CHAPTER

3

Modelling brain responses

K. Friston and K. Stephan

INTRODUCTION

In the previous chapter, we focused on the practical issues encountered in the analysis of neuroimaging data. In this chapter, we look at modelling in a more principled way; placing statistical parametric mapping in the larger context of modelling distributed brain responses. Inferences about the functional organization of the brain rest on models of how measurements of evoked responses are caused. These models can be quite diverse, ranging from conceptual models of functional anatomy to mathematical models of neuronal and haemodynamics. The aim of this chapter is to introduce the key models used in imaging neuroscience and how they relate to each other. We start with anatomical models of functional brain architectures, which motivate some of the fundaments of neuroimaging. We then reprise basic statistical models (e.g. the general linear model) used for making classical and Bayesian inferences about *where* neuronal responses are expressed. By incorporating biophysical constraints, these basic models can be finessed and, in a dynamic setting, rendered causal. This allows us to infer *how* interactions among brain regions are mediated. The chapter serves to introduce the themes covered in the final three parts of this book.

We will review models of brain responses starting with the general linear model of functional magnetic resonance imaging (fMRI) data discussed in the previous chapter. This model is successively refined until we arrive at neuronal mass models of electroencephalographic (EEG) responses. The latter models afford mechanistic inferences about how evoked responses are caused, at the level of neuronal subpopulations and the coupling among them.

Overview

Neuroscience depends on conceptual, anatomical, statistical and causal models that link ideas about how the brain works to observed neuronal responses. Here we highlight the relationships among the sorts of models that are employed in imaging. We will show how simple statistical models, used to identify *where* evoked brain responses are expressed (cf. neo-phrenology) can be elaborated to provide models of *how* neuronal responses are caused (e.g. dynamic causal modelling). We will review a series of models that cover conceptual models, motivating experimental design, to detailed biophysical models of coupled neuronal ensembles that enable questions to be asked at a physiological and computational level.

Anatomical models of functional brain architectures motivate the fundaments of neuroimaging. In the first section, we review the distinction between functional *specialization* and *integration* and how these principles serve as the basis for most models of neuroimaging data. The next section turns to simple statistical models (e.g. the general linear model) used for making classical and Bayesian inferences about functional specialization in terms of where neuronal responses are expressed. By incorporating biological constraints, simple observation models can be made more realistic and, in a dynamic framework, causal. This section concludes by considering the biophysical modelling of haemodynamic responses. All the models considered in this section pertain to regional responses. In the final section, we focus on models of distributed responses, where the interactions among cortical areas or neuronal subpopulations are modelled explicitly. This section covers the distinction between *functional and effective connectivity* and reviews dynamic causal modelling of functional

Statistical Parametric Mapping, by Karl Friston et al.
ISBN–13: 978-0-12-372560-8 ISBN–10: 0-12-372560-7

integration, using fMRI and EEG. We conclude with an example from ERP (event-related potential) research and show how the mismatch negativity (MMN) can be explained by changes in coupling among neuronal sources that may underlie perceptual learning.

ANATOMICAL MODELS

Functional specialization and integration

From a historical perspective, the distinction between functional specialization and functional integration relates to the dialectic between *localizationism* and *connectionism* that dominated thinking about brain function in the nineteenth century. Since the formulation of phrenology by Gall, who postulated fixed one-to-one relations between particular parts of the brain and specific mental attributes, the identification of a particular brain region with a specific function has become a central theme in neuroscience. Somewhat ironically, the notion that distinct brain functions could, at least to some degree, be localized in the brain was strengthened by early scientific attempts to refute the phrenologists' claims. In 1808, a scientific committee of the Athénée at Paris, chaired by Cuvier, declared that phrenology was an unscientific and invalid theory (Staum, 1995). This conclusion, which was not based on experimental results, may have been enforced by Napoleon Bonaparte (who, allegedly, was not amused after Gall's phrenological examination of his own skull did not give the flattering results he expected). During the following decades, lesion and electrical stimulation paradigms were developed to test whether functions could indeed be localized in animal models. Initial lesion experiments on pigeons by Flourens gave results that were incompatible with phrenologist predictions, but later experiments, including stimulation experiments in dogs and monkeys by Fritsch, Hitzig and Ferrier, supported the idea that there was a relation between distinct brain regions and certain cognitive or motor functions. Additionally, clinicians like Broca and Wernicke showed that patients with focal brain lesions in particular locations showed specific impairments. However, it was realized early on that, in spite of these experimental findings, it was generally difficult to attribute a specific function to a cortical area, given the dependence of cerebral activity on the anatomical connections between distant brain regions; for example, a meeting that took place on August 4th 1881 addressed the difficulties of attributing function to a cortical area, given the dependence of cerebral activity on underlying connections (Phillips *et al*., 1984). This meeting was entitled 'Localisation of function in the cortex cerebri'. Goltz (1881), although accepting the results of electrical stimulation in dog and monkey cortex, considered that the excitation method was inconclusive, in that the movements elicited might have originated in related pathways, or current could have spread to distant centres. In short, the excitation method could not be used to infer functional localization because localizationism discounted interactions, or functional integration among different brain areas. It was proposed that lesion studies could supplement excitation experiments. Ironically, it was observations on patients with brain lesions some years later (see Absher and Benson, 1993) that led to the concept of *disconnection syndromes* and the refutation of localizationism as a complete or sufficient explanation of cortical organization. Functional localization implies that a function can be localized in a cortical area, whereas specialization suggests that a cortical area is specialized for some aspects of perceptual or motor processing, and that this specialization is anatomically *segregated* within the cortex. The cortical infrastructure supporting a single function may then involve many specialized areas whose union is mediated by the functional integration among them. In this view, functional specialization is only meaningful in the context of functional integration and vice versa.

Functional specialization and segregation

The functional role of any component (e.g. cortical area, sub-area or neuronal population) of the brain is defined largely by its connections. Certain patterns of cortical projections are so common that they could amount to rules of cortical connectivity. 'These rules revolve around one, apparently, overriding strategy that the cerebral cortex uses – that of functional segregation' (Zeki, 1990). Functional segregation demands that cells with common functional properties be grouped together. This architectural constraint necessitates both convergence and divergence of cortical connections. Extrinsic connections among cortical regions are not continuous but occur in patches or clusters. This patchiness has, in some instances, a clear relationship to functional segregation. For example, V2 has a distinctive cytochrome oxidase architecture, consisting of thick stripes, thin stripes and inter-stripes. When recordings are made in V2, directionally selective (but not wavelength or colour selective) cells are found exclusively in the thick stripes. Retrograde (i.e. backward) labelling of cells in V5 is limited to these thick stripes. All the available physiological evidence suggests that V5 is a functionally homogeneous area that is specialized for visual motion. Evidence of this nature supports the notion that patchy connectivity is the anatomical infrastructure that mediates functional segregation and specialization. If it is the case that neurons in a given

cortical area share a common responsiveness, by virtue of their extrinsic connectivity, to some sensorimotor or cognitive attribute, then this functional segregation is also an anatomical one.

In summary, functional specialization suggests that challenging a subject with the appropriate sensorimotor attribute or cognitive process should lead to activity changes in, and only in, the specialized areas. This is the anatomical and physiological model upon which the search for regionally specific effects is based. We will deal first with models of regionally specific responses and return to models of functional integration later.

