INTERACTIONS AMONG NEURONAL SYSTEMS ASSESSED WITH FUNCTIONAL NEUROIMAGING

CHRISTIAN BÜCHEL AND KARL FRISTON

In the late nineteenth century, the early investigations of brain function were dominated by the concept of functional segregation. This approach was driven largely by the data available to scientists of that era. Patients with circumscribed lesions were found who were impaired in one particular ability while their other abilities remained largely intact. Indeed, descriptions of patients with different kinds of aphasia (an impairment of the ability to use or comprehend words), made at this time, have left a permanent legacy in the contrast between Broca's and Wernicke's aphasia. These syndromes were thought to result from damage to anterior or posterior regions of the left hemisphere, respectively. In the first part of the twentieth century, the idea of functional segregation fell into disrepute and the doctrine of "mass action" held sway, according to which higher abilities depended on the function of the brain "as a whole" (1). This doctrine was always going to be unsatisfactory. However, with the resources available at the time, it was simply not possible to make any progress studying the function of the "brain as a whole." By the end of the twentieth century, the concept of functional segregation had returned to domination.

The doctrine is now particularly associated with cognitive neuropsychology and is enshrined in the concept of double dissociation (2). A double dissociation is demonstrated when neurologic patients can be found with "mirror" abnormalities. For example, many patients have been described who have severe impairments of long-term memory but whose short-term memory is intact. In 1969, Warrington and Shallice (3) described the first of a series of patients who had severe impairments of phonologic shortterm memory but no impairments of long-term memory. This is a particularly striking example of double dissociation. It demonstrates that different brain regions are involved in short- and long-term memory. Furthermore, it shows that these regions can function in a largely independent fashion. This observation caused major problems for theories of memory, extant at the time, according to which inputs to long-term memory emanated from short-term memory systems (4).

Functional brain imaging avoids many of the problems of lesion studies, but, here too, the field has been dominated by the doctrine of functional segregation. Nevertheless, it is implicit in the subtraction method that brain regions communicate with each other. If we want to distinguish between brain regions associated with certain central processes, for example, then we design an experiment in which the sensory input and motor output are the same across all conditions. In this way, activity associated with sensory input and motor output will cancel out. The early studies of reading by Petersen et al. (5) and Posner et al. (6) are still among the best examples of this approach. The design of these studies was based on the assumption that reading goes through a single series of discrete and independent stages; visual shapes are analyzed to form letters, letters are put together to form words, the visual word form is translated into sound, the sound form is translated into articulation, and so on. By a comparison of suitable tasks (e.g., letters vs. false font, words vs. letters), each stage can be isolated and the associated brain region identified. Although subsequent studies have shown that this characterisation of the brain activity associated with reading is a considerable oversimplification, the original report still captures the essence of most functional imaging studies; a number of discrete cognitive stages are mapped onto discrete brain areas. Nothing is revealed about how the cognitive processes interact or how the brain regions communicate with each other. If word recognition really did depend on the passage of information through a single series of discrete stages, we

Christian Büchel: Cognitive Neuroscience Laboratory, Department of Neurology, Hamburg University, Hamburg, Germany.

Karl Friston: Wellcome Department of Cognitive Neurology, London, United Kingdom.

would at least like to know the temporal order in which the associated brain regions are engaged. Some evidence comes from encephaloelectrographic and myoelectrographic studies. In fact, we know that word recognition depends on at least two parallel routes—one of meaning and one of phonology (7). Given this model, we would like to be able to specify the brain regions associated with each route and have some measure of the strengths of the connections between these different regions.

In this chapter, we show how new methods for measuring effective connectivity allow us to characterize the interactions between brain regions that underlie the complex interactions among different processing stages of functional architectures.

DEFINITIONS

In the analysis of neuroimaging time series (i.e., signal changes in a set of voxels, expressed as a function of time), functional connectivity is defined as the temporal correlations between spatially remote neurophysiological events (8). This definition provides a simple characterization of functional interactions. The alternative is effective connectivity, the influence one neuronal system exerts over another (9). These concepts originated in the analysis of separable spike trains obtained from multiunit electrode recordings (10,11). Functional connectivity is simply a statement about the observed correlations; it does not comment on how these correlations are mediated. For example, at the level of multiunit microelectrode recordings, correlations can result from stimulus-locked transients, evoked by a common afferent input*, or reflect stimulus-induced oscillations, phasic coupling of neural assemblies mediated by synaptic connections (12). Effective connectivity is closer to the notion of a connection, either at a synaptic (cf synaptic efficacy) or cortical level. Although functional and effective connectivity can be invoked at a conceptual level in both neuroimaging and electrophysiology, they differ fundamentally at a practical level. This is because the time scales and nature of neurophysiologic measurements are very different (seconds vs. milliseconds and hemodynamic vs. spike trains). In electrophysiology, it is often necessary to remove the confounding effects of stimulus-locked transients (that introduce correlations not causally mediated by direct neural interactions) to reveal an underlying connectivity. The confounding effect of stimulus-evoked transients is less problematic in neuroimaging because propagation of signals from primary sensory areas onward is mediated by neuronal connections (usually reciprocal and interconnecting). However, it should be remembered that functional connectivity is not necessarily a consequence of effective connectivity (e.g., common neuromodulatory input from ascending aminergic neurotransmitter systems or thalamocortical afferents), and when it is, effective influences may be indirect (e.g., polysynaptic relays through multiple areas). In this chapter, we focus only on effective connectivity. More details about functional connectivity can be found in Friston et al. (8).

