A Mathematical Framework for the Global Brain

JG Taylor
Department of Mathematics, King's College, Strand, London WC2R2LS, UK
Email: john.g.taylor@kcl.ac.uk

Abstract
We develop a mathematical framework of continuum neural field theory as the beginning point for an analysis of the global brain. Arguments for and against this approach are developed. Applications to cortical dynamics and learning in somato-sensory cortex, visual cortex and motor control are briefly reveiewed. Extensions to a broader range of brain systems, including attention and emotions, are outlined. Open problems are listed to conclude the paper.

1. Introduction

The brain is an amazingly complex dynamical system. This has been found to be true at all levels of its investigation: from overall dynamics of the global brain down to the functioning of synapses, involved with a wealth of different neuro-chemicals. However I want to contrast those studies that get down to what is correct to call the 'nitty-gritty' of such micro-processing as compared to those that take a more global view The former tasks are involved with systems of variables more controllable than in the global case, as are the possibilities of performing experiments to test predictions of models. That is even possible at a still higher level, as for example in recent careful 'bottom-up' models of the cerebellum in its role in conditioned learning [1]. However there are only few detailed models of the dynamical interactions of modules across the vast reaches of the brain. Even these tend to be based on simulations, and not any mathematical principles. In this paper we take a step towards proposing a program to attempt to lift our modelling and analysis sights higher to the global brain. We do that in terms of recent functional proposals for overall control systems in the brain: of attention [2, 3], of motor control [4, 5], of emotions [6, 7] and even of consciousness [8].

To advance in this task of analysing the global brain, we have still to consider the level we will use to model it. Do we take it lobe by lobe, or module by module, or column by column, or neuron by neuron, or even down at synaptic level, or even lower at molecular level? The principle I propose to start with to attack the brain globally is to take it as composed of the simplest sorts of components, and try to determine what these could achieve. I will use functional guides coming from control theory to indicate how the different modules are expected to function together, either in attention, motor response or in emotion processing.

I start this program in the next section, where I present the basic components of the approach. This is then formulated in detail mathematically in the following section, and expected results from the mathematical analysis are developed in section 4. The paper concludes with a conclusion in section 5.

2. The Brain's Basic Components

We have already discussed the main problem we face in such a program: it possesses too much complexity everywhere we look. To underline this point, we have to face up to the hierarchy composed of:

- Chemical components of synapses
- Channel variable dynamics for each synapse
- Overall dynamics of each compartment
- Overall dynamics of each neuron
- Overall dynamics of each column
- Overall dynamics of each module
- Overall dynamics of the overall brain.

If each level were attempted to be modelled faithfully we would have enormous complexity of the overall system by the time we arrive at the overall brain itself. But then that is the brain in reality; we have to try to face up to it.

One way to tackle such complexity is by making a sequence of ever more complex approximations. The starting point would then reduce to a problem which looks as if it has a chance of being solved reasonably well. Further additions of complexity, going back up the bullet points above, would bring as back to the brain in all its complexity. However we would at least have some idea of what the powers were of the basic 'first approximation', and what further powers are needed, and might be added, by the further steps going up the ladder of bullet points.

The first step – of formulating the basic or simplest approximation–is the most important. It must be so structured as to lead us to expect interesting features that would be able to lead to some of the true powers of the brain in reality. But it must be simple enough to be relatively soluble (where that means existence theorems but not necessarily analytic forms of solution are obtainable).

I want to propose a particular first approximation which has drawn considerable interest, but only at the single (or few) module level. It is that of continuum neural field theory (CNFT). This was started in 1977 by Amari [9] who proved some remarkable features of a certain class of CNFT in 1-dimension: the existence of long-term solutions, or bubbles of restricted neural activity. This was extended more recently [10] to the 2-dimensional case. A little later Amari and colleagues developed a mathematical analysis of learning of afferents in the CNFT framework, in terms of the distribution of inputs [11]. That has since been applied to the brain in a variety of ways: showing the development of orientation selectivity in a structured form in V1 [12], supporting various illusion in visual perception [13}, the unmasking features of somato-sensory cortex when certain components of the inputs were damaged (and agreement with experimental results on this unmasking) [14, 15], various applications to psychological experiments [16], applications in motor control [17, 18, 19].

All of these applications show the value of CNFT for modelling local processing, by one or two modules, in the brain. However I want to extend this approach to a framework of global brain processing. The simplest way to proceed on this is to take hard-wired interacting CNFTs for the modules across the brain, and attempt to solve the resulting dynamics, or at least obtain general features of this dynamics. The next step is to include learning as done by Amari and his colleagues in their move from [9] dynamics to afferent learning [11].