STATISTICAL MODELS

Statistical parametric mapping

Functional mapping studies are usually analysed with some form of statistical parametric mapping. As described in the previous chapter, statistical parametric mapping entails the construction of continuous statistical processes to test hypotheses about regionally specific effects (Friston *et al*., 1991). Statistical parametric mapping uses the general linear model (GLM) and random field theory (RFT) to analyse and make classical inferences. Parameters of the GLM are estimated in exactly the same way as in conventional analysis of discrete data. RFT is used to resolve the multiple-comparisons problem induced by making inferences over a volume of the brain. RFT provides a method for adjusting *p*-values for the search volume of a statistical parametric map (SPM) to control false positive rates. It plays the same role for continuous data (i.e. images or time-series) as the Bonferroni correction for a family of discontinuous or discrete statistical tests.

We now consider the Bayesian alternative to classical inference with SPMs. This rests on conditional inferences about an effect, given the data, as opposed to classical inferences about the data, given the effect is zero. Bayesian inferences about effects that are continuous in space use posterior probability maps (PPMs). Although less established than SPMs, PPMs are potentially very useful, not least because they do not have to contend with the multiple-comparisons problem induced by classical inference (see Berry and Hochberg, 1999). In contradistinction to SPM, this means that inferences about a given regional response do not depend on inferences about responses elsewhere. Before looking at the models underlying Bayesian inference, we briefly review estimation and classical inference in the context of the GLM.

The general linear model

Recall from Chapter 2 that the general linear model:

$$y = X\beta + \varepsilon \qquad \textbf{3.1}$$

expresses an observed response y in terms of a linear combination of explanatory variables in the design matrix X plus a well-behaved error term. The general linear model is variously known as analysis of variance or multiple regression and subsumes simpler variants, like the *t*-test for a difference in means, to more elaborate linear convolution models. Each column of the design matrix models a cause of the data. These are referred to as explanatory variables, covariates or regressors. Sometimes the design matrix contains covariates or indicator variables that take values of zero or one to indicate the presence of a particular level of an experimental factor (cf. analysis of variance – ANOVA). The relative contribution of each of these columns to the response is controlled by the parameters β. Inferences about the parameter estimates are made using *t* or *F*-statistics, as described in the previous chapter.

Having computed the statistic, RFT is used to assign adjusted *p*-values to topological features of the SPM, such as the height of peaks or the spatial extent of blobs. This *p*-value is a function of the search volume and smoothness. The intuition behind RFT is that it controls the false positive rate of peaks corresponding to regional effects. A Bonferroni correction would control the false positive rate of voxels, which is inexact and unnecessarily severe. The *p*-value is the probability of getting a peak in the SPM, or higher, by chance over the search volume. If sufficiently small (usually less than 0.05) the regional effect is declared significant.

Classical and Bayesian inference

Inference in neuroimaging is restricted largely to classical inferences based upon statistical parametric maps. The statistics that comprise these SPMs are essentially functions of the data. The probability distribution of the chosen statistic, under the null hypothesis (i.e. the null distribution) is used to compute a *p*-value. This *p*-value is the probability of obtaining the statistic, or the data, given that the null hypothesis is true. If sufficiently small, the null hypothesis is rejected and an inference is made. The alternative approach is to use Bayesian or conditional inference based upon the posterior distribution of the activation given the data. This necessitates the specification of priors (i.e. the probability distribution of the activation). Bayesian inference requires the posterior distribution and therefore rests upon a posterior

density analysis. A useful way to summarize this posterior density is to compute the probability that the activation exceeds some threshold. This represents a Bayesian inference about the effect, in relation to the specified threshold. By computing posterior probability for each voxel, we can construct PPMs that are a useful complement to classical SPMs.

The motivation for using conditional or Bayesian inference is that it has high face validity. This is because the inference is about an effect, or activation, being greater than some specified size that has some meaning in relation to underlying neurophysiology. This contrasts with classical inference, in which the inference is about the effect being significantly different from zero. The problem for classical inference is that trivial departures from the null hypothesis can be declared significant, with sufficient data or sensitivity. From the point of view of neuroimaging, posterior inference is especially useful because it eschews the multiple-comparisons problem. In classical inference, one tries to ensure that the probability of rejecting the null hypothesis incorrectly is maintained at a small rate, despite making inferences over large volumes of the brain. This induces a multiple-comparisons problem that, for spatially continuous data, requires an adjustment or correction to the p-value using RFT as mentioned above. This correction means that classical inference becomes less sensitive or powerful with large search volumes. In contradistinction, posterior inference does not have to contend with the multiple-comparisons problem because there are no false positives. The probability that activation has occurred, given the data, at any particular voxel is the same, irrespective of whether one has analysed that voxel or the entire brain. For this reason, posterior inference using PPMs represents a relatively more powerful approach than classical inference in neuroimaging.

Hierarchical models and empirical Bayes

PPMs require the posterior distribution or conditional distribution of the activation (a contrast of conditional parameter estimates) given the data. This posterior density can be computed, under Gaussian assumptions, using Bayes' rule. Bayes' rule requires the specification of a likelihood function and the prior density of the model parameters. The models used to form PPMs and the likelihood functions are exactly the same as in classical SPM analyses, namely the GLM. The only extra information that is required is the prior probability distribution of the parameters. Although it would be possible to specify those using independent data or some plausible physiological constraints, there is an alternative to this fully Bayesian approach. The alternative is *empirical Bayes* in which the prior distributions are estimated from the data. Empirical Bayes requires a *hierarchical observation model* where the parameters and hyperparameters at any particular level can be treated as priors on the level below. There are numerous examples of hierarchical observation models in neuroimaging. For example, the distinction between fixed- and mixed-effects analyses of multisubject studies relies upon a two-level hierarchical model. However, in neuroimaging, there is a natural hierarchical observation model that is common to all brain mapping experiments. This is the hierarchy induced by looking for the same effects at every voxel within the brain (or grey matter). The first level of the hierarchy corresponds to the experimental effects at any particular voxel and the second level comprises the effects over voxels. Put simply, the variation in a contrast, over voxels, can be used as the prior variance of that contrast at any particular voxel. Hierarchical linear models have the following form:

$$\begin{aligned} y &= X^{(1)}\beta^{(1)} + \varepsilon^{(1)} \\ \beta^{(1)} &= X^{(2)}\beta^{(2)} + \varepsilon^{(2)} \\ \beta^{(2)} &= \ldots \end{aligned} \qquad \textbf{3.2}$$

This is exactly the same as Eqn. **3.1**, but now the parameters of the first level are generated by a supraordinate linear model and so on to any hierarchical depth required. These hierarchical observation models are an important extension of the GLM and are usually estimated using expectation maximization (EM) (Dempster *et al*., 1977). In the present context, the response variables comprise the responses at all voxels and $\beta^{(1)}$ are the treatment effects we want to make an inference about. Because we have invoked a second level, the first-level parameters embody random effects and are generated by a second-level linear model. At the second level, $\beta^{(2)}$ is the average effect over voxels and $\varepsilon^{(2)}$ its voxel-to-voxel variation. By estimating the variance of $\varepsilon^{(2)}$ one is implicitly estimating an empirical prior on the first-level parameters at each voxel. This prior can then be used to estimate the posterior probability of $\beta^{(1)}$ being greater than some threshold at each voxel. An example of the ensuing PPM is provided in Figure 3.1 along with the classical SPM.

In summary, we have seen how the GLM can be used to test hypotheses about brain responses and how, in a hierarchical form, it enables empirical Bayesian or conditional inference. Next, we deal with dynamic systems and how they can be formulated as GLMs. These dynamic models take us closer to how brain responses are actually caused by experimental manipulations and represent the next step towards causal models of brain responses.