EFFECTIVE CONNECTIVITY

A Simple Model

Effective connectivity depends on two models: a mathematical model, describing "how" areas are connected, and a neuroanatomic model, describing "which" areas are connected. We shall consider linear and nonlinear models. Perhaps the simplest model of effective connectivity expresses the hemodynamic change at one voxel as a weighted sum of changes elsewhere. This can be regarded as a multiple linear regression, in which the effective connectivity reflects the amount of rCBF (regional cerebral blood flow) variability, at the target region, attributable to rCBF changes at a source region. As an example, consider the influence of other areas M on area VI. This can be framed in a simple equation:

$$V1 = Mc + e$$
[1]

where VI is an $n \times 1$ column vector with *n* scans, *M* is an $n \times m$ matrix with *m* regions and *n* observations (scans), *c* is an $m \times 1$ column vector with a parameter estimate for each region, and *e* is a vector of error terms.

Implicit in this interpretation is a mediation of the influence among brain regions by neuronal connections with an effective strength equal to the (regression) coefficient *c*. This highlights the fact that the linear model assumes that the connectivity is constant over the whole range of activation and does not depend on input from other sources.

Experience suggests that the linear model can give fairly robust results. One explanation is that the dimensionality (the number of things that are going on) of the physiologic changes can be small by experimental design. In other words, the brain responds to simple and well-organized experiments in a simple and well-organized way. Generally, however, neurophysiologic interactions are nonlinear, and the adequacy of linear models must be questioned (or at least qualified). Consequently, we focus on a nonlinear model of effective connectivity (13).

Structural Equation Modeling

The simple model above is sufficient to analyze effective connectivity to one region at a time (e.g., V1 or V2). We will now introduce structural equation modeling as a tool allowing for more complicated models comprising many

^{*} That is, signal input into the neural system as a result of external stimulation.

regions of interest and demonstrate how nonlinear interactions are dealt with in this context. The basic idea behind structural equation modeling differs from the usual statistical approach of modeling individual observations. In multiple regression or ANCOVA (analysis of covariance) models, the regression coefficients derive from the minimization of the sum of squared differences of the predicted and observed dependent variables (i.e., activity in the target region). Structural equation modeling approaches the data from a different perspective; instead of variables being considered individually, the emphasis lies on the variance-covariance structure.* Thus, models are solved in structural equation modeling by minimizing the difference between the observed variance-covariance structure and the one implied by a structural or path model. In the past few years, structural equation modeling has been applied to functional brain imaging. For example, McIntosh et al. (14) demonstrated the dissociation between ventral and dorsal visual pathways for object and spatial vision by using structural equation modeling of positron emission tomographic (PET) data in the human. In this section, we focus on the theoretic background of structural equation modeling and demonstrate this technique with the use of functional magnetic resonance imaging (fMRI).

In terms of neuronal systems, a measure of covariance represents the degree to which the activities of two or more regions are related (i.e., functional connectivity). The study of variance–covariance structures here is much simpler than in many other fields; the interconnection of the dependent variables (regional activity of brain areas) is anatomically determined, and the activation of each region can be directly measured with functional brain imaging. This represents a major difference from "classic" structural equation modeling in the behavioral sciences, in which models are often hypothetical and include latent variables denoting rather abstract concepts, such as intelligence.

As mentioned above, structural equation modeling minimizes the difference between the observed or measured covariance matrix and the one that is implied by the structure of the model. The free parameters (path coefficients or connection strengths; c above) are adjusted to minimize the difference between the measured and modeled covariance matrix.† (See ref. 15 for details.)

An important issue in structural equation modeling is the determination of the participating regions and the underlying anatomic model. Several approaches to this issue can be adopted. These include categoric comparisons between different conditions, statistical images highlighting structures of functional connectivity, and nonhuman electrophysiologic and anatomic studies (16).

