Neural Dynamics In Brain Networks During The Resting State And Visual Word Recognition

Andrew J. Quinn

Doctor of Philosophy

University Of York

Department of Psychology

September 2014

Abstract

This thesis investigates the dynamics of information flow within brain networks during the resting state and visual word recognition. Functional connectivity within brain networks has become increasingly prominent across cognitive neuroscience and neuroimaging in recent years and conventional approaches for identifying instantaneous interactions within brain network and across the whole head are now commonplace. Magnetoencephalography (MEG) recordings have a very high temporal resolution which allows for the characterisation of delayed interactions between distant brain regions such as those caused by limited conduction speeds along white matter fibres. This thesis presents an approach to characterising such time-delayed interactions and critically, inferring the direction of information flow. This approach is used to demonstrate the existence of statistically significant differences in the information flow in each direction of a connection between two nodes in a resting state network. A Hidden Markov Model is then used to characterise dynamic changes in this directionality. Task driven directional connectivity is then investigated in the context of visual word recognition. A complex and rapidly evolving pattern of connectivity arises during visual word recognition, with specific connections modulated by the psycholinguistic properties of the stimulus. Critically, the influence from the Left Inferior Frontal Gyrus is shown to transfer more information to visual regions when reading a challenging stimulus.

Contents

List of Figures

List of Tables

Acknowledgements

Many thanks to my supervisor, Gary Green, for countless pieces of excellent advice in critical moments and for providing a discerning direction in the not-so-critical moments. Your passion for science in all its forms has been an inspiration.

Thank you to my first supervisor, Piers Cornelissen, for starting me off on this path four years ago and for four years of patience whilst I walked it.

Thanks to the MEG team at YNiC: Mark, Sam, Garreth, Becky and Michael for many instances of support and advice. I don't want to think where I would be without it.

Thanks to my friends in York (past and present) for all the good company over the past few years, and for many well timed distractions from work.

Tori, thank you for your strength and encouragement. Your smile has kept me going more times than I have said out loud.

Finally, thank you to my family. For all the years of caring, support and advice. Without you, I wouldn't have grown into who I am today.

Declaration

The present thesis contains original work completed by Andrew J. Quinn under the supervision of Professor Gary G.R. Green at the University of York. This research was supported by an ESRC White Rose Doctoral Training Centre Scholarship awarded to Andrew J. Quinn. This work has not previously been presented for an award at this, or any other, University.

Some of the empirical results from this thesis have previously been presented at the following conference:

Quinn, A.J., Hymers, M., Johnson, S., Lobier, M., Wheat, K., Hansen, P., Green, G.G.R. & Cornelissen, P.L. (2014) Neural Dynamics Of Visual Word Recognition. Poster presented at 19th International conference on Biomagnetism, Halifax, CAN.

Quinn, A.J., Hymers, M., Johnson, S., Cornelissen, P.L. & Green, G.G.R. Reconstructing Rapid Dynamics In Endogenous Brain Networks Reveals Frequency Specific Directional Interaction. Poster presented at 19th International conference on Biomagnetism, Halifax, CAN.

Chapter 1

Outline

Spatio-temporal patterns of neural activity evolve rapidly both at rest and in response to stimuli in the natural environment. This spatiotemporal structure has been widely studied since the advent of neuroimaging, building upon decades of neuropsychological research into the structure and functional organisation of the brain. More recently, increasing attention has been paid to the functional relevance of connections between brain regions.

This thesis investigates the dynamics of neural systems with an emphasis on directed functional connectivity. In other words, how information is transferred within brain networks and how this is modulated both spontaneously over time and during visual word recognition. Spontaneous brain activity has long been considered "noise" in investigations into brain function and is often subtracted off or averaged out during neuroimaging analysis. Yet, as already indicated, spontaneous brain activity has a robust spatiotemporal structure comprising several brain networks. Of these, the default mode network (DMN, containing principally the medial prefrontal cortex and posterior cingulate cortex) is unique in being discovered purely from neuroimaging analyses. However the network has been consistently identified in both fMRI and MEG and across very large datasets. Despite the large literature on the DMN, little is known about the dynamics of information flow between its parts or how this may change over time.

Recent evidence suggests that there is a correspondance between the brain's functional networks during the resting state and task situations (Smith et al., 2009). A critical difference is that while resting brain function may evolve on a relatively slow time-scale (from hundreds of milliseconds up to minutes), cognitive processes related to a specific task can evolve over milliseconds. As such, the analysis of a the reading network as it processes a word is highly challenging. Yet interrogating these very rapid task related dynamics may be criticallly important.

Reading plays a vital role in modern society; thoughts and ideas abstracted into writing can be transported great distances and preserved for long periods of time with relative ease. As such, writing systems have had a profound impact on human society and development in the 5000 years since their invention. A skilled reader can effortlessly extract the meaning and

name of a word within a few hundreds of milliseconds, yet this is a difficult process which can take many years to perfect. Continuous reading of sentences is a highly complex process, about which we known very little. Visual word recognition is the process that allows a person to efficiently identify a written word. It plays a vital role in the fluent reading of continuous writing and has been the subject of research for over a century. This thesis investigates the mechanisms within the brain associated with the recognition of single written words, an essential component to contextual reading.

Early visual processing starts with a representation of the inputs to the visual system and culminates with a description depending on both the inputs and the purpose for which they are viewed; the key problem is determining the processes between (Marr, 1976). In the case of reading, abstract visual word forms are detected and processed in such a way as to enable integration of visual information with the language networks of the brain. From these processes and interactions the phonological and semantic associations of the word form can be elicited.

In contrast to the DMN, the brain network associated with reading and word recognition has been investigated long before the introduction of neuroimaging. They are common in that, though the brain regions involved are well known, relatively little is known about the dynamics of directional information flow within these networks.

1.1 Chapter Outlines

This section provides a brief outline of the motivation behind each chapter and it's principal aims.

1.1.1 Literature Review

Previous research into the neural substrate underlying visual word recognition is outlined, starting with early neurology and cognitive models. The key contributions of neuroimaging are then introduced alongside the emerging resting state literature.

1.1.2 Estimating Directed Functional Connectivity From Dynamics In An Embedding Space

The mathematical context for the analysis used in the thesis is introduced. Critically, the change of a dynamical system over time can be reconstructed with a time-delay embedding. This can then be used as the basis of a multivariate autoregressive model which, in turn, provides an estimate of the linear temporal dependencies within a network of brain regions.

1.1.3 Realistic Simulations Of Causal Brain Network

Several practical aspects of the analysis outlined in chapter 2 are tested with multivariate simulations. Principally, the importance of having sufficient data for accurate parameter estimates and the conditions determining the frequency resolution of a model are investigated.

1.1.4 Directed Functional Connectivity In An Endogenous Brain Network

The analysis above is applied to identify whether the functional interactions between nodes in the DMN are symmetrical, ie equivalent in both directions. Eyes open resting state MEG data is acquired and source activity time-series estimated within the DMN and several associated brain regions. The directed functional connectivity is estimated and each pair of connections tested identify any asymmetries in functional influence.

1.1.5 Dynamic Changes In Within Network Directed Connectivity

Temporal changes in directed functional connectivity in the DMN from the previous chapter are characterised with a Hidden Markov Model (HMM). Four functional states are identified each with specific directional interactions within the network.

1.1.6 Neural Dynamics Of Visual Word Recognition

The analysis pipeline is adapted to estimate dynamics over very short timescales. This is applied to MEG data collected while participants viewed Words, Consonant Strings and False Font stimuli. Each stimulus category is shown to be associated with different patterns of directional connectivity within a reading network identified from beamformer power contrasts. Critically the Left Inferior Frontal Gyrus is shown to provide more top-down influence on visual brain regions during the unrecognisable stimuli (consonant strings and false fonts). This suggests that the LIFG may be most critical when viewing stimuli whose orthographical and phonological structure is challenging to decode, if not impossible.

1.1.7 Multiple Routes To Recognition

In a final experiment, the functional role of the early interactions between the LIFG and visual regions is probed using a contrast between the processing of high and low frequency words. It is shown that low frequency words elicit greater top-down influence from LIFG adding support for its role in decoding challenging word-like stimuli.

1.1.8 Discussion

The overall themes of the thesis are discussed and linked to specific results throughout the thesis.

Chapter 2

Literature Review

2.1 Visual Word Recognition

Many writing systems are in use by humans across the globe, however the majority of these use one of three strategies. A morpho-syllabic system, such as Chinese Hanzi or Egyptian Hieroglyphics, encodes a whole word with one symbol whereas a syllabic system, such as Japanese Kana, encodes a syllable or combination of sounds into one letter. Finally an alphabetic system, such as English written in the Latin script, represents a word as a series of phonemes, each represented by one or more letters. This thesis will be restricted to the discussion and analysis of the alphabetic writing system of English. In an alphabetic system such as English, there exists an orthography defining the mapping from letter shapes into phonemes or sounds. This representation of the sound of a word is known as its phonology, and both the orthographic and phonological information can play a role in accessing the meaning or semantics of a word. These mappings are a vital part of learning to read. Once learnt, they can be applied to read and pronounce novel words or pseudo-words. There is not necessarily a one-to-one mapping between a given visual word form or letter and a sound, English vowels in particular may change their sound depending on the surrounding letters and the identity of the word itself. For instance, the letter 'i' may refer to an /aɪ/ sound in 'pint' or an /ɪ/ sound in the word 'mint'.

Word recognition is further complicated by variability in the written script. Writing can appear in a wide range of sizes, fonts, positions and colours; yet a skilled reader is able to read a known language in a huge range of styles, even those they are naive to. The speed and effortlessness with which a reader is able to extract information from such a variety of possible inputs suggests a flexible and highly efficient system is present in the skilled reader. Moreover, it provides a considerable challenge to scientists looking to model and understand this system. The following section outlines previous work towards understanding visual word recognition in the human brain.

2.1.1 A Historical Perspective

During the nineteenth century Carl Wernicke suggested that only primary perceptual functions can be localised in specific regions of cortex and that any higher functions, such as thought or consciousness, rest on the fibre bundle connecting disparate regions of cortex (Wernicke 1874 as cited in Eggert, 1977). Specifically, Wernicke suggested that reading is learned by developing a strong association between the visual and acoustic 'images' of the word. The acoustic image could then be used to activate the motor image through interaction with Broca's region in the inferior frontal lobe so the word can be spoken aloud. Thus the union of the visual and acoustic images constitute the concept of the letter rather than a letter necessarily having its own representation (Eggert, 1977). Consequently, damage to a connection between regions may not just lead to lost communication but loss of higher level representations themselves.

In 1892 Jules Déjérine reported the case of patient who presented with a left hemianopia as well as an inability to read but with preserved writing ability and auditory language. Post-mortem examination of the patient revealed damage to the left central white matter leading to disconnection of the occipital radiation to the left visual cortex and disconnection to the bilateral visual cortices to the angular gyrus (Déjérine 1891,1892 as cited in Catani and Ffytche, 2005). Déjérine proposed that these disconnections might lead to pure alexia as it prevented integration of information between the primary visual cortices and communication between visual cortex and the left hemisphere Angular gyrus which was thought to act as a visual verbal centre, a store for visual word forms. (Déjérine 1891,1892 as cited in Catani and Ffytche, 2005). This was in contrast to Wenickes suggestion that the connection itself contained the representation of the word, yet the idea that cognitive function can be impaired by disconnecting two brain regions remains central.

Norman Geschwind outlined a novel account of disconnection syndromes (Geschwind, 1965a,b). Geschwind's innovation was based on the assumption that there were no connections between primary sensory areas in man

(Flichsig 1901). Geschwind suggested that the development in association cortex in humans allowed for cortical connections between unimodal processing in the primary sensory cortices, whereas in Monkeys and lower mammals these associations are mediated via the Limbic system. Thus in man, the existence of this higher/multimodel association cortex (such as the angular gyrus) allows for the sensory association cortices to interact in more complex ways (Geschwind, 1965b). In this sense, visual word recognition may arise from early visual processing in primary and associative visual cortex, before the visual association cortex engages the rest of the language network via higher association cortex (angular gyrus). More modern consideration of this theory suggests that the link between pure alexia and damage to left inferior occiptotemporal cortex (iOTC) involving the surrounding white matter 'cannot be doubted' (Bub et al., 1993) however there is less support for a visual verbal store in the left angular gyrus.

The classical neuropsychological evidence has provided increasingly detailed accounts of how damage to specific brain structures may lead to specific cognitive impairment. Though the precise model of how this might arise changes, the idea that disconnection between brains regions leading to cognitive deficits has remained prominent for over 100 years.

2.1.2 Models Of Word Recognition

The early part of the twentieth century saw a move towards the study of groups of participants in conjunctions with the development novel of statistical techniques. These developments played a part in what is known as the Cognitive Revolution (Gardner, 2008), during which cognition was conceptualised as a series of statistical processes with complex behaviours arising from combinations of simpler component processes. Cognitive science therefore places an emphasis on mental states and processes rather than considering the underlying biology. The focus was placed on the goals of visual word recognition and tested with observations of typical and disordered reading behaviour. What appropriate steps must be taken? and what strategies would be appropriate for undertaking these steps?

Early models of psychological phenomena such as visual word recognition tended to outline the sequence of component processes verbally (Jacobs and Grainger, 1994). These models typically contained several specialised components which process aspects of the stimulus. One influential model assumed the existence of a device called a *logogen* which receives linguistically relevant information from sensory processes and yields a response when a this input reaches a threshold (Morton, 1969). In the case of visual word recognition, a logogen might receive information about the visual features of a presented word and respond when sufficient information has accrued to confirm the identity of the word. Other models have a wider scope and describe how word representations interact with phonological and semantic information (LaBerge and Samuels, 1974) or how the duration of eye fixations while reading relates to several levels of cognitive processes (Just and Carpenter, 1980).

Many models of reading have been based on the idea that there are two or more routes from a seeing a word to recognising it (Davelaar et al., 1978; Coltheart et al., 1993; Ellis and Young, 1996). Of these a dual route architecture with two routes to recognition has been most influential (Coltheart et al., 1993). One route maps from vision to semantics and the other from vision to phonemes. The grapheme to phoneme (letters to sounds) is thought to be predominantly used for when learning to read or reading unfamiliar words. This allows for the possibility of 'indirect' recognition facilitated by the auditory lexicon which is activated by the conversion to phonemes. This model can be extended to incorporate speech input and written output (Ellis and Young, 1996), an example architecture can be see in figure 2.1.

Early theories emphasised the word as a whole perceptual unit (Cattell, 1886). Though having a holistic perception of a whole word would be efficient, even for skilled readers the fundamental unit of perception for written words is the letter (Pelli et al., 2003). This suggests that letter identities constitute a major part of the representation of a word. Beyond letters themselves, the order of the letters within a word must be encoded in order to distinguish between anagrams such as *able* and *bale*. Slot based coding schemes may be able to explain how we can disambiguate anagrams. These

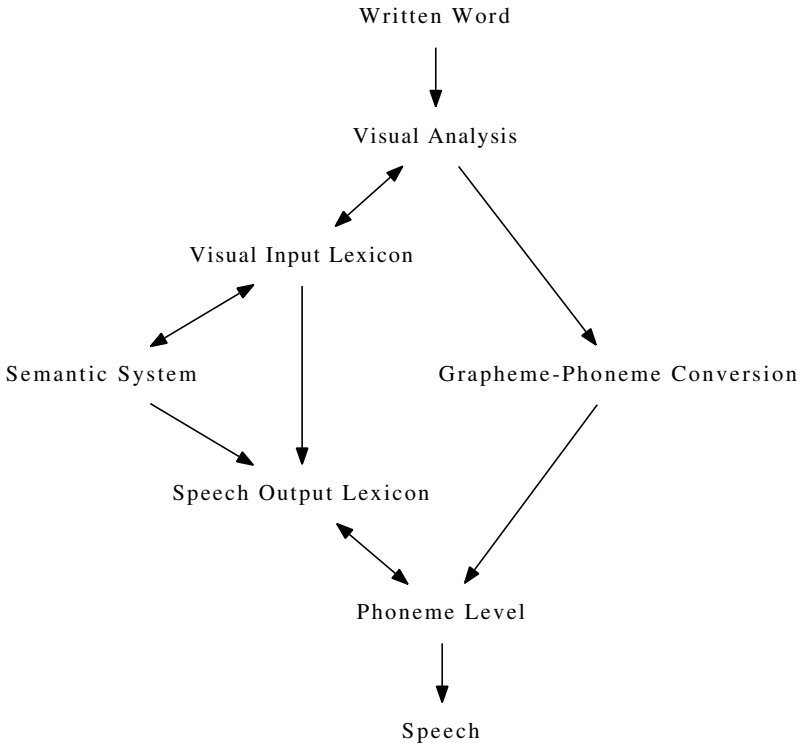

Figure 2.1: A model for recognising and speaking written words adapted from Coltheart et al. (1993) and Ellis and Young (1996). This features a highly modular dual route processing stream. Two routes from writing to speech are proposed: a direct route via the semantic system and an indirect route via grapheme to phoneme conversion.

state that there are a series of slots each representing letter positions within a word, perceived letters are sequentially added to successive slots to form a word level representation. For instance, the word brain might be represented with B in the first slot, R in the second and so on, forming a full representation of the word ($B_1R_2A_3I_4N_5$)

These slot models imply that letter position and identity are encoded simultaneously, as a consequence the R_2 in BRAIN would have a completely separate representation to the R_5 in its anagram ABNIR. This assumption has been criticised on the grounds that it is possible for a non-word to prime a target real word containing similar letters, the extent of the priming increasing with orthographic overlap (Humphreys et al., 1990; Peressotti and Grainger, 1999). Current theories that may account for these effects are open bigram coding, the SERIOL models and the SOLAR model.

Open bigrams represent words as a series of letter pairs, allowing for the letters in the pair to be separated by another letter as long as the relative order of the pair is maintained. This concept suggests that the word HEAD might be represented as the bigrams HE HA HD EA ED AD, and has become quite influential as evidence that the brain uses a relative letter position encoding scheme (Schoonbaert and Grainger, 2004). The SERIOL model presents a mechanistic account of how open bigrams might be used to activate a representation of the entire word (Whitney, 2001). Bigrams are very specialised mechanisms, tailored to one specific purpose. As such, models containing bigram representations imply that a totally specific processing pathway for reading must exist within the brain. Given that reading is learnt rather than innate and has only been in existence for the last few thousand years, it seems unlikely that such a mechanism might have evolved (Dehaene and Cohen, 2007).

Often models of reading tend to be underspecified, as they may be defined verbally or diagrammatically it is possible that critical details about its operation are not defined sufficiently or neglected all together (Jacobs and Grainger, 1994; Coltheart et al., 2001). For instance, all of the models referenced above include some reference to the encoding of visual information about a word. The inputs to a logogen are "visual attributes" which arise

from earlier visual processing (Morton, 1969), an "extract physical features" process is identified in Just and Carpenter (1980) and LaBerge and Samuels (1974) suggest that visual processing proceeds along a hierarchy such that features → letters → spelling → words. These processes are defined in varying detail, yet none of these models provide an explicit mechanism for the visual feature processing in word recognition. It is therefore challenging to empirically evaluate the performance of these models, making it difficult to objectively improve a theory or explicitly compare the performance of two models.

Computational Models

Computational models are computer programs capable of simulating a cognitive task via a mechanism specified in a theory of how this task is processed in a human (Coltheart et al., 2001) (Figure 2.2A). These models provide a way to systematically test and develop theories of human behaviour. The design of a computational model is very flexible and allows for several philosophical approaches to be directly tested. For instance the Dual Route Cascaded (DRC) model of visual word recognition is explicitly a computational realisation of the dual-route model of reading (Coltheart et al., 2001). Its design is motivated by the anatomical models from the nineteenth century and cognitive models of the twentieth, specifically, that language is best modelled as a multicomponent modular cognitive information processing system. In contrast the Parallel Distributed Processing (PDP) model is derived from a set of general principles concerning the nature of neural computations (Seidenberg, 2006) (Figure 2.2B).

There has been a series of models developed from these two frameworks and their relative performance is hotly debated (Coltheart et al., 1993, 2001; Zevin and Seidenberg, 2006). Yet, a direct comparison between the two approaches may not be appropriate. The DRC has a bottom-up data-fitting approach designed to account for maximum amount of variance in data collected from a series of lab experiments. A such they are tailored to a discrete range of specific tasks and may struggle to explain data from subtly

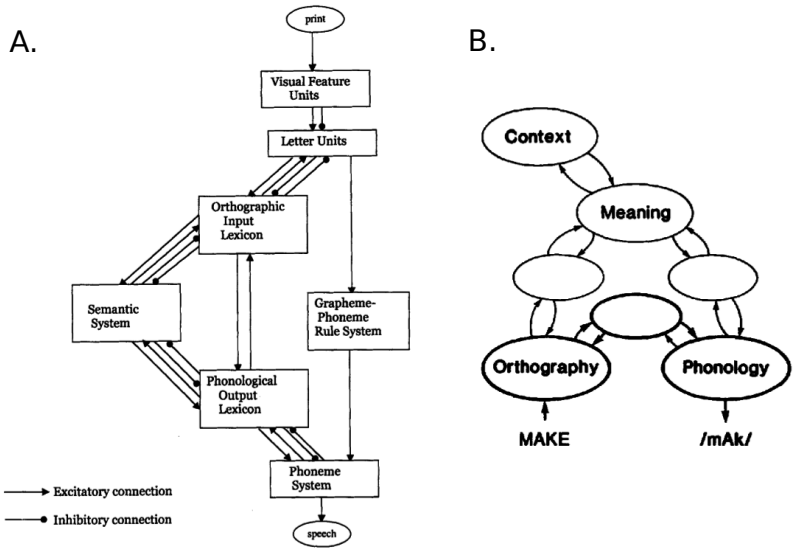

Figure 2.2: A - The Dual Route Cascaded (DRC) model (Coltheart et al., 2001). B - The Parallel Distributed Processing (PDP) model (Seidenberg and McClelland, 1989)

different tasks (Seidenberg, 2006). In contrast, the PDP approach is less able to explain data in specific instances, however its goal is to use evidence provided by a principled model along with evidence from other sources to converge on the correct theory Seidenberg (2005).

These computational models have a broad scope, tackling the whole process from seeing a word to reading it out loud. As such they may simplify subcomponent processes such as how letter position is encoded. Both the DRC and PDP suggest that letter position is encoded via the relatively simple slot based mechanism rather than bigrams (McClelland and Rumelhart, 1981; Coltheart et al., 2001). This is a critical question in visual word recognition, what is the precise nature of the representation of a word form in the brain?

Bayesian Models

Models which incorporate open bigram coding schemes can account for a wide range of behavioural findings and have provided a successful mechanism for several computational models, if this is true the brain has a highly ordered and specialised representation of letter strings. Moreover the SOLAR and SERIOL models propose a highly complex series of specialised mechanisms geared towards word recognition. Finally these models all assume that all the visual input occurs simultaneously.

This assumption may not be necessary. Many of the behavioural effects that inspired the development of these models can be accounted for by treating the word recognition process as a the categorisation of the outputs of a noisy channel rather than a specific orthographic code. In other words, the mechanisms underlying word recognition are best understood though analysis of the probabilistic computations that must be performed rather than the nature of its representations (Norris and Kinoshita, 2012).

The Bayesian Reader (BR) is a computational model assuming that a human acts as an optimal Bayesian decision maker when recognising a word (Norris, 2006). This model has successfully replicated several major behavioural word recognition effects such as the relationship between word

frequency to reaction time and identification thresholds (Norris, 2006, 2009). Words are represented as simple letters strings rather than the more complex open bigrams, suggesting that some of the more complex reading phenomena do no require a stage of representation beyond the letter string.

The Bayesian reader assumes that all possible words can be represented in one multidimensional perceptual space, perceptual inputs are then a method for sampling evidence about a stimulus from the environment, over successive samples this evidence can be used to compute the conditional probability of observing the perceptual input (I) for any given word (W). Bayes law in this context states that:

$$P(W|I_n) = P(W) * \frac{P(I_n|W)}{\sum_{j=0}^{j=n}(P(I_j) * P(I_j|W))} \tag{2.1}$$

Bayes theorem can be used to calculate the posterior probability of a particular word (W) being presented given the evidence sampled so far (I). This posterior probability is computed from the a prior probability ($P(W)$) that the word occurred and the likelihood that the evidence is consistent with that word being presented ($P(I_j|W)$) (Norris et al., 2008). The BR uses metrics such as the frequency of a word in a given language to determine its prior probability which imply the meta-features of words are represented alongside the words themselves.

Summary

A variety of reading models have been developed since the nineteenth century, yet the approaches to modelling have shown some consistencies. Broadly, the modeller either takes a data driven stance and tries to explain some observed data with a specific mechanism or computational framework (logogens,the bigram, the DRC) or they take a more theoretical stance based on some principles of the system (PDP,BR). Both the data-driven and principled approaches attempt to make the workings of an unobserved system explicit, and converge on an optimum explanation for how the brain com-

pletes a specific task. Several keys findings and consistencies can be draw from across aspects of this literature. The majority of models encompass a division of labour across distinct yet interactive modules or nodes.

The modelling literature attempts to explain how particular tasks might be achieved by specifying component processes in a computer program. While this evidence is often compelling, it remains abstracted from the brain itself. The development of neuroimaging techniques in recent decades has made it possible measure neural activity directly, providing a source of evidence about how the brain itself processes different aspects of a visual word form. The next section will outline this emerging literature on how and where different processes occur within the brain.

2.1.3 The Brain Basis Of Word Recognition

Modern neuroimaging methods allow the opportunity to investigate how brain responses measured *in vivo* correlate with perceptual and psychological events. Measures based on cerebral blood flow such as Positron Emission Tomography (PET) and Functional Magnetic Resonance Imaging (fMRI) are indirectly sensitive to neuronal activity, the assumption being that an increase in the activity of a population of neurons will lead to an increase in blood flow in that region. These methods typically make measurements every 1-3 seconds and can accurately localise activity in the brain within millimetres. Electrophysiological methods such as Electroencaphalography (EEG) and Magnetoencephaolography (MEG) measure neural activity more directly by observing the electrical and magnetic fields on the scalp originating from current flow in populations of neurons. These methods are able to take hundreds or thousands of samples per second. Inverse solutions for determining the brain region which generated a given measurement are in continuous development and are able to localise activity, under optimal conditions, with centimetre (EEG) or millimetre (MEG) resolution. Finally, invasive neuronal measure record activity from a single neuron or populations of neurons directly with a temporal resolution similar to EEG and MEG, however these methods require surgical implantation of electrodes

and as such opportunities to carry out such measurements are relatively rare.

All of these methods have been recruited to investigate the brain basis of visual word recognition. The functional and anatomical evidence they provide is typically used to localise cognitive functioning to a region in the brain and /or a time during a behavioural task. The following section provides an overview of contemporary literature seeking to investigate the neural basis of visual word recognition. Moreover the potential of neuroimaging to investigate localist versus distributed approaches to word recognition is discussed and related to the cognitive and computational models in the previous section.

Localising Word Recognition

Cohen et al. (2000) used fMRI and EEG to investigate the location and timing of brain responses to words presented in left and right visual fields in patients with posterior callosal lesion and healthy controls. They found early visual activation in controls to be contralateral to the visual field the stimulus was presented in, however a more anterior site was active in the left hemisphere only for both conditions. In contrast, the patients with posterior callosal lesions only showed this lateralised activation in words presented in the right visual field. Cohen et al. (2000) suggest that the visual word form system is localised in this anterior region and label it the Visual Word Form Area (VWFA). This anterior site in the ventral occipito-temporal cortex (vOT) has further been shown to respond preferentially to real words and pronounceable non-words over checkerboards or consonant strings (Cohen et al., 2002). Moreover there is evidence that it only responds to visual rather than spoken words (Dehaene et al., 2002) and that these representations are invariant to the retinal location of the visual input (Dehaene et al., 2004). Due to the apparent equivalence of words and pronounceable non-words, it has been suggested that the VWFA contains an abstracted sub-lexical representation of a visual word form. Many of these fMRI studies use an averaged BOLD signal change contrast to identify populations of

neurons which selectively respond to a given stimulus category, however a BOLD adaptation approach may be a more sensitive way to identify selectivity. Indeed, this approach has shown evidence that the VWFA is tuned to representations of whole, real words rather than a sub lexical visual word form (Glezer et al., 2009). This evidence suggests that the the VWFA in the vOT is tuned for the representation of words itself, this is in contrast with Déjérine's suggestion that the AG was the seat of representations of a visual words.

The assignment of a function to a discrete region in the brain can be problematic. In the case of left inferior occipito-temporal cortex, activity in the VWFA can be modulated by several tasks including visual motion discrimination, tactile word and object recognition and spoken word recognition (Price and Friston, 2005). This is hard to reconcile with cognitive models of reading which often present the components of a cognitive process as separate, discrete processes. Several avenues of investigation may help to resolve these difficulties and build upon this work on localisation. Firstly, due to the many to one mapping between function and region, it would be beneficial to consider multiple brain regions in such as way that a cognitive process is represented through the connection between these brain regions (Wernicke 1874 as cited in Eggert, 1977). Moreover it may be that this wider network context drives a regions function rather than any inherent specialisation (McIntosh, 2000). Secondly, activity within the brain unfolds on millisecond timescale. A region may perform subtly different roles at different points in the processing of a word, perhaps influenced by the wider state of the network. The next section discusses the timings associated with visual word recognition.

Dynamics Of Word Recognition

A common method for investigating the temporal evolution of brain activity elicited by a word is the event related potential (ERP) which is a measure of phase locked or evoked activity. Neural responses to a word are measured over many trials and averaged to remove inter-trial variability. Furthermore

source estimation techniques such as minimum norm estimates or beamforming can be used to reconstruct source activity as it changes over time. This section outlines how findings from these methods can be used to differentiate serial and parallel processing in the brain and how these findings relate to models of visual word recognition.

Pulvermüller et al. (2009) draw a distinction between serial and parallel models of word recognition (an outline can be seen in Figure 2.3). Serial models outline a chain of subprocesses which may have onset asynchronies of hundreds of milliseconds, in contrast parallel models allow several perceptual and cognitive processes to occur at or around the same time. Dien (2009) identify seven reading components from within the first 400ms of reading a single word. Neural generators for these components are suggested and the two broad stages of processing identified. The first 'estimation' phase involves early visual and inferior occipito-temporal areas processing the word and attempting a grapheme to phoneme conversion based on local information. The second 'resonance' stage involves refinement of the estimate by coordinating information across the whole reading network. A further literature review suggested that the early ERP responses reflect near parallel processing of several aspects of word recognition (Pulvermüller et al., 2009). These findings are more in line with a parallel distributed mechanism supporting visual word recognition.

Direct recordings of electrical activity on the cortex of patients undergoing treatment prior to surgery to relieve severe epilepsy allow for recordings with high spatial and temporal resolution. This method was used by Nobre et al. (1994) who found several large activations to words in the inferior temporal lobes within 500ms of stimulus onset. Two distinct regions were found; a posterior region responded equally to words and non-words was active around 200ms and was unaffected by context. Subsequently a more anterior region was sensitive to the difference between words and non words around 400ms and modulated by context. These regions were shown to be dissociated from regions responsive to other complex visual stimuli suggesting that they comprise a specialised processing stream within the ventral stream. This result provides a temporal scale to the cortical response to

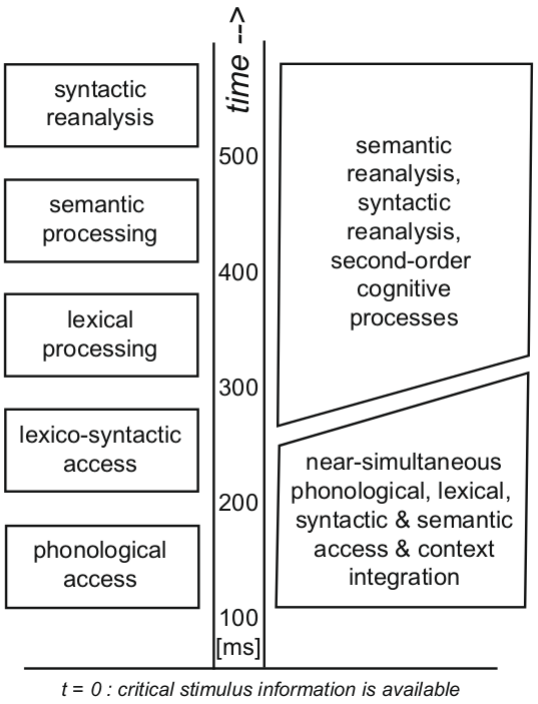

Figure 2.3: Pulvermüller et al. (2009)'s description of serial/modular theories of visual word recognition and interactive parallel accounts

words and is in general agreement with suggestions that the inferior temporal lobes contains regions which processes visual aspects of written words (Petersen et al., 1990; Pugh et al., 1996; Cohen et al., 2000).

Magnetoencephaolography (MEG) measures the modulation of magnetic fields generated by cortical sources outside the head, and provides a measurement technique with a high temporal resolution, and in conjunction with source estimation algorithms, a reasonable spatial resolution. MEG source analysis has shown that posterior visual region responses equivalently to both words and consonant strings around 130ms after stimulus onset, however, more anteriorly, the vOTC response does differentiate between the two (Tarkiainen et al., 1999; Cornelissen et al., 2003). More exploratory whole brain, analyses with MEG data have provided more direct evidence for the posterior to anterior processing direction by showing the evolution of the cortical response across the cortex over time in healthy individuals (Marinkovic et al., 2003; Pammer et al., 2004). Though these biomagnetic and blood flow based methods both measure signals coupled with neuronal responses, there is evidence that they reveal different neuronal functions in the occipito-temporal cortex during reading (Vartiainen et al., 2011).