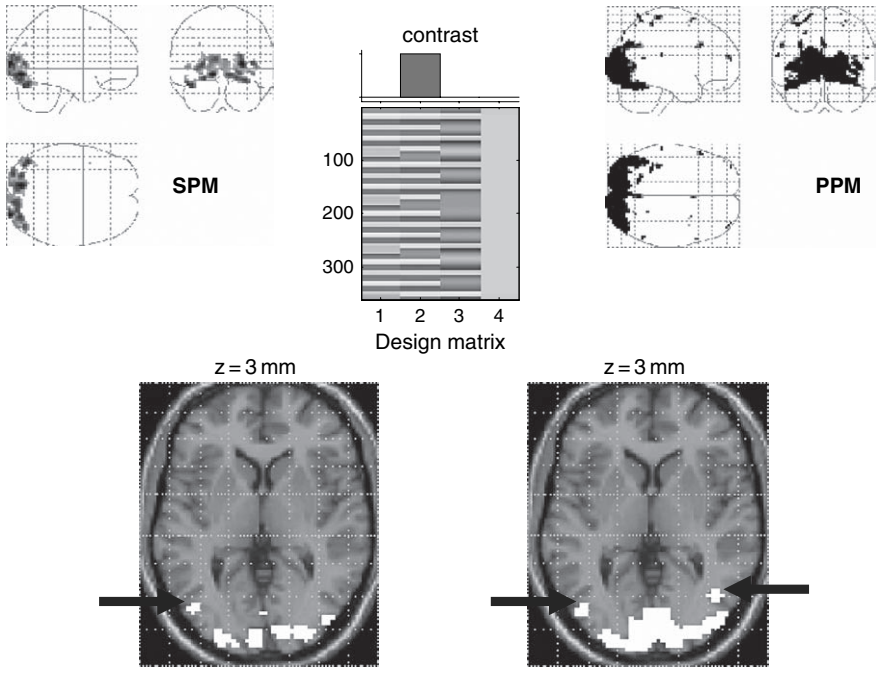

FIGURE 3.1 SPM and PPM for an fMRI study of attention to visual motion. The display format in the lower panel uses an axial slice through extrastriate regions but the thresholds are the same as employed in the maximum intensity projections (upper panels). Upper right: the activation threshold for the PPM was 0.7 a.u., meaning that all voxels shown had a 90 per cent chance of an activation of 0.7 per cent or more. Upper left: the corresponding SPM using an adjusted threshold at $p = 0.05$. Note the bilateral foci of motion-related responses in the PPM that are not seen in the SPM (grey arrows). As can be imputed from the design matrix (upper middle panel), the statistical model of evoked responses comprised boxcar regressors convolved with a canonical haemodynamic response function. The middle column corresponds to the presentation of moving dots and was the stimulus attribute tested by the contrast.

Dynamic models

Convolution models and temporal basis functions

In Friston *et al*. (1994) the form of the haemodynamic impulse response function (HRF) was estimated using a least squares de-convolution and a linear time invariant model, where evoked neuronal responses are *convolved* or smoothed with an HRF to give the measured haemodynamic response (see also Boynton *et al*., 1996). This simple linear convolution model is the cornerstone for making statistical inferences about activations in fMRI with the GLM. An impulse response function is the response to a single impulse, measured at a series of times after the input. It characterizes the input-output behaviour of the system (i.e. voxel) and places important constraints on the sorts of inputs that will excite a response.

Knowing the form of the HRF is important for several reasons, not least because it furnishes better statistical models of the data. The HRF may vary from voxel to voxel and this has to be accommodated in the GLM. To allow for different HRFs in different brain regions, temporal basis functions were introduced (Friston *et al*., 1995) to model evoked responses in fMRI and applied to event-related responses in Josephs *et al*. (1997) (see also Lange and Zeger, 1997). The basic idea behind temporal basis functions is that the haemodynamic response, induced by any given trial type, can be expressed as the linear combination of (basis) functions of peristimulus time. The convolution model for fMRI responses takes a stimulus function encoding the neuronal responses and convolves it with an HRF to give a regressor that enters the design matrix. When using basis functions, the stimulus function is convolved with each basis function to give a series of regressors. Mathematically, we can express this model as:

$$\begin{aligned} y(t) &= X\beta + \varepsilon & \qquad y(t) &= u(t) \otimes h(t) \\ & & \Leftrightarrow & \\ X_i &= T_i(t) \otimes u(t) & \qquad h(t) &= \beta_1 T_1(t) + \beta_2 T_2(t) + \ldots \end{aligned} \tag{3.3}$$

where $\otimes$ means convolution. This equivalence shows how any convolution model (right) can be converted into a GLM (left), using temporal basis functions. The parameter estimates β_i are the coefficients or weights that determine the mixture of basis functions of time $T_i(t)$ that models $h(t)$, the HRF for the trial type and voxel in question. We find the most useful basis set to be a canonical HRF and its derivatives with respect to the key parameters that determine its form (see below). Temporal basis

functions are important because they provide a graceful transition between conventional multilinear regression models with one stimulus function per condition and finite impulse response (FIR) models with a parameter for each time point following the onset of a condition or trial type. Plate 3 (see colour plate section) illustrates this graphically (see plate caption). In short, temporal basis functions offer useful constraints on the form of the estimated response that retain the flexibility of FIR models and the efficiency of single regressor models.

Biophysical models

Input-state-output systems

By adopting a convolution model for brain responses in fMRI, we are implicitly positing a dynamic system that converts neuronal responses into observed haemodynamic responses. Our understanding of the biophysical and physiological mechanisms that underpin the HRF has grown considerably in the past few years (e.g. Buxton and Frank 1997; Mandeville *et al*., 1999). Figure 3.2 shows some simulations based on the haemodynamic model described in Friston *et al*. (2000). Here, neuronal activity induces some autoregulated vasoactive signal that causes transient increases in regional cerebral blood flow (rCBF). The resulting flow increases dilate a venous balloon, increasing its volume and diluting venous blood to decrease deoxyhaemoglobin content. The blood oxygenation-level-dependent (BOLD) signal is roughly proportional to the concentration of deoxyhaemoglobin and follows the rCBF response with about a one second delay. The model is framed in terms of differential equations, examples of which are provided in the left panel.

Notice that we have introduced variables, like volume and deoxyhaemoglobin concentrations, that are not actually observed. These are referred to as the *hidden states* of input-state-output models. The state and output equations of any analytic dynamical system are:

$$\begin{aligned} \dot{x}(t) &= f(x, u, \theta) \\ y(t) &= g(x, u, \theta) + \varepsilon \end{aligned} \qquad \textbf{3.4}$$

The first line is an ordinary differential equation and expresses the rate of change of the states as a parameterized function of the states and inputs. Typically, the inputs u(t) correspond to designed experimental effects (e.g. the stimulus function in fMRI). There is a fundamental and causal relationship (Fliess *et al*., 1983) between the outputs and the history of the inputs in Eqn. **3.4**. This relationship conforms to a Volterra series, which expresses the output as a generalized convolution of the input, critically without reference to the hidden states x(t). This series is simply a functional Taylor expansion of the outputs with respect to the inputs (Bendat, 1990). The reason it is a functional expansion is that the inputs