With respect to anatomic connectivity in humans, the advent of new MR techniques promises a better characterization of neuronal connectivity in humans. Diffusion tensor imaging measures the anisotropy of diffusion in the brain. The main anisotropy exists in the white matter because the orientation of neuronal fibres (axons) allows molecules to diffuse more easily along the fiber than in other directions. Therefore, the main direction of the diffusion tensor reflects the underlying orientation of white matter tracts. Through tracing algorithms, it is now possible to infer the connectivity of individual regions (e.g., activations derived from an fMRI study) in an individual brain (17) (Fig. 29.1).

A model is always a simplification of reality; exhaustively correct models either do not exist or are too complicated to understand. In the context of effective connectivity, one has to find a compromise between complexity, anatomic accuracy, and interpretability. Mathematical constraints on the model also exist; if the number of free parameters exceeds the number of observed covariances, the system is underdetermined and no single solution exists.

Each estimated model can be analyzed to give an overall



FIGURE 29.1. Axial diffusion tensor image, obtained by using a TurboSTEAM diffusion sensitized pulse sequence on a Siemens Vision 1.5T MR scanner. Voxel size 3 imes 3 imes 3 mm. Average of 20 replications. Needles in each voxel show the largest eigenvector of the tensor (i.e., the main orientation of diffusion within this voxel). In white matter, the major axis of diffusion is constrained by the orientation of white matter tracts and therefore provides a good estimate of the direction of fiber bundles (17). As expected, the corpus callosum in the center of the image shows predominantly horizontal fibers connecting both hemispheres. In the occipital cortex, parts of the optic radiation with a predominantly anterior-posterior fiber orientation can be seen. The precision of the method is highlighted by the demonstration of corticocortical U fibers, magnified in the small image. (From Nolte U, Finsterbusch J, Frahm J. Rapid whole brain diffusion mapping without susceptibility artifacts using diffusion-weighted singleshot STEAM MRI. Proceedings of the eighth annual meeting of the International Society of Magnetic Resonance in Medicine, Denver, 2000:807.)

^{*} The variance-covariance structure describes in detail the dependencies between the different variables (in this case, the measured regional responses to stimulation).

[†]The free parameters are estimated by minimizing a function of the observed and implied covariance matrix. To date, the most widely used objective function in structural equation modeling is the maximum likelihood (ML) function.

goodness-of-fit measure for use when different models are compared with each other. A "nested model" approach can be used to compare different models (e.g., data from different groups or conditions) in the context of structural equation modeling. A so-called null model is constructed in which the estimates of the free parameters are constrained to be the same for both groups. The alternative model allows free parameters to differ between groups. The significance of the differences between the models is expressed by the difference of the goodness-of-fit statistic. Consider the following hypothetical example. Subjects are scanned under two different conditions (e.g., attention and no attention). The hypothesis might be that within a system of regions *A*, B, C, and D, the connectivity between A and B is different under the two attentional conditions. To determine whether the difference in connectivity is statistically significant, we estimate the goodness-of-fit measure for two models. Model 1 allows the connectivity between A and B to take different values for both conditions. Model 2 constrains the path coefficient between A and B to be equal for attention and no attention. If the change of connectivity between *attention* and *no attention* for the connection of A and B is negligible, the constrained model (model 2) should fit the data as well as the free model (model 1). We can now infer whether the difference of the two goodness-offit measures is significant. Nonlinear models can also be accommodated in the framework of structural equation modeling by introducing additional variables containing a nonlinear function (e.g., $f(x) = x^2$) of the original variables (18). Interactions of variables can be incorporated in a similar fashion, wherein a new variable, containing the product of the two interacting variables, is introduced as an additional influence. We will now demonstrate these ideas with an example. More details of structural equation modeling, including the operational equations, can be found in ref. 15.

Example: Learning

In the first example, we were interested in changes in effective connectivity over time as expected during paired-associates learning (19). In the case of object-location memory, several functional studies have demonstrated activation of ventral occipital and temporal regions during the retrieval of object identity and, conversely, increased responses in dorsal parietal areas during the retrieval of spatial location (20). These results suggest domain-specific representations in posterior neocortical structures that are closely related to those involved in perception, a finding that accords with the segregation of ventral and dorsal pathways in processing categoric or spatial stimulus features, respectively. Another phenomenon observed in some learning studies is a decrease of neural responses (i.e., adaptation) to repeated stimulus presentations. This repetition suppression has been replicated consistently in primate electrophysiologic and human functional imaging studies (21). For object-location learning, it is intuitively likely that two specialized systems need to interact to establish an association. Domain-specific representations or repetition suppression is not sufficient to account for this associative component. In other words, functional segregation and localized response properties cannot account for associative learning alone.