2.1.4 Brain Models Of Word Recognition

The cognitive models outlined in section 2.1.2 provide a detailed account of the sub-processes that support word recognition. In recent decades new models of word recognition inspired by functional anatomy have been developed in light of growing neuroimaging literature. The following sections outline two such models.

Hierarchical Models

The Local Combination Detector (LCD) model (Dehaene et al., 2005) is an influential hierarchical model of orthographic encoding. The LCD model describes how different parts of the ventral visual stream might be tuned into different letter and word features in a hierarchical configuration. The premise is that as visual information about a word travels down the visual

stream both the type of information encoded by the cortex and its receptive field size changes to reflect the integration of simple visual features into size and position invariant representations of letters, bigrams and complete words. Early in the model simple features such as contrast in the LGN and orientation in V1 are encoded retinotopically before V2 encodes letter fragments or combination of lines. Beyond this case-specific letters are encoded in V4 and abstract case-invariant letters in V8. All of these representations are bilateral. The first lateralised stage in the model is bigram encoding thought to reside in the left OTS posterior to the cortex encoding small words and morphemes.

This model would predict more selective responses to words over word-like stimuli in anterior over posterior parts of the left hemisphere ventral visual stream as has previously been found in the VWFA literature (Cohen et al., 2000, 2002). This prediction was tested in fMRI by (Vinckier et al., 2007) who found posterior occipital regions responded to all stimulus categories with more anterior cortex responding successively more selectively to the more word like stimuli. The LCD model can be considered within the wider framework of cultural recycling of cortical maps (Dehaene and Cohen, 2007), this is a mechanism by which regions of cortex which have evolved specialisations can be recruited to support relatively recent cultural innovations such as reading.

The LCD model contains bigram detectors similar to those in several cognitive and computational models and as such is open to the same criticisms. Primarily, the inclusion of a very specific and highly specialised mechanism in the brain for reading is unlikely to have developed in the 5000 years since writing systems began to develop. Moreover, little consideration is given for how feedback both within the visual hierarchy and between the visual system and the rest of the brain might fit into the process. The emphasis is firmly on how a simple-to-complex hierarchy of representations might be instantiated. This experiment is designed and justified purely in terms of orthographic processing, however the progression from letter strings to pronounceable non-words and finally words leads to a increasingly realistic orthographic structure which is increasingly pronounceable. Thus the find-

ings of Vinckier et al. (2007) might be equally attributable to phonological or orthographic processes, indeed same the pattern of differences found in the vOT was also seen in the LIFG.

There is also evidence against a the existence of areas specifically specialised for written words. Büchel et al. (1998) tested the reading of visual word forms and braille in sighted and blind participants. They found that anterior iOTC was activated in a word over consonant string reading in both blind people reading braille and sighted people reading visual words. The authors suggest that the role of this region is not linguistic per se but may promote activity in other regions which them leads to lexical access. This view is supported by the suggestion that there are no cases with exclusive reading deficit from a left mid fusiform lesion, and that this same region has been associated with a range of tasks (Price and Devlin, 2003).

The LCD model has the advantage that it makes explicit statements about where in the brain different types of information are extracted and integrated. In addition several aspects of the model are based on successful heuristics from other visual processing models, such as the visual processing hierarchy working from posterior to anterior regions idea from object recognition (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992; Riesenhuber and Poggio, 1999). It also follows directly from several very influential cognitive and computational approaches proposing increasingly abstract serial sub-processes underlying reading.

Interactive Models

Price and Devlin (2011) advocate an interactive account of ventral occipitotemporal function in reading. This is based on the premise that perception depends on interactions between sensory cortex and higher order processing regions via reciprocal connections. vOT activation is explained in terms of three things: 1) The bottom up connections communicating visual inputs 2) the top-down connections communicating predictions made by higher order cortical regions and 3) the mismatch between these inputs and predictions. Fundamentally, according to this approach vOT activity during reading is

modulated by top-down predictions and non visual properties of words, furthermore interaction between the ventral visual stream and higher order language areas optimised for comparing top-down predictions with bottom-up sensations is key to reading (Twomey et al., 2011). A key difference with hierarchical models is that the interactive account does not require a specialist mechanism to represent words in the visual cortex. They suggest that a the generic representation of a word form in the visual system is sufficient for recognition, this arises as higher regions of cortex are tuned to estimate the sensory input that would exist if a given word were present. This approach is similar to the BR computational model and implies that specialisation for language exists in non-visual parts of the brain but that these areas contain abstracted representations that can generate sensory predictions.

The Bayesian Brain Hypothesis (BBH) suggests that the brain has an internal representation of the external world which it uses to predict and explain sensory inputs (Knill and Pouget, 2004). In contrast to the hierarchical approach, the BBH states that sensory inputs are encoded probabilistically. In this sense, a letter or word would not be discretely represented by a single detector or similar, rather a probability density or likelihood function would represent the relative probability that a given letter or word generated the available sensory information. This is analogous to the BR (Norris, 2006) which models word recognition using a probabilistic letter representation to cope with noisy sensory input. When this representation is considered with a prior belief, Bayes Law can be used to calculate the posterior probability of a word given the sensory input so far. One criticism of these approaches is the source of the prior belief in these models, which is typically taken from an abstract, task relevant source such as relative word frequencies (Norris, 2006). The 'Free Energy Principle' (Friston, 2010) builds upon this idea by introducing the idea of an empirical prior, which provides a framework for optimising the priors on a model from the data rather than making *a priori* assumptions about what they are.

The vOT may act as the brain region in which the visual sensation of letters and words and the predictions from higher areas are evaluated. In this sense it can been seen how one region of cortex might assume different

functions in difference contexts as top-down modulations arrive from different networks in different context. This provides a framework for describing how the brain achieves visual word recognition in terms of prediction, yet no specific statements are made about the nature of the information represented in the cortex. The free energy approach might suggest that visual cortex represents high dimensional probability density functions based on sensory input, this representation is then modulated by a prior expectation or prediction from attentional or language circuits in the brain leading to the final representation of the word in cortex. In contrast to the LCD model, the interactive and bayesian approaches do not make specific statements about the nature of representation in the cortex.

The interactive model presents an approach which could explain how visual word recognition can be achieved by the interaction between the bottom-up visual process and a top-down predictive mechanism. Though this is a powerful framework, the precise nature of this interaction and the neural representations necessary for it to occur are underspecified. Furthermore, the Bayesian models imply that the brain deals directly with complex probability functions across a space which theoretically encompasses any possible word form. It may not be necessary for the brain to encode this entire function, just its sufficient statistics (for instance the mean, and variance of a normal distribution). Yet even the sufficient statistics of a rapidly evolving probability function with many peaks could potentially be very challenging to derive and encode efficiently.

Bayesian models of reading provide emphasis several important features of a successful reading system. Principally, any perceptual model must account for noise in the system. Noise can arise at any point in visual processing and probabilistic models such as the BR are able to demonstrate how a reader might be able to cope with it. Furthermore the emphasis on the interface between internal predictions about the world and incoming sensory information is a stark departure from the almost deterministic mechanisms outlined in the LCD. A significant difficulty with interactive models is their underspecification, though they present a relatively simple mechanism by which word recognition might occur under some very general assumptions,

they require the brain to process several abstracted probability functions. In short, hierarchical models such as the LCD make precise statements about how aspects of a word form can be represented in the brain, therefore the representations are simple. A feature or bigram is either present or not. In contrast the interactive models remove the necessity for detailed and specialised representational schemes but these then imply that the brain can represent abstract probability functions.

2.1.5 Overview

Broadly speaking, many cognitive, computational and brain models of word recognition fall into one of two camps. Firstly the cognitive models that dominated psychology from the 1960s are focuses on abstract symbolic computations. Key to these models are the structure and processes underlying mental representation and how these lead to complex behaviours (Gardner, 2008). Many models explain visual word recognition through these abstracted mental states (Morton, 1969; Just and Carpenter, 1980; Ellis and Young, 1996). In contrast, more recent computational PDP models have sought to explain word recognition from a few simple principles of the nervous system (Seidenberg, 2005). Models containing large numbers of interconnected and parallel computational units which behave like neurons are able to learn and perform many tasks remarkably similarly to humans. The implication being that the traditional view on abstracted cognitive representation may be over-conceptualised and that relatively simple biologically inspired mechanisms might explain many behavioural findings.

Neuroimaging research to date has primarily focused on the localisation of cognitive representations and processes. This approach provides more readily testable hypotheses for current neuroimaging methods to investigate the functioning of different brain regions. Moreover models such as the LCD (Dehaene et al., 2005; Dehaene and Cohen, 2011) provide explicit statements about how information is encoded in the brain. In contrast, the Interactive account of word recognition Price and Devlin (2011) explains the process in more general terms derived from lower level properties of neurons. This

explanation is related to the PDP approach in its attempt to build processes from simple principles and the Bayesian literature in its focus on internal models of the world. Yet it remains underspecified, these models do not currently state explicitly *what* neurons must represent or *how* they might represent it.

The nature of representation and information flow in the early visual system might be able to disambiguate between these approaches. Is there more evidence to support a symbolic processing hierarchy in which information is interactively processed in a simple-to-complex series of representational levels, or a noisy probabilistic process in which internal representations about the cause of sensations are dynamically updated across the whole processing stream as the visual system provides pieces of evidence?

2.2 Early Visual System

A general principle for the global function of visual cortex would have to account for both the specialisations and interactions within the system, and remains controversial (Zeki, 2001; Wandell et al., 2007). One influential model in object recognition makes a distinction between a dorsal 'where' stream and a ventral 'what' stream specialised in locating and identifying stimuli in space respectively (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992), though this model provides a framework for visual processing it stops short of a mechanistic account of how specialisation and interaction within the system give rise to visual percepts. Grill-Spector and Malach (2004) propose another model in which visual processing is described as a hierarchy of visual regions starting in V1 and ascending to higher level specialised regions such as the fusiform face are and PPA, within each stage of the hierarchy there is a specialisation gradient, in V1 this is a specialisation for the foveal visual inputs and in higher cortex specialisations from central stimuli such as faces to more peripheral-biased stimuli such as buildings and scenes.

In contrast to these hierarchical views, it has been suggested that the visual system operated in a reverse hierarchy (Hochstein and Ahissar, 2002).

Initial vision at a glance identifies general categories in the visual scene before a top-down process scrutinises the sensory input in more detail. The idea is that very rapid automatic visual processes are able to activate high level representations such as an object category or a word within a few hundred milliseconds, and only then are the fine grained details investigated in detail. Source estimates from MEG recordings have demonstrated that the orbito-frontal cortex is recruited during object recognition around 50ms before semantic/recognition related regions in the temporal lobes, moreover this activation is modulated by the low spatial frequencies in the image (Bar et al., 2006). This distinction between a rapid feed-forward early process and later interactive processing has been demonstrated during object recognition in the Macaque (Lamme and Roelfsema, 2000) who suggest that the early process reflects pre-attentive vision and that the later recurrent processing supports the attentive processing required for awareness.

Overall, the visual system begins processing in the eye. A range of information including the local contrasts and spectral content of light in the visual field is extracted a passed up several parallel routes to processing in the cortex. Very early in the cortex, several interactive processing streams further extract information from the visual scene, possibly guided by a prior belief or prediction about the world or the result of an automatic and rapid interpretation of the visual scene. Different types of visual information are available at different latencies with information from the Magnocellular pathway reaching cortex around 20ms before information from the Parvocellular pathway. Finally there is no clear end point to visual processing, top-down, bottom-up and lateral connections between many areas imply that recurrent feedback/feedforward processes, rather than a bottom-up hierarchy, may be key.

In relation to models of visual word recognition, evidence from the visual system would suggest a great deal more rapid interactive processing than has been suggested by models such as the LCD. Moreover, it seems likely that different information is available at different timescales with rapid processing related to low spatial frequencies in the orbitofrontal cortex around 130ms after stimulus onset, possibly facilitated by the conduction times

within the M pathway. This information must be integrated with the slower more detailed processes that come later. The visual processes at the start of written word recognition support a model of reading that emphasises interaction and can account for information about the word form becoming available dynamically over the first few hundred milliseconds after stimulus onset.

2.3 The Language Network

In 1865 Paul Broca stated that articulate speech is located in 'the third frontal convolution of the left hemisphere' (Berker, 1986). This now classic paper presented evidence that lesions in the left inferior frontal gyrus (LIFG) were solely responsible for the loss of articulate speech, and in doing so suggested that a single brain region might be responsible for a highly complex behaviour such as speech. This idea of functional localisation of specific processes has been influential ever since. More recent cognitive models of reading identify three major subprocesses, orthography, phonology and semantics, relating to the symbology, sound and meaning of a written word respectively. The localisation and function of these components within the brain has been the focus of a large neuropsychology and neuroimaging literature over the last century.

2.3.1 Language Areas And Recognition

Section 2.3.1 of this review outlined how lesion and neuroimaging evidence can build a case for the localisation of a function such as word recognition. Efforts to elucidate how phonology and semantics are accessed and influence the recognition process have taken a similar path. Early work was dominated by the deficits caused by lesions to different parts of the brain until advances in medical imaging allowed for *in vivo* localisation of function using activation or adaptation paradigms. This section outlines recent work on word processing outside the visual cortices, in particular, the brain basis of phonology and semantics.

The overall pattern of cortical recruitment over time during word recognition can be estimated from MEG data. Source estimates of phase-locked cortical activity during visual word recognition from anatomically constrained MEG source analysis data have found a posterior to anterior progression of activation (Marinkovic et al., 2003). Beamforming is another source estimation technique which can recover the total, as opposed to evoked, activity in the brain's source space. Studies using this method have suggested a more complex pattern of activity, including early activation of the LIFG, though the general posterior to anterior trend in processing remains (Pammer et al., 2004; Cornelissen et al., 2009).

A wealth of localisation studies in fMRI have been reported since the 1990s and this has lead to several formal meta-analyses looking for consistent findings across this literature. Turkeltaub et al. (2002) conducted a meta-analysis of 11 PET studies of single word reading which were compared to a single fMRI study. Several regions showed consistent activation, including bilateral motor and superior temporal cortices, the cerebellum and the left hemisphere fusiform gyrus. These regions were validated with a novel fMRI reading experiment which showed substantial correspondence with the results of the PET meta analysis. Consistent activation of the left fusiform supports findings outlined in section 2.3 about the relevance of this area to the processing of visual aspects of words. The authors further suggest a phonological role for the IFG and a semantic role for anterior fusiform cortex. A further meta-analysis focusing on cross cultural differences expands on this suggestion, phonological processes are thought to be seated in superior temporal region, inferior frontal, inferior parietal and pre-motor regions, whereas semantic processes occur in anterior fusiform cortex and the middle temporal gyrus (MTG) (Bolger and Perfetti, 2005). Several of these regions have been shown to interact during the resting state (Koyama et al., 2010). Finally (Jobard et al., 2003) conducted a meta-analysis of 35 fMRI studies with a view to evaluating cognitive dual-route models of reading. They suggest that a graphophonological route, similar to the indirect pathway rests on superior temporal gyrus, supramarginal gyrus and the posterior part of the IFG. In contrast, the lexicosemantic route arises

from co-activation of the VWFA with inferior temporal regions, the MTG and the IFG *pars triangularis*.

Phonology

The dual-route model of reading suggests a critical role for a mapping between a visual letter form and a sound, particularly while learning to read (Coltheart et al., 2001). Originally the transformation to a phonological code was thought to occur after the meaning had been extracted (Kleiman, 1975), yet this may occur before lexical access and provide a means for lexical access when orthographic information is insufficient (Besner, 1987).

Broca's area is the inferior frontal gyrus (IFG) of the left hemisphere and was initially thought to be involved in producing speech. Recent neuroimaging evidence has linked sections of it with phonological processes (Fiez, 1997) though it has also been implicated in language-related functions such as syntax, action perception, working memory (Grabowski et al., 1998) and selection from competing alternatives (Moss et al., 2005). Broca's area has a complex cytoarchitecture (Amunts et al., 1999; Keller et al., 2009) and structural connectivity profile (Anwander et al., 2007). As such, it may serve several parallel functions within the reading network. Furthermore, different subregions of the IFG are preferentially activated depending on whether a linguistic or task-demand based BOLD contrast is used in the analysis (Wright et al., 2011).

MEG evidence supports a role for the IFG in phonological processing (Cornelissen et al., 2009; Wheat et al., 2010). Of particular interest in these results is the early timing of the posterior IFG response; it is preferentially activated within 100ms when a word is phonologically, as opposed to orthographically, primed (Wheat et al., 2010) suggesting that this sub-region is involved in rapid phonological access. Whether this early response is necessary for recognition remains controversial, Wheat et al. (2013) used TMS to show that disruption to the IFG around 200-300ms after stimulus onset was most disruptive to articulatory-phonological processes rather than word recognition itself.

The superior temporal gyrus has also been implicated in a conversion from orthography to phonology though it is typically associated with processing of speech (Simos et al., 2002; Jobard et al., 2003). In contrast, a meta analysis of 120 fMRI studies suggests that it is associated with perceptual auditory processes (Binder et al., 1997, 2009). Furthermore, though the STG is associated with successful reading aloud its response is not correlated with word/bigram frequency, imagability or consistency (Graves et al., 2010), again suggesting a low-level perceptual rather than a more linguistic role. A meta analysis of 36 studies showed that the STG was only present in couple of studies using a word minus pseudoword or *vice versa* contrast (Taylor et al., 2012). In summary, there is little evidence from the fMRI literature that the STG is directly related to word recognition.

The evidence for phonological processes being important in visual word recognition is controversial. The lack of consistency may be as skilled readers only recruit the phonological pathway during recognition of difficult or infrequent words.

Semantics

Semantic processing refers to concepts or meaning. In visual word recognition this process is thought to be necessary for access to and comprehension of the meaning of a written word.

Semantic dementia patients typically show deteriorated semantic memory with relatively preserved linguistic function after damage to the temporal pole (Hodges et al., 1992). This suggestion that the temporal pole may subserve semantic processing is generally supported by the imaging literature (Hotton and Yoshimi, 2011), though there are some inconsistencies possibly based on the signal dropout around the temporal pole in fMRI (Visser et al., 2012). In general there is strong support for the recruitment of the temporal pole and ventral anterior temporal lobe in amodal semantic processing, as such it may play an under appreciated role in semantic aspects of reading.

The superior temporal gyrus has long been accredited with an essential role in language function, the posterior parts forming 'Wernicke's Area' along

with the Angular Gyrus (AG) (Wernicke 1874 as cited in Eggert, 1977) This region has been associated with language comprehension, with lesions leading to receptive aphasias. In contrast to this, more recent evidence presented in the section above suggests it is more associated with perceptual processing of the sound of a word.

Jules Déjérine proposed that the AG near the temporoparietal junction may act as a visual verbal word form store (Déjérine 1891,1892 as cited in Catani and Ffytche, 2005). The AG is densely interconnected with other relevant regions in the reading network (Catani et al., 2005; Seghier and Friston, 2013) and appears in contemporary reading models (Price, 2012). The precise role of the AG remains unclear, the early models of Déjérine suggest a role in the initial identification of a stimulus as a word. The consequences from a lesion to AG suggest a more general cognitive role with damage potentially leading to a wide range of deficits from alexia, acalculia, apraxia and dementia (Binder et al., 2009). As such, it may act as an amodal high level conceptual integration area (Geschwind, 1965a).

The connections of the AG in relation to reading, present a clearer picture. The gray and white matter associated with the AG has been shown to be different between early and late literate readers (Carreiras et al., 2009), suggesting that rather than a mere association, the structure and connections of the AG may be shaped by reading. In addition, the angular gyrus of people with developmental dyslexia has been shown to be relatively disconnected from the reading network in comparison to typical readers (Horwitz et al., 1998). In summary, there is strong evidence to suggest that the AG is relevant to the reading network, possibly acting as a visual lexical store.

Overall, several brain regions are associated with semantic processes relating to the meaning of a word. This meaning must be access to fully recognise a word suggesting that visual word recognition is very likely to engage semantic regions such as the AG and temporal pole. The nature of the interactions between vision and semantics remains unclear. A hierarchical model such as the LCD might suggest a bottom-up information flow in which a representation of a word in vOT engages a semantic representation in the TP, in contrast interactive accounts would suggest that top-down

interactions are necessary to recognise a word.

2.3.2 Summary

Despite the inputs to the reading system being visual, the processes that follow recruit a widely distributed and highly interactive network of brain regions. Phonological processes are associated with regions such as the superior temporal lobe and Broca's area in the inferior frontal gyrus, while semantic processes are associated with the temporal pole, anterior fusiform and middle temporal gyus. As previously discussed with the localisation of word recognition, it can be problematic to make one-to-one structure-to-function mappings McIntosh (2000); Price and Friston (2005). Though the precise mapping of cognitive functions to brain regions is difficult, there remains considerable evidence for specialisation of function within the reading network.

Overall, there is considerable evidence in support of a distributed, dynamical network supporting the processing of visual words in the brain. There are some difficulties mapping functions to specific regions within this network, however these can be tackled by modelling the network as a whole. A burgeoning connectivity literature in the field of neuroimaging provides several approaches for the modelling of interdependencies and interactions between brain regions. Understanding these interactions will be a critical step towards understanding how we are able to recognise a visual word form.

2.4 Modelling The Visual Word Recognition System

The importance of interactions between brain regions in reading, has been established through a consideration of contemporary theories of visual word recognition, the distributed and interactive processing in the early visual system and complex network of higher brain regions recruited as we read. There is more evidence to support a distributed processing model such as the PDP rather than a strictly hierarchical framework. Despite this evidence,

the majority of contemporary neuroimaging research in this area still only interrogates the brain using mass-univariate models which are optimal for localisation. This is a powerful approach for localising isolated brain regions associated with specific tasks, however different methods are needed to begin to investigate the connections and interactions between brain regions.

Several approaches for estimating connections and information flow in the brain are available. These are forms of multivariate time series analysis, which enables us to quantify dependencies and causal relationships between brain regions. There are two main approaches to multivariate modelling of neural interactions. Data-driven methods draw on classical time-series analysis commonly found in the physical sciences and engineering, whereas dynamic causal modelling provides a hypothesis testing framework which is explicitly designed and optimised for modelling the brain and neuronal interactions. Both of these approaches are outlined in this section along with examples of their application with respect to visual word recognition. More mathematical detail on the following measures will be provided in Chapter 2 of this thesis.

The most common measures of association between two time series are the correlation and cross-correlation (David et al., 2004). The coherence function is the normalised Fourier transform of the cross-correlation and expresses linear dependencies in the frequency domain. The cross-correlation and coherence are measures of association only, characterising how related two time series are. No causal inferences can be drawn due to the symmetry of these metrics. That is the correlation of A and B is the same and the correlation of B and A. Some directionality can be inferred from the cross-correlation, which is asymmetric, however this does not necessarily imply a causal relationship and may be mediated by an unmeasured third source.

Granger Causality (GC) is a test for statistical dependencies between activity in one area and the *past history* of activity in a second. A time series x can be said to be Granger Causal to a second time series y if prediction of y given the past history of x and y is more accurate that the prediction of y from the past history of y alone (Granger, 1969). This is intimately related to the cross correlation function between x and y and has subsequently been

extended into the frequency domain (Geweke, 1982). GC can be thought of as a measure of directed functional connectivity (Friston et al., 2012) in that the GC of A and B will not necessarily be the same as the GC from B to A.

Multivariate auto-regressive (MVAR) models, provide a method for characterising the multivariate dependencies between time-series. This is achieved by fitting a model which predicts a subsequent data point in one node from the past history of the node, and every other node in the network. These models are derived from observed data and can be fitted with the appropriate selection of two free parameters, the delay and the model order. The delay determines the time between observations used as predictors in the model and the model order the number of past history samples to include from each node.

After a MVAR model has been fitted to the observed data, the parameter matrix can be manipulated in a number of ways to express the influences between the time series in the model. The parameters can be used in a bivariate GC test to establish information flow between pairs of nodes in the network. In addition, several multivariate causality measures have been developed which express the relative strength of particular connections as a ratio of the total information flow in a network. The most widely used methods are are the Directed Transfer Function (Kaminski and Blinowska, 1991) and the Partial Directed Coherence (Baccala and Sameshima, 1998; Baccalá and Sameshima, 2001). Both of these metrics can reveal information about the structure of the dependencies between the modelled time-series, however only PDC is able to discriminate between direct and indirect connections.

The temporal resolution of MEG is sufficient to allow estimation of MVAR models over time windows as short as 50ms (Ding et al., 2000; Sun et al., 2009a). This creates the potential to evaluate measures such as the DTF and PDC on a very short time-scale using a sliding window approach. This approach has been applied to object recognition (Supp et al., 2007) and auditory word repetition (Korzeniewska et al., 2008), demonstrating a highly dynamic and complex pattern of information flow.

2.4.1 Application to Visual Word Recognition

In application to visual word recognition, a common analysis pathway is to identify relevant brain nodes using a traditional activation paradigm before estimating the connectivity between the nodes (Salmelin and Kujala, 2006). In this way connectivity information can be used to directly extend the existing literature.

Dynamic imaging of coherence sources is a metric applied to source estimates of brain activity designed to reveal information about coherence between regions in the brain (Gross et al., 2001). It is able to establish associations between brain regions in an exploratory way and as such is able to naturally extend previous work charting the evolution of cortical responses to words (Pammer et al., 2004). Applied to MEG collected as a participant performs a reading task many brain regions have been found to be coherent in the alpha band, moreover coherence and GC patterns across the whole left hemisphere have been estimated in this manner (Kujala et al., 2008). Both metrics reveal a complex pattern of interconnections within the left hemisphere. There have been several recent reports of directional links between nodes in the reading network. Simos et al. (2002) used multiple autoregressions to probe the time lagged dependencies between several reading nodes. They found that early activity in fusiform contributed to responses in AG around 200-250ms after stimulus onset, furthermore the AG and STG activity contributes to later IFG responses between 350 and 450ms after stimulus onset.

GC analysis of the interactions between STG and vOT during a visual word priming task showed that despite reduced local activation as a result of priming, connection strength between the superior temporal cortex influences ventral occipito-temporal cortex increased between 150 and 290ms (Kujala et al., 2012). This finding suggests top-down modulation of visual responses possibly based on the phonological content of a word. More generally, the increased information flow on when reading a primed word highlights the potential functional relevance of connections on cognitive functioning.

Given the pattern dynamic changes in source activation during reading (Marinkovic et al., 2003; Pammer et al., 2004) we might expect to see dynamic changes in connectivity as well. The few studies that have applied these methods to reading in a time-sensitive manner have found dynamic shifts but have stopped short of a providing a detailed account of the interactions across the whole network. Instead presenting connectivity over relatively broad consecutive time windows or restricting analysis to very few nodes. A principled investigation into the dynamic changes in connectivity across the reading network has the potential to disambiguate between the hierarchical and interactive accounts of visual word recognition by providing a framework to test for early or transient information flow during word recognition.

2.4.2 Summary

Overall, data driven MVAR/GC methods provide a powerful metric for estimating information flow within the brain. They have been successfully applied to several other domains in neuroscience Supp et al. (2007); Kaminski and Blinowska (1991) and have demonstrated the modulation of connections within a reading network by both experimental manipulations and dyslexia. These methods model dependencies between observations of a system, that is dependencies between observed responses rather than neural systems *per se* (Friston et al., 2012). As such there is no explicit modelling of the neural system generating the observations.

2.5 Generative Models Of Visual Word Recognition

Hinton et al. (1995) makes the distinction between recognition and generative processes in neural network learning. Recognition is the transformation from raw inputs to the model to some internal representation, while generative processes work to reconstruct the raw inputs from an internal state. In the context of cognitive neuroscience, generative models seek to predict ob-

served brain responses from a model of the brains hidden dynamics (Friston and Price, 2001). Generative models use both a neuronal model of the underlying dynamics and an observation model describing how these internal states produce observed data in contrast to the more data driven approached which only model dependencies between observed data. In the case of visual word recognition, the internal representation might be of a written word form. The recognition process would then be the transformation of a visual input into this abstracted word representation while the generative process would reconstruct the sensory input associated with a word from.

As generative models account for internal states explicitly, they are able to model causal effects between these states that might be hidden from the observed data. Moreover, the distinction between recognition and generative processes may provide a mechanism for learning and prediction in the brain. Internal representations of perceptual inputs are are updated and optimised in order to predict the sensory inputs with minimal error.

A Generative model can be fitted to optimally predict several brain responses. The majority of studies fit these models to explain raw brain response data from neuroimaging experiments, however it is also possible to optimise the model to explain more complex observations such as the cross-spectral or even the PDC/DTF of the observed responses.

Dynamic causal modelling is the most prominent generative modelling approach used in neuroimaging (Friston et al., 2003). Originally formulated for fMRI data it has since been extended for EEG and MEG (Kiebel et al., 2008). Observed data are explained by fitting parameters to one or more biologically motivated neural mass or neural field models located in the brains source space. Typically, several DCMs with differing assumptions about the nature of interactions within the network are investigated. Comparing the ability of each of these models to explain the raw data provides a framework for testing hypotheses about the interactions between the neuronal assemblies underlying the responses we observe from neuroimaging.

There has been a strong debate in the literature about the ability of DCM to accurately select the correct model given many alternatives and the practicalities of comparing large numbers of models or models with large

numbers of nodes (Lohmann et al., 2012; Friston et al., 2013; Lohmann et al., 2013). Moreover, despite being grounded in the biophysics of the brain (Daunizeau et al., 2011) there is as yet no evidence that the canonical microcircuit model and other source models used in DCM accurately reflect the working of the brain.

DCM And The Inferior Frontal Gyrus

The role of the inferior frontal gyrus in reading is hotly debated, it has a presence in many brain models of language and as such provides a rich source of potential connectivity hypotheses. The direction of influence between the IFG and associated language areas is vital in determining its role. The winning DCM which explains most of the data from an fMRI single letter detection task Liu et al. (2011) contains top down connections originating from the IFG to the right middle occipital gyrus. This implies the IFG has a top-down role in VWR, highlighting relevant features for further processing. A more complex spelling and rhyming task in fMRI has found converging evidence for this top-down influence. The winning DCM suggested that task modulated the target region for top-down influence from the IFG (Bitan et al., 2005). The authors suggest that the IFG enhances sensitivity to task dependent information. These findings support reading models in which the IFG has an attentional role, modulating activity in earlier areas depending on the task or context of the experiment.

Other language properties have been associated with connections between the IFG and the reading network. Heim et al. (2009) compared several models involving two areas in the LIFG and the inferior temporal gyrus, during visual presentations of words and pseudowords. The stimuli were assumed to directly excite the ITG which would subsequently activate either BA44 (more posterior), BA45 (more anterior) or both. The winning model showed that the task modulated the ITG to BA45 connection, implying that BA45 supports explicit lexical decision making. Mechelli et al. (2005) conducted a DCM to establish how information is passed between the inferior frontal gyrus, fusiform gyrus and superior temporal gyrus. They found that

pseudoword reading was associated with a stronger connection from posterior fusiform gyrus to the dorsal premotor cortex (just posterior to the IFG). This may be indicative of a greater reliance on phonological processing in the dorsal premotor cortex while reading pseudowords.

The above DCM studies all use fMRI data which typically samples BOLD every two seconds. While this is sufficient data to fit a DCM, there is no possibility of establishing how causal structure within a network changes over time. EEG and MEG data do provide the temporal resolution to establish rapid changes in connectivity, however DCM for M/EEG has been less widely applied to date. Yvert et al. (2012) used DCM of ERPs to look at how effective connectivity within a fronto-temporal language network is modulated by phonological and semantic processing. Multiple pathways in language processing are established, in particular frontal areas are suggested to exert a top-down control on temporal regions. The timing of any influence of the IFG on the reading network was investigated by Woodhead et al. (2012) who conducted a MEG DCM study. The findings establish that the LIFG influences vOT in a top-down manner within 200ms of stimulus onset.

2.5.1 Summary

DCM provides a powerful mechanism for testing hypotheses about how neural regions interact with each other. In its application to VWR a number of important theories have been questioned. As proposed by several cognitive and anatomical theories of reading, the DCM literature has found evidence for a number of mechanisms for reading. These include number of posterior to anterior routes as well as evidence for top-down influence of regions such as the IFG. Finally, the Putamen has been implicated in reading out loud though again, this connection may be one of several possible mechanisms in the reading process.

Recently information about the timing of interactions between regions has been found using DCM on MEG data (Woodhead et al., 2012). Of particular interest in this result is support for the early influence of the IFG, within 200ms of stimulus onset. Similarly to the data-driven methods,

the majority of applications of DCM present a static view of the brain.

Most current applications of DCM are focused on testing hypotheses. VWR has a rich history in psychology and neuroscience and as such provides a wealth of potential hypotheses about how the brain processes words; however many of these posit multiple processing routes within the VWR network and suggest complex patterns of interaction, even before consideration of the timing and frequency content of such interactions. Given this level of complexity, it is hard to be certain that the selection of models given to DCM to test between will contain all possibilities. Moreover, for anything more than a simple network an exhaustive search of the model space can become computationally expensive if not impossible. (Lohmann et al., 2012)

Generative models can provide insight into the internal workings of a system that might be hidden from observations. Biologically plausible generative models describing how layers of neurons in cortex might interact provide the basis for the very influential DCM method. Currently formulated as a hypothesis testing framework, DCMs can provide evidence for the ability of a given model to explain observed data. The parameters of the optimal model can then be interrogated to provide some insight to how the neurons generating the data behave. Generative models have considerable potential to predict and explain very complex interactions in observed data, including cross-spectra and phase-coupling however this has been a very recent development and has not yet been applied to visual word recognition.

2.6 General Summary

There has been a over a century of research into the brain basis of visual word recognition, with contributions from a diverse range of fields such as neurology, psychology and computer science. In general there have been two broad approaches to uncovering how the system works. The cognitive symbolic approach seeks to break down word recognition into discrete but interactive subprocesses with a focus on the nature of how words are represented at different stages in word recognition. Models such as the LCD suggest a brain basis for this kind of explanation. In contrast, the PDP

style approach looks to explain word recognition from a simple set of principles. This approach is most apparent in PDP and BR computational models and focuses on the nature of the processes that must occur rather than the specifics of the representation used.