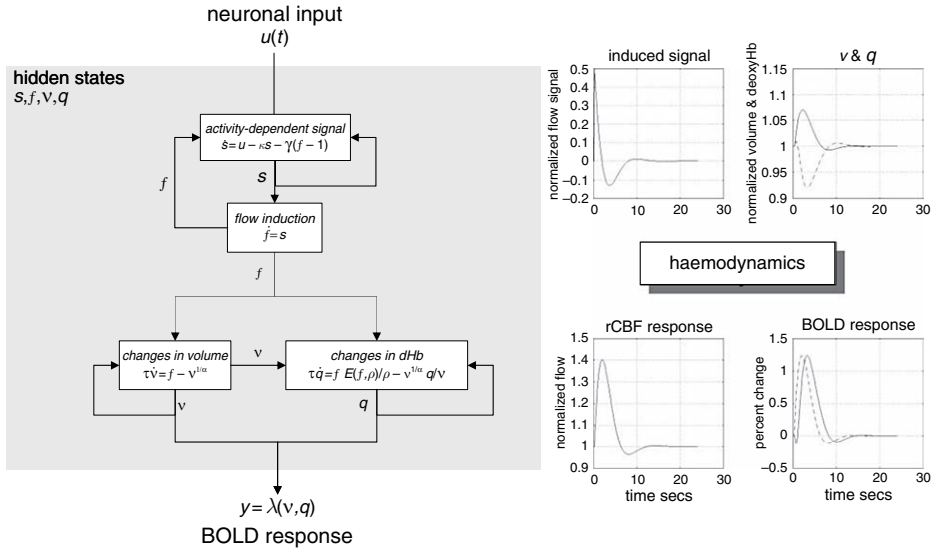

FIGURE 3.2 Right: haemodynamics elicited by an impulse of neuronal activity as predicted by a dynamical biophysical model (left). A burst of neuronal activity causes an increase in flow-inducing signal that decays with first order kinetics and is downregulated by local flow. This signal increases rCBF, which dilates the venous capillaries, increasing volume v. Concurrently, venous blood is expelled from the venous pool decreasing deoxyhaemoglobin content q. The resulting fall in deoxyhaemoglobin concentration leads to a transient increase in BOLD (blood oxygenation-level-dependent) signal and a subsequent undershoot. Left: haemodynamic model on which these simulations were based (see Friston *et al*., 2000 and Chapter 27 for details).

are a function of time. (For simplicity, here and in and Eqn. **3.7**, we deal with only one experimental input.)

$$y(t) = \sum_i \int_0^t \cdots \int_0^t \kappa_i(\sigma_1, \ldots, \sigma_i) \times u(t-\sigma_1), \ldots, u(t-\sigma_i) d\sigma_1, \ldots, d\sigma_i$$

$$\kappa_i(\sigma_1, \ldots, \sigma_i) = \frac{\partial^i y(t)}{\partial u(t-\sigma_1), \ldots, \partial u(t-\sigma_i)} \qquad \textbf{3.5}$$

where $\kappa_i(\sigma_1, \mathrm{K}, \sigma_i)$ is the i-th order kernel. In Eqn. **3.5** the integrals are restricted to the past. This renders the system causal. The key thing here is that Eqn. **3.5** is simply a convolution and can be expressed as a GLM, as in Eqn. **3.3**. This means that we can take a neurophysiologically realistic model of haemodynamic responses and use it as an observation model to estimate parameters using observed data. Here the model is parameterized in terms of kernels that have a direct analytic relation to the original parameters θ of the biophysical system. The first-order kernel is simply the conventional HRF. High-order kernels correspond to high-order HRFs and can be estimated using basis functions as described above. In fact, by choosing basis functions according to:

$$A(\sigma)_i = \frac{\partial \kappa(\sigma)_1}{\partial \theta_i} \qquad \textbf{3.6}$$

one can estimate the biophysical parameters because, to a first-order approximation, $\beta_i = \theta_i$. The critical step we have taken here is to start with a dynamic causal model of how responses are generated and construct a general linear observation model that allows us to estimate and infer things about the parameters of that model. This is in contrast to the conventional use of the GLM with design matrices that are not informed by a forward model of how data are caused. This approach to modelling brain responses has a much more direct connection with underlying physiology and rests upon an understanding of the underlying system.

Non-linear system identification

Once a suitable causal model has been established (e.g. Figure 3.2), we can estimate second-order kernels. These kernels represent a non-linear characterization of the HRF that can model interactions among stimuli in causing responses. One important manifestation of the non-linear effects, captured by the second-order kernels, is a modulation of stimulus-specific responses by preceding stimuli that are proximate in time. This means that responses at high stimulus presentation rates saturate and, in some instances, show an inverted U behaviour. This behaviour appears to be specific to BOLD effects (as distinct from evoked changes in cerebral blood flow) and may represent a *haemodynamic refractoriness*. This effect has important implications for event-related fMRI, where one may want to present trials in quick succession. (See Figure 2.8 in the previous chapter for an example of second-order kernels and the implications for haemodynamic responses.)

In summary, we started with models of regionally specific responses, framed in terms of the general linear model, in which responses were modelled as linear mixtures of designed changes in explanatory variables. Hierarchical extensions to linear observation models enable random-effects analyses and, in particular, empirical Bayes. The mechanistic utility of these models is realized though the use of forward models that embody causal dynamics. Simple variants of these are the linear convolution models used to construct explanatory variables in conventional analyses of fMRI data. These are a special case of generalized convolution models that are mathematically equivalent to input-state-output systems comprising hidden states. Estimation and inference with these dynamic models tells us something about *how* the response was caused, but only at the level of a single voxel. The next section retains the same perspective on models, but in the context of distributed responses and functional integration.

MODELS OF FUNCTIONAL INTEGRATION

Functional and effective connectivity

Imaging neuroscience has firmly established functional specialization as a principle of brain organization in humans. The integration of specialized areas has proven more difficult to assess. Functional integration is usually inferred on the basis of correlations among measurements of neuronal activity. Functional connectivity is defined as statistical dependencies or correlations *among remote neurophysiological events*. However, correlations can arise in a variety of ways: for example in multiunit electrode recordings, they can result from stimulus-locked transients evoked by a common input or reflect stimulus-induced oscillations mediated by synaptic connections (Gerstein and Perkel, 1969). Integration within a distributed system is usually better understood in terms of effective connectivity; effective connectivity refers explicitly to *the influence that one neural system exerts over another*, either at a synaptic (i.e. synaptic efficacy) or population level. It has been proposed that 'the [electrophysiological] notion of effective connectivity should be understood as the experiment- and time-dependent, simplest possible

circuit diagram that would replicate the observed timing relationships between the recorded neurons' (Aertsen and Preißl, 1991). This speaks of two important points: effective connectivity is dynamic, i.e. activity-dependent and it depends upon a model of the interactions. The estimation procedures employed in functional neuroimaging can be divided into linear non-dynamic models (e.g. McIntosh and Gonzalez-Lima, 1994) or non-linear dynamic models.

There is a necessary link between functional integration and multivariate analyses because the latter are necessary to model interactions among brain regions. Multivariate approaches can be divided into those that are inferential in nature and those that are data-led or exploratory. We will first consider multivariate approaches that look at functional connectivity or covariance patterns (and are generally exploratory) and then turn to models of effective connectivity (that allow for inference about their parameters).

Eigenimage analysis and related approaches

In Friston *et al*. (1993), we introduced voxel-based principal component analysis (PCA) of neuroimaging time-series to characterize distributed brain systems implicated in sensorimotor, perceptual or cognitive processes. These distributed systems are identified with principal components or *eigenimages* that correspond to spatial modes of coherent brain activity. This approach represents one of the simplest multivariate characterizations of functional neuroimaging time-series and falls into the class of exploratory analyses. Principal component or eigenimage analysis generally uses singular value decomposition (SVD) to identify a set of orthogonal spatial modes that capture the greatest amount of variance expressed over time. As such, the ensuing modes embody the most prominent aspects of the variance-covariance structure of a given time-series. Noting that covariance among brain regions is equivalent to functional connectivity renders eigenimage analysis particularly interesting because it was among the first ways of addressing functional integration (i.e. connectivity) with neuroimaging data. Subsequently, eigenimage analysis has been elaborated in a number of ways. Notable among these is canonical variate analysis (CVA) and multidimensional scaling (Friston *et al*., 1996a, b). Canonical variate analysis was introduced in the context of MANCOVA (multiple analysis of covariance) and uses the generalized eigenvector solution to maximize the variance that can be explained by some explanatory variables relative to error. CVA can be thought of as an extension of eigenimage analysis that refers explicitly to some explanatory variables and allows for statistical inference.