In our fMRI experiment, decreases in activation during learning, indicative of repetition suppression, were observed in several cortical regions in the ventral and dorsal visual pathway. Within the framework of repetition suppression, it has been hypothesized that decreases in neural responses are a secondary result of enhanced response selectivity (22). By analogy to the development and plasticity of cortical architectures, this refined selectivity is likely to be a consequence of changes in effective connectivity within the system at a synaptic level. We explicitly addressed this notion by characterising time-dependent changes in effective connectivity during learning.

The experiment was performed on a 2-tesla (T) MRI system equipped with a head volume coil. fMRI images were obtained every 4.1 seconds with echo-planar imaging (48 slices in each volume). Six subjects had to learn and recall the association between 10 simple line drawings of real-world objects and 10 locations on a screen during fMRI. Each learning trial consisted of four conditions: *encoding, control, retrieval,* and *control* (Fig. 29.2A). The behavioral data acquired during *retrieval* demonstrated that all six subjects were able to learn the association between object identity and spatial location, for all 10 objects, within eight learning blocks, as indicated by the ensuing asymptotic learning curves (Fig. 29.2B).

The structural model used in the analysis embodies connections within and across ventral and dorsal visual pathways and was based on anatomic studies in primates (Fig. 29.2C). Primary visual cortex was modeled as the origin of both pathways. In addition to "interstream" connections between dorsal extrastriate cortex and the fusiform region and between the posterior parietal cortex and the posterior inferotemporal cortex, we included direct connections based on a hierarchic cortical organization. Given our hypothesis relating to changes in effective connectivity between dorsal and ventral pathways, the path analysis focused on the connection between posterior parietal cortex (PP, dorsal stream) and posterior inferotemporal cortex (ITp, ventral stream). We divided each learning session into EARLY (first part) and LATE (second part) observations and estimated separate path coefficients for each partition.

The path coefficient between PP and ITp increased significantly during learning in the group (p < .05) and was confirmed by an analysis of individual subjects showing an increase in effective connectivity between PP and ITp of 0.27. In contrast to the connections between streams, connections within the dorsal pathway decreased over time.

The estimated change in connectivity from PP to ITp



FIGURE 29.2. Changes in effective connectivity over time in paired-associates learning. **A**: Design of the study. Blocks of *encoding* and *retrieval* were alternated with control conditions. Subjects had to complete three individual learning sessions to avoid the confounding effect of time. **B**: Behavioral performance data for each of the six subjects averaged across all three learning sessions. **C**: Anatomic model. Processing of object identity is mainly a property of the ventral visual pathway, whereas object location is a property of the dorsal stream. We focused on the interstream connections (mainly posterior parietal cortex to posterior inferotemporal cortex) based on the hypothesis that learning the association of object identity and spatial location leads to an increase in effective connectivity between the ventral and dorsal streams. (From Büchel C, Coull JT, Friston KJ. The predictive value of changes in effective connectivity for human learning. *Science* 1999;283: 1538–1541, with permission.)

clearly depended on the cutoff point between EARLY and LATE. To establish unequivocally a relationship between neurophysiologically mediated changes in connectivity and behavioral learning, we examined the relationship between the temporal pattern of effective connectivity changes and learning speed for all sessions and subjects. We estimated the differences in effective connectivity for seven EARLY and LATE partitions by successively shifting the cutoff. The cutoff time at which the connectivity change peaked was used as a temporal index of changes in effective connectivity (i.e., plasticity). The significant regression of k, a measure of learning speed*, on this plasticity index indicated that for sessions showing fast learning (i.e., high value of k), the maximum difference in path coefficients between PP and

ITp was achieved earlier in the session (i.e., EARLY comprises fewer scans relative to LATE) (Fig. 29.3). In other words, the temporal pattern of changes in effective connectivity strongly predicted learning or acquisition.

Example: Attention

Electrophysiologic and neuroimaging studies have shown that attention to visual motion can increase the responsiveness of the motion-selective cortical area (V5) (23,24) and the PP (25). Increased or decreased activation in a cortical area is often attributed to attentional modulation of the cortical projections to that area. This leads to the notion that attention is associated with changes in connectivity.

Here we present fMRI data from an individual subject, scanned under identical visual motion stimulus conditions while only the attentional component of the tasks employed was changed. First, we identify regions that show differential activations in relation to attentional set. In the second stage,

^{*} All individual behavioral learning curves were well approximated by the function $1 - e^{-kx}$, where 0 < k < .1 indexes learning speed. Small values of *k* indicate slower learning.