Evidence from the early visual processing suggests that visual word recognition is supported by a dynamic system with several parallel processing streams. Furthermore many regions within the frontal, temporal and parietal lobes are recruited rapidly and in parallel when reading a word. There are interactive models of visual word recognition that explain how sensory inputs might interact with higher level representations and predictions, however these tend to underspecify the nature of the representation of hte word form in cortex.

A growing literature is investigating the connectivity within the language network, however the results are often oversimplified and it is difficult to find an consensus. Moreover, despite the problems with localising cognitive processes to specific brain structures, much of the current literature still takes this approach. DCM takes a more principled stance in its modelling, building from simple structures to explain complex behaviours, however again the results to many applications of DCM boil down to an attempt to localise a cognitive process to a discrete connection. Moreover, current implementations of DCM are dependent on hypothesis testing. Though visual word recognition has a rich theoretical framework from which to draw hypotheses, little exploratory work has been done to map pattern of connectivity in an unbiased, data driven way.

This review has outlined evidence that visual word recognition is supported by a dynamic, distributed and highly interactive network of brain regions. Convergent evidence from abstract cognitive and computational models and neuroimaging imply that the connections between brain regions are highly important. Data driven analysis techniques allow for a detailed investigation into the the connectivity within the reading network and how it changes over time. This approach has the potential to identify the critical connections and characterise how they are modulated by information content of the written word.

2.7 Research Question

Given the evidence presented above, this thesis will specifically address several questions. Firstly, a set of analysis techniques which allow for robust and principled estimation of directional connectivity within a brain network are described. These methods are validated through simulations before application to brain data. The methods are applied to resting state brain dynamics on a timescale of seconds before the more rapid dynamics associated with visual word recognition are probed. This will extend the work of studies such as Kujala et al. (2007) by providing a temporally rich and fully multivariate description of information flow within the reading network. Once this method has been shown to be effective, further work will investigate how manipulations of the orthographic structure of a word form can modulate the strength and timing of these connections.

Chapter 3

Estimating Directed Functional Connectivity From Dynamics In An Embedding Space

3.1 Introduction

The human brain is a complex system with a dynamic spatio-temporal structure. Many consistent functional structures within the brain have been observed on the basis of functional neuroimaging data in recent years, however we still know relatively little about the dynamics of the system we are observing. A critical issue is that the internal state of the brain is not directly observable and as such we are limited to making inferences from external observations of its behaviour. Several approaches are available to tackle this problem. Dynamic causal modelling (DCM) makes use of an explicitly defined model of the neural generators of the observed response which is then parameterised by the data. Such models have a rich history in neuroscience and although they have biologically plausible neural generators these represent a considerable simplification of the neuronal assembled in the brain. Moreover the structure of neurons in gray matter is not uniform across the brain (Mesulam, 2000). An alternative approach based on state-space reconstruction allows the investigator to remain agnostic about the nature of the internal states generating our observations. This is achieved through a reconstruction of the internal state-dynamics directly from the observed data, under certain assumptions this has been shown to recreate the internal dynamics in a new coordinate space completely characterised from the data. This state-space approach is the basis for the analysis pipeline outlined within this chapter and is applied throughout this thesis.

3.2 Introduction To MEG

Magnetoencaphalography (MEG) is a non-invasive neuroimaging technique which measures the magnetic fields naturally produced from neuronal activity within the brain (Cohen, 1972). The fields are recorded from a distance of several centimetres from the scalp using a super conducting quantum interference device (SQUID). These SQUIDs are used to measure current in a pickup coil generated from electromagnetic flux. The SQUIDs are immersed in liquid helium at temperatures close to absolute zero in a cryogenic dewar,

which in turn, is place inside a magnetically shielded room. These steps are required to ensure that the SQUID is sensitive enough to record the tiny magnetic fluctuations arising from neuronal activity. This is challenging as the biological signal is several orders of magnitude smaller than signals caused from the environment or even the earth's magnetic field (Hämäläinen et al., 1993).

MEG has a very fine grained temporal resolution, being able to sample in the kilohertz range. However, the growth in the use of MEG in recent years has arisen from developments in source localisation techniques. Modern methods such as minimum norm estimation (Hämäläinen and Ilmoniemi, 1994; Vrba and Robinson, 2001) and beamforming (Hillebrand et al., 2005; Van Veen et al., 1997; Huang et al., 2004; Johnson et al., 2011) are able to offer a good spatial resolution in addition to the excellent temporal resolution. This makes MEG a good compromise between the detailed spatial resolution of functional magnetic resonance imaging (fMRI) and the rapid temporal resolution of EEG. The details of MEG function and source localisation algorithms are outside of the scope of this thesis, for a more detailed introduction see: (Vrba and Robinson, 2001; Baillet et al., 2001)

The methods outlined in this chapter begin with a set of virtual electrodes; these are virtual sensors at a specific location in the brain whose recordings are constructed from a weighted sum of the sensor recordings.

3.3 Dynamical Systems

A dynamical system is a set of rules which describe how a point in a coordinate space depends on time. We can consider the brain to be a dynamical system with three critical features; a state-space (S), set of times (T) and a rule (R) governing the evolution of the systems state ($s \in S$) over time with initial conditions s_0 at t_0 (Meiss, 2007). A system's internal state may not be directly observable, therefore an additional rule (O) is defined which relates the current state to the observable outputs of the system. The coordinates of a state in S describe the values of the internal components of the system at a given time which can be related to both the change in s over

time and the observable outputs of the system x by the following system of two equations:

$$\dot{s} = R[t_0, t, s_0, s_t, u_t] \tag{3.1a}$$

$$x = O[t, s_t] \tag{3.1b}$$

in which $t \in T$ determines the time over which the system evolves and u defines any inputs to the system. $\dot{s}$ refers to the differential of s, which is it's change over time. In the case that R and O are linear combinations of s and u we can rewrite the equations in matrix form

$$\dot{s} = Fs_t + Gu_t \tag{3.2a}$$

$$x = Hs_t \tag{3.2b}$$

Equation 3.2a is the state equation which is an explicit definition of R that governs how s changes over time. F is the system matrix which relates the current state to the subsequent change and G is the control matrix relating inputs (u) to change in state. Equation 3.2b is the output equation which determines the observable outputs of the equation to the s through the output matrix H. This is the output equation in the case that O is considered to be linear. Together these state-space equations form a model of the dynamical evolution of the hidden state of a system and its observable outputs.

Any system which can be described or approximated by a set of ordinary differential equations (ODEs) can be described using the state-space equations above. As such they provide a powerful means for describing a system governed by an unknown system of ODEs.

The state-space equations can be reformed for a discrete time system in which t is now an integer multiple of the sampling rate of observations through T.

$$s_{t+1} = Fs_t + Gu_t \tag{3.3a}$$

$$x = Hs_t \tag{3.3b}$$

In an application to brain dynamics, s is unobservable and moreover, its physical interpretation is unclear. In addition though u is likely to be related to physical stimuli and sensory input, its precise form is very difficult to quantify. As such, we are not able to directly solve the state equation as both s and the differential equations which govern its motion are unknown. For such cases there are methods for reconstructing the dynamics of s directly from observed data. A highly influential example is time-delay embedding and the method of delays.

3.4 Time Delay Embedding

A systems trajectory through S is determined through a system of unknown and unobservable ODEs, however several methods exist for reconstruction of these dynamics directly from the data.

Time delay embedding (TDE) is one such method for reconstructing the dynamics governed by the state equation in a system from a sequence of observations of the output. S is an d-dimensional manifold M or topological space defining all possible states of a system. Though this cannot be observed directly, a smooth and invertible mapping can be made from M to space with a different coordinate system (Q) (Broomhead and King, 1986). The advantage of such as step is that though M cannot be directly observed, Q can be defined such that the coordinate of s within Q can be determined directly from the observed data. As such the dynamics within M can be realised within the new and observable coordinate space Q. A d-dimensional manifold may be embedded within a new space of sufficient dimensionality $(2d + 1)$ (Whitney, 1936). This finding has led to several techniques for reconstructing state dynamics from time-series (Takens, 1981; Packard et al., 1980)

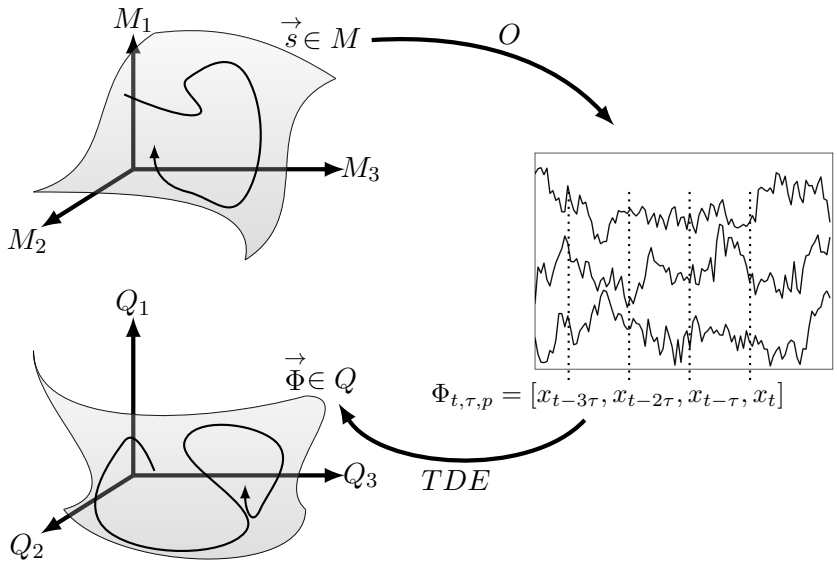

Figure 3.1: A schematic depicting the relation between the hidden state space S to the embedding space Q through a time-delay embedding.

The method of Takens (1981) states that an embedding Φ, which maps between M and Q, can be defined with coordinates from a sequence of observations x of the system over time. The critical parameters for this embedding are p - the dimensionality of the embedding and τ the time-delay between observations to include.

$$\Phi_{t,p,\tau} = \left[\mathbf{x}_t, \mathbf{x}_{(t-\tau)}, \mathbf{x}_{(t-2\tau)}, \ldots, \mathbf{x}_{(t-p\tau)}\right] \tag{3.4}$$

where $\Phi_{t,p,\tau}$ defines an embedding window, a vector containing p elements of the time series. If $\tau = 1$ these observations are consecutive, otherwise when τ is an integer multiple of the sampling frequency it defines the delay in between observations in the embedding (see figure 3.1 for a graphical depiction). When this embedding window is advanced stepwise along the time-series it yields a sequence of vectors ($\Phi_{t,p,\tau}$ as defined in equation 3.5) in the embedding space Q. Under the hypothesis/assumptions of

Takens (1981) this trajectory will maintain the dynamics of the evolution of s through M within the new coordinate space defined by the embedding. The transformation between M and the embedding space may be highly non-linear however a point-to-point mapping between the coordinate systems should be preserved.

$$\Phi_{t,p,\tau} = \left[\Phi_{t,p,\tau}, \Phi_{t+1,p,\tau}, \ldots, \Phi_{t,p,\tau}\right] \tag{3.5}$$

The vector in equation 3.5 can then be used as an approximation to the state equation in equation 3.3a. Finally the link between this approximation to the internal states and the system outputs can be quantified by substituting Φ into the output equation in equation 3.3b. This use of a Multivariate Autoregressive (MVAR) model to quantify this interaction is discussed in a later section.

$$x = H\Phi_{t,p,\tau} \tag{3.6}$$

3.4.1 State Space Reconstruction Parameters

We want to ensure that the space of past states to use in the reconstruction are optimally sampled for the prediction of future states, conditional on the avoidance of experimental noise. There are a range of methods for objectively establishing optimal values for the delay and dimensionality of the system (Ragwitz and Kantz, 2002; Cellucci et al., 2003) which have been applied to non-linear analyses of neuroimaging data (Stam, 2005). An inappropriate value for either p or τ can lead to a distorted representation of the dynamics of the observed system potentially leading to the detection of spurious dynamical relations or the omission of real ones. As such, a large literature on the selection of optimal parameters has arisen. The fact that the reconstruction parameters of critical importance to ensure a good representation of the underlying dynamics is very clear, however the methods for their selection often come with a health warning. Various methods have

been shown to provide good reconstructions however they are often based on rather arbitrary criteria. The performance of a given method depends on the data being analysed (Kugiumtzis, 1996). Moreover, it is not the case that there is 'correct' reconstruction. The goal of TDE is to establish an embedding which preserves the topological properties of the unknown underlying dynamical system. In a very complex system there will be many possible embeddings which approximate this objective. As such, we must approach the selection of the state space reconstruction parameters as an optimisation problem tuned to our specific needs. That is, 'which selection of p and τ gives the simplest embedding which allows for accurate prediction of future observations? '

In the case of the neural data in this thesis the final objective in the analysis is to establish the predictive power of a systems state on subsequent data observations. As such the methods for optimising the embedding of a data set will be utilised to optimise the prediction of future neural behaviour.

Estimating Time Delay (τ)

With observations of a system in the presence of experimental or environmental noise, the information between adjacent observations may become highly associated. This can mask the systems true trajectory through state space. Under the assumption that the noise will be locally correlated in time and does not vary systematically with the observed systems state, this confound can be minimised through the selection of an appropriate time-delay between observations to be included in the TDE.

In the analysis of chaotic time-series and non-linear attractors there are two widely accepted criteria for selecting an appropriate time-delay (Kugiumtzis, 1996):

- The reconstructed attractor must be expanded from the diagonal.
- The components of the delay vector must be uncorrelated.

The first of these criteria refers specifically to the unfolding of a chaotic attractor and becomes problematic to find a meaningful value where $p > 2$

(Kugiumtzis, 1996). As a neural data is likely to arise from a complex and potentially very high order system and any useable reconstruction is likely to have more than two dimensions, evidence from unfolding is not considered in this thesis.

The second criterion is more tractable for high order systems and is usually assessed using one of two methods. The first zero-crossing of the autocorrelation function (equation 3.7) of the observed data provides a simple metric for establishing τ where adjacent observations are decorrelated. The autocorrelation function $R_{t,\tau}$ can be calculated as:

$$R_{t,\tau} = \frac{E[(x_t - \mu)(x_{t-\tau} - \mu)]}{\sigma^2} \tag{3.7}$$

in which E denotes a sample expectation, μ is a sample mean and σ^2 is the sample variance.

While the autocorrelation function characterises linear associations the non-linear associations can be characterised by assessing the Mutual Information (MI) function (equation 3.8) as suggested by Fraser and Swinney (1986). These two approaches can yield different and possibly contradictory results. The autocorrelation function provides the simplest solution yet when empirically assessed, the optimal value for τ only equals the first zero crossing when the signal contains very little or no noise. The optimal value decreases monotonically with increasing noise until the point where a highly stochastic system will have an optimal delay of one sample. (Ragwitz and Kantz, 2002). As such the autocorrelation method may overestimate τ in noisy systems, in contrast, the plateau of the MI function tends to provide a value in a moderately noisy system in between these extremes. The mutual information between two series of observations x_1 and x_2 can be computed with:

$$MI(\tau) = \sum_{x_1,x_2} P(x_1, x_2) \log_2 \frac{P(x_1, x_2)}{P(x_1)P(x_2)} \tag{3.8}$$

in which $P(x_1)$ is the probability distribution of values in x_1, $P(x_1, x_2)$ is the

joint probability density function of both series and $P(x_1)P(x_2)$ the product of the two single probability density functions. It may be expected that in a highly noisy system, such as an externally observed brain response, the optimal delay will be very short (perhaps only a few samples). In practice, such a short delay might lead to a restrictive resolution in the spectral representation of the reconstruction particularly if the sampling rate is very high. This may present a problem for applications in which the spectral content of a system is very important.

Both of these metrics can be complicated by periodicity in the observed data which could lead to a case where x_t and $x_{t-\tau}$ are decorrelated but x_t and $x_{t-2\tau}$ show some association.

For anything beyond a very simple system it is unlikely that a value for τ that will satisfy each of the above criteria while avoiding all of the pitfalls. Moreover, the sampling rate is likely to have been determined by practical rather than *a priori* experimental reasons to optimise sampling of the system in question. In addition the validity and results from any TDE are specific to the dataset from which it was estimated, and as such might not provide useful heuristics for optimising state reconstructions in unseen data.

Estimating Dimensionality (p)

Typically the embedding dimension is chosen using a method such as False Nearest Neighbours (FNN) which establishes the optimal dimensionality as the point at which increasing dimensionality preserves the relative spatial relationships between the observations (Kennel et al., 1992). This procedure is highly computationally intensive, particularly for large datasets, as it involves the computation of the distance between each observation and all others in an increasingly high dimensional space. In addition it is explicitly designed to minimise intersections in the reconstructed attractor rather than optimise future predictions.

The criterion for selecting the optimal dimensionality for a system depends on the objective of the study. For instance, in the reconstruction of a

non-linear attractor we may prioritise the preservation of topological properties. In contrast, a study seeking to predict future states from past ones might optimise dimensionality as the value which allows the best next step prediction (Maus and Sprott, 2011). Such an estimate can be established empirically by finding the value of p which minimises prediction error or with a more sophisticated measure such as Akaike's Information Criterion. These approaches will be discussed in detail further on.

Time-Delay Embedding and Causality

The process of objectively time-delaying a set of observations naturally leads to the estimation of predictive causality such as G-Causality (Granger, 1969). G-Causality asserts that if the prediction error in signal X_1 from its past history can be reduced by the inclusion of the past of X_2, then X_2 can be said to G-Cause X_1. This is typically assessed with methods such as MVAR models which fit parameters indicating the predictive power of a point in a TDE on subsequent data points. In the context of the state-equations, such a model can be seen to parameterise the predictive power of each dimension in the embedding space on the observed outputs. This pipeline represents an objective, data-driven framework for estimating predictive causality within a brain network, in contrast to model based methods such as Structural Equation Modelling and Dynamic Causal Modelling which require prior statements about causal structure.

3.5 MultiVariate Autoregressive Modelling

The output equation defined in equation 3.3b relates a system's inferred state to it's outputs. This can be reformulated such that the model parameterises the predictive strength of each dimension in the embedding space on future observations. In other words, which points in a systems state are most predictive of future outputs.

3.5.1 Formulation

Statistical dependencies between multivariate time series such as those described in equation 3.6 can be described with a MultiVariate AutoRegressive (MVAR) model. Each vector in the TDE includes a subset of past history points, and so a subset MVAR model can be used to parameterise the predictive power of each point onto a subsequent value. The predicted value of an output $\hat{x}$ (where the hat denotes a prediction/estimate) can be estimated with an MVAR of the form:

$$\hat{\mathbf{x}}_t = -\sum_{j=1}^{p} A_j \mathbf{x}_{(t-j\tau)} \tag{3.9}$$

A vector $\mathbf{x}$ containing the samples from n locations at time t is modelled as a linear combination of p past history values of the n time series. A_j is a n x n MVAR coefficients matrix at lag j in which a_{1k} represents the interaction/influence of $x_j(t-k)$ has on $x_1(t)$. τ represents the delay between past history values used as predictors in the model. In this form A is a fitted output matrix H and values from $x_{t-j\tau}$ can be seen to be vectors within $\Phi_{t,p,\tau}$. MVAR equations have been successfully applied to many areas of science and engineering and are very well understood - as such formulating the output equation as an MVAR allows us to draw on this highly successful methodology in subsequent stages of the analysis.

The residual error from the model fit can be determined with:

$$\mathbf{e}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t \tag{3.10}$$

given the residual prediction error after model fitting. Finally the error covariance is:

$$\Sigma = \frac{1}{t-1} \sum_{t} (\mathbf{e}_t - \hat{\mathbf{e}}_t)(\mathbf{e}_t - \hat{\mathbf{e}}_t)^T \tag{3.11}$$

3.5.2 Estimating A

Many methods for fitting the parameter matrix A to an observed dataset have been developed. These include the Normal Equations and the Yule-Walker equations. This thesis uses the Vieira-Morf (VM) algorithm, which is a modification of the multivariate Yule-Walker method which has been shown to perform well (Schlögl, 2006). The VM algorithm has been implemented in several toolboxes including the Time Series Analysis toolbox in BioSig (http://biosig-consulting.com/matlab/tsa/) and the Source Information Flow Toolbox (http://sccn.ucsd.edu/wiki/SIFT). This thesis uses an implementation developed in Python in conjunction with the Neuroimaging Analysis Framework (http://vcs.ynic.york.ac.uk/docs/naf/).

The Vieira-Morf algorithm is an implementation of the Levinson-Durbin recursion (Levinson, 1946; Durbin, 1960) which uses an estimate of the partial correlation matrix based on the available data (Marple, 1987, chapter 15). This is the critical step in the model fit recursion, if we are not able to get a good estimate of the lagged correlation matrix then the parameters will not be well estimated either. This places a limitation on the amount of data observations that must be used to ensure a good model fit. This issue is explored in more detail in chapter 3.

The Vieira-Morf algorithm estimates the parameters based on the partial correlation matrix computed between channels of observations at increasing lags. It is a recursive algorithm which computes the parameters at order p based on the residuals of the model at $p-1$ before updating all previous coefficients depending those just computed. See appendix A for full derivation of this algorithm.

3.6 Model Validation

3.6.1 Stability/Stationarity

MVAR modelling assumes that the time-series being modelled is a stationary process i.e. that its mean and variance do not change as a function of time. This assumption may be tested by way of the Perron-Frobenius theorem

(Perron, 1907; Frobenius, 1912) which states that for a real-valued square matrix A, there exists an eigenvalue r which is strictly larger than any other eigenvalue associated with A. The magnitude of r is closely linked to the amount of growth or decay in the system, in that where $-1 > r$ or $r > 1$ the system will grow or decay over time, reaching values of $\pm\infty$ in the limit, in contrast where $-1 < r < 1$ the system can be regarded as stable.

This has been formulated as a formal stability condition (Lütkepohl, 1991) which states that the reverse characteristic polynomial must have no roots in or on the unit circle for a system to be stable. A simpler test based on this is that the modulus of the largest eigenvalue of the A_1 matrix should be less than 1. This indicates that we have a stable solution which will never reach plus or minus infinity. This condition implies stationarity, a critical assumption for the fitting of an MVAR model.

$$|r| < 1 \tag{3.12}$$

3.6.2 Durbin Watson

The Durbin-Watson statistic (Durbin and Watson, 1950, 1951) tests for serial autocorrelation in the model residuals. Though serial autocorrelation in the residuals will not affect the consistency of the estimated parameters it may lead to an under or over estimation of the standard errors, biasing any subsequent statistical tests.

$$DW = \frac{\sum_{t=2}^{N}(e_t - e_{t-1})^2}{\sum_{t=1}^{N} e_t^2} \tag{3.13}$$

in which N denotes the total number of observations.

As a consequence of using a subset autoregressive model which sparsely samples the past history of the observed system, the model is explicitly not parameterising part of the autocorrelation in the data. This is a design choice with the implication that we are not interested in parts of the auto-

correlation function, which may lead to the Durbin-Watson test suggesting that this unmodelled correlation in the residuals is characteristic of a poor model. To correct for this we can change the delay in the DW to our value for τ.

$$DW(\tau) = \frac{\sum_{t=\tau}^{N}(e_t - e_{t-\tau})^2}{\sum_{t=1}^{N} e_t^2} \tag{3.14}$$

3.6.3 R^2

This is the coefficient of determination, which is the square of the Pearson product-moment correlation coefficient. It tells us how much of the variance in the data we are explaining. This is taken as the scaled ratio between the explained and unexplained variance.

$$R^2 = 100 * \left(1 - \frac{var(resid)}{var(data)}\right) \tag{3.15}$$

3.6.4 Percent Consistency

Percent Consistency (Ding et al., 2000) assesses the amount of the auto and cross correlation structure in the data captured by the model. This is expressed by comparing the the auto/cross correlations in the real data and the model's fit.

$$PC(t) = 100 * \left(1 - \frac{|\mathbf{R}_{fit} - \mathbf{R}_{real}|}{|\mathbf{R}_{real}|}\right) \tag{3.16}$$

Where $\mathbf{R}_{fit}$ refers to the correlation vector of the model's fit to the data and $\mathbf{R}_{real}$ the correlation vector of the real data. Similarly to the Dubin-Watson test this metric might be biased when applied to subset MVAR models as it is testing the entire autocorrelation series when it is only sparsely modelled.

3.7 Model Selection

3.7.1 Multivariate Likelihood Estimation

The likelihood of a set of model parameters gives the probability that the parameters explain observed data. More formally, the likelihood function of a set of model parameters (θ) given a set of observed outcomes (x) is equal to the probability of outcomes given the model parameters.

$$L(\theta|x) = p(x|\theta) \tag{3.17}$$

The multivariate likelihood function is:

$$L(\mathbf{y}|\mathbf{x}, \mathbf{b}, \Sigma) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} e^{(-\frac{1}{2}(\hat{\mathbf{x}}'\Sigma^{-1}\hat{\mathbf{x}})} \tag{3.18}$$

In which x' denotes the transpose of x. This computes the probability of each point within the dataset on a multidimensional Gaussian distribution. The likelihood of the dataset is then the product of the probability of each point. For large datasets this value can be below numerical precision as the function involves the product of many values bounded between 0 and 1. As such, the log likelihood function is often used (See Appendix B for derivation).

$$ln(L) = -\frac{N^2}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \sum_{i=1}^{N}(\frac{1}{2}(\hat{\mathbf{x}}'\Sigma^{-1}\hat{\mathbf{x}})) \tag{3.19}$$

When comparing the likelihood of several models we can ignore the first term which is constant across models with the same number of observations of the dataset (N).

$$ln(L) \propto -\frac{N}{2}\ln(|\Sigma|) - \sum_{i=1}^{N}(\frac{1}{2}(\hat{\mathbf{x}}'\Sigma^{-1}\hat{\mathbf{x}})) \tag{3.20}$$

A Likelihood can provide a method for evaluating the fit of a single model, however it is often the case that we will want to compare multiple models and identify the optimal one.

3.7.2 Kullback-Leibler Divergence

The Kullback-Leibler (KL) divergence or information provides a directed measure of the discrepancy between two probability distributions (Kullback and Leibler, 1951). It provides a measure of the information lost when one function is used to approximate another, or in application to model fitting, the divergence between the distributions of the residuals from different models. The K-L information between two functions f_1 and f_2 applied to set of observations x can be written as:

$$I_{KL}(1:2) = \int_{-\infty}^{\infty} f_1(x) ln\left(\frac{f_1(x)}{f_2(x)}\right) dx \tag{3.21}$$

note that $I_{KL}(1:2) \neq I_{KL}(2:1)$ and as such the K-L divergence is not strictly a measure of distance, rather it is a directed discrepancy. Equation 3.21 is valid in cases in which the value of $f(x)$ comes from a continuous distribution, though I_{KL} can also be defined for discrete distributions (Burnham and Anderson, 2010).

A logical, if implausible, method for identifying an ideal model would be to compare each candidate to the absolute 'truth'. If we consider $f_0(x)$ to denote a 'true' model which produces the observed data and $f_n(x|\theta)$ where $n = [1, 2, \ldots, n]$ to be a set of candidate models dependent on parameters θ fitted to the observed data, this comparison could be written as:

$$I_{KL}(0:n) = \int_{-\infty}^{\infty} f_0(x) ln\left(\frac{f_0(x)}{f_n(x|\theta)}\right) dx \tag{3.22}$$

This would provide a measure of the extent to which the information in the true model f_0 is preserved within the model. The optimal approximating model would then be the one with the smallest K-L divergence. This can

be further simplified by rewriting equation 3.22 as a subtraction.

$$I_{KL}(0:n) = \int_{-\infty}^{\infty} f_0(x) ln(f_0(x)) dx - \int_{-\infty}^{\infty} f_0(x) ln(f_n(x|\theta)) dx \quad (3.23)$$

It is typically the case that θ is unknown and has been estimated from the data (let $\hat{\theta}$ denote an estimate). This introduces uncertainty into the estimation of I_{KL}. As such, we should deal with the expectations of the right hand side terms. The expectation operator is a weighted average of a set of observations, in this case the weights are taken from the probability distribution of f_0.

$$I_{KL}(0:n) = E_0[ln(f_0(x))] - E_0[ln(f_n(x|\hat{\theta})] \quad (3.24)$$

in which E_0 denotes the expectation operator with respect to the distribution of f_0. The first term in equation 3.24 depends on the unknown ideal distribution alone and will be constant across a comparison between several approximating models. This can be dropped to allow for a measure of relative divergence.

$$I_{KL}(0:n) - constant = -E_0[ln(f_n(x|\hat{\theta})] \quad (3.25)$$

The right hand side of this equation quantifies the goodness of fit of $f_n(x|\hat{\theta})$ as an estimator of $f_0(x)$. The goodness of the estimation procedure can then be measured by

$$E_n E_0[ln(f_n(x|\hat{\theta})] \quad (3.26)$$

which is the expected log likelihood of the model with respect to the absolute truth (Akaike, 1978). This value is critical to model selection approaches based on the KL divergence (Burnham and Anderson, 2010), however it cannot be estimated without knowledge of the unknown true distribution

$f_0(x)$

3.7.3 Akaike's Information Criterion

Akaike 1973 showed that the maximised log likelihood of the model in question is an upwardly biased estimator of the critical value in equation 3.26, and with certain assumptions the difference between the two is around k, the number of freely estimable parameters in the model. Therefore an unbiased estimator of the relative expected K-L divergence is:

$$ln(L(\hat{\theta}|x)) - k \tag{3.27}$$

This relation between the log likelihood of a model and the relative expected K-L divergence allows for an estimation of the quality of a model without an unknown 'ideal' model from just the number of parameters and the log likelihood of the fitted model, both of which are relatively straightforward to obtain.

For "historical reasons" An Information Criterion (AIC) was then defined by multiplying by -2 (Burnham and Anderson, 2010)

$$AIC = -2ln(L(\hat{\theta}|x)) + 2k \tag{3.28}$$

thus the model with the lowest AIC can be taken as optimal as it is thought to be closest (in terms of the K-L divergence) to some unknown 'ideal' model which actually generated the data.

Interpretation

There are two terms in the AIC calculation. The first term containing L provides a measure of how well a model fits the data. The second term can be seen as a complexity penalty for the model. It is likely that a more complex model will be more able to fit the data, however increasingly complex models will begin to fit unsystematic variation in the data. The second term seeks to prevent this over fitting by penalising increasingly complex values.

The optimal model will be the one with the lowest AIC estimate. Taking the two terms as *accuracy* and *complexity* respectively, the AIC becomes:

$$AIC = -accuracy + complexity \tag{3.29}$$

In this form, it can be seen that the lowest AIC value will arise from the model which is most accurate without being too complex.

In terms of information theory and the K-L divergence the AIC is the estimated expected relative K-L divergence. It is relative in the sense that we are computing the divergence between a model and some unknown generating mechanism, when done multiple times the first term of equation 3.23 drops out as a constant and as such the AIC can be calculated without reference to the unknown generating mechanism.

3.8 Connectivity Measures

Several informative metrics can be estimated from the fitted A matrix once the optimal model from a set has been identified, fitted and validated. The spectral content of the model is of particular interest in neural data. The coefficients matrix A_j of a fitted MVAR model can be expressed in the frequency domain through a discrete time Fourier transform.

$$A(f) = \sum_{n=1}^{p} A_i e^{-ifn} \tag{3.30}$$

in which $i = \sqrt{-1}$. $A(f)$ is then a [signals by signals by frequency] matrix. $A(i, j, f)$ then contains the amount of variance in each vector $A(i, j)$ which can be found in the spectral component at frequency f. Where $i \equiv j$ this is the power spectral density and where $i \neq j$ it contains the cross spectral density. Critically, the cross spectral density contains the extent to which the parameters which predict i from the history of j oscillate at frequency f. The inverse of this A matrix give the transfer function of the system.

$$T(f) = A(f)^{-1} \tag{3.31}$$

The power spectrum of the system can be calculated from the transfer function and the covariance of the model's residuals (Priestley, 1981; Sameshima and Baccalá, 1999).

$$PSD(f) = T(f)\Sigma T(f) \tag{3.32}$$

in which Σ is the residual covariance matrix as defined in equation 3.11.

3.8.1 Coherence

The coherency at a frequency (f) can be estimated from this power spectrum of the coefficients.

$$C_{ij}(f) = \frac{PSD_{ij}(f)}{PSD_{ii}(f)PSD_{jj}(f)} \tag{3.33}$$

This coherency is a complex valued function from which a magnitude and phase can be derived. A more commonly used metric is the real-valued magnitude squared coherence.

$$MSC_{ij}(f) = \frac{|PSD_{ij}(f)|^2}{PSD_{ii}(f)PSD_{jj}(f)} \tag{3.34}$$

This yields a real value between 0 and 1 where 0 indicates no association and 1 is a perfect association. The MSC is undirected, in that $MSC_{ij} \equiv MSC_{ji}$ and unable to distinguish direct connections from spurious connection generated by a common third source.

3.8.2 Directed Transfer Function

Directed associations between channels in the system can be calculated from the Directed Transfer Function.

$$DTF_{ij}(f) = \frac{|T_{ij}(f)|^2}{\sum_{k=1}^{n} |T_{ik}(f)|^2} \tag{3.35}$$

The DTF is able to resolve directionality ($DTF_{ij} \not\equiv DTF_{ji}$) however due to the mixing of information between channels when taking the inverse of the A matrix it is still unable to differentiate between direct and indirect connections.