In fMRI, eigenimage analysis (e.g. Sychra *et al*., 1994) is generally used as an exploratory device to characterize coherent brain activity. These variance components may, or may not, be related to experimental design. For example, endogenous coherent dynamics have been observed in the motor system at very low frequencies (Biswal *et al*., 1995). Despite its exploratory power, eigenimage analysis is limited for two reasons. First, it offers only a linear decomposition of any set of neurophysiological measurements and second, the particular set of eigenimages or spatial modes obtained is determined by constraints that are biologically implausible. These aspects of PCA confer inherent limitations on the interpretability and usefulness of eigenimage analysis of biological time-series and have motivated the exploration of non-linear PCA and neural network approaches.

Two other important approaches deserve mention here. The first is independent component analysis (ICA). ICA uses entropy maximization to find, using iterative schemes, spatial modes or their dynamics that are approximately *independent*. This is a stronger requirement than *orthogonality* in PCA and involves removing high-order correlations among the modes (or dynamics). It was initially introduced as *spatial* ICA (McKeown *et al*., 1998) in which the independence constraint was applied to the modes (with no constraints on their temporal expression). More recent approaches use, by analogy with magneto- and electrophysiological time-series analysis, *temporal* ICA where the dynamics are enforced to be independent. This requires an initial dimension reduction (usually using conventional eigenimage analysis). Finally, there has been an interest in cluster analysis (Baumgartner *et al*., 1997). Conceptually, this can be related to eigenimage analysis through multidimensional scaling and principal coordinate analysis.

All these approaches are interesting, but they are not used very much. This is largely because they tell you nothing about how the brain works nor allow one to ask specific questions. Simply demonstrating statistical dependencies among regional brain responses or endogenous activity (i.e. demonstrating functional connectivity) does not address how these responses were caused. To address this one needs explicit models of integration or more precisely, effective connectivity.

Dynamic causal modelling with bilinear models

This section is about modelling interactions among neuronal populations, at a cortical level, using neuroimaging time-series and dynamic causal models that are informed by the biophysics of the system studied. The aim of dynamic causal modelling is to estimate, and make inferences about, the coupling among brain areas and how

that coupling is influenced by experimental changes (e.g. time or cognitive set). The basic idea is to construct a reasonably realistic neuronal model of interacting cortical regions or nodes. This model is then supplemented with a forward model of how neuronal or synaptic activity translates into a measured response (see previous section). This enables the parameters of the neuronal model (i.e. effective connectivity) to be estimated from observed data.

Intuitively, this approach regards an experiment as a designed perturbation of neuronal dynamics that are promulgated and distributed throughout a system of coupled anatomical nodes to change region-specific neuronal activity. These changes engender, through a measurement-specific forward model, responses that are used to identify the architecture and time constants of the system at a neuronal level. This represents a departure from conventional approaches (e.g. structural equation modelling and autoregression models; McIntosh and Gonzalez-Lima, 1994; Büchel and Friston, 1997), in which one assumes the observed responses are driven by endogenous or intrinsic noise (i.e. innovations). In contrast, dynamic causal models assume the responses are driven by designed changes in inputs. An important conceptual aspect of dynamic causal models pertains to how the experimental inputs enter the model and cause neuronal responses. Experimental variables can elicit responses in one of two ways. First, they can elicit responses through direct influences on specific anatomical nodes. This would be appropriate, for example, in modelling sensory evoked responses in early visual cortices. The second class of input exerts its effect vicariously, through a modulation of the coupling among nodes. These sorts of experimental variables would normally be more enduring; for example attention to a particular attribute or the maintenance of some perceptual set. These distinctions are seen most clearly in relation to particular forms of causal models used for estimation, e.g. the bilinear approximation:

$$\begin{aligned} \dot{x} &= f(x, u) \\ &= Ax + uBx + Cu \\ y &= g(x) + \varepsilon \\ A &= \frac{\partial f}{\partial x} \quad B = \frac{\partial^2 f}{\partial x \partial u} \quad C = \frac{\partial f}{\partial u} \end{aligned} \qquad \textbf{3.7}$$

where $\dot{x} = \partial x / \partial t$. This is an approximation to any model of how changes in neuronal activity in one region x_i are caused by activity in the other regions. Here the output function $g(x)$ embodies a haemodynamic convolution, linking neuronal activity to BOLD, for each region (e.g. that in Figure 3.2). The matrix A represents the coupling among the regions in the absence of input $u(t)$. This can be thought of as the latent coupling in the absence of experimental perturbations. The matrix B is effectively the change in latent coupling induced by the input. It encodes the input-sensitive changes in A or, equivalently, the modulation of coupling by experimental manipulations. Because B is a second-order derivative it is referred to as *bilinear*. Finally, the matrix C embodies the extrinsic influences of inputs on neuronal activity. The parameters $\theta = A, B, C$ are the connectivity or coupling matrices that we wish to identify and define the functional architecture and interactions among brain regions at a neuronal level.

Because Eqn. **3.7** has exactly the same form as Eqn. **3.4**, we can express it as a GLM and estimate the parameters using EM in the usual way (see Friston *et al*., 2003). Generally, estimation in the context of highly parameterized models like DCMs requires constraints in the form of priors. These priors enable conditional inference about the connectivity estimates. The sorts of questions that can be addressed with DCMs are now illustrated by looking at how attentional modulation is mediated in sensory processing hierarchies in the brain.

DCM and attentional modulation

It has been established that the superior posterior parietal cortex (SPC) exerts a modulatory role on V5 responses using Volterra-based regression models (Friston and Büchel, 2000) and that the inferior frontal gyrus (IFG) exerts a similar influence on SPC using structural equation modelling (Büchel and Friston, 1997). The example here shows that DCM leads to the same conclusions but starting from a completely different construct. The experimental paradigm and data acquisition are described in Figure 3.3. This figure also shows the location of the regions that entered the DCM. These regions were based on maxima from conventional SPMs testing for the effects of photic stimulation, motion and attention. Regional time courses were taken as the first eigenvariate of 8 mm spherical volumes of interest centred on the maxima shown in the figure. The inputs, in this example, comprise one sensory perturbation and two contextual inputs. The sensory input was simply the presence of photic stimulation and the first contextual one was presence of motion in the visual field. The second contextual input, encoding attentional set, was one during attention to speed changes and zero otherwise. The outputs corresponded to the four regional eigenvariates in (Figure 3.3, left panel). The intrinsic connections were constrained to conform to a hierarchical pattern in which each area was reciprocally connected to its supraordinate area. Photic stimulation entered at, and only at, V1. The effect of motion in the visual field was modelled as a bilinear modulation of the V1 to V5 connectivity and attention was allowed to modulate the backward connections from IFG and SPC.

