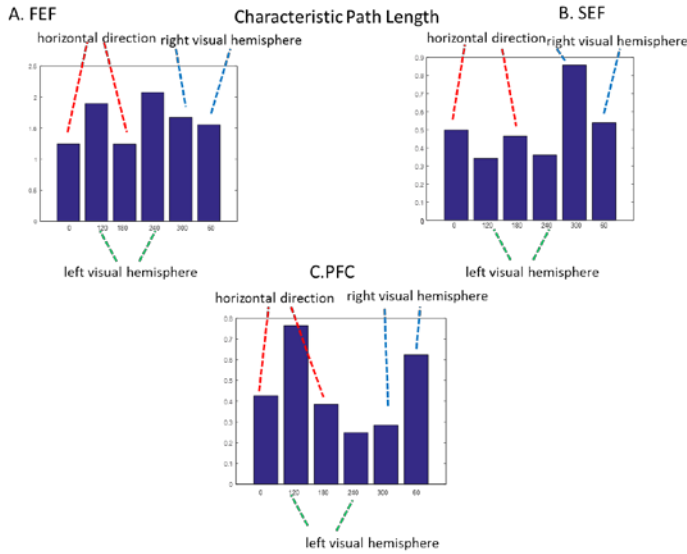

Figure 2. Characteristic path length estimates for all cued locations and corresponding FEF, SEF and PFC neural ensembles.

Maximum Likelihood Optimization and Deep Networks

The two approaches for the analysis of brain data presented above are based on maximum likelihood optimization algorithms. The first approach focused on studying representations at different cortical layers. This approach used ideas from nonlinear dynamical systems, Bayesian Model Comparison (Kass and Raftery, 1995) and Expectation Maximization (EM; Dempster et al., 1977). Model fitting to real data was performed using a Fisher Scoring ascent on the objective function (Friston et al., 2003).

The second approach focused on describing the structure of memory representations. This was based on training a neural field model as a particular type of a deep network called an auto-encoder. This neural field can learn and

maintain its inputs. We trained this model using an EM algorithm, similar to a variational auto-encoder (Kingma and Welling, 2013).

Biophysical Modeling

To explain data obtained using microelectrodes, we fitted a four population neural mass model. These populations are thought to occupy different cortical layers. Our model included two pairs of pyramidal cell populations and inhibitory interneurons in the superficial and deep cortical layers according to the cortical architecture suggested by (Bush and Sejnowski, 1993). Firing rates within each subpopulation provided inputs to other populations and convolution of presynaptic activity produced postsynaptic depolarization. We considered separate time series of activity from superficial and deep populations as opposed to past work in the literature that uses weighted sums of activity from all subpopulations.

We described high spatial and temporal resolution data from memory networks in frontal areas using a model of wave dynamics known as a neural field. This allowed us to use patterns of LFP activity across recording sites to infer the underlying functional connectivity for each of the cued locations. The neural field model described each ensemble's network interactions and predicted patterns of activity that correspond to different attractor states, see also (Durstewitz et al., 2000).

Conclusions

We reviewed two recent approaches for the analysis of electrophysiology data of high spatial and temporal resolution. These combine ideas from biophysical modeling and machine learning. We wanted to show how marrying these two fields can shed new light to questions of importance in cognitive neuroscience. As an illustration, we showed how to answer questions about physiological processes like cholinergic neuromodulation and psychophysics like the oblique effect.

Acknowledgments

References

Buschman, T.J., Denovellis, E.L., Diogo, C., Bullock, D., and Miller, E.K. (2012). Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron 76, 838–846.

Bush, P.C., & Sejnowski, T.J. (1993). Reduced compartmental models of neocortical pyramidal cells. J. Neurosci. Methods 46, 159–166.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. Nature neuroscience, 3, 1184-1191.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 1-38.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. Neuroimage, 19(4), 1273-1302.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the american statistical association, 90(430), 773-795.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

Pinotsis, D. A., Loonis, R., Bastos, A. M., Miller, E. K., & Friston, K. J. (2016). Bayesian Modelling of Induced Responses and Neuronal Rhythms. Brain Topography, 1-14.

Pinotsis, D.A., Geerts, J.P., Pinto, L., FitzGerald, T.H.B., Litvak, V., Auksztulewicz R., & Friston, K.J. (in press) Linking canonical microcircuits and neuronal activity: Dynamic Causal Modelling of Laminar Recordings, Neuroimage.

Pinotsis, D.A., Brincat S., & Miller, E.K. (preprint). On memories, neural ensembles and mental flexibility.

Pinto, L., Goard, M. J., Estandian, D., Xu, M., Kwan, A. C., Lee, S. H., & Dan, Y. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. Nature neuroscience, 16(12), 1857-1863.