3.8.3 Partial Directed Coherence

The $A(f)$ matrix can also be used to compute the Partial Directed Coherence (PDC)(Sameshima and Baccalá, 1999; Baccalá and Sameshima, 2001) which provides a frequency domain measure of the directed influence each time series has on the others.

$$PDC_{ij}(f) \triangleq \frac{A_{ij}(f)}{\sqrt{\sum_{k=1}^{n} A_{kj}^{H}(f) A_{kj}(f)}} \tag{3.36}$$

PDC describes the relative strength of information flow between a source and target node compared to the total influence of the source node on all other nodes.

3.9 Statistics

Once the desired metric for characterising network structure has been computed from our fitted model, the statistical significance of it's findings must be considered. This section outlines several methods for assessing significance both within a single epoch and for a contrast between two epochs. This section only considers the confidence of a PDC as this will be the focus of the analyses within this thesis.

3.9.1 Single Epoch

There are two methods for estimating confidence intervals from a PDC estimate from one epoch. The first is a method based on non-parametric permutations at the level of the time series and the second is an analytic method. Critically, both of these methods test against the null hypothesis that there is no PDC within a connection.

Permutations

The random permutations technique first builds a surrogate or null distribution of PDC estimates from time-series with the observations randomly scrambled in time (Kaminski et al., 2001; Florin et al., 2011). For a delayed parameter within an MVAR model to significantly predict future behaviour, there must exist some non-zero auto or cross correlation at the time delay in question. As a consequence, a non-zero PDC estimate is also critically dependent on the whole auto/cross correlation having non-zero spectral power at a given frequency. Scrambling the experimental observations in time will hopefully remove these delayed correlations from the data, therefore any MVAR or PDC estimate made from this scrambled data can be attributed to chance alone.

The data are scrambled a large number of times (typically > 500) and the PDC from each scrambled sample added to the null distribution. The experimental estimate can then be compared to this null distribution and declared significant if it is larger than the $1 - \alpha$ percentile.

Though effective, this procedure can be very computationally expensive. Each iteration in building the null distribution restarts the analysis procedure from before the MVAR model fitting. In cases where a confidence limit is required for a single PDC estimate (with no condition or time window contrasts) the methods for estimating analytic confidence intervals defined in (Schelter et al., 2006) is used. This procedure is described in the appendices.

3.9.2 Contrasts

It will often be the case that differences between PDC estimates in different windows or conditions will be the metric of interest rather than it's absolute value at one time point. There are no analytical methods for estimating the significance of a difference between PDC estimates in two conditions, however the question is well suited for a permutation approach.

We take the null hypothesis to be that there is no difference between the distributions of PDC estimated from two conditions λ_1 and λ_2. Prior to collecting any data, certain parts of our recording over time will be labelled as one of these conditions. If the null hypothesis is true, these labels are considered exchangeable. This means that the distribution of data under a given sequence of labels will be equivalent to the distribution from any other sequence of labels.

A difference between conditions defined in our experimental label sequence can then be considered significant if the distributions it produces are sufficiently different to the distributions from a random label sequence. A null distribution is built by randomly swapping the labels assigned to each time-point and calculating a t-statistic based on the difference between the distributions the randomised labels produce. Once complete the observed t-statistic from our experimental label sequence is compared to the null distribution and can be considered significant if it is larger than the $1 - \alpha$ percentile of the null distribution.

Multiple Comparisons

When computing differences between PDC in different conditions, we are computing the same metric across many connections and frequencies. The total number of comparisons is $m^2 f$ in which m is the number of channels and as such we may expect several if not many of our comparisons to appear significant by chance. This problem can be alleviated by taking a maximum statistic approach when creating the null distribution.

The critical issue is that each shuffling of the condition labels is applied to the entire network rather than just a single estimate in one connection at

one frequency. The null hypothesis at the level of an entire network is the omnibus hypothesis, that there is no experimental effect anywhere in the network. This can be accepted or rejected if the maximum statistic within the whole network is less than or greater than a critical threshold. The critical threshold is then the value at which the probability that the maximal statistic within the whole network exceeds it is $1 - \alpha$. As a consequence, we need a distribution of the maximal statistics within the whole network across our permutations (Nichols and Holmes, 2002).

This may be realised by taking PDC estimates across the whole network for each epoch in the experiment, scrambling the condition labels, computing the test-statistic and add the maximum test-statistic from the whole network to a null distribution. These steps are then repeated many times and the experimental test-statistic compared to the null distribution.

3.10 Summary & Conclusion

This chapter has outlined an analysis pipeline designed to accurately reconstruct brain dynamics from a set of observed responses. Such a dynamical system may be modelled with the state-space equations which relate the current state of a system with change in state and the observed responses. A critical issue is that the state transitions which characterise a system may not be directly observable. Delay embedding methods provide a means for approximating the dynamics of a system's state from the observed data alone. This method requires the somewhat subjective selection of two key parameters (p and τ), however with the objective that our reconstruction of the systems state must maximise our prediction of future observations we may obtain a good approximation. A multivariate autoregressive model may the be used to directly parameterise the predictive power of the coordinates of each state on future observations. Subsequently these fitted parameters can be used to calculate a metric such as partial directed coherence which characterises the causal structure within a network in the spectral domain. Finally, these estimates are subjected to significance testing.

Overall, this pipeline represents a theoretically justified method for the

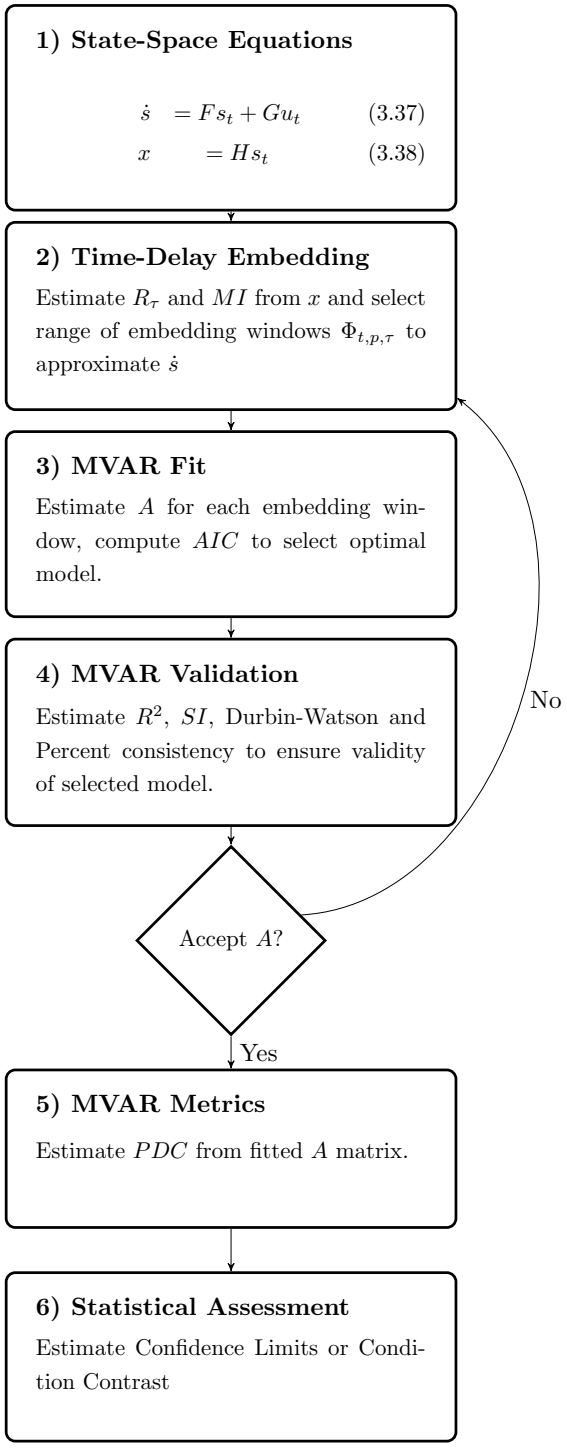

Figure 3.2: A schematic representing the analysis pipeline outlined by this chapter. The decision whether A is either accepted or not is made based on the MVAR validation metrics described in step 4. These are objective measure of the quality of the model and independent of any experimental outcomes.

interrogation of the brain's hidden states and its relation to observed responses and behaviour. No assumptions about the nature or dynamics of the hidden states need to be made and the critical parameters may be directly estimated from the data.

3.11 Notation

Most of the mathematical notation in this chapter is defined in the surrounding text, this section is intended as a quick reference.

3.11.1 General

i	$\sqrt{-1}$
I	Identity matrix
$\|a\|$	The determinant of a matrix a

3.11.2 Dynamical Systems

S	A state-space containing all possible states
s	A single state defining a systems internal parts where $s \in S$
T	A set of times over which a state evolves
t	A single time-point, where $t \in T$
R	A rule governing how state evolves through S over T
u	The input to a system
x	An observed outcome
O	An output rule governing how s relates to x
F	System matrix
G	Control matrix
H	Output matrix

3.11.3 Time-Delay Embedding

M	A manifold on which S evolves
Q	A diffeomorphic space to M
d	A space's dimensionality
p	Model order
τ	The time-delay between samples in an embedding
Φ	A delay-embedding

μ	A sample mean
σ^2	Sample variance
$P(x)$	The probability of x
$P(x\|y)$	The probability of x given y

3.11.4 Multivariate Autoregressive Modelling

$\hat{x}$	An estimate of x
A	Matrix of forward MVAR parameters
B	Matrix of backward MVAR parameters
e	Model prediction error
Σ	The covariance of the model prediction errors
$\mathbf{P}_p^f$	Forward lagged covariance matrix at order p
$\mathbf{P}_p^b$	Backward lagged covariance matrix at order p
$\hat{\Lambda}_{p+1}$	Estimate of partial correlation matrix

3.11.5 Model Validation

r	The largest eigenvalue of a square matrix as defined by the Perron-Frobenius theorem
N	The number of observations of a system
m	The number of channels of observations of a system
R	Pearson's product moment correlation coefficient
R^2	The coefficient of determination
$\mathbf{R}$	A vector containing the auto and cross correlations in a system

3.11.6 Model Selection

$L(x)$	The likelihood of x
$L(x\|y)$	The likelihood of x given y
θ	A set of parameters
$\hat{\theta}$	A estimation of a set of parameters
f_0	A hypothetical 'true' model which produced the observed observations
f_n	An approximating model

k	The number of parameters in a model
E_n	Expectation operator with respect to F_n
ln	Natural logarithm

3.11.7 Metrics

T	The transfer matrix of a system
PSD	The power spectral matrix of a system
C	Coherency
MSC	Magnitude squared coherence
DTF	Directed transfer function
PDC	Partial directed coherence

Chapter 4

Realistic Stimulations Of Causal Brain Networks

4.1 Introduction

The previous chapter outlined one method for constructing a model of brain interactions. This chapter more directly assesses the performance of this modelling approach through simulating dynamical systems as defined by known systems. Through this process we can establish the validity of our modelling approach and gain an insight into the practical issues associated with the data analysis.

There are three key features of a connectivity analysis which are interrogated with simulations: 1. Directionality 2. Direct/Indirect connections 3. Spectral Resolution. Directionality is the identification of asymmetries between the possible directions of information flow through a connection i.e. the ability to disambiguate between $x_1 \rightarrow x_2$, $x_1 \leftarrow x_2$ and $x_1 \leftrightarrow x_2$. Directness is the ability to identify when a connection between two nodes is entirely mediated through a third party. For example, if x_1 influences both x_2 and x_3, a direct measure will be able to correctly show that there is no connection between x_2 and x_3. Finally the spectral response is the extent to which a metric can identify the frequencies at which a connection might be occurring.

4.1.1 Simulation In Connectivity Research

As in the previous chapter, the dynamics of a brain network are characterised as a trajectory through an unobservable state space S. This trajectory is governed by a rule R which determined the change of $s \in S$ over a set of times T. These dynamics are then reconstructed with a time delay embedding Φ in an embedding space U which can be established directly from the observed data. Φ is then an approximation of the internal rule R through a coordinate space established from the observations.

$$\Phi \approx \dot{s} = R \tag{4.1}$$

Observations (x) from the generating state space are determined by the

output or observer matrix H.

$$x = HR \tag{4.2}$$

There is also an observer for Φ which relates the internal states to observed outcomes, however as Φ is based in a different coordinate system to H, a different observer may be required. This observer A may be related to H however it is not necessarily the same. Moreover A may often be unknown and estimated from the data, estimates of x and A will be denoted $\hat{x}$ and $\hat{A}$.

$$\hat{x} = \hat{A}\Phi \tag{4.3}$$

It is this relationship which is simulated as both the outcome on the left and the approximation of the internal dynamics on the right are formed from observations of the system and do not require knowledge of the internal states of the system. Once simulated, the relationship between the approximation of the current state in the embedding space and forthcoming observations are parameterised with a multivariate autoregressive model. Finally, several metrics for describing the interactions between nodes in the brain network are simulated form the fitted MVAR parameters.

The fitted $\hat{A}$ matrix from an MVAR model can be used to estimate several connectivity metrics as outlined at the end of the previous chapter. These differ in their ability to detect frequency, directness and directionality as outlined above, these features of Magnitude Squared Coherence, Directed Transfer Function and Partial Directed Coherence are investigated in this chapter.

4.1.2 Simulation Methods

In neuroimaging applications, there are three main approaches employed to generate time series with a known relationship between the past and future observations, equivalent to a known H matrix in equation 4.3. 1. Delayed Signals 2. Time-Domain Difference Equations 3. Time-Domain Differential Equations. Many other methods including Volterra and Wiener kernels may

also be used, however these are not frequently used in simulations of linear dynamics.

Delayed Signals

The delayed signals method defined in Kus et al. (2006) utilises a real M/EEG signal to ensure that the signal is similar to real situations. The interactions between different nodes are generated in steps in which the previous signal is delayed by one sample and adds more noise. This propagates information from the original time-series through a network of time-series at a set of known time delays as illustrated in figure 4.1.

This scheme has been widely applied through the multivariate signal processing literature (Blinowska, 2011). The use of a real electrophysiological recording ensures a realistic signal however there are disadvantages. Firstly the delayed interactions are specified in the time-domain with the delayed signal propagating through the system, as such the spectral content of the interactions is difficult to control. In addition this method is not generative in that the number of observations in the system is strictly limited to the number of observations in the initial M/EEG signal. Finally, it is difficult to add feedback or reciprocal connections into a network with this method.

Difference Equations

Many simulations explicitly define a set of MVAR parameters and driving noise sources. The simulated signal is then established by multiplying the driving noise by the known MVAR parameters. This method has been widely used across the literature, examples including the original validation of Partial Directed Coherence (Baccalá and Sameshima, 2001), in application to short-window spectral analysis of ERPs (Ding et al., 2000), investigation into the effect of signal-to-noise and number of epochs on PDC/DTF estimation (Astolfi et al., 2008) and in the derivation of confidence limits for PDC (Schelter et al., 2006).

Though effective, this method does not replicate many of the features of the neural responses as observed with EEG or MEG recordings. Firstly,

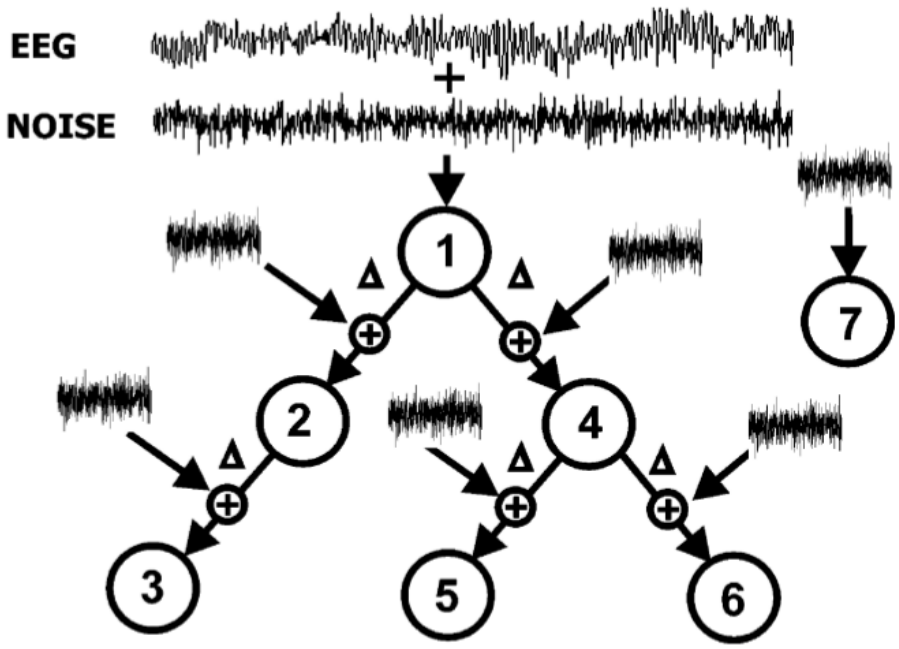

Figure 4.1: The simulation methodology used in Kus et al. (2006). The signal in channel 1 is generated through the addition of white noise to an EEG signal recorded from a scalp electrode. Signals $2-6$ are generated by sequential delays of this signal (where Δ denotes a time delay of one sample) with additive noise Kus et al. (2006).

the driving noise is typically white (flat spectrum), however the nature of any driving noise in the brain is unknown. The observed outputs from EEG/MEG signals typically have a $\frac{1}{f}$ spectrum, though this may arise from the driving noise or the autoregressive coefficients themselves. In addition, as the MVAR signal is defined in the time domain, it is very difficult to create an interaction between channels in a specific frequency, though there is a spectral response.

Differential Equations

More complex coupled systems such as coupled oscillators have also been used for simulating neural systems. Single oscillators have a cyclical output with a single degree of freedom, however more complex cases in which the behaviour of each oscillator can influence the others. The coupling between these nodes is often highly non-linear and many very complex systems can be described in this way. Such oscillators are defined using a series of simultaneous differential equations.

When fitting to a linear system, an autoregressive model can out perform a more complex non-linear model due to it's higher statistical robustness, however in a non-linear system it may make significant errors in predicting non-linearities in the observations (Ragwitz and Kantz, 2002). As these systems can be highly complex, a part of their interactions may not be linear and as such, not approximated with an MVAR model. The linear part of the interactions is visible to MVAR though the proportion of interactions that are linear will vary considerably from system to system.

4.1.3 MVAR Connectivity Metrics

A wide variety of connectivity estimators based on the decomposition of a fitted multivariate autoregressive model have been developed in recent years. The previous chapter introduces the mathematical form of several of these metrics while the more practical questions of their application and reliability are discussed here.

Overview

The properties of four commonly used connectivity metrics are summarised in table 4.1. Perhaps the most commonly used connectivity metric is the coherence and the related Magnitude Squared Coherence (Nolte et al., 2004; Faes et al., 2012). The complex valued-coherency is the normalised cross spectral density of two signals, the squared magnitude of which is the real valued Magnitude Squared Coherence (MSC). As outlined in table 4.1, though the MSC is a frequency domain measure which is able to locate and discriminate between interactions at different frequencies, it is not able to reject indirection connections or identify dominant directionality. This first criticism is alleviated with the partial coherence (pCOH) which estimates the coherence between two time series factoring out the influence of one or more other channels in the system. This value has directness, in that the pCOH characterises the unique coherence between two time-series that cannot be accounted for by influence from other mediating channels.

The Directed Transfer Function (DTF) is a spectral measure which is able to estimated directionality within connections, however it not able to disambiguate between direct and indirect connections (Kaminski and Blinowska, 1991; Kaminski et al., 2001). Finally the Partial Directed Coherence (PDC) is able to identify only direct and directional interactions in the frequency domain from a fitted MVAR parameter matrix (Baccalá and Sameshima, 2001).

Many reviews of these metrics have been published (Astolfi et al., 2007; Blinowska, 2011; Florin et al., 2011) and in general, there is agreement about many of their general properties illustrated in table 4.1.

Astolfi et al. (2006) introduce a modification to the PDC, the squared PDC (sPDC) which provides a better description of network interactions due to a greater weighting of the larger PDC estimates. They go on to show that both the PDC and sPDC are able to accurately reconstruct known connectivity patterns in a simulated multivariate dataset, provided with a sufficiently long sample recording. The recommended number of samples was 6750 when fitting an MVAR model on 5 channels with order 10. This

Metric	Abbreviation	Spectral	Directness	Directional
Magnitude Squared Coherence	MSC	✓	×	×
Partial Coherence	pCoh	✓	✓	×
Directed Transfer Function	DTF	✓	×	✓
Partial Directed Coherence	PDC	✓	✓	✓

Table 4.1: Review of MVAR connectivity estimators and their properties. *Spectral* indicates whether the metrics operates in the time or frequency domain, *Directness* indicates whether a metric can partial out common influence on two independent nodes from a third source and *Directional* indicates whether the metrics produces symmetrical estimates for each direction of a connection.

corresponds to a ratio of 27/1 samples to parameters fitted in the model. Astolfi et al. (2007) replicate this result with DTF and dDTF, reinforcing the need for a high number of samples relative to the number of parameters to be fitted in the model. Blinowska (2011) provide further evidence that the DTF and PDC are able to reconstruct known connectivity patterns using the delayed signals method of data simulation. They further state that as the directionality in these metrics arise from the phase differences in between the time-series they are very robust in respect to noise.

The review by Florin et al. (2011) investigated the reliability of each of these measures with different data lengths, noise levels, connections strengths and model orders. It was found that a variant of the PDC Astolfi et al. (2006) used was most robust across these tests while both the DTF and dDTF were unable to distinguish direct from indirect connections. Moreover this validation was performed using both the Delayed Signals and Defined Equations methods of data simulation.

Relation To Driving Sources And G-Causality

A test for causality or driving influence between signals in a multivariate time-series can be defined based on the partial Coherences within the network (Gersch and Goddard, 1970; Gersch, 1972). This concept, known as Gersch-Driving, suggests that a time series x can be said to be causal relative

to time series y and z if three conditions are met.

- There is a non-zero ordinary coherence between each pair of time-series at a given frequency f
- The partialed coherence in between y and z at frequency f is zero.
- The partialed coherence in-between x and y and x and z at frequency f is non-zero.

A recent assessment for this method has suggested that it may be highly sensitive to signal-to-noise ratio in that the signal within a system with the highest noise may often be found to 'drive' the other nodes (Albo et al., 2004).

The notion of causality rests on cause preceding effect, however the concept of Gersch-Driving as outlined above is insensitive to temporal precedence in time-series interactions and as such cannot be considered a true measure of causality (Baccala and Sameshima, 2006). In contrast, a definition of causal interactions within a system of time-series can be given purely in terms of temporal prediction. This concept of G-causality (Granger, 1969) states that a time series x can be said to be G-causal to y if prediction future observations of y from past observations of y can be significantly improved by including past observations of x in the model. In other words, x is g-causal to y if the past activity of x has information about the future of y that is not contained in the past of y itself.

There have been claims that the DTF and PDC measures identify causality in terms of g-causality (Granger, 1969) by several authors (Sameshima and Baccalá, 1999; Baccalá and Sameshima, 2001; Blinowska et al., 2004; Blinowska, 2011). This has been stated most strongly for the DTF which has been described as an extension of g-causality to an arbitrary number of channels (Blinowska et al., 2004). These claims have been interrogated in detail by Eichler (2006) who states that the PDC at frequency f reaches 0 if and only if the autoregressive parameters for that connection equal 0 for all lags. As such, the PDC can be viewed as a frequency domain extension of G-causality as it meets the critical condition that the causality estimate

can only be zero in the case that the past of another signal has no predictive power on future observations of another. In contrast, the DTF is computed from the transfer function $T = A(f)^{-1}$ of the autoregressive model and there is no simple relationship between T and $A(f)$ in which the element of the transfer function for the connection $i \rightarrow j$ can equal zero ($T_{ij} = 0$) if and only if $A(f)_{ij} = 0$. As such the DTF is a measure of the total influence of a time series on another in contrast to multivariate g-causality which would be only related to the direct effect of one time-series on another (Eichler, 2006).

As such, the work in this thesis will concentrate on the magnitude squared PDC estimate and the complex valued PDC from which it is estimated. These metrics have been shown to be highly robust across many critical factors and is sensitive to both directness and directionality within connection in a network. Moreover the PDC has been shown to outperform a more traditional frequency domain test of granger causality in tests into correct rejection of fictitious causal interactions (Fasoula et al., 2013).

4.1.4 Volume Conduction

The question of field spread or volume conduction is a critical one in the estimation of connectivity in brain networks. This is particularly the case with instantaneous measures such as the MSC which may be particularly sensitive to spurious connections arising from signal mixing when source activity in the brain projects to distant M/EEG sensors (Nolte et al., 2004). This problem is thought to be alleviated when analysing network interactions in source space (Schoffelen and Gross, 2009) as the signal mixing should have a much smaller impact when working with source projections.

Several more complex measures of coherence are said to be invariant to volume conduction either though consideration of only the imaginary part of coherency (Nolte et al., 2004) or through orthogonolising the signals at each source to discount any spurious connection which may have zero-phase differences (Hipp et al., 2012). This second approach is inherent in the DTF and PDC estimations as they require a phase difference between the

two channels for the connectivity estimate to be non-zero (Kaminski and Blinowska, 2014).The propagation of a signal from a source to a sensor is the spread of an electromagnetic wave and is assumed to happen almost instantly. As such the DTF and PDC estimates should not be effected by volume conduction. The authors of the DTF caution that this only holds true if a phase difference is not induced by short-sighted data preprocessing such as ICA, Laplace transforms or projection to the brain's source space (Kaminski and Blinowska, 2014). This last point is in contrast to previous advice to only consider connectivity in source space (Schoffelen and Gross, 2009).

The critical issue is in preserving the phase in the signal, if the phase is preserved then the connectivity estimates given by the PDC will be correct. Beamforming is a commonly used technique for estimating neural activity in source space (Huang et al., 2004; Van Veen et al., 1997). A beamformer is a spatial filter across the EEG/MEG sensors which is optimised to maximise the power at one point in the brain while minimising the power in the signal elsewhere. Phase-based estimates at the source level are very commonly used, with little evidence in the literature that the beamforming process produces distorted phase information in source space. Beamformers can underperform when spatially separate but temporally correlated sources are present (Brookes et al., 2007). This means that the beamforming algorithm might be unable to supress a distant signal in the brain which could cause phase distortions in the reconstruted signal.

4.2 Time Domain Simulation And Fitting

This first set of simulations uses a set of simultaneous difference equations which produce a set of time series with known interactions. This approach is used as it allows for simple generation of many realisations of the time series of arbitrary length. In addition, the equations chosen are used throughout the literature and have served as validation examples for several metrics.

The initial question asked with this system is how many observations are necessary to obtain a good reconstruction of the known equations? This

has been interrogated previously using both DTF and PDC (Astolfi et al., 2006, 2007) however these studies did not directly relate the number of observations in the sample to the number of parameters to be estimated in the model. The equations 4.4a-4.4e define a system which would require a $5 * 5 * 4$ A matrix to be fitted in order to fully characterise the interactions. 1000 realisations of each of 5 different data lengths are generated and used to fit an MVAR model of order 4 before the variance explained in the time series and the correlation between the estimated and known parameters are computed.

Secondly, a well fitted parameter matrix A can be used to compute several connectivity metrics which are able to characterise the interactions within the system. Three metrics defined in the previous chapter are tested: 1. Magnitude Squared Coherence (MSC) 2. Directed Transfer Function (DTF) 3. Partial Directed Coherence (PDC) . The ability of these measures to correctly identify directionality and directness is then discussed. All the connections show directionality, though the connection between x_4 and x_5 is reciprocal. In addition there is no direct connection between x_1 and x_5, a metric which is not sensitive to directness in interactions should be able to correctly estimate no interaction between these nodes.

4.2.1 Methods

The critical system analysed in the first part of this chapter is outlined in figure 4.2 and defined in Baccalá and Sameshima (2001). This system is used for both the data ratio testing and the MVAR metric validation.

Data Ratio

Five simultaneously observed time series were generated with equations 4.4a to 4.4e as defined in (Baccalá and Sameshima, 2001). The system has 5 channels and is of order 4, therefore 100 parameters will have to be fitted to model the system if we only know the order, not the specific delays and connections with non-zero parameters. 5 realisations of the system were generated with increasing numbers of samples.

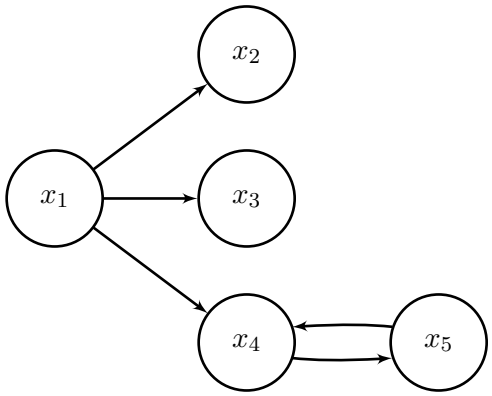

Figure 4.2: The vector autoregressive system used in the first set of simulations. The interactions are governed by equations 4.4a to 4.4e below, as defined in Baccalá and Sameshima (2001)

$$x_1(t) = 0.95\sqrt{2}x_1(t-1) - 0.9025x_1(t-2) + w_1(t) \tag{4.4a}$$
$$x_2(t) = 0.5x_1(t-2) + w_2(t) \tag{4.4b}$$
$$x_3(t) = -0.4x_1(t-3) + w_3(t) \tag{4.4c}$$
$$x_4(t) = -0.5(t-2) + 0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + w_4(t) \tag{4.4d}$$
$$x_5(t) = -0.25\sqrt{2}x_4(t-1) + 0.25\sqrt{2}x_5(t-1) + w_5(t) \tag{4.4e}$$

	Realisation				
	1	2	3	4	5
№Parameters	100	100	100	100	100
№Observations	100	200	500	1000	5000
Ratio	1	2	5	10	50

Table 4.2: Number of parameters and data observations generated in each realisation of the system.

1000 MVAR models of order 4 were fitted to each realisation and the consistency of the accuracy and explanatory power of the models assessed. The accuracy was assessed by estimating Pearson's product moment correlation coefficient between the known coefficients and the estimate of each model. Finally the explanatory power was the correlation between the generated time series of observations and the model's fit.

MVAR Metric Validation

Five simultaneously observed time series were generate with equations 4.4a to 4.4e as defined in (Baccalá and Sameshima, 2001). 10000 observations were generated before an MVAR model was fitted. The system is of order 4 but the MVAR model was fitted up to order 8 to allow a replication of the frequency resolution seen in (Baccalá and Sameshima, 2001) without interpolating the coefficients. The parameters from the fitted model are then used to calculate the magnitude squared coherence and directed transfer function. Finally the partial directed coherence with analytic confidence limits with $\alpha = 0.001$ were computed.

4.2.2 Results

Data Ratio

The R^2 value indicating the variance in the signals explained by the fitted models has a wide and low distribution with a low ratio between the number of parameters and number of data observations. This increases to around .8 for the higher ratio. The simulated system has noise added, as such this value is unlikely to reach 1.0. In addition the maximum value for R^2 is likely to be considerably different between systems, depending on noise in the interactions and measurement and any non-linearities in the interactions between nodes

The correlation between the fitted MVAR coefficients and the known parameters defined in equations 4.4a-4.4e can be seen in figure 4.3. There is a wide distribution of r values which crosses 0 for data ratio 1, indicating a poor and inconsistent model fit in this condition. The distribution of r

increases dramatically for the larger ratios until realisation 5 in which there is a tight distribution indicating a consistently high quality model fit.

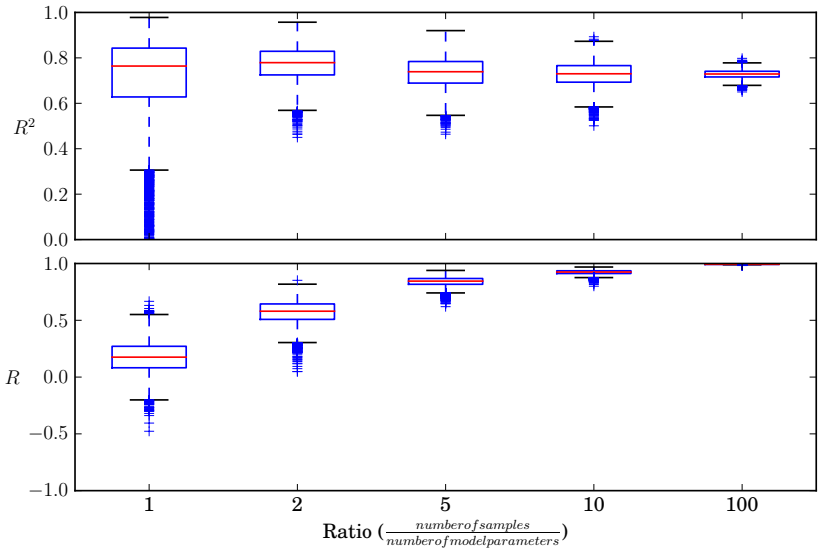

Figure 4.3: MVAR models were fitted to different realisations of the process defined in figure 4.2. Realisations were of five different lengths were used. These were calculated as the ratio between the number of samples in the realisation and the number of parameters estimated in the model. The x-axes denotes this ratio. *top:* The distribution of the coefficient of determination (R^2) over each of the $10,000$ model fits for each data length. *bottom:* The distribution of the correlation coefficient between the known parameters (see equations 4.4a-4.4e) and the estimated parameters for each data length.

MVAR Metric Validation

Several connectivity metrics were estimated from a signal realisation of the system. Firstly the Magnitude Squared Coherence (figure 4.4) characterises both the direct and indirect interactions within the network and as such estimates a spurious connection between $x_1 \rightarrow x_5$. In addition, MSC is a symmetrical measure in that $MSC_{ij} \equiv MSC_{ji}$ which leads to the connectivity matrix being symmetrical about the diagonal.