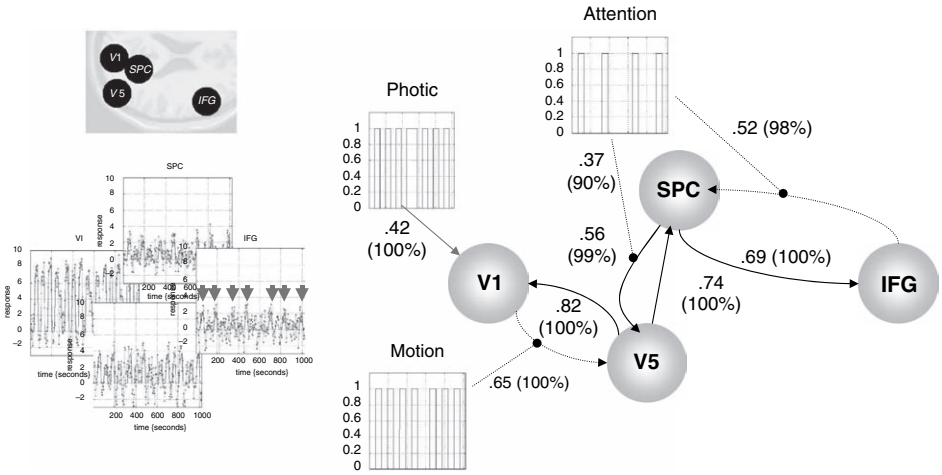

FIGURE 3.3 Results of a DCM analysis of attention to visual motion with fMRI. Right panel: functional architecture based upon the conditional estimates shown alongside their connections, with the per cent confidence that they exceeded threshold in brackets. The most interesting aspects of this architecture involve the role of motion and attention in exerting bilinear effects. Critically, the influence of motion is to enable connections from V1 to the motion-sensitive area V5. The influence of attention is to enable backward connections from the inferior frontal gyrus (IFG) to the superior parietal cortex (SPC). Furthermore, attention increases the influence of SPC on V5. Dotted arrows connecting regions represent significant bilinear effects in the absence of a significant intrinsic coupling. Left panel: fitted responses based upon the conditional estimates and the adjusted data are shown for each region in the DCM. The insert (upper left) shows the location of the regions.

Subjects were studied with fMRI under identical stimulus conditions (visual motion subtended by radially moving dots) while manipulating the attentional component of the task (detection of velocity changes). The data were acquired from a normal subject at 2 Tesla. Each subject had four consecutive 100-scan sessions comprising a series of 10-scan blocks under five different conditions, D F A F N F A F N S. The first condition (D) was a dummy condition to allow for magnetic saturation effects. F (Fixation) corresponds to a low-level baseline where the subjects viewed a fixation point at the centre of a screen. In condition A (Attention), subjects viewed 250 dots moving radially from the centre at 4.7 degrees per second and were asked to detect changes in radial velocity. In condition N (No attention), the subjects were asked simply to view the moving dots. In condition S (Stationary), subjects viewed stationary dots. The order of A and N was swapped for the last two sessions. In all conditions subjects fixated the centre of the screen. During scanning there were no speed changes. No overt response was required in any condition.

The results of the DCM are shown in Figure 3.3 (right panel). Of primary interest here is the modulatory effect of attention that is expressed in terms of the bilinear coupling parameters for this input. As expected, we can be highly confident that attention modulates the backward connections from IFG to SPC and from SPC to V5. Indeed, the influences of IFG on SPC are negligible in the absence of attention (dotted connection). It is important to note that the only way that attentional manipulation can affect brain responses is through this bilinear effect. Attention-related responses are seen throughout the system (attention epochs are marked with arrows in the plot of IFG responses in the left panel). This attentional modulation is accounted for, sufficiently, by changing just two connections. This change is, presumably, instantiated by instructional set at the beginning of each epoch.

The second thing this analysis illustrates is how functional segregation is modelled in DCM. Here one can regard V1 as 'segregating' motion from other visual information and distributing it to the motion-sensitive area, V5. This segregation is modelled as a bilinear 'enabling' of V1 to V5 connections when, and only when, motion is present. Note that, in the absence of motion, the latent V1 to V5 connection was trivially small (in fact the estimate was -0.04). The key advantage of entering motion through a bilinear effect, as opposed to a direct effect on V5, is that we can finesse the inference that V5 shows motion-selective responses with the assertion that these responses are mediated by afferents from V1. The two bilinear effects above represent two important aspects of functional integration that DCM is able to characterize.

Structural equation modelling as a special case of DCM

The central idea behind dynamic causal modelling is to treat the brain as a deterministic non-linear dynamic system that is subject to inputs and produces outputs. Effective connectivity is parameterized in terms of coupling among unobserved brain states (e.g. neuronal activity in different regions). The objective is to estimate these parameters by perturbing the system and measuring the response. This is in contradistinction to established methods for estimating effective connectivity from neurophysiological time-series, which include structural equation modelling and models based on multivariate

autoregressive processes. In these models, there is no designed perturbation and the inputs are treated as unknown and stochastic. Furthermore, the inputs are often assumed to express themselves instantaneously such that, at the point of observation the change in states is zero. From Eqn. **3.7**, in the absence of bilinear effects we have:

$$\begin{aligned}\dot{x} &= 0 \\ &= Ax + Cu \\ x &= -A^{-1}Cu\end{aligned} \qquad \textbf{3.8}$$

This is the regression equation used in structural equation modelling (SEM) where $A = D - I$ and D contains the off-diagonal connections among regions. The key point here is that A is estimated by assuming $u(t)$ is some random innovation with known covariance. This is not really tenable for designed experiments when $u(t)$ represent carefully structured experimental inputs. Although SEM and related autoregressive techniques are useful for establishing dependence among regional responses, they are not surrogates for informed causal models based on the underlying dynamics of these responses.

In this section, we have covered multivariate techniques ranging from eigenimage analysis that does not have an explicit forward or causal model to DCM that does. The bilinear approximation to any DCM has been illustrated through its use with fMRI to study attentional modulation. The parameters of the bilinear approximation include first-order effective connectivity A and its experimentally-induced changes B. Although the bilinear approximation is useful, it is possible to model coupling among neuronal subpopulations explicitly. We conclude with a DCM that embraces a number of neurobiological facts and takes us much closer to a mechanistic understanding of how brain responses are generated. This example uses responses measured with EEG.

Dynamic causal modelling with neural-mass models

Event-related potentials (ERPs) have been used for decades as electrophysiological correlates of perceptual and cognitive operations. However, the exact neurobiological mechanisms underlying their generation are largely unknown. In this section, we use neuronally plausible models to understand event-related responses. Our example shows that changes in connectivity are sufficient to explain certain ERP components. Specifically, we will look at the MMN, a component associated with rare or unexpected events. If the unexpected nature of rare stimuli depends on learning which stimuli are frequent, then the MMN must be due to plastic changes in connectivity that mediate perceptual learning. We conclude by showing that advances in the modelling of evoked responses now afford measures of connectivity among cortical sources that can be used to quantify the effects of perceptual learning.

Neural-mass models

The minimal model we have developed (David *et al.*, 2006) uses the connectivity rules described in Felleman and Van Essen (1992) to assemble a network of coupled sources. These rules are based on a partitioning of the cortical sheet into supra-, infra-granular layers and granular layer (layer 4). Bottom-up or forward connections originate in agranular layers and terminate in layer 4. Top-down or backward connections target agranular layers. Lateral connections originate in agranular layers and target all layers. These long-range or extrinsic cortico-cortical connections are excitatory and arise from pyramidal cells.

Each region or source is modelled using a neural mass model described in David and Friston (2003), based on the model of Jansen and Rit (1995). This model emulates the activity of a cortical area using three neuronal subpopulations, assigned to granular and agranular layers. A population of excitatory pyramidal (output) cells receives inputs from inhibitory and excitatory populations of interneurons, via intrinsic connections (intrinsic connections are confined to the cortical sheet). Within this model, excitatory interneurons can be regarded as spiny stellate cells found predominantly in layer 4 and in receipt of forward connections. Excitatory pyramidal cells and inhibitory interneurons are considered to occupy agranular layers and receive backward and lateral inputs (Figure 3.4).