3. Mathematics of the Simple Brain

A CNFT model is based on the approximation of a module of neurons as composed of a two-dimensional sheet of neurons, with the neuron at position **r** on the sheet having membrane potential V(**r**). The simplification in this 'sheet-like' assumption allows us to consider many neurons at once, although we can relatively easily reduce the sheet back to a finite number of neurons by using the localised distributions of neurons at a finite set of points on the sheet.

The dynamics of the neurons is also greatly simplified by assuming a graded response pattern for each neuron output (although this can be extended to spiking neurons if needed). Thus the output of each neuron is taken to be some function f( ) of its potential; the simplest case is a step function, although results have been obtained for more general sigmoid functions. Thus the simplest dynamics is taken as

$$\tau dV(\mathbf{r})/dt = -V(\mathbf{r}) + w * f(V)(\mathbf{r}) + I(\mathbf{r}) - h \qquad (1)$$

where I is the external input to the module at that point, w(**r, r'**) denotes the lateral connection strength between the neuron at **r'** and that at **r**, -h is a constant inhibitory bias to all neurons to assure stability and suitable competition between neurons, and * is the usual symbol for the convolution product taken over the positions of the module.

We now extend (1) to a set of interacting modules as

$$? \, d\mathbf{V}(\mathbf{r})/dt = -\mathbf{V}(\mathbf{r}) + W * f(\mathbf{V})(\mathbf{r}) + \mathbf{I}(\mathbf{r}) - H \qquad (2)$$

where the extension of (1) to (2) is achieved by taking **V** to be a vector-valued field of membrane potentials (each component denoting the neural field for a given module),
W(**r', r**) now denoting the matrix of field connections, with diagonal connections being the lateral one w in (1), the off-diagonal ones being those connecting different modules, -H is now a diagonal matrix, with constant values in each entry for a given module although with possible differences across modules to allow different levels of overall inhibition, and I denoting a vector field of external inputs, each component again being associated with a given neural field module; the matrix **?** denotes a diagonal matrix of time constants (where also in (1) this can be extended to different time constants for different neuron positions if so desired). We note that we are simplifying by taking the same co-ordinates for each module; again that can be generalised. Further we are only taking neurons of the same type in (1) and (2); again that can be extended to those of inhibitory and excitatory form or of different sub-populations, by suitably extending the notation in these equations.

We now indicate some of the features expected from (2), extending those possessed by (1) – bubble existence, dynamics of bubbles, learning structures.

a) Basic Features of the Dynamics

A Liapunov function can be derived for the dynamics of (2) extending that for (1) in the case of symmetric connection matrix W (across modules as well across the lateral connections in each module), as in [17], thus providing general stability arguments.

b) Existence of Bubbles

The one- and two-dimensional bubble analyses of [9] and [10], and the many simulations in the references [12 – 19], lead to the expectation that there will exist multiple bubbles, across a range of modules. It is possible to develop equations for coupled bubbles from (2); that will be described elsewhere.

c) Dynamics of Bubbles

Bubbles have been found [9 – 19] to driven to flow to regions of highest input. A similar situation is expected to occur for the coupled bubbles in the expanded version (2) of NFT; the nature of these bubbles will be decsribed.

d) Learning Structures

The original one-dimensional work of [11] was extended in [10] to two dimensions and to applications to specific brain modules in [12 – 19]. The most crucial feature of this study was the presence and exploitation of instability in the learning law dynamics, producing discontinuous periodic structures mapping higher dimension input spaces down to the two-dimensional sheet in a clearly defined, even analytic, manner. Similar structures are to be expected in the extended case of (2). How this applies when variations of the lateral connection weight functions are included (as described below to encode genetic memory) will be discussed in detail.

4. Insertion of Control Structures

So far we have only extended the standard CNFT model of a single module to that of several such modules, without any understanding of how functional differentiation can be included in the model. We now turn to that important aspect. To justify our approach we need to accept that we cannot expect our extended model to learn its feed-forward and feedback connections all on its own, without any use of genetic memory. This will have been built up over many millions of years by pressure of the environment. It has led to crucial differences between modules that allow them to be differentiated into input processing modules, higher level control modules and response modules. The first and third of these modules have already been discussed in the brain context in [12-19]. Here we turn specifically to the second class of modules, those for attention control, by suitable assumptions on the lateral connection matrix W: by depth and width of the lateral connection matrix internal to a module. This affects the size of bubbles, and the overall level of the WTA nature of the module. The higher level modules in parietal lobe will therefore be allocated large values of inhibitory connections so as to provide a strong bias towards competition and hence generation of attention control signals. The feedback signals will also be sculpted initially by hand so as to provide excitation to lower level sites, as well as more distributed inhibition.