FIGURE 29.3. Changes in effective connectivity predict learning. This graph shows the correlation between the temporal index of changes in effective connectivity and learning. The temporal index is defined as the time of a maximum increase in effective connectivity between posterior parietal cortex and posterior inferotemporal cortex. For example, a temporal index of 3 indicates that the maximum increase in effective connectivity occurred between the third and fourth blocks. The numbers denote the subject from which this temporal index of effective connectivity was obtained. Each subject was scanned during three independent learning sessions; therefore, each number appears three times. A negative slope means that the maximum increase in effective connectivity occurs earlier in fast learning. (From Büchel C, Coull JT, Friston KJ. The predictive value of changes in effective connectivity for human learning. Science 1999;283:1538-1541, with permission.)

changes in effective connectivity to these areas are assessed with structural equation modeling. Finally, we show how these attention-dependent changes in effective connectivity can be explained by the modulatory influence of parietal areas by using a nonlinear extension of structural equation modeling. The specific hypothesis we addressed was that parietal cortex could modulate the inputs from VI to V5.

The experiment was performed on a 2-T MRI system equipped with a head volume coil. fMRI images were obtained every 3.2 seconds with echo-planar imaging (32 slices in each volume). The subject was scanned during four different conditions: fixation, attention, no attention, and stationary. Each condition lasted 32 seconds to give 10 volumes per condition. We acquired a total of 360 images. During all conditions, the subjects looked at a fixation point in the middle of a screen. In this section, we are interested only in the two conditions with visual motion (attention and no attention), in which 250 small white dots moved radially from the fixation point, in random directions, toward the border of the screen at a constant speed of 4.7 degrees per second. The difference between attention and no attention lay in the explicit command given to the subject shortly before the condition: just look indicated no attention, and detect changes indicated the attention condition. Both visual motion conditions were interleaved with *fixation*. No response was required.

Regions of interest were defined by categoric comparisons with use of an output statistical image $(SPM\{Z\})$ comparing *attention* with *no attention* and comparing *no attention* with *fixation*. As predicted, given a stimulus consisting of radially moving dots, we found activation of the lateral geniculate nucleus, primary visual cortex (VI), motion-sensitive area (V5), and posterior parietal complex. For the subsequent analysis of effective connectivity, we defined regions of interest with a diameter of 8 mm centered around the most significant voxel as revealed by the categoric comparison. A single time series, representative of this region, was defined by the first eigenvector of all the voxels in the region of interest (15).

Our model of the dorsal visual stream included the lateral geniculate nucleus, *V1*, *V5*, and the PP. Although connections between regions are generally reciprocal, for simplicity we modeled only unidirectional paths.

To assess effective connectivity in a condition-specific fashion, we used time series that comprised observations during the condition in question. Path coefficients for both conditions (*attention* and *no attention*) were estimated by using a maximum likelihood function. To test for the impact of changes in effective connectivity between *attention* and *no attention*, we defined a free model (allowing different path coefficients between *V1* and *V5* for *attention* and *no attention*) and a constrained model (constraining the *V1* \rightarrow *V5* coefficients to be equal). Figure 29.4 shows the freemodel and estimated path coefficients. The connectivity between *V1* and *V5* increases significantly during attention. Note also a significant difference in connectivity between *V5* and the PP.

The linear path model comparing attention and no atten-



FIGURE 29.4. Structural equation model of the dorsal visual pathway, comparing *attention* and *no attention*. Connectivity between right primary visual cortex (V1) and motion-sensitive area (V5) is increased during *attention* relative to *no attention*. This is also shown for the connection between V5 and the posterior parietal cortex. (From Büchel C, Friston KJ. Effective connectivity in functional brain imaging. *Neural Networks* 2000;13:871–882, with permission.)



FIGURE 29.5. Structural equation model of the dorsal visual pathway incorporating the interaction effect of right posterior parietal cortex on the connection from right primary visual cortex (*V1*) to motion-sensitive area (*V5*). (From Büchel C, Friston KJ. Effective connectivity in functional brain imaging. *Neural Networks* 2000;13:871–882, with permission.)

tion revealed increased effective connectivity in the dorsal visual pathway in relation to attention. The question that arises is, which part of the brain is capable of modulating this pathway? Based on lesion studies (26) and the system for directed attention described in ref. 27, the PP is hypothesized to play such a modulatory role.

We extended our model accordingly to allow for nonlinear interactions, testing the hypothesis that the PP acts as a moderator of the connectivity between V1 and V5. Assuming a nonlinear modulation of this connection, we constructed a new variable, V1PP, in our analysis. This variable, mediating the interaction, is simply the time series from region V1 multiplied (element by element) by the time series of the right posterior parietal region.

The influence of this new variable on V5 corresponds to the influence of the PP cortex on the connection between V1 and V5 (i.e., the influence of V1 on V5 is greater when activity in the PP is high). The model is shown in Fig. 29.5. Because our nonlinear model could accommodate changes in connectivity between *attention* and *no attention*, the entire time series was analyzed (i.e., attention-specific changes are now explicitly modeled by the interaction term).