To model event-related responses, the network receives inputs via input connections. These connections are exactly the same as forward connections and deliver inputs to the spiny stellate cells in layer 4. In the present context, inputs $u(t)$ model sub-cortical auditory inputs. The vector C controls the influence of the input on each source. The lower, upper and leading diagonal matrices A^F, A^B, A^L encode forward, backward and lateral connections respectively. The DCM here is specified in terms of the state equations shown in Figure 3.4 and a linear output equation:

$$\begin{aligned}\dot{x} &= f(x, u) \\ y &= Lx_0 + \varepsilon\end{aligned} \qquad \textbf{3.9}$$

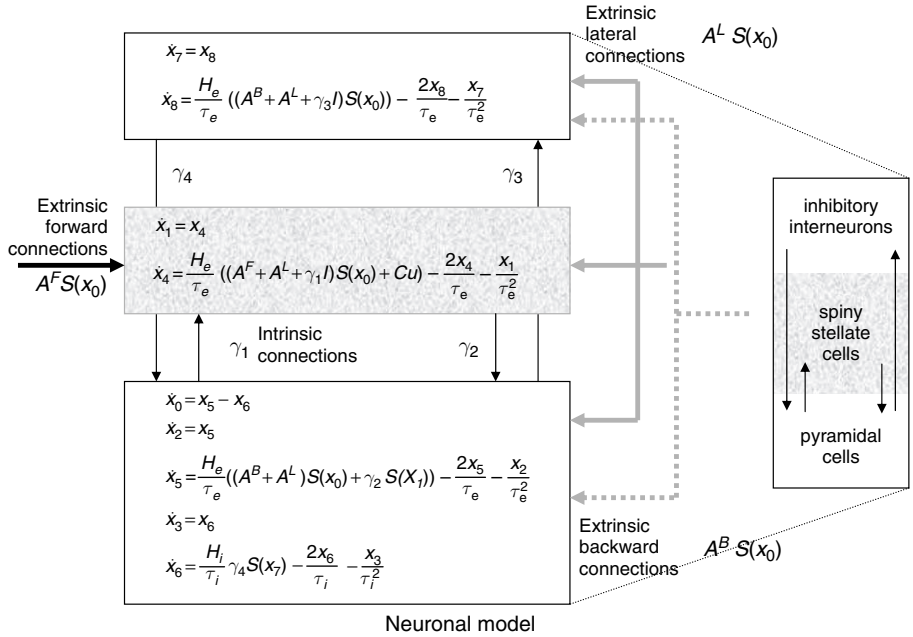

FIGURE 3.4 Schematic of the DCM used to model electrical responses. This schematic shows the state equations describing the dynamics of sources or regions. Each source is modelled with three subpopulations (pyramidal, spiny stellate and inhibitory interneurons) as described in Jansen and Rit (1995) and David and Friston (2003). These have been assigned to granular and agranular cortical layers which receive forward and backward connections respectively.

where x^0 represents the transmembrane potential of pyramidal cells and L is a lead field matrix coupling electrical sources to the EEG channels. This should be compared to the DCM above for haemodynamics; here the equations governing the evolution of neuronal states are much more complicated and realistic, as opposed to the bilinear approximation in Eqn. **3.7**. Conversely, the output equation is a simple linearity, as opposed to the nonlinear observer used for fMRI. As an example, the state equation for the inhibitory subpopulation is:

$$\begin{aligned}\dot{x}_7 &= x_8 \\ \dot{x}_8 &= \frac{H_e}{\tau_e}((A^B + A^L + \gamma_3 I)S(x_0)) - \frac{2x_8}{\tau_e} - \frac{x_7}{\tau_e^2}\end{aligned} \qquad \textbf{3.10}$$

Propagation delays on the extrinsic connections have been omitted for clarity here and in Figure 3.4.

Within each subpopulation, the evolution of neuronal states rests on two operators. The first transforms the average density of presynaptic inputs into the average postsynaptic membrane potential. This is modelled by a linear transformation with excitatory and inhibitory kernels parameterized by $H_{e,i}$ and $\tau_{e,i}$. $H_{e,i}$ control the maximum postsynaptic potential and $\tau_{e,i}$ represent a lumped rate-constant. The second operator S transforms the average potential of each subpopulation into an average firing rate. This is assumed to be instantaneous and is a sigmoid function. Interactions among the subpopulations depend on constants $\gamma_{1,2,3,4}$, which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation. In Eqn. **3.10**, the top line expresses the rate of change of voltage as a function of current. The second line specifies how current changes as a function of voltage, current and presynaptic input from extrinsic and intrinsic sources. Having specified the DCM in terms of these equations, one can estimate the coupling parameters from empirical data using EM as described above.

Perceptual learning and the MMN

The example shown in Figure 3.5 is an attempt to model the MMN in terms of changes in backward and lateral connections among cortical sources. In this example, two [averaged] channels of EEG data were modelled with three cortical sources. Using this generative or forward model, we estimated differences in the strength of these connections for rare and frequent stimuli. As expected, we could account for detailed differences in the ERPs (the MMN) by changes in connectivity (see Figure 3.5 for details). Interestingly, these differences were expressed selectively in the lateral connections. If this model is a sufficient approximation to the real sources, these changes are a non-invasive measure of plasticity, mediating perceptual learning, in the human brain.

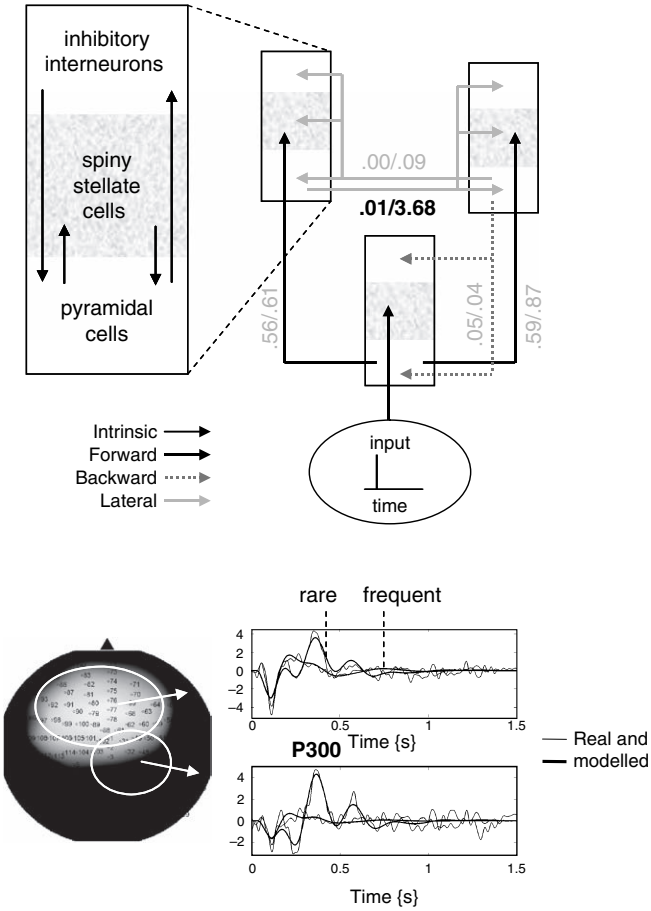

FIGURE 3.5 Summary of a DCM analysis of event-related potentials (ERPs) elicited during an auditory oddball paradigm, employing rare and frequent pure tones. Upper panel: schematic showing the architecture of the neuronal model used to explain the empirical data. Sources were coupled with extrinsic cortico-cortical connections following the rules of Felleman and van Essen (1992). The free parameters of this model included intrinsic and extrinsic connection strengths that were adjusted best to explain the data. In this example, the lead field was also estimated, with no spatial constraints. The parameters were estimated for ERPs recorded during the presentation of rare and frequent tones and are reported beside their corresponding connection (frequent/rare). The most notable finding was that the mismatch response could be explained by a selective increase in lateral connection strength from 0.1 to 3.68 Hz (highlighted in bold). Lower panel: the channel positions (left) and ERPs (right) averaged over two subsets of channels (circled on the left). Note the correspondence between the measured ERPs and those generated by the model (see David *et al*., 2006 for details).