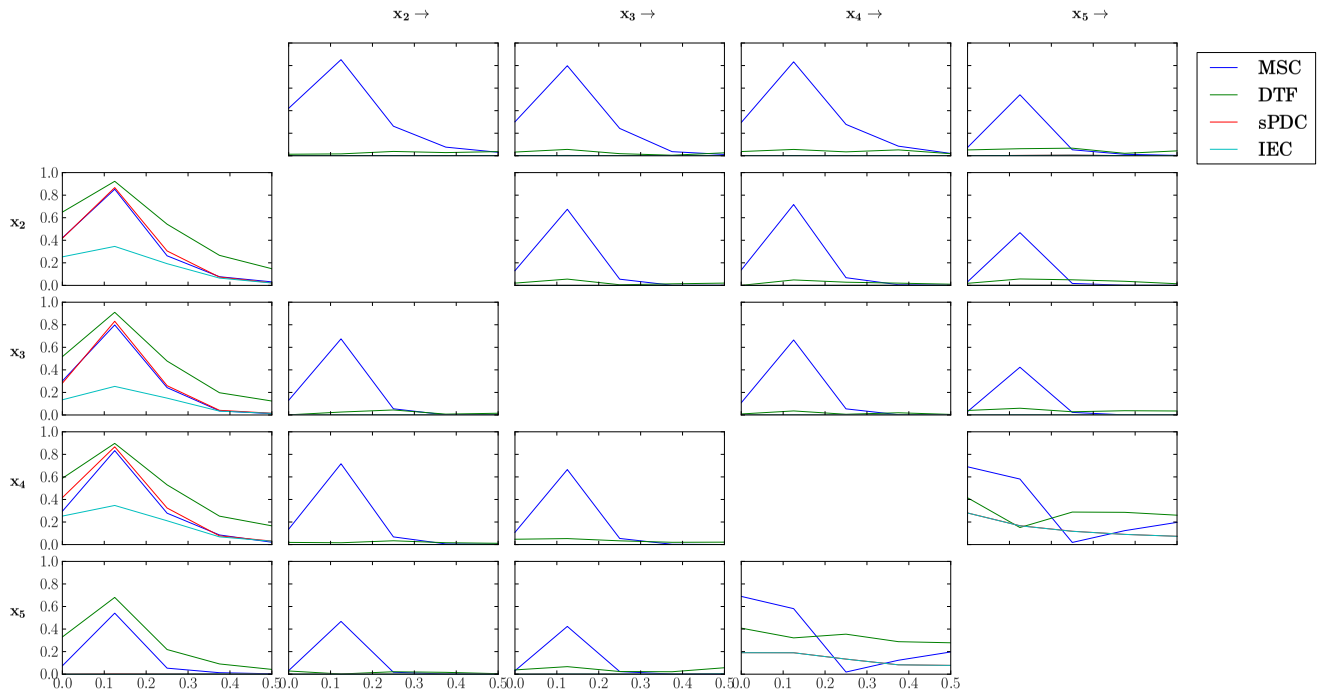

Figure 4.4: Four connectivity metrics estimated from the system defined in equations 4.4a to 4.4e. The plot can be read such that each plot represents the connectivity in the connection in which the node in the column is driving the node in the row. The x-axis denotes frequency as a proportion of the sampling frequency and the y-axis represents the estimate of each metric. The four connectivity metrics are represented in the coloured lines. The Magnitude Squared Coherence (blue) is symmetrical across the diagonal and unable to disambiguate between direct and indirect connections. The directed transfer function (green) is able to resolve directionality yet erroneously suggests a connectivity in the $x_1 \rightarrow x_5$ connection. sPDC (red) and IEC (can) are both able to identify the correct directionality and reject connectivity in the $x_1 \rightarrow x_5$

The Directed Transfer Function (green line in figure 4.4) is able to identify directional connections, however it may still identify indirect connections. The DTF value is normalised across each row in the coefficients matrix, in which the square of the value at a given frequency in all the connections along a row sum to 1.

Finally the Partial Directed Coherence (red line in figure 4.4) is able to identify directional connections and disambiguate between direct and indirect connections. Unlike the DTF, PDC is normalised within each frequency down each column of the parameter matrix. This normalisation can be seen on the diagonal of figure 5.4 wherein an increase in the PDC within the off-diagonal connection is associated with a decreased PDC estimate on the diagonal.

4.2.3 Interim Summary

The optimal ratio between the number of fitted parameters is likely to vary from dataset to dataset and potentially very difficult to identify in a real system with unknown dynamics. The results from the data length simulations provide an illustration of the importance of a high (> 5) ratio between the number of fitted parameters and number of observations. The potential consequences of having too few observations is also illustrated. When there are as many observations as parameters there is a high variability in the variance explained by the fitted A matrix. Critically many realisations had an R^2 of close to 0 or 1. In the case that our experimental realisation has a very low R^2 we may correctly reject this model or attempt an alternative approach. In contrast with a large R^2 it is likely that we may accept this model. These simulations demonstrate that fewer observations may lead to an inflated possibility that we arrive at a good model through chance alone. This is a type 1 error or false positive, the chance of which may be unusually high when our model is under powered, that is there are insufficient observations to arrive at a robust model.

Overall the PDC provides the best method for characterising network structure from the elements of a fitted MVAR parameter matrix. It pro-

vides a directional measure of information flow which is not susceptible to spuriously identifying indirect connections in a network. One caveat to this second point is that the PDC can only disambiguate between $x_1 \rightarrow x_2 \rightarrow x_3$ and $x_1 \rightarrow x_3$ if all three nodes are included in the MVAR estimate.

The arbitrary normalisation seen in both the DTF and PDC can lead to some problems in interpreting results, particularly change in DTF/PDC estimates over time or condition. As both values are fractions in which the denominator is consistent across a whole row or column of the spectral representation of the fitted parameter matrix. As such, a change in either of these metrics could arise from either the denominator of numerator of the metric. Moreover, a change in one connection within the row/column can lead to a change in the value of another connection which actually remains constant over the modulations

4.3 Frequency Domain System

The frequency resolution of an MVAR system is determined by the delay and order parameters and the frequencies observed in the interactions determined by the cross spectral density of the autoregressive parameters. Together these factors control the frequencies seen in the connectivity within the modelled system. The spectral resolution has attracted much attention in the literature alongside growing interest in oscillatory neural activity.

The MVAR model is a powerful spectral density estimator and has been widely used in digital signal processing for this purpose (Marple, 1987), moreover there is evidence that it might outperform multi-taper methods in short-window spectral analysis (Nalatore and Rangarajan, 2009).

The frequency resolution Δf is determined by the TDE parameters p and τ. Critically, the sampling rate in the autoregressive parameters (determined by τ) and the total number of past observations in the autoregression (p).

$$\Delta f = \frac{SR}{p} \tag{4.5}$$

The frequency resolution for a system is determined by equation 4.5 can be seen for a wide range of models in figure 4.5.Very low orders and delays can lead to a physiologically uninformative frequency resolution of $\Delta f > 100Hz$ or more, however this rapidly decreases to around $10Hz$ in the dark green colour. This implies that for a physiologically plausible frequency resolution, MVAR parameters from around this region or to the bottom-right of the plot must be used.

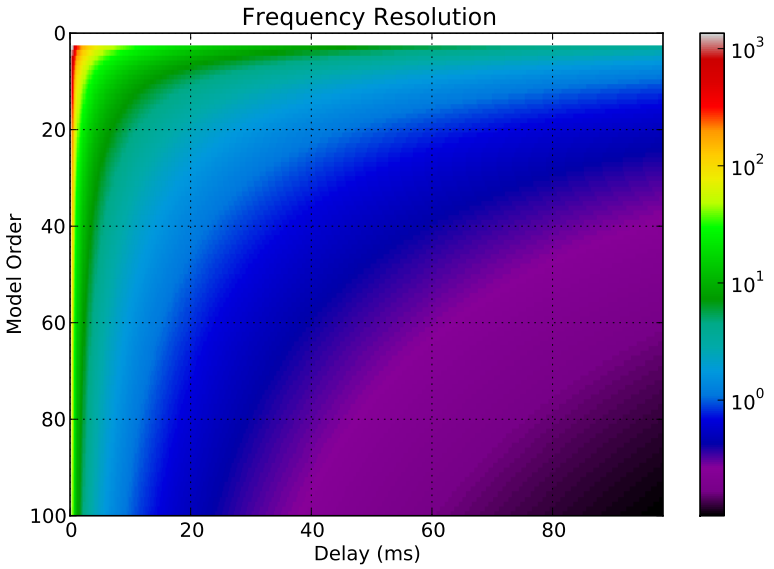

Figure 4.5: The frequency resolution ($\triangle f$) of MVAR models with different delays and model orders. Δf is displayed in a log scale in the colour bar. The estimates for model orders > 3 have been set to zero as these systems will only have one frequency component in addition to DC.

This principle is not strictly adhered to in the literature, though there is acknowledgement of the limited frequency resolution of a model of a system (Schlögl and Supp, 2006), in practice it is common for $A(f)$ to be estimated for an arbitrary set of frequencies. For instance, though many papers state that p is importance for spectral resolution, spectral connectivity metrics may be estimated over a wide frequency band (Brovelli et al., 2004; Pascual-Marqui et al., 2014) or more commonly the precise frequencies estimated are

unclear ambiguous (Korzeniewska et al., 2008; Supp et al., 2007; Sun et al., 2009a).

Moreover, a debate into the power of the frequency response in PDC has recently arisen. Pascual-Marqui et al. (2014) state that PDC may misinterpret the spectral representation of the connectivity in a system. To counter this problem, the authors define isolated effective coherence (iCOH). The issue with PDC and the solution given by the iCOH is illustrated with simulations from known differential equations of order 2 and applied to real EEG data. The results demonstrate that the iCOH produces more defined peaks that the PDC at the expected frequencies, however the PDC and iCOH are estimated at 127 separate frequencies from the order 2 autoregressive parameters. This is an interpolation in the spectral domain and as such most of the PDC and iCOH estimates will be smoothed versions of the small number of frequencies which naturally arise from the MVAR model. Moreover, the iCOH computation scales the $A_{ij}(f)$ estimate by the inverse of the error covariance, or the precision matrix. This will prioritise the connections with lower residuals giving greater weight to these connections. Critically the iCOH still uses the $A(F)$ matrix in a manner completely analogous to the PDC, as such, when the natural frequencies from the MVAR model are used, neither measure will have more or less frequency information that the other. Where the frequency response of the MVAR model is interpolated to a larger number of frequencies, the higher response to strong/more precise connections in the iCOH could lead to a larger peak in the estimate and an apparently more accurate frequency response.

This critical issue is tested using the $x_1 \to x_2$ connection in the system with the known spectral density in the autoregressive parameters. MVAR models from 3 time delay embeddings are fitted and the PDC and iCOH estimated at the resolution given by equation 4.5. No spectral interpolation is used, moreover the critical test is the two interactions within the $x_1 \to x_2$ connection, rather than a localisation of one interactions. The ability of both measures to correctly reconstruct the known connection at different resolutions is discussed.

4.3.1 Methods

Though the signals generated by known parameter matrices in the time domain do have a spectral profile, it is difficult to exert fine control over the cross spectral densities between signals. This is particularly the case with the very low order systems commonly used in connectivity simulation. As such, the frequency resolution of the approach used in this thesis was investigated by simulating systems with known auto and cross spectral densities in the autoregressive parameters. First the order and sample rate for the system were chosen. The sample rate was 4069 which is consistent with the MEG sampling frequency used in this thesis. A large order was chosen to ensure that the small differences in frequency can be embedded in the system. In this example a order of 1000 was selected which, with the sampling frequency, allows for a frequency resolution of $10Hz$.

A known set of cross spectral densities were defined to create the interactions seen in figure 4.6. These were then converted into the time domain using an inverse Fourier transform and used to generate a set of four simultaneously observed time-series with 100000 observations. This gives the system a parameter to sample ratio of just over 6.

These time-series were modelled using an MVAR model which was then used to estimate the Partial Directed Coherence of system with analytic confidence limits at $\alpha = 0.001$. Additionally, the complex form of PDC is plotted in a z-plane to establish whether the phase of the known CSD could be reconstructed from the PDC of the model.

Frequency Resolution and Time Delay Embedding

The p and τ parameters in the TDE and MVAR model are critical in determining the frequency resolution of our model. The Connection $x_1 \rightarrow x_3$ is used as a test case for the frequency resolution of different TDEs. MVAR models are fitted to different frequency resolutions and the ability of the model to disambiguate the two separate connections evaluated. two separate connections evaluated.

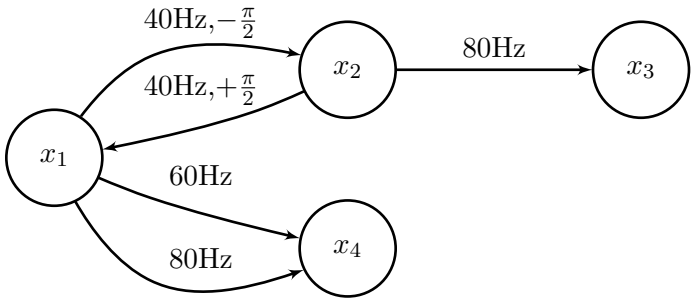

Figure 4.6: A system who's parameters are designed in the frequency domain in order to establish connections with known interactions at specific frequencies. Two main challenges arise from this model. 1) the connection between x_1 and x_2 contains reciprocal directional connections but with a known phase shift. 2) x_1 and x_4 have two driving influences at distinct frequency bands.

4.3.2 Results

Frequency Validation

The absolute value of a Fourier transform of the estimated MVAR parameters for system 4.6 can be seen alongside the known value in figure 4.7. Critically, there are no large peaks in spectral power for any frequencies other than the known frequencies defined in figure 4.6. These estimates form the numerator of the PDC estimation, and the denominator comes from the sum within each frequency for each column.

The PDC estimated from the MVAR model fitted to the spectrally defined system can be seen in figure 4.8. The PDC correctly identifies the reciprocal interaction between $x_1 \leftrightarrow x_2$ at $40Hz$ and the directional connection $x_2 \rightarrow x_3$ at $80Hz$. Critically, the PDC is able to disambiguate between the $x_1 \rightarrow x_3$ connections at 60 and $80Hz$. All of these interactions are above the threshold defined by the analytic confidence limit.

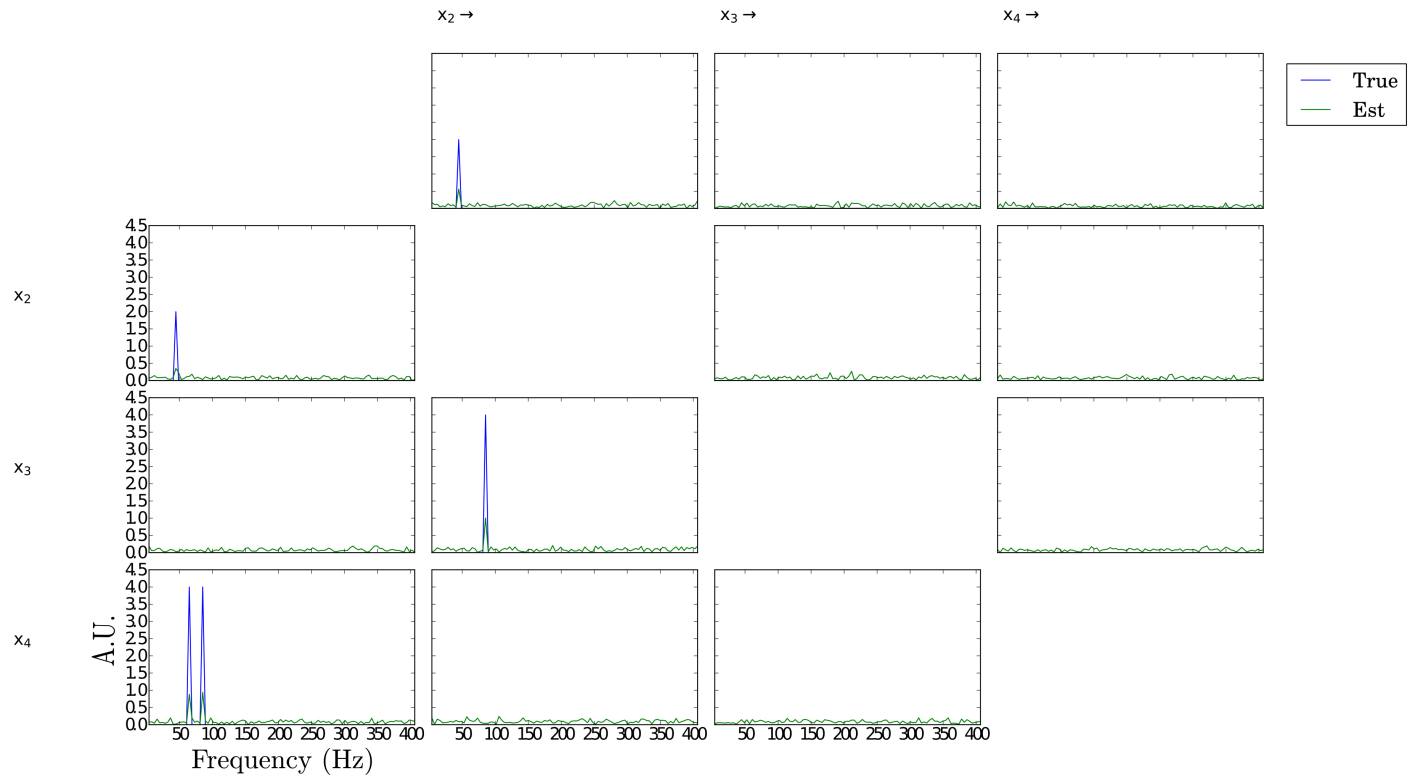

Figure 4.7: $|A(f)|$ estimates for the MVAR model fitted to the system defined in figure 4.6. This is the power spectral density of the autoregressive parameters for each direction of each connection in the system. Clear peaks can be seen the 4 connections defined in the system. Each peak is very narrow, only having a bandwidth of around $10Hz$.

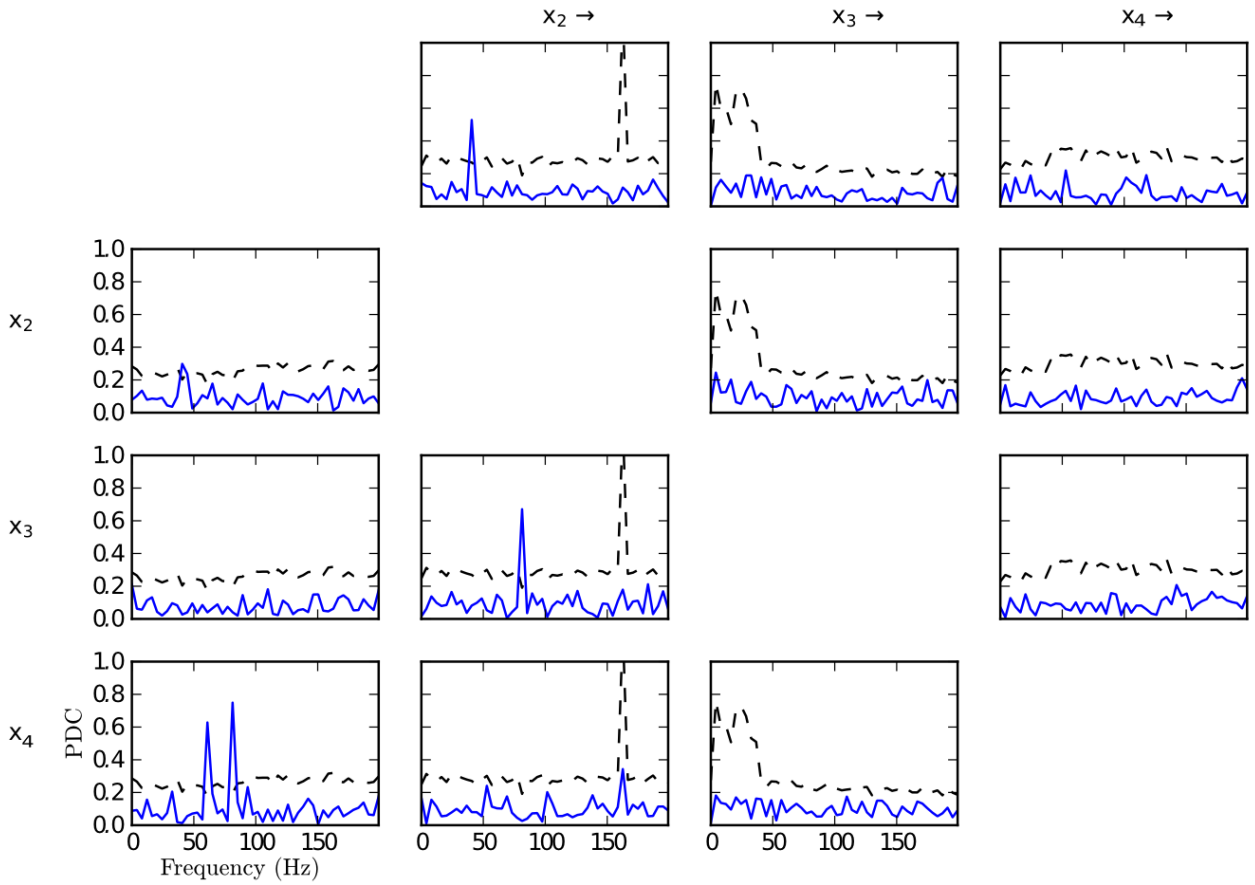

Figure 4.8: Partial Directed Coherence estimates from the system defined in system 4.6. The blue lines refer to the PDC estimate and the black dashed lines are the analytic confidence intervals.

Phase Validation

PDC uses a modulus in the numerator in it's definition, however if this is not done the PDC estimate is complex valued. This form of PDC retains information about the phase of the interactions between nodes in the system. The complex valued PDC estimate is plotted on a z-plane in figure 4.9. In this form the phase difference between directions in the reciprocal interaction $x_1 \leftrightarrow x_2$ can clearly be seen as a rotation of the significant value around the origin. The remaining significant connections have a value of 0 on the imaginary axis and therefore contain phase delay.

The diagonal plots in figure 4.9 represent the PDC influence of a signals own past onto it's future. This is likely to be the largest value down each column, and as such will make the largest contribution to the denominator of the PDC. As a consequence, many of the diagonal PDC estimates are very large and significant.

Figure 4.9 shows one spurious interaction in the $x_3 \rightarrow x_2$ connection. This is a chance finding and the only miss in the whole system, the rest of the estimates either being correctly identified as significant or correctly rejected as chance. One false alarm in the many tests seen in figure 4.9 is not unexpected and well within acceptable number given the alpha value used.

Frequency Resolution

As the frequency resolution of an MVAR model is critically dependent on p and τ, the ability of three different TDEs to disambiguate the two interactions in the $x_1 \rightarrow x_3$ connection. Figure 4.10 shows the PDC results with confidence limits on the right hand side and the iCOH estimate on the left. Both metrics are clearly able to separate the two interactions at the two higher resolutions (6 and 10Hz), in addition, the signal to noise ratio appears to be higher for the peaks in the iCOH. This is due to the scaling of the parameters by the inverse of the residual covariance prioritising the larger values in $A(f)$. Neither metric is able to separate the two interactions at the lower resolution (20Hz).

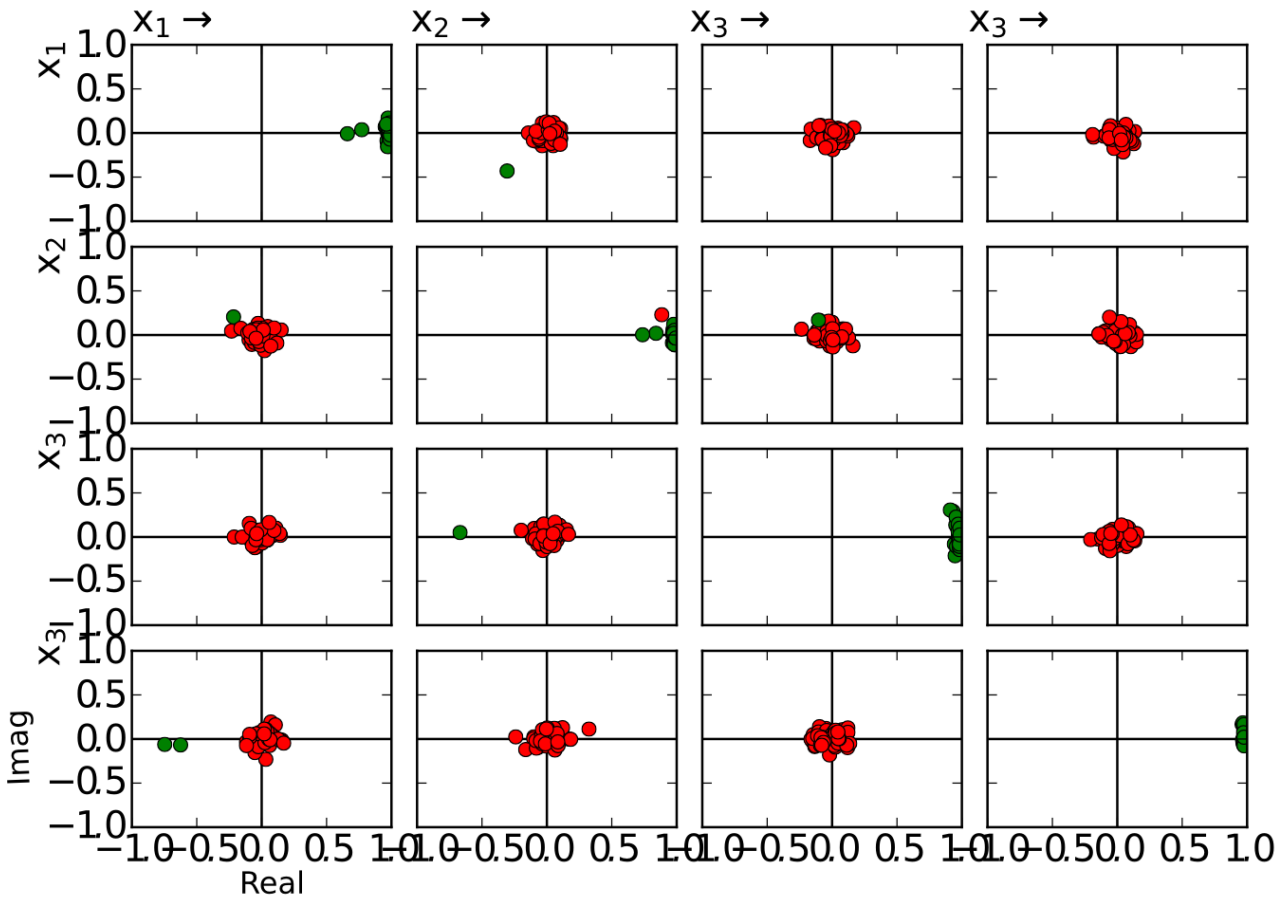

Figure 4.9: The complex value from which the Partial Directed Coherence is established by taking the absolute value. The system is the one defined in figure 4.6. Each dot is the PDC estimated at one frequency, the red dots are estimates lower than the confidence limit and the green dots are above the confidence limit.

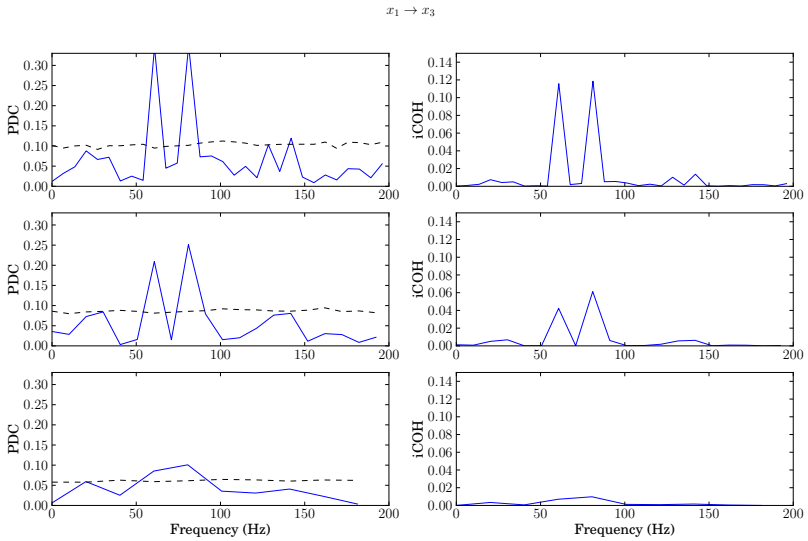

Figure 4.10: The PDC across $1-200Hz$ estimated from MVAR models with different time delay embeddings. The system at the top has a resolution of $\Delta f \approx 6Hz$, the middle a resolution of $\Delta f \approx 10Hz$ and the bottom system a resolution of $\Delta f \approx 20Hz$. Only the top two systems are able to disambiguate between the two interactions in this connection, with the resolution of the bottom model being so wide that it integrates across the interactions and only presents one significant frequency. To date, there are no methods for estimating analytic confidence intervals for the IEC, as a consequence, the confidence limits are only presented for the PDC.

4.3.3 Interim Summary

The spectral resolution of an MVAR model and any metrics computed from it is completely determined by the TDE parameters p and τ. A novel simulation which constructs a system of time-series with parameters with known power spectral density in the off-diagonal autoregressive coefficients was used to test this principle. With a very large order and very small delay, both the frequency and phase of the known interactions can be reconstructed validating the simulation method. Moreover, the TDE parameters are shown to critically determine the frequency resolution of the MVAR model rather than the properties of a given connectivity metric. There is some indication that a measure such as iCOH may provide a greater signal to noise ratio that PDC, however this effect is similar to the principle behind the use of the squared PDC.

4.4 Discussion

4.4.1 Time-Domain System

An investigation into the number of samples required to minimise type 1 errors in a model fit revealed that a minimum ratio of $5:1$ between number of observations and number of parameters produced reliable results. This result builds on previous literature highlighting the importance of signal to noise ratio and the data length (Astolfi et al., 2006, 2007; Fasoula et al., 2013).

The present findings and along with the previous literature highlight the importance of sample size when fitting a linear model. In this thesis the linear model in question takes the form of an output or observer on an approximation to a set of hidden states, which leads to the question of data sufficiency for estimating the TDE parameters p and τ. As highlighted in the previous chapter, there may be no such thing as an abstractly correct or optimal TDE reconstruction (Kugiumtzis, 1996; Cellucci et al., 2003). Moreover the results of any optimisation procedure are strictly only valid for the data on which they were estimated. As such, there may be no sample

size at which these parameters may be more or less robustly estimated as results may not be comparable across different sets of input data. This may not be an issue as the TDE simply provides a reconstruction of the observed data rather than an abstracted model of it, moreover the only viable measure of the quality of this reconstruction is whether is provides a basis for prediction of future observations. This brings us back to the question of optimising our observer model as discussed above.

Many connectivity metrics can be estimated from a fitted MVAR parameter matrix and much research has been done into their relative performance and merits (Fasoula et al., 2013; Astolfi et al., 2007; Blinowska, 2011; Florin et al., 2011). Critically these metrics differ in their ability to disambiguate direct and directional connections as outlined in table 4.1. This section replicated these findings and established that the PDC provides the most viable metric for identifying structure within a network of time-series. This is because the PDC is a spectral estimator which is sensitive to directionality and able to disambiguate direct from indirect connections. Moreover, unlike the Isolated Effective Connectivity, much work has been done in characterising the asymptotic statistics of the PDC and establishing methods for computing confidence limits for PDC estimates (Schelter et al., 2006).

4.4.2 Frequency-Domain System

The spectrally defined simulations in this chapter seek to validate the phase and frequency information in the PDC estimate based on two conditions. Firstly, can the PDC correctly estimate the frequency with a known magnitude or phase in the spectral density of the autoregressive parameters? And secondly, is the PDC able to disambiguate two interactions at different frequencies within the same connection? The PDC was able to complete both of these validation steps.

A critical question in neuroimaging literature concerns the spectral content of directed interactions in the brain. This interest is born out of a wide understanding of the importance of oscillatory activity in the brain suggesting that the same oscillations may be seen in functional connections. The

frequency resolution of an MVAR model is completely determined by the TDE parameters p and τ. Along with the sampling frequency in the parameters, these values determine the natural frequency resolution of the system. Despite this limit, it is common to see frequency interpolated connectivity estimates in the literature (Baccalá and Sameshima, 2001; Pascual-Marqui et al., 2014; Korzeniewska et al., 2008; Supp et al., 2007). Estimating connectivity at frequencies beyond the natural resolution of the MVAR model does not increase spectral resolution of a system, rather these results will simply be interpolations of the natural frequencies arising from the system. In addition, we must be cautious when comparing the frequency response of different metrics when using such an interpolation as we are not able to attribute any differences to the metric or the smoothing itself. This issue is addressed in the final part this chapter which shows that the spectral response or selectivity of a system is determined by the MVAR parameters alone and not the metric computed from it as suggested by some authors (Fasoula et al., 2013; Pascual-Marqui et al., 2014).

4.5 Conclusions

There are several methods for creating systems of time-series with known time-delayed interactions. Through the use of the known-difference equations method, this chapter has established the effectiveness of the Partial Directed Coherence connectivity metric. A classic set of equations from Baccalá and Sameshima (2001) were first used to illustrate the importance of the ratio between the number of observed data points and the number of parameters fitted in a model. Though the precise value might vary from system to system a minimum ratio of around $5:1$ was shown to be highly desirable. Secondly the ability of PDC to rule out indirect interactions and establish whether there are asymmetries in the dominant direction of information flow in a connection.

A critical issue in MVAR based connectivity method is the spectral resolution of the model. There are suggestions in the literature that some connectivity metrics estimated from a fitted MVAR model may provide better

reconstructions of the frequency of any interactions. The known differential equation method was extended to create a simulated dataset in which the spectral density of the off diagonal autoregressive coefficients were known. The PDC and iCOH estimates from this dataset from MVAR models with different resolutions were used to demonstrate that though the signal-to-noise from the connectivity metric may vary, this does not effect the frequency resolution when the spectral estimates are not interpolated.