As in the linear model, we tested for the significance of the interaction effect by comparing a restricted and a free model. In the restricted model, the interaction term (i.e., path from *V1PP* to *V5*) was set to zero. Omitting the interaction term led to a significantly reduced model fit (p < .01), which indicated the predictive value of the interaction term.

The presence of an interaction effect of the PP on the connection between V1 and V5 can also be illustrated by a simple regression analysis. If the PP shows a positive modulatory influence on the path between V1 and V5, the influ-

ence of V1 on V5 should depend on the activity of the PP. This can be tested by splitting the observations into two sets, one containing observations in which the PP activity is high and another one in which the PP activity is low. It is now possible to perform separate regressions of V5 on V1 by using both sets. If the hypothesis of positive modulation is true, the slope of the regression of V5 on V1 should be steeper under high values of PP.

Variable Parameter Regression

As demonstrated in the previous sections, the basic linear model can be seen as a linear regression. The regression coefficient is then interpreted as a measure of the connectivity between areas. This interpretation of course implies that the influence is mediated by neural connections with an effective strength equal to the regression coefficient. Using this approach, one immediately makes the assumption that the effective connectivity does not change over observations because only a single regression coefficient for the whole time series is estimated. This is unsuitable for the assessment of effective connectivity in functional imaging because the goal in some experiments is to demonstrate changes in effective connectivity-for instance, as a function of different conditions (e.g., attention and no attention) or simply time itself. In the framework of regression analysis, there are three ways around this problem. Firstly, one can split the data in different groups according to the experimental condition (e.g., attention and no attention) and then test for the difference of the regression coefficients. However, we may not know a priori the time course of the changes that allow us to split the data in this way. A second, more general solution is to expand the explanatory variable in terms of a set of basis functions to account for changes in connectivity. Here, we present another alternative, variable parameter regression, that allows one to characterize the variation of the regression coefficient by using the framework of state-space models and the Kalman filter (28,29).

Mathematical Background

Consider the classic regression model

$$y = x\beta + u \qquad [2]$$

where *y* is the measured data vector, *x* is a vector of explanatory variables, and β is the unknown parameter. Usually, β is estimated as

$$\hat{\beta} = pinv(x)y$$
 [3]

However, β can also be estimated recursively with the advantage that inversion of a smaller matrix is necessary. This approach is known as *recursive least squares* (30). This basic model is now extended to allow β to evolve over time.

Variable parameter regression assumes *T*-ordered scalar observations (y_1, \ldots, y_T) generated by the following model:

$$y_t = x_t \beta_t + u_t, t = 1, \dots, T,$$
 [4]

$$u_t \sim N(0, \sigma^2)$$
 [5]

where x_t is an *n*-dimensional row vector of known regressors and β_t is an *n*-dimensional column vector of unknown coefficients that corresponds to estimates of effective connectivity. u_t is drawn from a gaussian distribution. All observations are expressed as deviations from the mean.

A recursive algorithm known as the Kalman filter (29) can now be applied to estimate the state variable (β) at each point in time and also allows one to estimate the loglikelihood function of the model. A numeric optimization algorithm is then employed to maximize the likelihood function with respect to *P*. As the Kalman filter is a recursive procedure, the estimation of β_t is based on all observations up to time *t*. Therefore, the filtered estimates will be more accurate toward the end of the sample. This fact is corrected for with the Kalman smoothing algorithm, which is used *post hoc* and runs backward in time, taking account of the information made available after time *t*. Details of the Kalman filter and smoothing recursions can be found in standard textbooks of time series analysis and econometrics (31,32).

Example: Attention to Visual Motion

To illustrate variable parameter regression, we use the singlesubject data set from the study of attention to visual motion. We concentrate on the effect of attention on the connection between the motion-sensitive area (V5) and the PP in the right hemisphere. Using structural equation modeling, we demonstrated that it is principally this connection, in the dorsal visual stream, that is modulated by attention (15). In the current analysis, we are interested in whether variable parameter regression is capable of reproducing these findings. We therefore have assessed the effective connectivity β_t by regressing PP on V5. An alternate direction search, numeric optimization, gave a χ^2 statistic of 56.4. We therefore had to reject the null hypothesis of no variation at the 5% level. P was estimated to be 0.074 and σ^2 was 0.23. The ordinary regression coefficient β for the model $\gamma =$ $x\beta$ + *u* was estimated at 0.73. Figure 29.6 A,B shows the trajectories of the smoothed and filtered estimates $\beta^{\mathbf{I}}_{t}(T)$ together with the associated standard errors. It is clearly evident that β_t is higher during the *attention* conditions than during the no attention conditions. Figure 29.6D relates our technique to an ordinary regression. In this analysis, we constrained the variance term P to zero and reestimated β_t . The trajectory of β_t now converges to β , the ordinary regression coefficient of the model $y = x\beta + u$. As expected, the smoothed estimates are simply a constant (i.e., $\beta = 0.73$).