Auditory stimuli, 1000 or 2000 Hz tones with 5 ms rise and fall times and 80 ms duration, were presented binaurally. The tones were presented for 15 minutes, every 2 s in a pseudo-random sequence with 2000 Hz tones occurring 20 per cent of the time and 1000 Hz tones occurring 80 per cent of the time. The subject was instructed to keep a mental record of the number of 2000 Hz tones (non-frequent target tones). Data were acquired using 128 EEG electrodes with 1000 Hz sample frequency. Before averaging, data were referenced to mean earlobe activity and band-pass filtered between 1 and 30 Hz. Trials showing ocular artefacts and bad channels were removed from further analysis.

CONCLUSION

In this chapter, we have reviewed some key models that underpin image analysis and have touched briefly on ways of assessing specialization and integration in the brain. These models can be regarded as a succession of modelling endeavours that draw more and more on our understanding of how brain-imaging signals are generated, both in terms of biophysics and the underlying neuronal interactions. We have seen how hierarchical linear observation models encode the treatment effects elicited by experimental design. General linear models based on convolution models imply an underlying dynamic input-state-output system. The form of these systems can be used to constrain convolution models and explore some of their simpler non-linear properties. By creating observation models based on explicit forward models of neuronal interactions, one can model and assess interactions among distributed cortical areas and make inferences about coupling at the neuronal level. The next years will probably see an increasing realism in the dynamic causal models introduced above. These endeavours are likely to encompass fMRI signals enabling the conjoint modelling, or fusion, of different modalities and the marriage of computational neuroscience with the modelling of brain responses.

REFERENCES

Absher JR, Benson DF (1993) Disconnection syndromes: an overview of Geschwind's contributions. *Neurology* **43**: 862–67

Aertsen A, Preißl H (1991) Dynamics of activity and connectivity in physiological neuronal networks. In *Non linear dynamics and neuronal networks*, Schuster HG (ed.). VCH Publishers Inc., New York, pp 281–302

Baumgartner R, Scarth G, Teichtmeister C *et al*. (1997) Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part 1: reproducibility. *J Mag Res Imaging* **7**: 1094–101

Bendat JS (1990) *Nonlinear system analysis and identification from random data*. John Wiley and Sons, New York

Berry DA, Hochberg Y (1999) Bayesian perspectives on multiple comparisons. *J Stat Plann Inference* **82**: 215–227

Biswal B, Yetkin FZ, Haughton VM *et al*. (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Mag Res Med* **34**: 537–41

Boynton GM, Engel SA, Glover GH *et al*. (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* **16**: 4207–21

Büchel C, Friston KJ (1997) Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. *Cereb Cortex* **7**: 768–78

Buckner RL, Koutstaal W, Schacter DL, *et al*. (1998) Functional-anatomic study of episodic retrieval. II. Selective averaging of event-related fMRI trials to test the retrieval success hypothesis. *NeuroImage* **7**: 163–75

Buxton RB, Frank LR (1997) A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *J Cereb Blood Flow Metab* **17**: 64–72

David O, Friston KJ (2003) A neural mass model for MEG/EEG: coupling and neuronal dynamics. *NeuroImage* **20**: 1743–55

David O, Kiebel SJ, Harrison LM *et al*. (2006) Dynamic causal modelling of evoked responses in EEG and MEG. Neuroimage. Feb 8; (Epub ahead of print)

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc Series* B **39**: 1–38

Felleman DJ, Van Essen DC (1992) Distributed hierarchical processing in the primate cerebral cortex. *Cerebr Cortex* **1**: 1–47

Fliess M, Lamnabhi M, Lamnabhi-Lagarrigue F (1983) An algebraic approach to nonlinear functional expansions. *IEEE Trans Circuits Syst* **30**: 554–70

Friston KJ, Büchel C (2000) Attentional modulation of effective connectivity from V2 to V5/MT in humans. *Proc Natl Acad Sci USA* **97**: 7591–96

Friston KJ, Frith CD, Liddle PF *et al*. (1991) Comparing functional (PET) images: the assessment of significant change. *J Cereb Blood Flow Metab* **11**: 690–99

Friston KJ, Frith CD, Liddle PF *et al*. (1993). Functional connectivity: the principal component analysis of large data sets. *J Cereb Blood Flow Metab* **13**: 5–14

Friston KJ, Jezzard P, Turner M (1994) Analysis of functional MRI time-series. *Human Brain Mapping* **1**(2): 153–71

Friston KJ, Frith CD, Turner R *et al*. (1995) Characterising evoked hemodynamics with fMRI. *NeuroImage* **2**: 157–65

Friston KJ, Poline J-B, Holmes AP *et al*. (1996a) A multivariate analysis of PET activation studies. *Hum Brain Mapp* **4**: 140–51

Friston KJ, Frith CD, Fletcher P *et al*. (1996b) Functional topography: multidimensional scaling and functional connectivity in the brain. *Cereb Cortex* **6**: 156–64

Friston KJ, Mechelli A, Turner R *et al*. (2000) Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. *NeuroImage* **12**: 466–77

Friston KJ, Harrison L, Penny W. (2003) Dynamic causal modelling. *NeuroImage* **19**:1273–302

Gerstein GL, Perkel DH, Taylor JG (2001) Neural modelling and functional brain imaging: an overview. *Neural Netw* **13**: 829–46

Gerstein GL, Perkel DH (1969) Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* **16**: 828–30

Jansen BH, Rit VG (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol Cybern* **73**: 357–66

Josephs O, Turner R, Friston K (1997) Event-related fMRI. *Human Brain Mapping* **5**(4): 243–8

Lange N, Zeger SL (1997) Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion) J *Roy Stat Soc Ser C* **46**: 1–29

Mandeville JB, Marota JJ, Ayata C *et al*. (1999) Evidence of a cerebrovascular postarteriole Windkessel with delayed compliance. *J Cereb Blood Flow Metab* **19**: 79–89

McIntosh AR, Gonzalez-Lima F (1994) Structural equation modelling and its application to network analysis in functional brain imaging. *Hum Brain Mapp* **2**: 2–22

McKeown M, Jung T-P, Makeig S *et al*. (1998) Spatially independent activity patterns in functional MRI data during the Stroop colour naming task. *Proc Natl Acad Sci* **95**: 803–10

Phillips CG, Zeki S, Barlow HB (1984) Localisation of function in the cerebral cortex. Past present and future. *Brain* **107**: 327–61

Staum M (1995) Physiognomy and phrenology at the Paris Athénée. *J Hist Ideas* **56**: 443–62

Sychra JJ, Bandettini PA, Bhattacharya N *et al*. (1994) Synthetic images by subspace transforms. I Principal component images and related filters. *Med Physics* **21**: 193–201

Zeki S (1990) The motion pathways of the visual cortex. In *Vision: coding and efficiency*, Blakemore C (ed.). Cambridge University Press, Cambridge, pp 321–45