This chapter has outlined several practical issues with network connectivity estimation from observed data including sample size and frequency resolution. These questions have been directly interrogated using objective data simulations and their validity and limits established. The next step is to apply these methods to a set of real data acquired from a human resting state MEG scan.

Chapter 5

Directed Functional Connectivity In An Endogenous Brain Network

5.1 Introduction

The idea that our cognition may depend on connections between brain regions rather than isolated processes within individual brain areas is not new (Wernike 1874 as cited in Eggert, 1977). There is now a wealth of neuropsychological and neuroimaging evidence which demonstrates how specific networks of brain areas map on to particular behavioural tasks and cognitive processes. However, even when participants have no particular task to perform, highly organised spatio-temporal structure has been identified within neural networks from the low frequency ($f < 0.1Hz$) fluctuations of their BOLD signal measured in fMRI (Biswal et al., 1995; Fox et al., 2005). This structure generally follows the known functional topography within sensory and motor regions whilst within association cortex, there is no clear hierarchical structure (Yeo et al., 2011). One such example is the Default Mode Network (DMN), a brain system which is preferentially active when a participant is not focused on the external environment (Raichle et al., 2001; Greicius et al., 2003; Buckner et al., 2008).

5.1.1 Default Mode Network

The DMN comprises the Medial Prefrontal Cortex (MPFC) and Posterior cingulate cortex (PCC) in addition to the bilateral Angular Gyri and the middle temporal gyrus (Raichle et al., 2001; Greicius et al., 2003; Buckner et al., 2008). These are all regions thought to act as 'hubs' which interconnect functionally specialised systems (Buckner et al., 2009). These hubs (particularly the PCC and MPFC) have a high network centrality in both structural and functional networks and are located in known heteromodal/paralimbic association areas (van Oort et al., 2014a). As such, the DMN has been linked to many cognitive processes including self-referential processing (Fair et al., 2008) and has been shown to be dysfunctional in several clinical disorders (Buckner et al., 2008). There is further evidence that functionally specialised sub-components of the DMN exist, a midline core (MPFC and PCC) relevant to self focused affective thought and a lateral temporal system involved in memory (Andrews-Hanna et al., 2010). The

interpretation of the functional relevance of the DMN and its subsystems are complicated by findings of a DMN in rats (Lu et al., 2012) and unconscious monkeys (Vincent et al., 2007), suggesting that its function may be quite low level, possibly not requiring conscious awareness. As each of the nodes within the DMN is associated with several functions both in and outside the context of endogenous brain networks - its functional role may best be identified by the state of the network as a whole in any given context (McIntosh, 2000; Price and Friston, 2005). If so, elucidating the direction of information flow within the DMN should help to characterise any putative role.

It is important to note that the 'hubs' within the DMN are distinct from the 'rich-hubs' which are found within the human connectome (Bullmore and Sporns, 2012; van den Heuvel and Sporns, 2011). There is some partial overlap, particularly in the superior frontal cortex (Yeo et al., 2011) however the medial frontal cortex, middle temporal gyrus and inferior parietal lobe are not considered to be in the 'rich-club' (van den Heuvel and Sporns, 2011). The precuneus is a 'rich-hub' and can be found directly anterior to the posterior cingulate, a critical hub in the DMN. As such the 'hubs' within the DMN are not defined by a general interconnectedness with the whole brain, rather they form a core within the DMN itself (Andrews-Hanna et al., 2010). The exception might be in the PCC which is both a major node within the DMN and a rich-hub potentially making it critical in interactions between the DMN and the rest of the brain.

G-causality analysis in whole brain resting fMRI data have shown that the nodes in the DMN are primarily driven by activity in other brain regions (Yan and He, 2011). Moreover, within the DMN, fMRI has shown that the ventral MPFC node exerts a driving influence on the PCC (Jiao et al., 2011; Uddin et al., 2009; Zhou et al., 2011), a finding confirmed in a recent DCM study (Di and Biswal, 2014). In contrast, a recent EEG connectivity study has shown that the PCC exerts driving influence on the vMPFC during resting state scans in which the participant has their eyes open, though this influence disappears when the participants eyes are closed (Pascual-Marqui et al., 2014). These results suggest that the direction of information

flow within the DMN are very consistent on the longer time-scales analysed in fMRI, however the more rapid dynamics may be more prone to larger changes depending on perceptual and task constraints.

5.1.2 Magnetoencephalography

The fine-grained temporal resolution of Magnetoencephalography (MEG) allows us to address the question of *whether* nodes interact and moreover, specifically *how* they interact. In addition, its spatial resolution is sufficient to reveal similar spatio-temporal structure to fMRI in ultra slow ($f < 0.1$Hz) spontaneous power modulations (Liu et al., 2010). Beyond a replication of the fMRI findings, two main types of coupling have been identified (amplitude and phase) which may reflect different functionalities within the brain (Engel et al., 2013). MEG has also identified non-stationarities in frequency-band limited networks (de Pasquale et al., 2010; Baker et al., 2014). This point is reinforced by the computational literature which suggests that over the longer time-scales considered by fMRI functional connections are largely related to structural connections, in contrast many configurations of sub-systems are possible over faster time-scales (Deco et al., 2011). This variance over shorter time scales may reflect an exploration of the possible functional architectures allowed by the structural-anatomical skeleton of the brain.

MEG has additionally demonstrated that network connectivity may occur within frequency bands that are known to be physiologically relevant(alpha-$8-13Hz$,beta-$13-20Hz$ and gamma-$20Hz+$). Moreover, ICA applied to band limited amplitude fluctuations has identified a DMN in MEG in the alpha band with no prior assumptions about its spatial configuration (Brookes et al., 2011b). Correlations in spontaneous fluctuations show frequency specific spatial structure, though most strongly within the alpha to beta range, however these methods (power envelope correlations) have not identified functional networks in the gamma band (Hipp et al., 2012).

Although nodes within the DMN are well established, the causal structure within the network is not clearly understood. The vast majority of evidence on endogenous brain networks are derived from correlational neu-

roimaging methods, limiting any interpretation to simple, instantaneous interactions as seen with measures such as Magnitude Squared Coherence. Another difficulty is the lack of neuropsychological evidence relating to the behavioural relevance of the DMN (Buckner et al., 2008). One recent paper directly investigated the causal structure within the DMN using a dynamic causal model (Di and Biswal, 2014). The winning model contains a driving influence from the MPFC to the PCC, suggesting that there may be hierarchical influences within and between the sub-components of the DMN. Such characterisation of the behaviour of nodes within the DMN is essential for establishing the network context of each node and subsequently, establishing the functionality of both individual nodes and the network as a whole. This is a critical step in establishing the functional relevance of the nodes within the DMN. By establishing the nature of information flow any hierarchy within the network can be established, furthermore we can estimate the modulation of this structure with cognitive processes providing several novel insights to both the functional structure and cognitive outputs of the DMN.

5.1.3 Conduction Delays

An increasing literature on the role of delay in large-scale brain networks has arisen in recent years. Several computational models simulating spontaneous fluctuations in brain function suggest that non-zero transmission delays may play a role in establishing the spatio-temporal structure of endogenous brain networks (Deco et al., 2011). Delays due to conduction speeds of around 5-20 m/s and 1-5m/s are seen for myelinated and unmyelinated axons in the adult primate (Ghosh et al., 2008). Physiologically detailed generative models of spontaneous activity are most effective when a realistic delay in this range is incorporated (Cabral et al., 2014b; Nakagawa et al., 2014). Delayed interactions may lead to a reduction in the stability of a system and are more influential when the delay is of the same order as the oscillation period of the dynamics and may be neglected altogether when considering very slow dynamics (Cabral et al., 2014a). The computational literature

has demonstrated that realistic simulations of network dynamics only arise with a proper scaling of connection strength and conduction delay resulting in meta-stable dynamics giving rise to the low frequency synchronisation observed with fMRI (Deco et al., 2011). Conventional correlation and coherence based connectivity measures are often insensitive to such delayed interactions, focusing instead on instantaneous synchrony. Several approaches are able to incorporate the interactions that might arise from the limited conduction times seen in the brain and emphasised by the literature in this paragraph. One such approach uses Time Delay Embedding (TDE) to create a representation of the activity in the brain across several nodes and time delays.

5.1.4 Time Delay Embedding

The present study seeks to characterise the predictive causal structure within the DMN though characterisation of the predictive strength of delayed interactions within observations of activity at each node. In this approach, the rapid dynamics of the DMN are reconstructed before the causal structure is estimated. The coordinates for the state of a dynamical system are established through a time delay embedding (TDE) of observations of the system and their time delays, this may be done over any time-scale and may establish either rapid or slow dynamics. A sufficiently large set of independent delays (dimensionality) has been shown to provide acceptable coordinates for the system (Takens, 1981). High data sampling rates are necessary to characterise the physiologically relevant conduction delays. There is a range of methods for objectively establishing optimal values for the delay and dimensionality of the system (Ragwitz and Kantz, 2002; Cellucci et al., 2003) which have been applied to non-linear analyses of neuroimaging data (Stam, 2005). These approaches allow us to establish an objective TDE before a multivariate autoregressive model is used to parametrise the strength of the influence in between delayed observations across all nodes in a network. In this way we are able to estimate the directionality of information flow within the system from the fitted regression model. This

can be done in a data-driven manner in which causality estimates can be made without a prior causal models being specified or tested. The interdependencies between nodes in a fitted MVAR model are then characterised in the frequency domain using Partial Directed Coherence (Sameshima and Baccalá, 1999; Baccalá and Sameshima, 2001).

5.1.5 Chapter Outline

In this chapter, the rapid dynamics within an example resting state network (RSN), the DMN, is characterised using TDE to explicitly account for delayed interactions between nodes. Critically, the inclusion of time delayed interactions in this approach naturally leads to estimation of causal influences within the networks. The optimal TDE parameters are objectively assessed and physiologically plausible values dimensionality and delay are established. These parameters are used as the basis of a subset MVAR model which in turn, provides a description of the causal structure within several RSNs. Directionality within the RSN as a function of frequency is further examined and its functional relevance discussed.

5.2 Methods

5.2.1 Data Acquisition

Data were continuously recorded at 4069.017Hz using a 4D Neuroimaging Magnes 3600 whole head system. The data were passed through a low-pass filter set at $1500Hz$. Participant head shape and reference coil location were recorded using a 3D digitiser (Polhemus Fastrak) allowing for co-registration with a high resolution anatomical T1 MRI image acquired using a GE 3.0T Signa Excite HDx (Kozinska et al., 2001).

Five participants completed 5 resting state recording sessions of 11 minutes each. Scans were taken across multiple days over a total time of 10 days. Participants were fully briefed prior to data acquisition and clearance to acquire data was obtained from the York Neuroimaging Centre research ethics committee.

5.2.2 Data Analysis

Artefact Rejection

Data from each participant were split into non-overlapping 2 second segments. These epochs were visually inspected for artefacts and trials containing muscle or eye movements were discarded from the analysis. The MEG data were then co-registered with the individual's structural T1 MRI scan and the MNI 152 standard brain to allow for source analysis to be conducted.

Data Preprocessing

A 1Hz high pass filter was used to remove drift from the reconstructed signals and line noise removed using tight notch filters at 50Hz and it's first 4 harmonics. Previous work has shown that excessive filtering time-series prior to g-causality estimation can distort the subsequent network structure, however a simple notch filter such as this to remove line-noise is recommended (Barnett and Seth, 2011). The samples within each epoch were normalised by demeaning each sample and dividing by the standard deviation. The epochs from each of the five runs were then concatenated.

Virtual Electrode Analysis

Nodes within the Default Mode network were taken from an analysis of 1000 fcMRI datasets (Yeo et al., 2011). Details of these nodes can be seen in Table 5.1 and Figure 5.1. A tangentially constrained LCMV beamformer (Van Veen et al., 1997) was used to generate source space estimates of the current flow at each of these locations.

The processed sensor recordings were projected into the brain's source space at the locations defined in table 5.1. The magnetic field b is measured from each of M sensors outside the head to form a t by M matrix B. An adaptive spatial filter called a beamformer is used to estimate neuronal activity Q at a specific source p in within the brain. The electrical activity at time t is estimated from a weighted sum of the sensor measurements B.

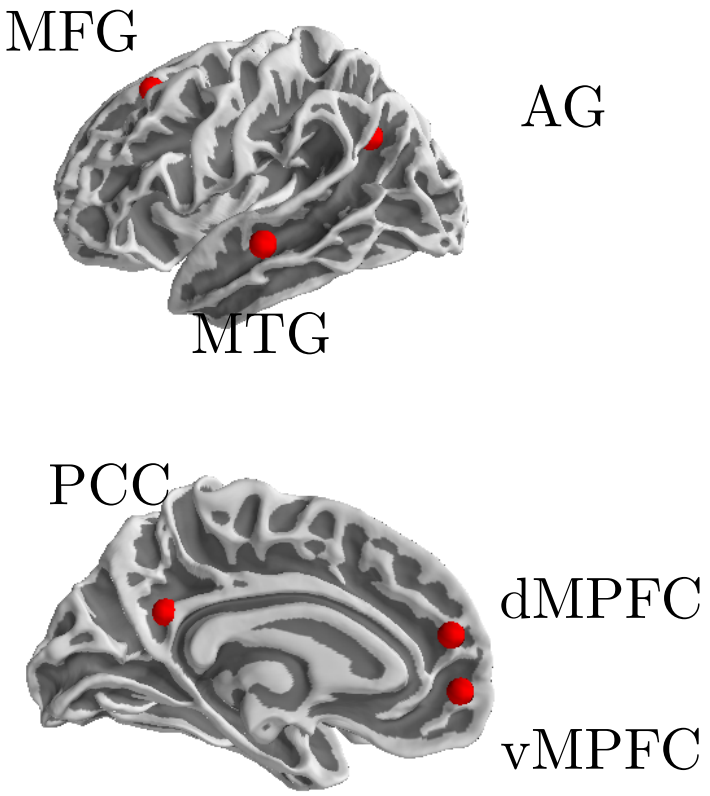

Figure 5.1: Schematic representation of the location of six nodes in the DMN projected onto the cortical surface of the left hemisphere. The nodes are: *MFG* - Middle Frontal Gyrus, *AG* - Angular Gyrus, *MTG* - Middle Temporal Gyrus, *PCC* - Posterior Cingulate Gyrus, *dMPFC* - dorsal Medial Frontal Cortex and *vMPFC* - ventral Medial Frontal Cortex

MNI coord				
x	y	z	**Abbreviation**	**AAL Location**
-27	23	48	MFG	Frontal Mid L
-41	-60	29	AG	Angular L
-64	-20	-9	STG	Temporal Mid L
-7	49	18	dMPFC	Frontal Supp Medial L
-6	52	-2	vMPFC	Frontal Supp Medial L
-7	-52	26	PCC	Cingulum Post L

Table 5.1: ROIs used in the connectivity analysis. These coordinates were taken from table 5 in Yeo et al. (2011) and anatomical locations estimated in fslview

$$Q = WB \tag{5.1}$$

A linearly constrained minimum variance (LCMV) beamformer was used to estimate the sensor weightings W (Van Veen et al., 1997; Huang et al., 2004) from an estimate of the covariance between all pair of MEG sensors. To optimise the orientation of the reconstruction of electrical activity a tangentially constrained search for the source orientation which maximises the total power in the signal at each node was conducted across all 5 acquisitions for each participant. The source reconstructions were reoriented into this optimal orientation per node, per participant.

5.2.3 Time Delay Embedding

Time delay embedding (TDE) is a method for reconstructing a dynamical system from a sequence of observations. TDE is described in detail in chapter 2, and so only a brief introduction is included here. Practically, TDE involves the conversion of a set of time-series into a sequence of vectors containing observations across a number of time points. The number of time-points included in the vector is the dimensionality or order of the embedding space and is denoted p. The time-delay between the observations

included in the embedding vector is known as lag or delay and denoted τ. A p dimensional TDE (Φ) can be written as

$$\Phi_{t,p,\tau} = \left[\mathbf{x}_t, \mathbf{x}_{(t-\tau)}, \mathbf{x}_{(t-2\tau)}, \ldots, \mathbf{x}_{(t-p\tau)}\right] \tag{5.2}$$

An inappropriate value for either of these terms can lead to a distorted representation of the dynamics of the observed system potentially leading to the detection of spurious dynamical relations or the omission of real ones.

The process of objectively time-delaying a set of observations naturally leads to the estimation of predictive causality such as G-Causality (Granger, 1969). G-Causality asserts that if the prediction error in signal X_1 from its past history can be reduced by the inclusion of the past of X_2, then X_2 can be said to G-Cause X_1. This is typically assessed with methods such as MVAR models which fit parameters indicating the predictive power of a point in a TDE on subsequent data points. This pipeline represents an objective, data-driven framework for estimating predictive causality within a brain network, in contrast to model based methods such as Structural Equation Modelling and Dynamic Causal Modelling which require prior statements about causal structure.

The TDE reconstruction can be thought of as a non-linear transformation of the dynamics of the underlying states in the brain (see chapter 2), however in anything more than a very simple system this transformation is likely to be highly complex. Moreover, we have no information about the "true" states producing the observable outputs. As such, the TDE process is treated as an optimisation process which selects p and τ based on how able we are to predict future observations on the basis of $\Phi_{p,\tau}$. This prediction is performed with a multivariate autoregressive model.

5.2.4 sMVAR Model

Statistical dependencies between multivariate time series were described with a MultiVariate AutoRegressive (MVAR) model. Each vector in the TDE includes a subset of past history points, and so a subset MVAR model can be used to parametrise the predictive power of each point onto a subse-

quent value. This can be written as:

$$\hat{\mathbf{x}}_t = -\sum_{j=1}^{p} A_j \mathbf{x}_{(t-j\tau)} \tag{5.3}$$

A vector $\mathbf{x}$ containing the samples from n locations at time t is modelled as a linear combination of p past history values of the n time series. A_j is a n x n MVAR coefficients matrix at lag j in which a_{1k} represents the interaction/influence of $x_j(t-k)$ has on $x_1(t)$. τ represents the delay between past history values used as predictors in the model.

Assessing p and τ

The TDE and MVAR parameters p and τ are critical for ensuring that any results are valid. As outlined in chapter 2, there are several methods for selecting these parameters, however it is not the case that there is an objectively correct or optimal choice, moreover any results are only valid for the data in question. As such, the TDE parameters are optimised to ensure that τ is sufficiently large to avoid including interdependent predictors in the MVAR model. In addition, the choice of p (and further refinement of τ) will come from identifying the state-space reconstruction that best allows for prediction of subsequent observations.

This second criterion was assessed using the autocorrelation and Mutual Information function (Stam, 2005; Abarbanel and Parlitz, 2006) which was assessed within each epoch at each node from $\tau = 0$ to $\tau = 200$. The delay at which the ACF and MI function reached $\frac{1}{e}$ of its initial value was used to indicate the preferable delay. The criteria have previously been shown to provide different results as the ACF is only sensitive to linear dependencies. The non-linear dynamics literature has suggested that the MI function may outperform the ACF (Fraser and Swinney, 1986). As MVAR models are inherently linear, this preference for the MI function may be less pronounced. As such, both functions will be investigated.

The MI and autocorrelation results were used to constrain a systematic

search of the parameter space across both τ and p specifically designed to identify models which best predict future observations. Akaike's Information Criterion (AIC) (Akaike, 1974) was calculated from the residuals of fitted sMVAR models with each combination of parameters. The AIC is a trade off between the accuracy and complexity of a model. The accuracy term is taken as the negative log likelihood of the residuals from the fitted model, while the complexity term comes from k - the number of parameters.

$$AIC = 2k - 2log(L) \tag{5.4}$$

An ideal model is typically taken as the one with the lowest AIC estimate, that is the model which is most accurate without becoming too complex. The mean and variance across epochs of these AIC estimates were used to identify an optimal set of parameters which provide a good, consistent model with a good frequency resolution ($\triangle f < 10$Hz).

Model Fitting and Validation

Four methods were used in model validation.

- Percent Consistency (Ding et al., 2000) assesses the amount of the auto and cross correlation structure in the data captured by the model. This is expressed by comparing the the auto/cross correlations in the real data and the model's fit.
- Stability Index (Lütkepohl, 1991) - The reverse characteristic polynomial has no roots in or on the unit circle. A simpler test based on this is that the modulus of the largest eigenvalue of the A_1 matrix should be less than 1. This indicates that we have a stable solution, which implies stationarity.
- The Durbin-Watson tests for serial autocorrelation in the model residuals. Though serial autocorrelation in the residuals will not affect the consistency of the estimated parameters it may lead to an under or

over estimation of the standard errors, biasing any subsequent statistical tests.

- R^2 - Coefficient of determination, this states how much variance of the data we are explaining.

5.2.5 PDC and Directionality

PDC was calculated from the parameters of the fitted model in each epoch. This yields a value for the directed influence in each direction of each connection. A paired-sample t-test was used to establish directionality and non-parametric shuffle permutations used to estimate significance thresholds (Nichols and Holmes, 2002).

5.3 Results

5.3.1 Parameter Search

To ensure that the system is not oversampled, the Autocorrelation function (ACF) and the MI were assessed (see Figure 5.2. The average first zero crossing of the ACF was around 1500 samples, corresponding to around $360ms$ at $4069.17Hz$. The first zero-crossing of the ACF is known to overestimate the optimal delay when reconstructing a dynamical system, particularly in the presence of noise (Ragwitz and Kantz, 2002), as an alternative the Mutual Information (MI) function is estimated as a function of delay as outlined in Fraser and Swinney (1986) with the optimal threshold given by the delay at which the MI/ACF estimate reaches $\frac{1}{e}$ of it's original value (Stam, 2005). Figure 5.2 shows the ACF and MI functions across participants. Both functions are highly variable in the very shortest delays until $\tau \approx 5ms$. The slope then becomes much smoother before reaching an elbow around $20ms$. The $\frac{1}{e}$ value of the MI/ACF estimates were highly inconsistent, with the optimal delay for the ACF being around $5ms$ and the MI around $200ms$.

Part of this inconsistency may be driven by the very large value of the first lag of the ACF, this one sample delay corresponds to approximately

.025ms and is around double the value seen for the following autocorrelations. In contrast the first lag of the MI function is much lower. This first lag is very short and unlikely to reflect delayed interactions due to conduction delays. The remaining parts of the ACF and MI function are quite similar and both show a slight elbow at a lag of around 20ms. Though this suggests a possible value for τ, neither the ACF or MI function directly address our key criterion for establishing the TDE parameters; "Which values of p and τ best allow us to predict future observations?".

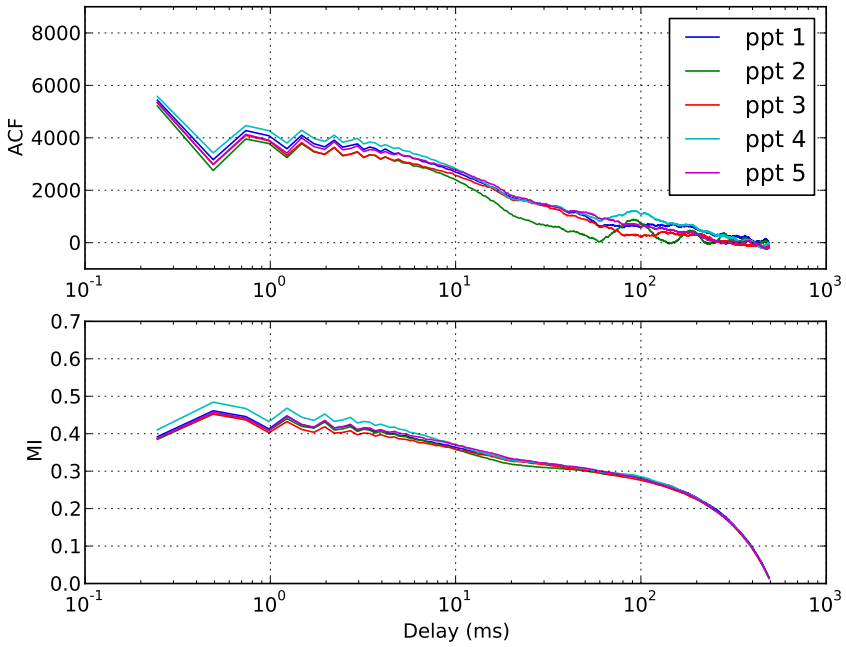

Figure 5.2: *top row* - The autocorrelation function for each participant. The ACF is calculated for each epoch for each node and the average per participant is plotted. *bottom row* - The mutual information as a function of delay.

To address this critical question more directly and build upon the ACF and MI findings, a wider search of the parameter space was conducted. An MVAR model was fitted to the data for a wide range of p and τ values.

Akaike's Information Criterion was estimated from each model and used to establish which model provided the reconstruction which allowed for optimal prediction of future observations. The optimal model was considered to be the model with the lowest but most consistent AIC estimate across all epochs.

The average and standard deviation of the AIC estimates across all epochs for each point in parameter space are shown in figure 5.3a and b respectively. Three main points can be seen from these plots.

- The very short delays produce unstable AIC estimates across trials. These delays ($\tau < 3ms$) are not physiologically plausible, as even a rapid $10m/s$ transmission speed would only be able to travel a couple of centimetres in this time. As such, it is likely that the low AIC estimates in this region arise from fitting only the very top of the autocorrelation function.

- Delays between $5 - 10ms$ have a higher AIC than where $\tau < 5ms$, however these models are are much more stable across trials and fall into a more plausible range. In addition the AIC descends monotonically with increasing model order, as has been previously reported in the literature (Brovelli et al., 2004; Supp et al., 2007; Gow et al., 2008; Gow and Segawa, 2009). Moreover this time-range contains the point at which the ACF reaches $\frac{1}{e}$ of it's initial value.

- longer delays ($\tau > 10ms$) begin to show local minima in the AIC function. These models are consistent across trial until the order passes the minima at which point we have considerable trial by trial variation.

The chosen TDE parameters were taken with a delay sufficient to minimise interdependence between observations with a sufficient order to allow for reasonable frequency resolution ($\triangle f < 10$Hz). A deeper discussion of the issue of frequency resolution can be found in chapter 3. The final values selected were $\tau = 5ms$ and $p = 20$.

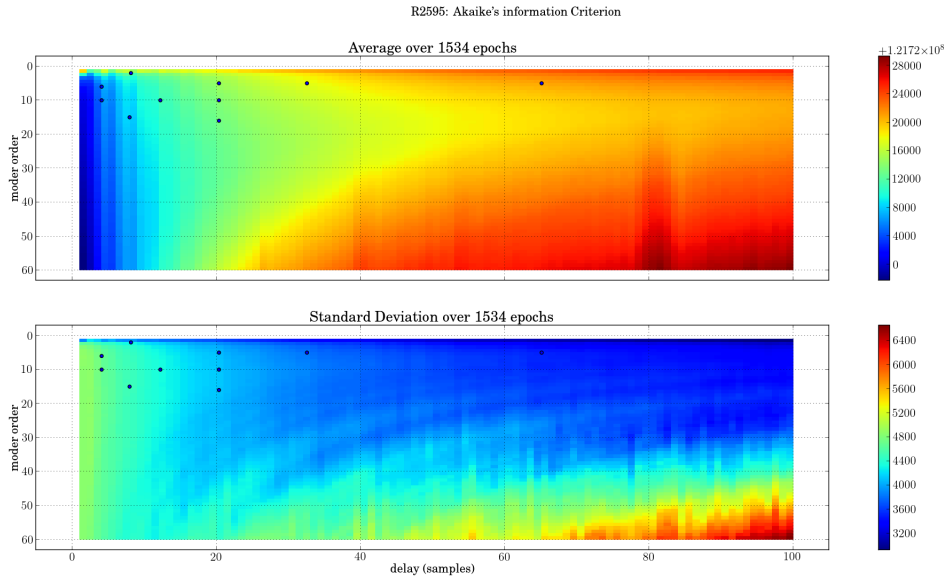

Figure 5.3: *Top* The average AIC estimate from an MVAR model fitted with the model order (p) indicated in the y-axis and the delay (τ) indicated in the x-axis. A lower AIC estimate indicates a "better" model (accurate without being too complex). *Bottom* The standard deviation of AIC estimates from MVAR models fitted with the model order (p) indicated in the y-axis and the delay (τ) indicated in the x-axis. A lower variance indicates that the quality (indicated by the absolute AIC estimate) of the MVAR model fitted to each epoch was consistent across all epochs. *Overall* The blue dots indicate TDE parameters that have been used in previous neuroimaging studies using MVAR methods, details of these studies are included in Appendix D.

Participant	Epochs	PC	D-W(τ)	SI	R^2
1	1159	42.6	1.89	0.42	0.18
2	1460	41.2	1.86	0.42	0.21
3	1353	45.4	1.9	0.39	0.19
4	1479	48.7	1.87	0.45	0.21
5	1482	51.4	1.88	0.4	0.24
Average	1386.6	45.86	1.88	.416	0.21

Table 5.2: Model assessment and validation measures for 5 participants. Averaged across $1300 - 1500$ epochs. *PC* - percent consistency (Ding et al., 2000), *SI* - stability index (Lütkepohl, 1991), *D-W* - Durbin Watson test.

5.3.2 Model Validation

Four measures of model fit and validity were applied to a fitted model from each participant (see table 5.2). These indicated that the model was able to explain around 21% of the variance in the data and 45% of its auto/cross correlation structure. The Durbin-Watson test indicated that there was no serial autocorrelation in the data at the τ used in the modelling and the stability condition indicated that a stable solution was found for all participants.

The values for the PC and R^2 reported here are lower than some recommended values in the literature (Ding et al., 2000). This is because many past studies have fitted models with delays of 1 and as such, they oversample the upper part of the autocorrelation function leading to high variance explained. The present approach sparsely parameterised the autocorrelation function, though this may lead to a more accurate reconstruction of the critical dynamics it explicitly and by design does not parameterise the very short delays, which may carry information about subsequent observations.

5.3.3 Directional Interactions Within The DMN

The fitted A matrices within each epoch were used as the basis for a partial directed coherence analysis. Statistical significance for each of the PDC estimates were computed using the analytic confidence limit defined in Schelter

et al. (2006). The average PDC estimate across epochs for each participant are presented in figure 5.4

The PDC seen in figure 5.4 has a distinctive $\frac{1}{f}$ profile. This power distribution can be seen in the signals arising from neural data in both EEG and MEG. The existence of this profile in the PDC estimates may arise from two factors. Firstly, as the lower frequency signals have a larger magnitude, the signal to noise ratio of any phase differences may be higher. As such, the MVAR is more likely to capture the relationship, resulting in higher PDC estimates. The second factor is of consistency within the analysis window. The MVAR model can only characterise relationships that are consistent across a two second analysis window. The lower frequency oscillations are likely to have slower dynamics that faster oscillations. We may therefore expect that the low frequency phase differences may be more consistent across a two second window than seen in a higher frequency which may show far more complex dynamics within the same time-window.

A further contrast between the two directions in each connection was performed to establish which connections show asymmetrical connectivity. A paired t-test between the PDC estimates in each direction of each connection was computed for each participant and significance levels assessed using non-parametric label permutations with a maximum statistic approach to control for type-1 errors. The results of this contrast can be seen in figure 5.5

DMN Hubs

Figure 5.5 shows that the two 'hub' nodes in the DMN, the Posterior Cingulate Cortex (PCC) and ventral medial prefrontal cortex (vMPFC) have very little consistency in the direction of interaction across participants at the lower frequencies ($< 20Hz$). Above $20Hz$, the vMPFC exerts a driving influence on the PCC for several participants with a peak around 70 in which three out of the five participants showed this directionality.

The PCC itself exerts a highly consistent driving influence on the middle frontal gyrus (MFG) and posterior middle temporal gyrus (pMTG). This effect is present for all five participants across the whole frequency spectrum.

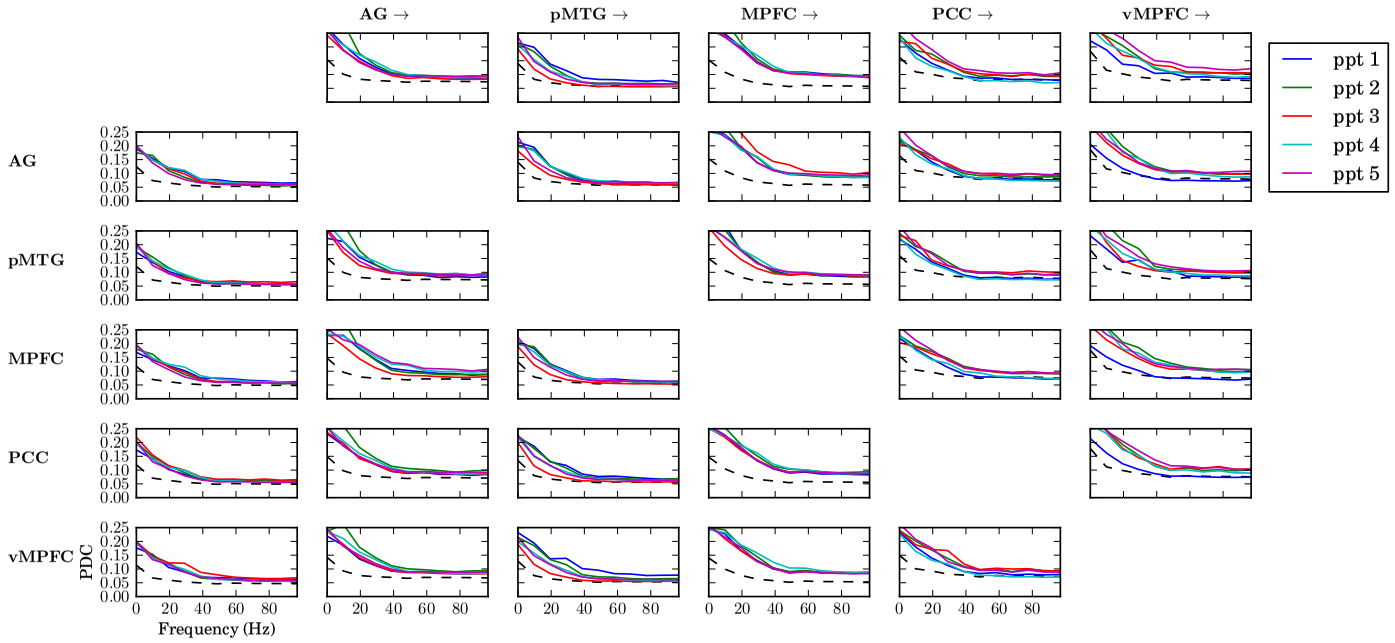

Figure 5.4: PDC within the default mode network for 5 participants. The x axis denotes frequency in Hz and the y-axis denotes the PDC estimate. Each sub-plot contains the PDC estimate for the connection in which the region in the column is driving the region in the row. The dotted line indicates the confidence limits for one exemplar participant as defined in (Schelter et al., 2006) at $\alpha = .001$.