We interpret $\beta_t^{\mathbf{z}}$ as an index of effective connectivity between area V5 and the PP. In our example, the connection between V5 and the PP resembles the site of attention modulation. This leads to an interesting extension, in which one might hypothesize that a third region is responsible for the observed variation in effective connectivity indicated by the trajectory of $\beta_{t}^{I}(T)$. In other words, after specifying the *site* and nature of attentional modulation, we now want to know the location of the *source*. We addressed this by using $\beta^{\mathbf{I}}_{t}(T)$ as an explanatory variable in an ordinary regression analysis to identify voxels that covaried with this measure of effective connectivity. Figure 29.6C shows the result of this analysis. Among areas with statistically significant (p < .001, uncorrected) positive covariation were the dorsolateral prefrontal cortex and the anterior cingulate cortex. This result confirms the putative modulatory role of the dorsolateral prefrontal cortex in attention to visual motion, as suggested by previous analyses (15).

Effective Connectivity versus Categorical Comparisons

One obvious advantage of the assessment of effective connectivity is that it allows one to test hypotheses about the integration of cortical areas. For example, in the presence of modulation, the categoric comparison between *attention* and *no attention* might reveal prestriate, parietal, and frontal activations. However, the only statement possible is that these areas show higher cortical activity during the *attention* condition as opposed to the *no attention* condition. The analysis of effective connectivity revealed two additional results. Firstly, attention affects the pathway from V1 to V5 and from V5 to PP. Secondly, the introduction of nonlinear interaction terms allowed us to test a hypothesis about how these modulations are mediated. The latter analysis suggested that the PP exerts a modulatory influence on area V5.

The measurements used in all examples in this chapter were *hemodynamic* in nature. This limits an interpretation at the level of *neuronal* interactions. However, the analogy between the form of the nonlinear interactions described above and voltage-dependent (i.e., modulatory) connections is a strong one. It is possible that the modulatory impact of PP on V5 is mediated by predominantly voltage-dependent connections. We know of no direct electrophysiologic evidence to suggest that extrinsic backward PP to V5 connections are voltage-dependent; however, our results are consistent with this. An alternative explanation for modulatory effects, which does not necessarily involve voltage-dependent connections, can be found in the work of Aertsen and Preissl (10). These authors show that effective connectivity varies strongly with, or is modulated by, background neuronal activity. The mechanism relates to the efficacy of subthreshold excitatory postsynaptic potentials in establishing dynamic interactions. This efficacy is a function of post-



FIGURE 29.6. A,**B**: The trajectory of the smoothed and filtered estimates $\beta_{t}^{2}(T)$ together with the associated standard errors for the variable parameter estimation of effective connectivity between motion-sensitive area (*V5*) and posterior parietal cortex (PP). It is evident that β_{t}^{1} (the dynamic regression coefficient) is higher during the *attention* conditions than during the *no attention* conditions. **C**: Areas that significantly covaried with the time-dependent measure of effective connectivity between *V5* and the PP [i.e., $\beta_{t}^{1}(T)$]. The output statistical image *SPM*/*Z*/ thresholded at p < .001 (uncorrected) overlaid on coronal and axial slices of the subject's structural MRI. The maximum under the cross-hairs was at 45, 21, 39 mm, Z = 4. **D**: The relationship between our technique and an ordinary regression analysis. In this analysis, the variance term *P* was set to zero (i.e., fixed regression model). The trajectory of β_{t}^{1} now converges to β (= 0.73), the regression coefficient of the model $y = x\beta + u$. (From Büchel C, Friston KJ. Dynamic changes in effective connectivity characterized by variable parameter regression and Kalman filtering. *Hum Brain Mapping* 1998;6:403–408, with permission.)

synaptic depolarization, which in turn depends on the tonic background of activity.

CONCLUSIONS

This chapter has reviewed the basic concepts of effective connectivity in neuroimaging. We have introduced several methods to assess effective connectivity—multiple linear regression, covariance structural equation modeling, and variable parameter regression. In the first example, structural equation modeling was introduced as a device that allows one to combine observed changes in cortical activity and anatomic models. An application of this technique revealed changes in effective connectivity between the dorsal and the ventral stream over time in a paired-associates learning paradigm. The temporal pattern of these changes was highly correlated with individual learning performance, and therefore changes in effective connectivity predicted learning speed. The second example of structural equation modeling focused on backward modulatory influences of high-order areas on connections among lower-order areas. Both examples concentrated on changes in effective connectivity and allowed us to characterize the interacting areas of the network at a functional level. Variable parameter regression was then introduced as a flexible regression technique that allows the regression coefficient to vary smoothly over time. Again, we confirmed the backward modulatory effect of higher cortical areas on those areas situated lower in the cortical hierarchy. Although this field is less than mature, the approach to neuroimaging data and regional interactions discussed above is an exciting endeavor that is starting to attract more and more attention.