At the same time the possibility of applying developmental knowledge to the learning process will also be considered. This will be achieved by including learning in an incremental fashion, so that lower level representations will be learnt, and stabilised, before further learning under top-down control will be analysed. Furthermore the manner in which goal representations in prefrontal cortex could arise will be considered.

Emotion will be added by addition of 'valence' modules (amygdale and orbito-frontal cortex), following the emotion brain architecture already presented elsewhere [7], but now represented in the CNFT framework.

5. Results of the Program

The basic results are of three sorts:

1) Coupled bubble formation and dynamics, under simple feed-forward- & feedback coupling assumptions, with sizes and expected influences of bubbles on each other determined by relative parameter choices and fan-in values in the various modules;

2) Learning of cortical representations, both of feedback and feed-forward form, supporting topographic spatial and localised object representations (using pre-specified fan-ins depending on the site of the CNFT module being considered.

3) Provision of a basis for addition of further complexity into the system, as well as applying other criteria, such as information maximisation, to constrain the approach.

6. Conclusions

A general framework has been developed to attack the brain. It allows stable state analysis as well as extension to the temporal dynamics of a set of interacting CNFT modules. Learning presents also dynamical features that allow the analysis of pattern structure of the synaptic weights. The nature of emotional modulation has yet to be properly inserted by use of reward learning, but this will be included in subsequent version of the over CNFT brain model.

Much work lies ahead, but general features will be obtainable that will indicate the value of the approach.

Acknowledgments

The author would like to thank his colleagues N Fragopanagos, N Taylor, M Hartley, and S Kasderides for numerous useful discussions on attention and emotions.

References

[1] Ohyama T, Nores WL, Murphy M & Mauk MD (2003) What the Cerebellum computes. Trends in Neuroscience 26:222-227

[2] Corbetta M & Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201-215

[3] Taylor JG & Fragopanagos N (2003)

[4] Miall RC & Wolpert D (1996) Forward models for physiological motor control. Neural Networks 9:1265-1279

[5] Wolper DM, Miall RC & Kawato M (1998) Internal models in the cerebellum. Trends in Cognitive Science 2:338-347
[6] Pessoa L, McKenna M, Gutierrez E & Ungerleider LG (2002) Nural processing of emotional faces requires attention. Proc Ntl Acad Sci USA99:11458-11463
[7] Taylor JG, Fragopanagos N, Cowie R, Douglas-Cowie E, Fotinea S-E & Kollias S (2003) An Emotional Recognition Architecture Based on Human Brain Structure. Proc ICANN/ICONIP 2003, pp 1133-1140. Berlin: Springer
[8] Taylor JG (2003) Prg Neurbiology 41:305-335
[9] Amari S-I (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. Biological Cybernetics 27:77-87
[10] Taylor JG (2000) Neural 'bubble' dynamics in two dimensions: foundations. Biological Cybernetics 80:393-409
[11] Takeuchi A & Amari S-I (1999) Formation of Topographic Maps and Columnar Microstructures in Nerve Fields.
[12]] Fellenz W & Taylor JG (2000) Establishing retinotopy by lateral-inhibition-type homogeneous neural fields. ESANN Proceedings ES 200-49 (http://www.dice.ucl.ac.be/esann/proceedings/esann2000/content.htm#ES2000-49)
[13]Taylor JG (1997) Perception by neural networks. Neural Network World 4, 5:363-395
[14] Petersen RS & Taylor JG (1996) Reorganization of somatosensory cortex after tactile training. pp 82-88 In Touretsky DS, Mozer MC, Hasselmo ME (eds) Advances in Neural Information Processing Systems, Cambridge Mass: MIT Press.
[15] Peteresen R (1997) Modelling Learning in the Brain. London Univ PhD thesis (unpublished)
[16] Trappenberg TP, Dorris M, Klein RM & Munroe DP (2001) A model of saccade intiaition based on hte competitive integration of exogenous and engoenous signals from the superior colliculus. J Cognitive Neuroscinece 13:256-271
[17] Giese MA (1998) Dynamic Neural Field Theory for Motion Perception. Boston: Kluwer Academic Publishers
[18] Stringer SM, Rolls ET, Trappenberg TP & de Araujo IET (2003) Self-organising continuous attractor networks and motor function. Neural Networks 16:161-182
[19] Stringer SM, Rollls ET & Trappenberg TP (2003) Self-organising continuous attractor networks with mulitple active packets, and the representation of space. Neural Networks (in press)
[20] Taylor JG (2003) Bubbles in the Brain? Trends in Cognitive Neuroscience 7:429-430