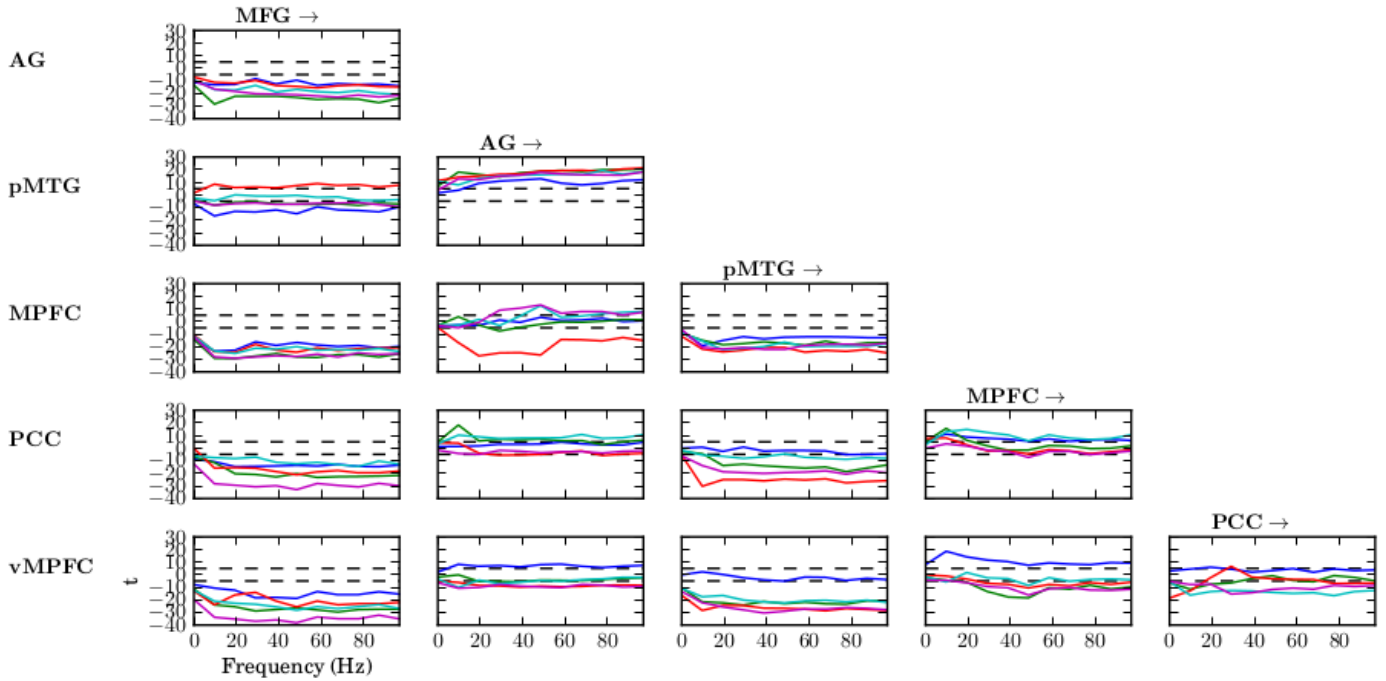

Figure 5.5: Directionality within the default mode network for 5 participants. The x axis denotes frequency in Hz and the y axis t-values from a independent samples t-test between the two directions in that connection. A positive value indicates that the regions in the column is exerting a driving influence on the region in row and a negative value indicates that the area in the row is driving. A value within the dotted lines indicates that there was no significant difference between the influence in the two directions of that connections. This does not indicate that the two regions are not communicating, rather than any communication is symmetrical in strength.

In addition, the PCC drives the dorsal MPFC node in four participants around $10Hz$. This effect is less consistent at higher frequencies ($> 10Hz$) where only one or two participants show significant directionality. There is little evidence for directionality in the interactions in between the PCC and AG. This does not imply that there is no integration between the two nodes, figure 5.4 shows that both directions within this connection demonstrate significant PDC. The lack of directionality implies that these influences are equally weighted.

The two MPFC nodes show a similar directionality profile to the PCC, including a strong driving influence on the MFG and pMTG across all tested frequencies. Moreover there is again little evidence for consistent directionality in d/vMPFCs interactions with the AG with most frequencies in these connections only reaching significance in one or two participants. The exception being the vMPFC driving the AG in three participants around $40Hz$.

The interactions between the dorsal and ventral MPFC show a slight preference for directional influence from the ventral node to the dorsal. This is most consistent across participants below $20Hz$ and above $70Hz$.

5.3.4 Angular Gyrus

As outlined above, there is little evidence for consistent directionality in the connections between the AG and the PCC/vMPFC and dMPFC. In contrast the AG exerts a highly consistent driving influence over the MFG and pMTG across all frequencies. The directional influence between AG and pMTG is most consistent across participants from $20Hz$ and above, with only three or four participants showing the effect at lower frequencies.

5.3.5 pMTG/MFG

The pMTG and MFG are very consistently driven by all of the other nodes. This directionality is broadly consistent across frequencies and present for four or five out of the five participants.

The interaction between pMTG and MFG shows some evidence for a driving influence coming from pMTG in two participants. This effect is

present in three participants around $40Hz$.

Several connections show significant directionality across the 5 participants. Critically the connection between the two 'hub' nodes in the DMN shows that the posterior cingulate exerts a driving influence over the medial prefrontal gyrus consistently across time, frequency and participant.

5.3.6 Between Subject Variability

Though several connections show very consistent effects across all participants (PCC→MFG, PCC→pMTG, d/vMPFC→MFG and AG→pMTG) many others show considerable variability. The connections between the AG and PCC/vMPFC/dMPFC are the most variable as seen in Figure 5.5. In each of these connections the average t-value from the directionality test is very small, though individual participant do show an inconsistent pattern of significant directionality. The connection between AG and PCC shows one participant with a strong directionality effect from PCC to AG, in contrast the opposite effect is seen in one participant around $30Hz$ and two participants at frequencies greater than $50Hz$. Similar patterns can be seen in the PCC - vMPFC and pMTG - MFG connections.

5.4 Discussion

This study set out to characterise the interdependencies between nodes in the DMN from a reconstruction of the rapid dynamics of the system. This novel approach considered the interplay between regions within the network on a much finer time-scale than previously reported in fMRI or MEG. We observed a complex parameter space and established a physiologically plausible set of parameters for the time delay embedding. MVAR models fitted with these embedding parameters yielded PDC estimates which revealed a complex set of directional interconnections.

The pattern of directional connections within the DMN is complex, however subsets of nodes within the DMN appear to have similar functional connectivity profiles. v/dMPFC, PCC and AG all exert a large influence on the

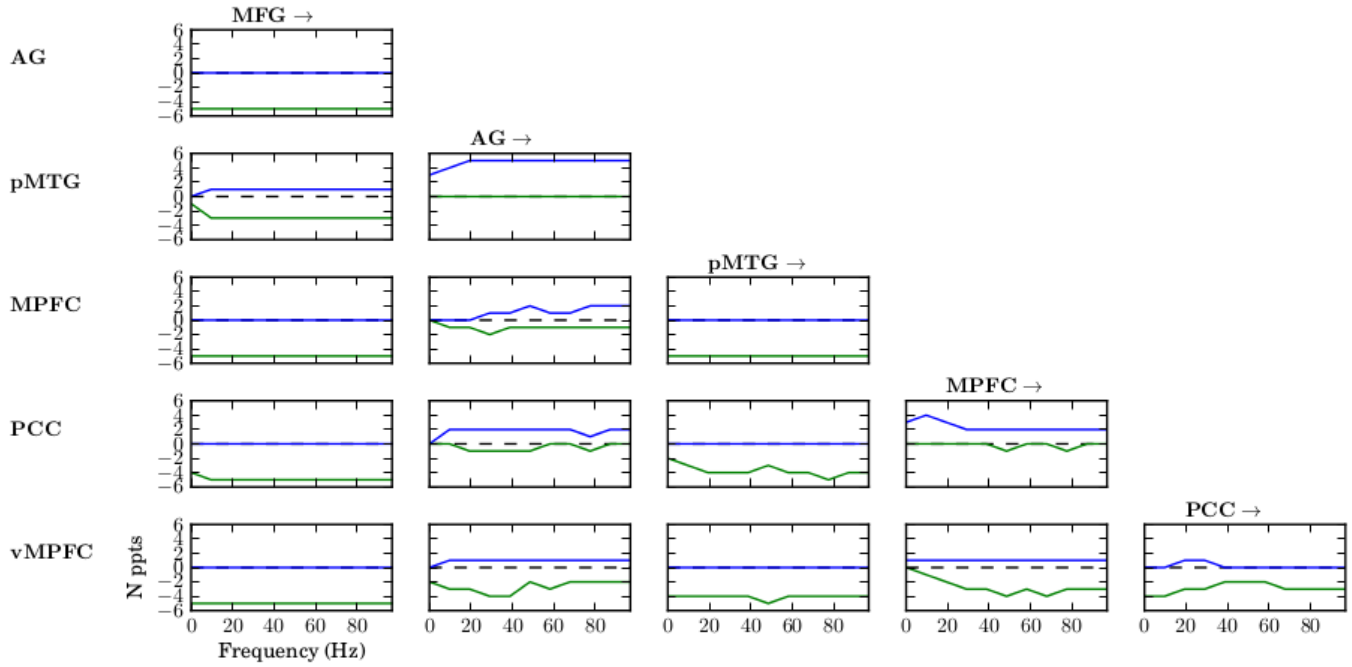

Figure 5.6: Directionality within the default mode network for 5 participants. The x axis denotes frequency in Hz and the y axis shows the number of participants showing a significant effect in that direction the two directions in that connection. Positive values in blue indicate the number of participants for whom the node in the column exerted a driving influence on the node in the row. Negative values in green indicate the number of participants for whom the node in row exerted a significant driving influence on the node in the column.

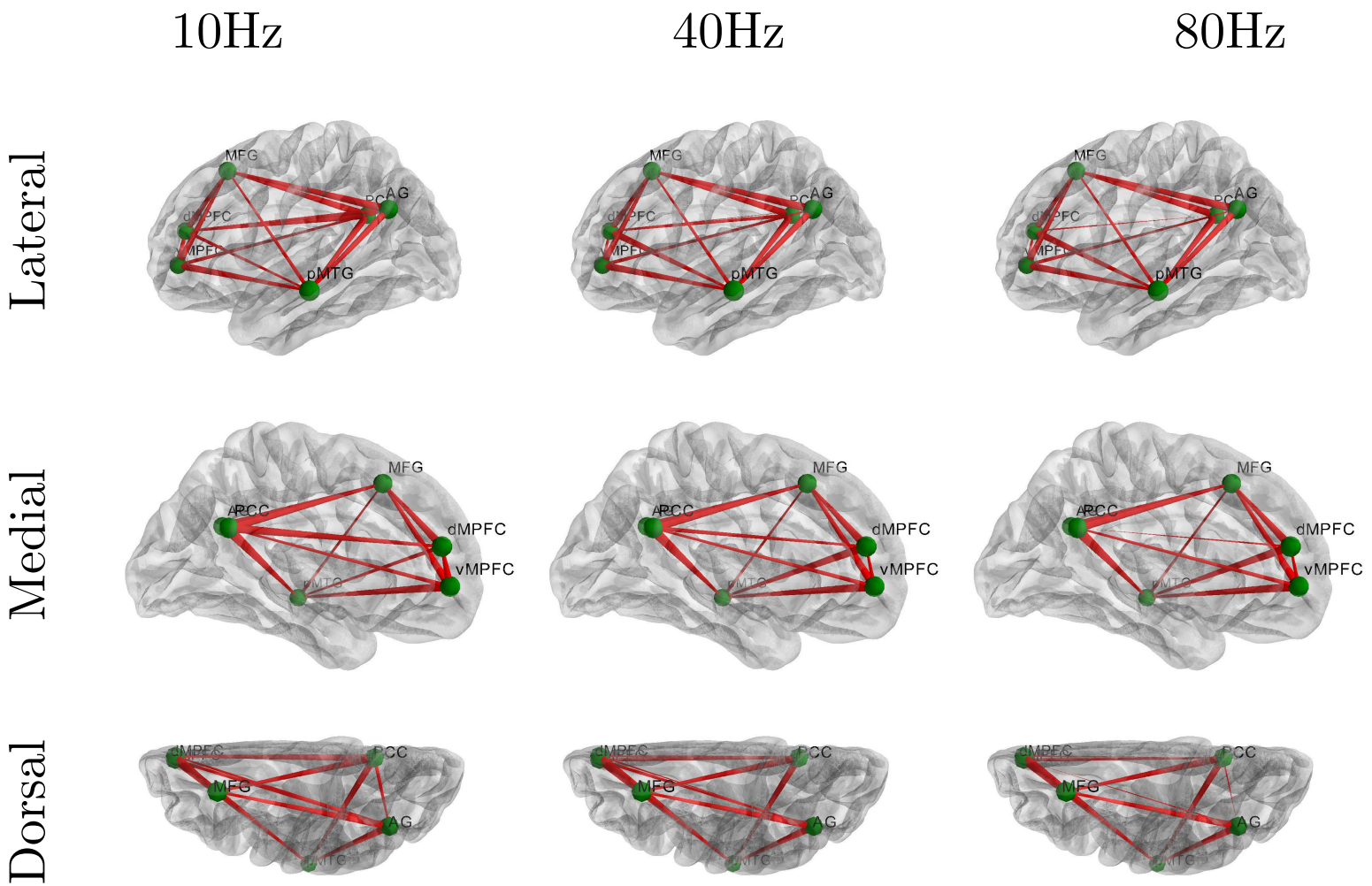

Figure 5.7: Schematic representation of the directional connections within the DMN which are most consistent across participants. The PDC estimates across all two second windows for both directions of each connection are contrasted. The thickness of the edges represent the number of participants for whom the connection was significantly directional, the tapering represents the directionality itself with the thicker end denoting the source and the thin end the target. The network is represented in 3D within a glass brain. Three different viewpoints on this are shown.

MFG and MTG. Moreover there are consistent driving influences between the two anterior driving nodes and the posterior driving nodes. The most variable directionalities occur in the connections between these anterior and posterior nodes.

5.4.1 Anterior Driving Nodes

The two nodes in the Medial Prefrontal Cortex exert a strong driving influence on both MFG and MTG in all five participants across all frequency bands. In addition, the interactions between the two medial frontal nodes shows a very strong tendency for the vMPFC to drive dMPFC, particularly at less than $40Hz$ and greater than $70Hz$. Though consistent, this interactions is quite weak with t-values only just reaching significance.

There is a weak driving influence from the ventral MPFC to the PCC is not significant in more than one participant in low frequency bands (<20Hz) but becomes very consistent at higher frequencies. A weak directionality in this connection (vMPFC $\rightarrow$ PCC) is consistent with previous fMRI findings (Di and Biswal, 2014; Jiao et al., 2011). These regions form the first key regions in the DMN (Buckner et al., 2008) and a clear picture of a driving influence from MPFC to PCC is building up.

5.4.2 Posterior Driving Nodes

The Angular Gyrus exerts a very consistent influence on pMTG and MFG, however the rest of it's connections are relatively inconsistent across participants. The exception being a driving influence from vMPFC to AG around $40Hz$. AG is structurally connected the pMTG by the arcuate fasisculus (Catani et al., 2005) in addition there is strong evidence that the posterior AG is functionally connected to the posterior cingulate and veteromedial prefrontal cortex in fMRI though there is less evidence for a direct structural connection (Uddin et al., 2010). These latter two connections form a crucial part of the DMN and are relatively symmetrical in terms of directional influences. However the strong driving influences seen here are inconsistent with previous fMRI results (Jiao et al., 2011). This inconsis-

tency may reflect the time-scale on which the interactions occur, sub-second influences from the AG to the rest of the DMN might be invisible to an fMRI analysis but may be readily apparent in MEG.

In contrast to the relationship between PCC and vMPFC discussed above, the PCC exerts influence over the dMPFC. This is seen most consistently across participants in the alpha band around 5-15Hz. Previous fMRI findings have suggested that the PCC is primarily driven by other regions within the DMN (Jiao et al., 2011; Yan and He, 2011). In contrast the current MEG evidence suggests that the PCC exerts a consistent influence on both MTG and MFG with a less consistent driving influence on dMPFC. Moreover, this finding is consistent with a recent EEG study showing that PCC exerts a driving influence on the anterior cingulate, primarily in the theta and alpha frequency band (de Pasquale et al., 2010).

The PCC and MPFC are structurally connected by the Cingulum; a white matter bundle which projects along the cingulate gyrus (van den Heuvel et al., 2008; van Oort et al., 2014b). Diffusion tensor imaging (DTI) data measuring the diffusion of water along white matter bundles has suggested that participants who show higher fractional anisotropy in the Cingulum also show greater DMN functional connectivity as estimated with fMRI (van den Heuvel et al., 2008). The present results build on this literature in two ways. Critically the interactions arising from this connection show frequency specific asymmetries. In addition, there is different directionality in the connections between the PCC and the d/vMPFC, suggesting that fine grained functional parcellation within the MPFC might be associated with differences in directional influence.

5.4.3 Driven Nodes

The MFG and pMTG are consistently driven by each of the other nodes in the network with the exception of each other. The interactions between these nodes shows a slight preference for pMTG to drive activity in MFG, most consistently in the low gamma range. The MFG has been identified as a predominantly ‘driven‘ hub in resting state fMRI (Yan and He, 2011).

5.4.4 Frequency Specific Connections

Two connections show large differences in directionality as a function of frequency. The vMPFC $\rightarrow$ PCC is weakly directional, however the directionality is most consistent at higher frequencies. In contrast the PCC $\rightarrow$ dMPFC connection is only significant around $10Hz$, though again this is a relatively weak directionality.

These two connections suggest that the interactions between the primary nodes in the DMN are critically dependent on frequency. There is growing evidence that alpha band oscillations may govern top-down modulatory processes while the bottom up processing is done in the gamma band (von Stein and Sarnthein, 2000). In this context the PCC $\rightarrow$ vMPFC connection might reflect the modulatory top-down influence or integrative processes such as decision making (Donner and Siegel, 2011), while the dMPFC $\rightarrow$ PCC connection reflects interactions more relevant to stimulus driven bottom up processes. Though this theory is very influential in neural oscillations related to specific tasks, it is less clear what "top-down" and "bottom-up" might mean in connection between two heteromodal brain regions such as the PCC and MPFC. Further investigation into the task modulation of the interactions between these brain regions might provide deeper insight.

5.4.5 Individual Differences

Many of the asymmetries identified in this study are consistent across 5 recording sessions for all 5 participants. In contrast, some connections show considerable individual differences, for instance connection between MFG and MTG has participants showing a significant directionality in both directions of the connection. A portion of this variance may reflect estimation/fitting error, however it is possible that variance in these connections may have a cognitive or perceptual correlate that reflects individual differences in state-of-mind in the scanner.

5.4.6 Methodological Statements

The reconstruction of rapid dynamics from very short time delays in this study allows for the inference of directional influence within RSNs. This has not been possible with previous methods based around instantaneous correlation/coherence based measures, which do not take delayed interactions into account and explicitly focus on dynamics over much larger time scales. Moreover it has been shown that data preprocessing methods such as bandpass filtering can distort measures of G-Causality (Barnett and Seth, 2011). Specifically, the smoothing effect of a filter can dramatically affect the empirical model order estimate making it an inappropriate step in estimating causal influences within a system (Barnett and Seth, 2011).

The time-delay embedding method used here is optimised to find the reconstruction of past observations that best predicts future observations. This provides a mechanism by which the TDE/MVAR parameters p and τ can be established relatively objectively. Moreover the inclusion of τ as an explicit parameter in the MVAR may allow for more consistency in the MVAR and neuroimaging literature in which different studies may use very difference sampling frequencies in their models.

In addition, though state-reconstruction through TDE can naturally lead to directional connectivity estimation with MVAR modelling it can be applied to other approaches. These include the estimation of Markov Chain models (Ragwitz and Kantz, 2002) and more traditional non-linear dynamical methods. MVAR modelling is a linear approach, while this provides many advantages in stability and simplicity the brain is a highly non-linear system. This limits the amount of variance that the model can account for. As such, the application of TDE state reconstruction to more complex bilinear or non-linear equations may provide a more complete account of the underlying neural dynamics. In addition, as time-delays are more widely incorporated into computational models of neural dynamics, TDE could potentially provide a mechanism for establishing these quantities from observed data.

5.4.7 Future Directions

There is considerable evidence that many critical brain networks are disrupted as a result of clinical or psychiatric disorders (Buckner et al., 2008). The present study has characterised network interactions in neurally typical participants however this method may provide a critical insight into disruption of directional information flow in clinical disorders.

The directionality of the connections within the DMN demonstrated in this chapter represent a step forward in our understanding of the structure of this network. Critically the idea that the MPFC and PCC are highly influential can now be explicitly modelled and found within several participants. This is, however, a static picture of these interactions. Much of the computational modelling done on resting state brain networks have emphasised the dynamic nature of the network structure within the brain. Even suggesting that the structure changes over time as the brain explores the dynamic repertoire allowable from the underlying anatomical skeleton. Further research looking to explicitly characterise these dynamic changes in brain connectivity would provide a link back to the computation modelling literature which currently works mainly on dynamics on an fMRI timescale. Moreover the long (2 second) time windows used here will reflect brain connectivity which is stable over a long time period for the brain. Critically, this window is too long for a direct investigation into the task related dynamics which are known to unfold on a millisecond timescale.

The cognitive/perceptual relevance of the connections identified in this chapter remain unclear. A critical avenue for future work could look for the cognitive/behavioural correlates of the individual differences in connection strength or directionality within brain networks. This would provide strong evidence for the role of different parts of the network. This may be simpler to achieve within an explicitly task related brain network, though the DMN is a compelling example of ongoing interactions within a brain network it has been linked to a great many cognitive processes. Moreover, these tend to be rather abstract processes such as self-referential thought which may be challenging to manipulate experimentally.

5.4.8 Conclusion

The critical novel contribution of this chapter is an illustration of how time delay embedding can naturally provide a principled foundation for MVAR models and subsequent connectivity and information flow analyses. Previous literature in this area has typically focused on instantaneous measures of neural synchrony. Moreover previous analyses incorporating delayed interactions, such as MVAR, have not investigated the choice of τ leaving it set implicitly by the data sampling rate. Finally, detailed computational models have begun to incorporate delayed interactions and emphasised their importance in neural function, however to date these models have been used to show that a system of delayed interactions between nodes may produce the slow-wave fluctuations in BOLD seen in fMRI. Though this represents an important step forward, these models demonstrate that such interactions are capable of producing the observations we make rather than inform us about the nature of the actual interactions in the brain. This work provides a middle ground between these three literatures, building upon the large literature on instantaneous synchrony in resting state networks by using TDE to make an informed choice about the delay and order in the MVAR model which characterises information flow within a network.

This study has provided the first insight into the directionality of information flow within the DMN during a resting MEG scan. Many of the connections within the DMN showed significant asymmetries which were consistent with previous fMRI and EEG studies, however these results challenge the claim that the PCC is a primarily driven node. Moreover the extent to which the PCC is driving or driven by the medial PFC may modulate as a function of frequency.

Chapter 6

Dynamic Changes In Within Network Directed Connectivity

6.1 Introduction

The previous chapter demonstrated that there are many asymmetrical connections within the DMN, in that there are significant differences in the directed functional connectivity between the two directions of a single connection. This builds upon previous work demonstrating the functional connectivity between these regions in multiple neuroimaging modalities, however we are still left with a static image of brain connectivity. The TDE does capture the dynamic interactions between brain regions, however the MVAR model is only able to parameterise any interactions which are consistent within a single epoch. This chapter uses Hidden Markov Modelling to characterise structure in the modulation of PDC on an epoch-by-epoch basis. In other words, attempting to model dynamic changes in the interactions within the default mode network.

6.1.1 Non-Stationarity

Fluctuations in band-limited power correlations within resting state networks have suggested that networks such as the DMN may form transiently and switch between periods of relatively high or low connectivity (de Pasquale et al., 2010). Such fluctuations in within network functional connectivity have been robustly identified in a number of contexts and modalities (de Pasquale et al., 2012; Chang and Glover, 2010) and have become a feature of many computational models of resting state brain function (Deco et al., 2013). Indeed, this non-stationarity in within network functional connectivity has been highlighted as a critical area for future work (Cabral et al., 2014a).

It may be the case that these non-stationarities reflect "noise" in the brain rather than structured dynamics, in which case such spontaneous or background activity would have no functional role. However this view has long been thought of as an over simplification (Pinneo, 1966). Recent computational evidence suggests that the spatio-temporal structure of the resting brain cannot be replicated without the presence of some neural noise to drive "exploration of the brain's dynamic repertoire" (Ghosh et al., 2008).

This idea may be pushed further when considering a Bayesian brain model in which the brain is constantly striving to generate predictions about future events and establishing optimal network structures to deal with them (Knill and Pouget, 2004). Within this context variance in predictions driven by noise could provide a mechanism which ensures that a wide range of hypotheses are tested (Deco et al., 2011).

Though the specific role of noise or non-stationarities in resting state network configurations is not well understood, current methodologies such as Hidden Markov Modelling used in conjunction with MVAR-PDC estimates are well placed to provide objective descriptions of which the dynamic repertoire of a functional network such as the DMN. In addition, this provides a critical step towards investigation of the highly dynamic brain responses to external stimuli.

6.1.2 Functional States

One very influential approach to accounting for non-stationary neural behaviour has been the identification of 'functional states'. In contrast to the abstracted states referred to in previous chapters, a functional state is a period in time in which the observed responses show homogeneous characteristics. For example, spatial configurations in EEG topologies have been identified and shown to be stable over short periods of time by segmenting EEG scalp recordings (Lehmann, 1984; Lehmann et al., 1998). Such "microstates" remain stable for around 100ms and transition in a rapid, almost step-wise way. Moreover information in the EEG microstates may be related to the resting state activation patterns seen in the much slower BOLD fluctuations in fMRI which are typically below $1Hz$ (Musso et al., 2010; Britz et al., 2010). Finally, microstates prior to onset of a visual stimulus have been show to predict subsequent perceptual awareness of the stimulus (Britz et al., 2014) suggesting that such states represent cognitively relevant neural configurations. It has further been suggested that such microstates are the "building blocks of spontaneous thinking" (Lehmann et al., 1998) in that chains of microstates might combine small steps to produce meaningful

cognitive processes.

In a practical sense, this allows for a simplification of the brain's highly complex dynamics by segmenting or clustering observed data into discrete functional states. This may be done over several time-scales from milliseconds to seconds, trading off time and frequency resolution (Koenig et al., 2005). In addition this process may be applied to either sensor or source space M/EEG data (Lefèvre and Baillet, 2009). This approach has been highly influential, however it seeks only to quantify the spatio-temporal structure of each state and does not consider the dynamics of the transitions between states themselves.

These transitions can be characterised with a Hidden Markov Model which provides a probabilistic description of transitions between states. This method has been applied to sensor MEG data to identify points in time with similar sensor covariance matrices providing a more finely tuned spatial filter for beamforming (Woolrich et al., 2013). Moreover, HMMs have been used to identify how brain networks reorganise on a sub-second time-scale (Baker et al., 2014). This approach has demonstrated that spatial topographies in the brains source space show transient functional states which remain stable for 100-200ms at a time. Moreover, the transitions between these states are not equally probable, with the brain being highly unlikely to transition directly between DMN and dorsal attention network (Baker et al., 2014). Critically, this builds on the EEG micro-state literature by providing a spatial map of each state in the brain rather than on the EEG sensors.

This chapter applies the HMM iethodology recently developed in application to MEG data (Woolrich et al., 2013; Baker et al., 2014) and apply it directly to the PDC estimates from the previous chapter. The fitted HMM will then probabilistically characterise dynamic transitions between functional states in which the directional connectivity within the DMN is relatively homogeneous. The directionality test outlined in the previous chapter will then be applied to the epochs within each functional state providing a description of how modulations in directional influence within the DMN change over time.

6.1.3 Markov Models

As outlined above, the dynamics of brain states can be characterised with a Markov model which describes the probabilities that a system moves through a series of discrete states. This section outlines the basis theory and structure of a Markov model.

Markov Chains

As described in chapter two, a system at a given point in time may be described by its state. That is, a complete description of the internal parts of the system. A completely stationary system may only occupy one state over time, whereas a non-stationary system may transition through many different states during the time it is observed.

A Markov model creates a probabilistic description of a systems change in state through time. For a discrete, first order Markov chain the probability that the system will by in one of N given states at time t is determined by the state of the system at time $t-1$.

$$P(s_t|s_1, \ldots, s_{t-2}, s_{t-1}) = P(s_t|s_{t-1}) \tag{6.1}$$

In which, $P(s_t|s_{t-1})$ is the conditional probability that the system is in each of the N states at time t given the state at time $t-1$. Equation 6.1 describes the Markov Property, and indicates that the process in question is *memoryless* in that its future is conditioned only on its present and not on its past. In contrast to the state space used in chapter 2 and the TDE-MVAR analyses in previous chapters, a Markov process uses a categorical distribution of states.

If we consider 3-state Markov process with states s_1, s_2 and s_3 . We may assume that brain activity at a given time point is well described by one of these states and estimate a set of transition probability from a sequence of observations of neuronal activity.

$$\eta = \begin{pmatrix} P(s_1|s_1) & P(s_2|s_1) & P(s_3|s_1) \\ P(s_1|s_2) & P(s_2|s_2) & P(s_3|s_2) \\ P(s_1|s_3) & P(s_2|s_3) & P(s_3|s_3) \end{pmatrix} \tag{6.2}$$

The transition matrix η then contains the conditional probabilities that given the weather in in one state, what is the probability that the weather is in the other state on the next day.

Hidden Markov Models

The examples and models discussed so far assume that the state of the system is directly observable, however this is not the case for many systems. such as the brain. For these systems, we may generalise the Markov Process to one in which the observations are some probabilistic function of a hidden (non-observable) state. This is known as a Hidden Markov Model (HMM). This process is embedded in two probabilistic functions, one hidden process governing the change in state over time and a second which produces an observation depending on the current state.

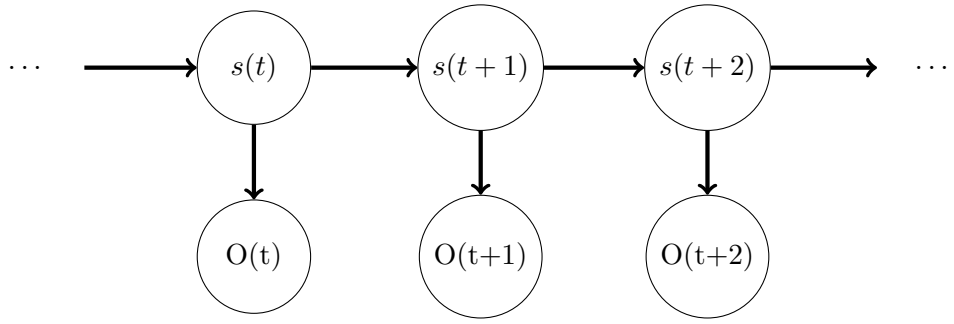

Figure 6.1: Schematic of a Hidden Markov Model. The top line shows the first probabilistic process describing how a state at a given time can condition the probability of which state is present at time $t+1$. The second probabilistic process is shown in the top to bottom arrows which illustrate the how the observation at time t is influenced by the current state alone.

An HMM is characterised by five parameters (Rabiner, 1989):

N The number of discrete hidden states.

M The number of possible observation symbols, the output of the system.

η The state transition probability matrix.

B The probability distribution of the observations.

π The initial state distribution.

The outputs of the system must have been observed and N and M must be defined prior to fitting the HMM. Once these are known the probability measures η, B and π may be computed. For clarity, the compact notation:

$$\lambda = (\eta, B, \pi) \tag{6.3}$$

is used to denote the full parameters of the model.

The first critical question concerns optimisation of the model. Given *a priori* expectations about λ and and observation sequence O, how can we optimise λ to maximise $P(O|\lambda)$. A Variational Bayesian inference as described in (Rezek and Roberts, 2005) and applied in (Baker et al., 2014; Woolrich et al., 2013) is used to train the HMM. This method is fully probabilistic and provides full posterior distributions for the model parameters. The priors for the model are chosen to be non-informative.

The probability of a given state at time t given the observed data ($P(s_t|O)$) after this training or optimisation phase is used to determine which state is active at a given time. This is done by Viterbi decoding, only the most probable *a posteriori* state at a given time point is considered to be 'on', the other less likely states are all classified as 'off'.