REFERENCES

- Lashley KS. Brain mechanisms and intelligence. Chicago: University of Chicago Press, 1929.
- Shallice T. From neuropsychology to mental structure. Cambridge: Cambridge University Press, 1988.
- Warrington EK, Shallice T. The selective impairment of auditory short-term memory. *Brain* 1969;92:885–896.
- 4. Atkinson RC, Shiffrin RM. Human memory: a proposed system

and its control processes. In: Spence KW, Spence JT, eds. *The psychology of learning and motivation: advances in research and theory* (vol 2). New York: Academic Press, 1968.

- Petersen SE, Fox PT, Snyder AZ, et al. Activation of extrastriate and frontal cortical areas by words and word-like stimuli. *Science* 1990;249:1041–1044.
- 6. Posner MI, Petersen SE, Fox PT, et al. Localization of cognitive operations in the human brain. *Science* 1988;240:1627–1631.
- Marshall JC, Newcombe F. Patterns of paralexia: a neurolinguistic approach. J Psycholinguist Res 1973;2:175–199.
- 8. Friston KJ, Frith CD, Liddle PF, et al. Functional connectivity: the principal component analysis of large (PET) data sets. *J Cereb Blood Flow Metab* 1993;13:5–14.
- Friston KJ, Frith CD, Frackowiak RSJ. Time-dependent changes in effective connectivity measured with PET. *Hum Brain Mapping* 1993;1:69–80.
- Aertsen A, Preissl H. Dynamics of activity and connectivity in physiological neuronal networks. New York: VCH Publishers, 1991.
- Gerstein GL, Perkel DH. Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science* 1969;164:828–830.
- Gerstein GL, Bedenbaugh P, Aertsen A. Neuronal assemblies. IEEE Trans Biomed Eng 1989;36:4–14.
- Friston KJ, Ungerleider LG, Jezzard P, et al. Characterizing modulatory interactions between V1 and V2 in human cortex with fMRI. *Hum Brain Mapping* 1995;2:211–224.
- McIntosh AR, Grady CL, Ungerleider LG, et al. Network analysis of cortical visual pathways mapped with PET. *J Neurosci* 1994; 14:655–666.
- Büchel C, Friston KJ. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modeling and fMRI. *Cereb Cortex* 1997;7:768–778.
- McIntosh AR, Gonzalez-Lima F. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapping* 1994;2:2–22.
- 17. Moseley ME, Cohen Y, Kucharczyk J, et al. Diffusion-weighted

MR imaging of anisotropic water diffusion in cat central nervous system. *Radiology* 1990;176:439–445.

- Kenny DA, Judd CM. Estimating nonlinear and interactive effects of latent variables. *Psychol Bull* 1984;96:201–210.
- Büchel C, Coull JT, Friston KJ. The predictive value of changes in effective connectivity for human learning. *Science* 1999;283: 1538–1541.
- Milner B, Johnsrude I, Crane J. Right medial temporal-lobe contribution to object-location memory. *Philos Trans R Soc Lond B Biol Sci* 1997;352:1469–1474.
- Desimone R. Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci USA* 1996;93:13494–13499.
- 22. Wiggs CL, Martin A. Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol* 1998;8:227-233.
- O'Craven KM, Savoy RL. Voluntary attention can modulate fMRI activity in human MT/MST. *Invest Ophthalmol Vis Sci* 1995; 36[Suppl]:S856.
- Treue S, Maunsell HR. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* 1996;382: 539–541.
- Assad JA, Maunsell JH. Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature* 1995;373:518–521.
- Lawler KA, Cowey A. On the role of posterior parietal and prefrontal cortex in visuo-spatial perception and attention. *Exp Brain Res* 1987;65:695–698.
- Mesulam MM. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol* 1990;28:597–613.
- Büchel C, Friston KJ. Dynamic changes in effective connectivity characterized by variable parameter regression and Kalman filtering. *Hum Brain Mapping* 1998;6:403–408.
- Kalman RE. A new approach to linear filtering and prediction problems. *Trans Am Soc Med Eng J Basic Eng* 1960;D82:35–45.
- Harvey AC. *Time series models*. London: Harvester & Wheatsheaf, 1993.
- 31. Chow GC. Econometrics. New York: McGraw-Hill, 1983.
- 32. Harvey AC. Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press, 1990.