6.1.4 Summary

In this chapter a HMM will be used to characterise the transitions between hidden states over time within the Default Mode Network. The PDC estimates calculated in the previous chapter will be used as the observations for

this model and the hidden states assumed to reflect some specific configuration of neuronal connectivity within the DMN.

This analysis will provide several informative statistics, critically we will be able to group time points during a resting state scan during which the network connectivity within the DMN is similar. Secondly, the total time and average length of time that the brain spends in each state can be estimated. These metrics will allow for a detailed characterisation of how stable the DMN is over time and whether all connections are equally variable. We will then be able to build on the results of the previous chapter by repeating the directionality analysis within each state identified by the HMM to identify any connections whose directionality is modulated by state. Finally a finer grained interrogation into which connections are modulated by the transitions between different Markov states will be done with a state time course regression. This regresses the epoch-by-epoch PDC estimate in each connection against the state-time course, providing a measure of which states predict modulation in PDC.

6.2 Methods

6.2.1 PDC Estimates

The input to the HMM process were the estimates of the Partial Directed Coherence within the Default Mode Network calculated in chapter 4. The PDC from all two second epochs for all participants were included in the analysis.

6.2.2 PDC Dimensionality Reduction

Each PDC estimate was unfolded into a $[1, nchannels \times nchannels \times nfrequencies]$ vector containing all of the estimates for all connections at all frequencies for that epoch. A total of ≈ 6500 vectors were unfolded from all participants and used as the input to the HMM. The large dimensionality of these vectors can lead to numerical instability in the estimation of the Kullback-Leibler divergences when estimating the free energy of the HMM. As such, PCA

was used to reduce the dimensionality of these PDC vectors.

PCA was performed over the PDC vectors for all epochs from all participants, 53 principle components explaining 66.174% percent of the variance in the PDC estimates. The weighting of each component for each epoch was then computed, these weightings formed the observation sequence O for the HMM.

6.2.3 Markov Processes

An HMM with four discrete hidden states was inferred from the PCA weights. The state with the most probable posterior after fitting was identified using the Viterbi algorithm (Rezek and Roberts, 2005). The state time-course (u) was defined to indicate which of the k states was most probable for each successive two second window. HMM analyses were performed using the HMMBox toolbox for MatLab which can be downloaded from www.fmrib.ox.ac.uk/~woolrich/HMMtoolbox.

6.2.4 Within State Directionality

The observation sequence used to fit the HMM contained the weightings from the first 53 principle components of the unfolded PDC vector. As such the multivariate normal distributions defining each state only span this reduced subspace. This means that the posterior distributions cannot be readily interrogated to identify connection and frequency specific asymmetries in the DMN as one component may load onto the PDC estimate from several connections. Therefore, the state time course was used to group epochs which contained the same winning state. The original, full-dimensionality, PDC estimates from these epochs were then used to compute the directionality test as described in the previous chapter Therefore, the state time course was used to group epochs which contained the same winning state. The original, full-dimensionality, PDC estimates from these epochs were then used to compute the directionality test as described in the previous chapter.

6.2.5 State Time Course Regression

For each state and connection in turn, the state time course u which indicates whether that particular state was "on" or "off" was used to predict the PDC estimates over time. This performed two critical tasks.

Firstly, the HMM is a probabilistic description of the PDC estimates, as such it is difficult to estimate a "goodness of fit" measure that would compare the difference between the HMM and some "true" model producing the data. In order to validate the HMM, we look to establish the extent to which the fitted HMM is able to predict our data observations. The tests the null hypothesis if there were no probabilistic structure to be captured within the observations the HMM would still return four states this is determined *a priori*. Moreover if this were the case, the state-time course would be random and unable to predict variance in the observations.

One possible complication is that the four states may not act equally on all connections within the DMN. Therefore, by estimating the extent to which the states predict the PDC for each connections at each frequency we may objectively determine which connections are modulate most in by the HMM state transitions. As a consequence of computing so many regressions the statistical significance of each will be subject to a Bonferroni correction. Moreover, as the PDC distribution is highly unlikely to be Gaussian, the p-value will be adjusted by a Greenhouse-Geisser correction. Any remaining significant state modulations of PDC estimates are likely to be highly robust.

The result of this analysis will be a regression β values for each connection (both directions and for each frequency) for each state which quantifies the extent to which the state is able to predict change in the PDC in that connection. In addition the coefficient of determination is calculated for the significant β estimates. It is important to note that this β value will indicate whether the state predicts something different in the coefficients relative all other states. That is, as positive value will indicate that the PDC estimate for that state was significantly larger than in the other states.

6.3 Results

6.3.1 Number Of States

Figure 6.2 shows the Free Energy estimates from HMMs with different assumed numbers of states ($k = 2, \ldots, 10$). As the HMM inference may be affected by variance in the initialisation, the distribution of free energy is shown for each number of states. The median of the free energy distribution drops from $k = 2$ to $k = 4$ before plateauing before increasing again after $k = 8$. The median is used as the estimation procedure is affected by the random initialisation conditions and as such, some iterations may produce very different results. The median of the distribution is less susceptible to bias from any outliers in the distribution potentially caused by noise in the initialisation.

The notches in figure 6.2 indicate whether the distributions of two boxes are likely to be significantly different. The overlapping notches in the distributions for 4,5 and 6 states indicate that these distributions are not significantly different. As such, the final number of states was chosen to be four, as this is the smallest number of states with the lowest free energy distributions and as such will produce the simplest model.

6.3.2 HMM Inference

Figure 6.3 shows the state time courses for the HMM inference with four states for the PDC windows from five participants concatenated in time. A close up of one section of this plot for one participant is shown in the lower part of the figure.

Differences between the five participants can clearly be seen in this plot, indicating that there may be high individual difference in the transition matrix between participants. For participant one State 4 is highly dominant with no time spend in any of the other states. State 4 is then only revisited regularly in participants 4 and 5. Participants 2-5 all show rapid transitions in and out of States 1,2 and 3. A close up of one section of the state time course for participant five can be seen in figure 6.3. This clearly shows the

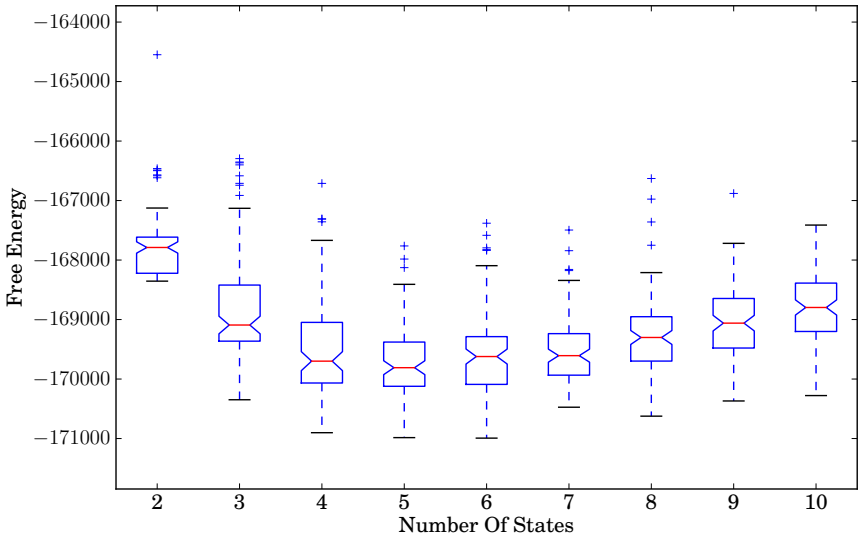

Figure 6.2: Distribution of free energy estimates from HMMs with different numbers of states fitted to the PDC estimates from the previous chapter. Lower free energy suggests a more stable solution. The red line indicates the median Free Energy for a given number of states, while the box extends from the lower to upper quartile values of the data. The whiskers show extend to 1.5 time the interquartile range ($1.5 \times (75\% - 25\%)$). The height of the notches indicate the standard error of the median, if the notches for two boxes overlap it is likely that the distributions they represent will not be significantly different.

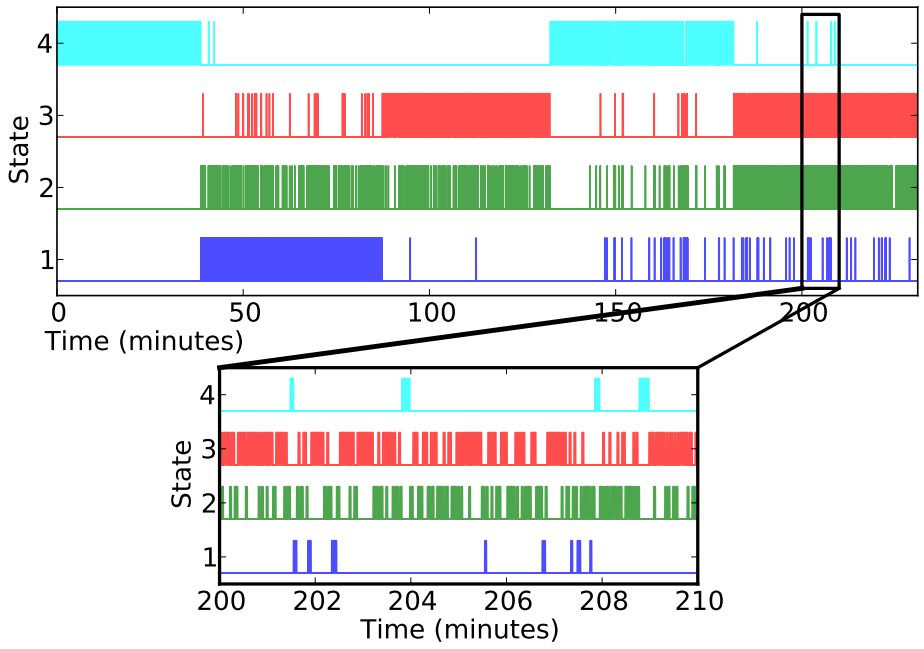

Figure 6.3: *top* - State time course for the five participants concatenated in time. Each time series represents whether one state was "on" or "off" as determined by the Viterbi decoding. Only one state may be on at a given time. The black box represents an example period of time which is shown in close up in the bottom figure. *bottom* - A close up on a ten minute segment of the state time course for one participant.

short-lived states and rapid transitions between windows.

6.3.3 State-Specific Directionality

The epochs from each of the four states were used for a directionality analysis identical to the one described in the previous chapter except that epochs from all five participants were included in the analysis. The t-values for a paired t-test between the PDC estimates for the different directions of each connection can be seen in figure 6.4.

Figure 6.4 clearly shows that the state transitions have a relatively subtle effect on the directionality within the DMN. There are no cases in which a state transition reverses the directionality in a clearly directional connection. However, several connections with no significant directionality in two or three of the state may show significant directionality in the remaining state.

Though there are differences in the directionality within the DMN as a function of state, we cannot establish whether these differences are significant from this picture. It may be the case that by dividing up the epochs from the state time course, we are simply one distribution differently in each state leading to an expected variance in the resulting directionality. The following section quantifies the strength of influence of the state transitions on the PDC estimates themselves and seeks to establish the extent to which state transitions predict connectivity within the DMN.

6.3.4 State Time Course Regression

Figure 6.5 shows the significant β estimates from the state time course regression. The first point is that despite the highly conservative Bonferroni and Greenhouse-Geisser correction to the p-values the states have been shown to predict the PDC values for several connections. This suggests that the HMM has been able to characterise some probabilistic structure within the PDC estimates. Critically, this is not the case for all connections. The MTG $\rightarrow$ dMPFC connection is not predicted significantly by the presence of any state. In addition, several other connections are only predict during the presence of one state. For instance, MTG $\rightarrow$ AG, MTG $\rightarrow$ PCC and

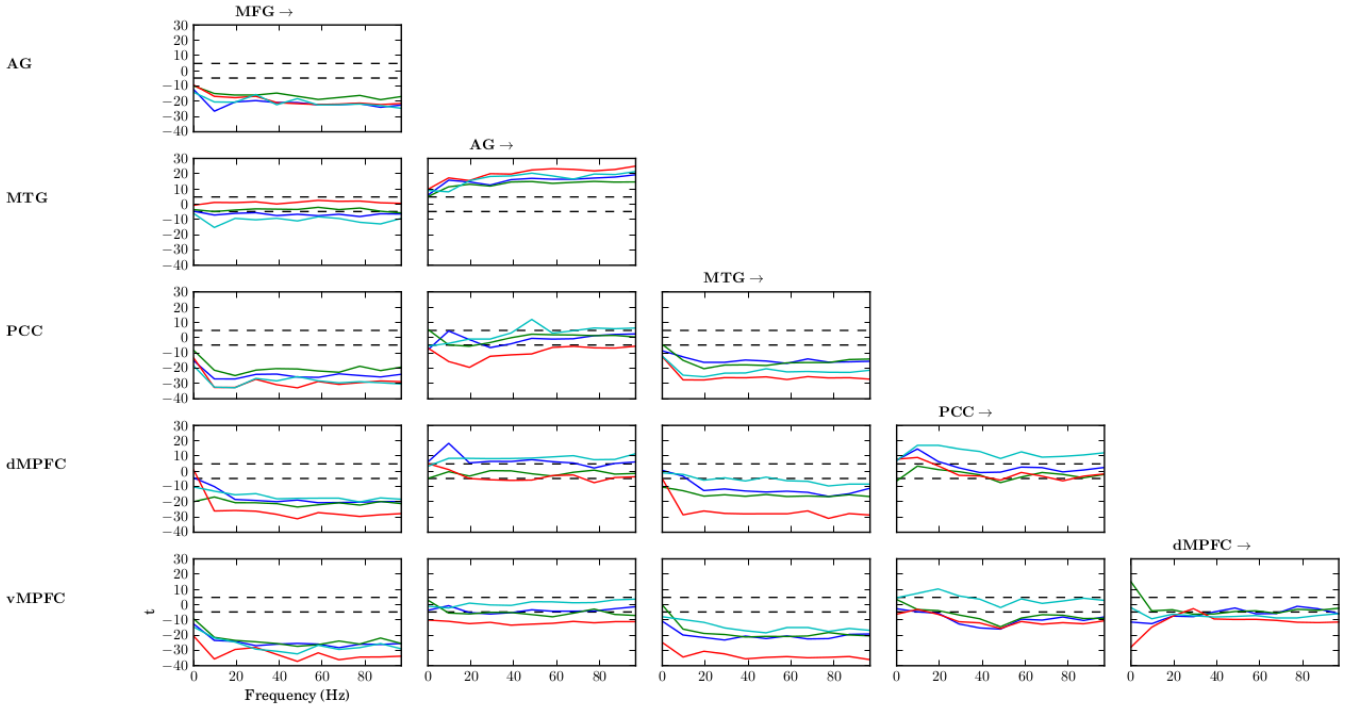

Figure 6.4: The result of the directionality test for all the epochs across participants within each of the four states, the dotted lines are the significance thresholds for $\alpha = 0.01$. The x-axis indicates frequency in Hz and the y-axis indicates t-value. A negative t-value indicates that the directional connection from the region in the row was larger than the influence from the area in the column to the row.

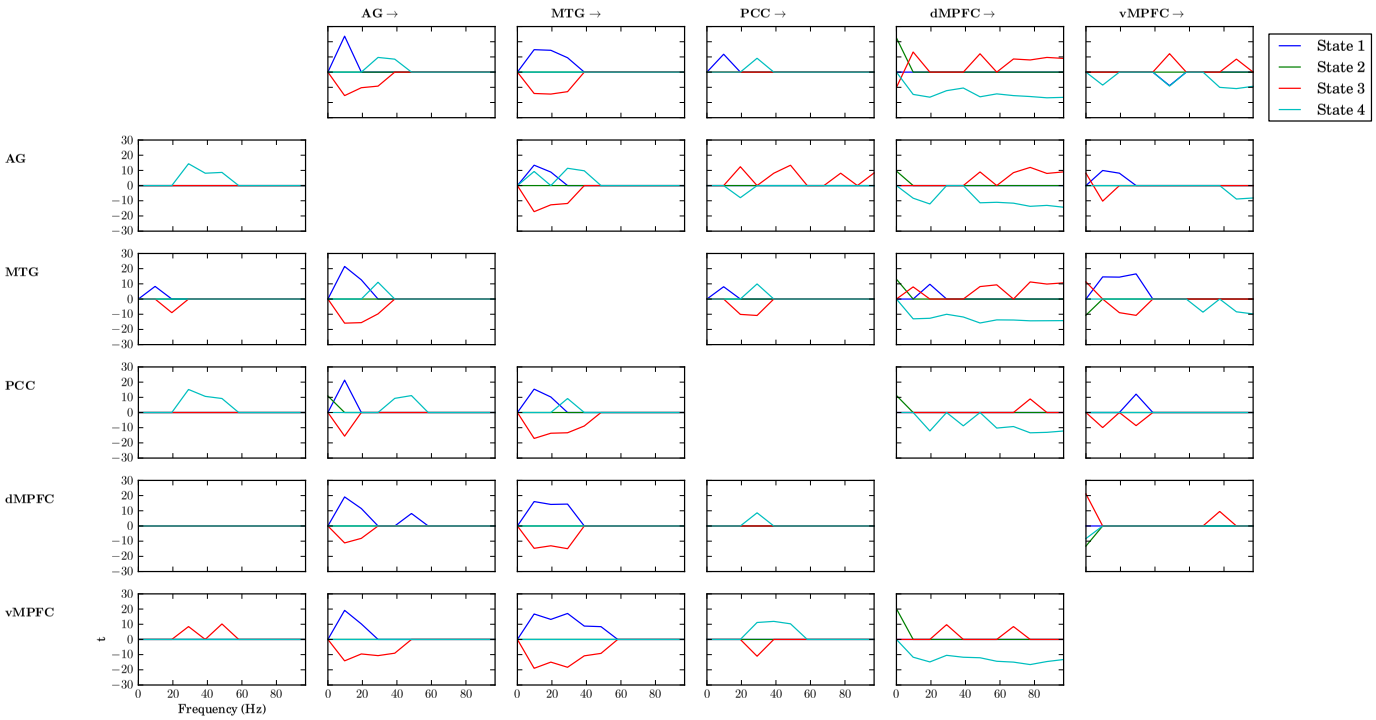

Figure 6.5: The significant β values from the state time-course regression analysis, non-zero values indicate that the presence of a given state significantly predicted epoch-by-epoch PDC estimates in that connection. Significance thresholds were $\alpha = 0.05$ with Bonferroni and Greenhouse-Geisser correction. A positive β value indicates that the presence of the state predicted a significant increase in the PDC estimate in that connection while a negative β would predict a decrease in PDC while that particular state was "on".

PCC → dMPFC are only predicted by state 4, even then only between 20 and 60Hz.

States 1 and 3 have opposing influences on many connections. Critically, during state 1 the PDC from AG and MTG to the rest of the network is significantly increased, whereas the same connections are decreased during state 3. Again this is a frequency specific effect, the modulation only occurring between 10 and 30Hz for most connections. Indeed, there are no connections for which state 1 predicts a decrease in PDC, suggesting that this state reflects the time points at which the DMN is maximally connected. In contrast, the decreased output from AG and MTG predicted in state 3 is accompanied with an increase in influence in the PCC → AG connection and from dMPFC to the rest of the network (particularly at higher frequencies, above 50Hz).

State 2 does not predict modulation in PDC for any connection from 10Hz and above. As the regression analysis for each state was done independently for each state against all other states, this does not mean that state 2 does not modulate connectivity at all. Rather it may reflect an "average" or "baseline" state in which none of the other states were able to characterise anything meaningful. In other words, in the other three states we are able to explain variance in the PDC on the basis of the Markov state which is associated with a multivariate Gaussian probability density function predicting the PDC. Each of these probability functions for states 1,3 and 4 have biases for larger PDC values in certain connections, in contrast the probability function for state 2 may not contain any such biases.

Finally state 4 completely dominates the dynamics in participant 1 and characterises much of the time for participant 4 as can be seen in figure 6.3. In contrast, this state is relatively rare in the remaining 3 participants. This relatively static pattern for participant 1 may arise as 4 states were not sufficient to unfold the dynamics for all participants, though figure 6.2 suggests that there is not sufficient evidence to move to an HMM with 5 or 6 states. This may be a shortcoming of the analysis being based on a small number of participants. In this case it maybe that with a large amount of data from relatively few participants, the HMM may not be able

to characterise some less common states well. As a result, the inclusion of these states does not modulate the free energy of the model enough to justify their inclusion, instead these dynamics may end up being represented one apparently static state as seen in participant 1.

6.3.5 Frequency Specific Modulation Of PDC By State

The previous section looks in general at the results of the state-time course regression. This section describes the state modulation of DMN connectivity within three frequency bands: theta (1-5Hz), alpha (5-15Hz) and one part of the gamma spectrum (55-65Hz).

Theta Modulation

The results from the theta band can be seen in figure 6.7. States 1 and 4 do not predict any significant modulation in PDC, as outlined above this may not reflect a lack of influence. Rather the contrasts in the regression were between each state and all other states. As such this may mean that any structured variance during this time may be common to all states. Critically, the theta band is the only frequency in which state 2 is predictive of variance. This state is associated with an increase in the PDC from the dMPFC to all other nodes in the network along with a decrease in influence from vMPFC to the MTG. In contrast, state 3 is associated with an increase in influence from vMPFC to AG,dMPFC and MTG and a decrease in influence from dMPFC to MFG. Finally, state 4 contains a slight decrease in influence from vMPFC to dMPFC.

Alpha Modulation

The results for the alpha modulations can be seen in figure 6.8. State 1 contains a large increase in influence from both AG and PCC over many of the other nodes in the network. Critically, while the AG's influence on v/dMPFC increases in this state, PCC's interactions with MPFC do not change. Moreover, AG's influence on the PCC also increases during state 1. State 2 is not predictive of any changes in PDC in the alpha band.

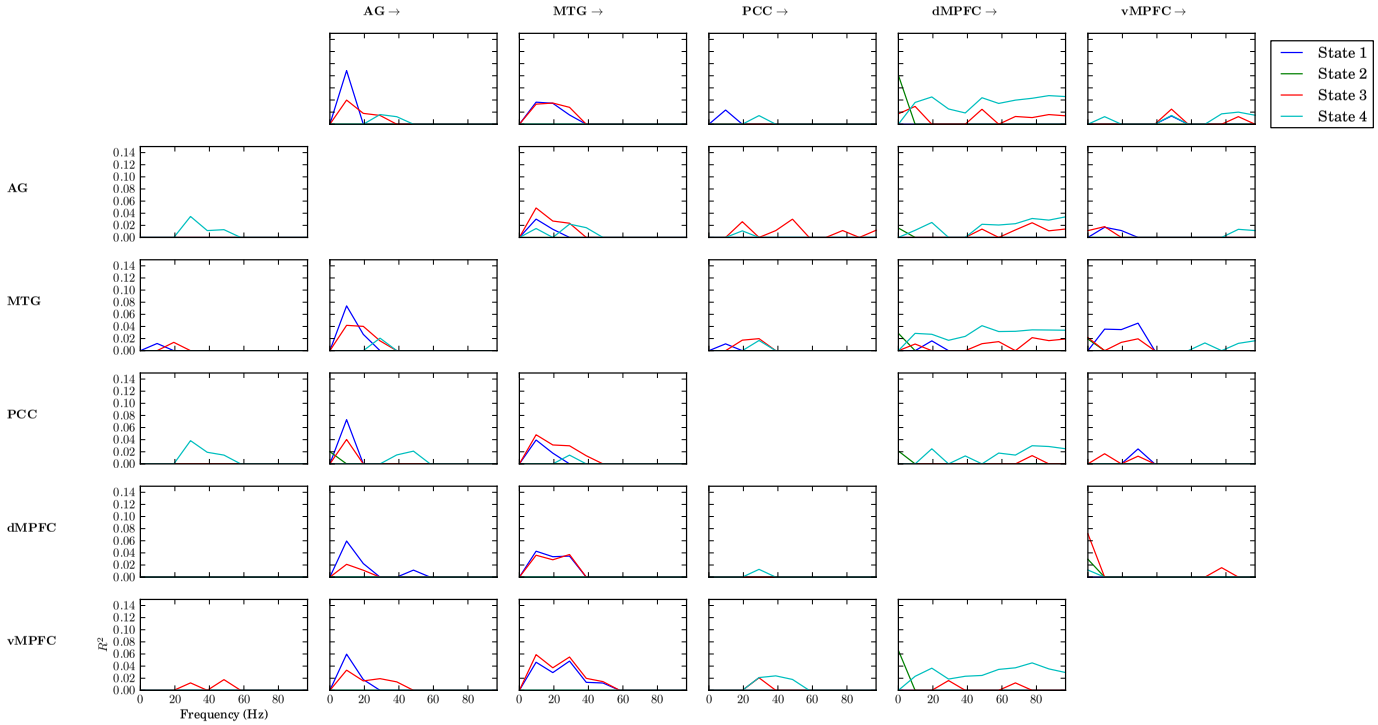

Figure 6.6: The variance explained in the connections with a significant result in the state time-course regression analysis, non-zero values indicate the proportion of the epoch-by-epoch variance explained by the presence or absence of each state in turn. Significance thresholds were $\alpha = 0.05$ with Bonferroni and Greenhouse-Geisser correction.

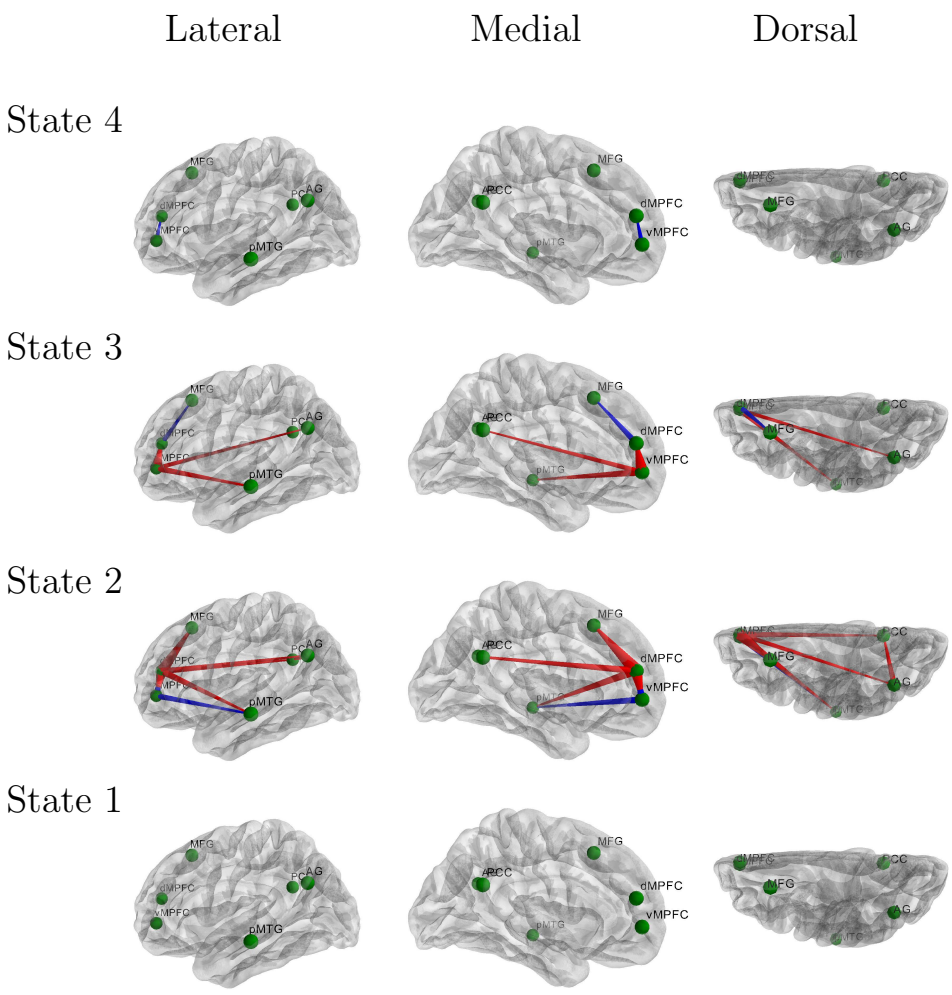

Figure 6.7: Representation of the state time course regression betas for the 1-5Hz frequency band within a glass brain. The edges indicate significant β estimates from the state changes indicated in the row. A red edge denotes a positive beta and a blue edge indicates a negative beta. The taper in the edge describes the direction of the connection in which the thick end of the edge is the source and the thin end the target, an untapered connection indicates that both directions were modulated by that state. The overall width of an edge indicates the magnitude of the beta estimate. Three views on the glass brain are shown in the columns, from left to right: looking at the lateral face of the left hemisphere, the medial wall of the left hemisphere and down on the dorsal part of the hemisphere.

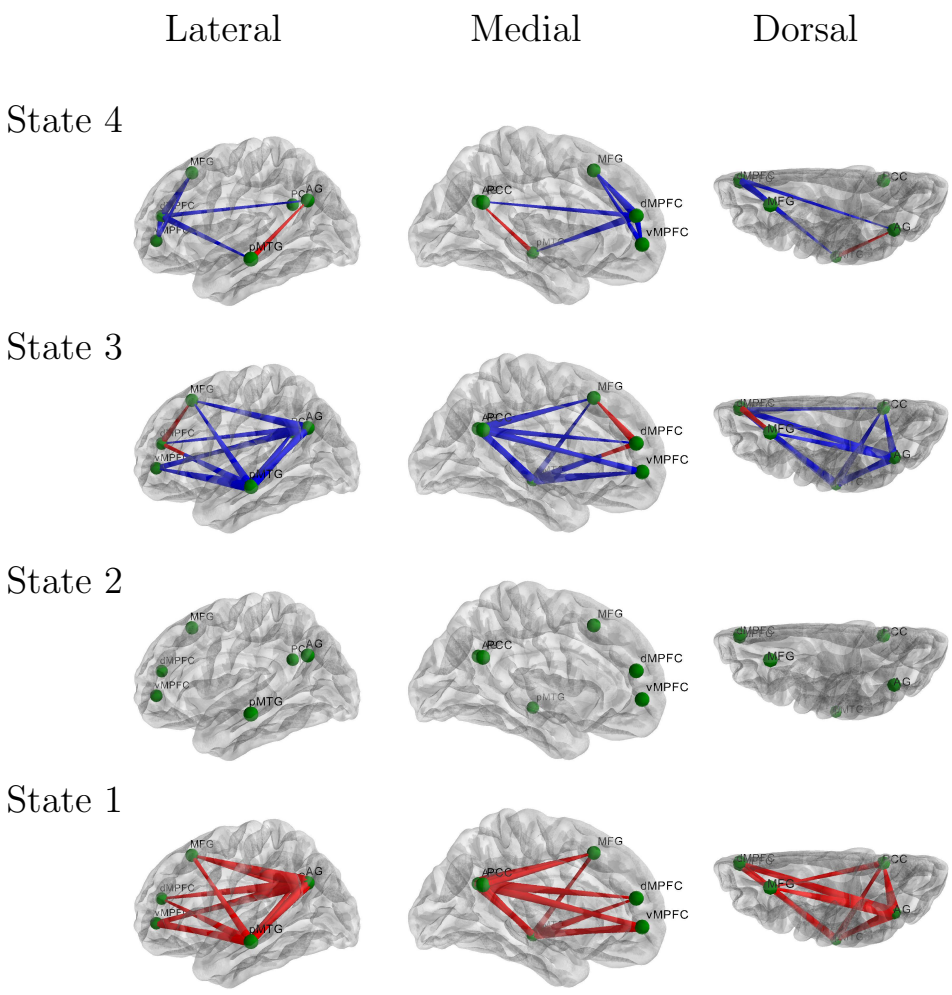

Figure 6.8: Representation of the state time course regression betas for the 5-15Hz frequency band within a glass brain. The edges indicate significant β estimates from the state changes indicated in the row. A red edge denotes a positive beta and a blue edge indicates a negative beta. The taper in the edge describes the direction of the connection in which the thick end of the edge is the source and the thin end the target, an untapered connection indicates that both directions were modulated by that state. The overall width of an edge indicates the magnitude of the beta estimate. Three views on the glass brain are shown in the columns, from left to right: looking at the lateral face of the left hemisphere, the medial wall of the left hemisphere and down on the dorsal part of the hemisphere.

In a large contrast to state 1, state 3 is characterised by a large decrease in influence from AG to the rest of the network. In addition, the influence of dMPFC on both MFG and MTG increases. Finally state 4 sees a decrease in connection strength between the frontal nodes in addition to an decrease in the connections from vMPFC to AG and MTG. MTG $\rightarrow$ AG is the only connection to show an increase in PDC in state 4.

Gamma Modulation

The results for the gamma modulation on DMN connectivity can be seen in figure 6.9. Neither state 1 or 2 predict any significant changes in PDC in this frequency band, while on the the dMPFC $\rightarrow$ MTG connection increases in state 3. Finally, state 4 shows a similar set of modulations to those seen in the alpha band. The influence of dMPFC on all other nodes decreases significantly. In contrast to state 4 in the alpha band there are no increases in PDC for this gamma band, moreover figure 6.5 shows that state 4 is only predictive of decreases in all frequencies above $60Hz$.

6.4 Discussion

The chapter has characterised the time-varying directional connectivity within the DMN using an HMM. Changes in connectivity were resolved within two second time windows to allow for a relatively fine grained frequency resolution. The inferred states revealed dynamic changes in several connections within the DMN. Moreover two of the states were associated with modulations in the directionality of two or more connections, suggesting that the directionality in several connections may co-vary in time. The modulation of the states on the PDC estimates was quantified by using the state time courses to predict epoch-by-epoch variance in the PDC.

6.4.1 Functional States Within The DMN

Four states were identified in the HMM, each of these states were fitted to principle components of the PDC to make the dimensionality of the Vari-

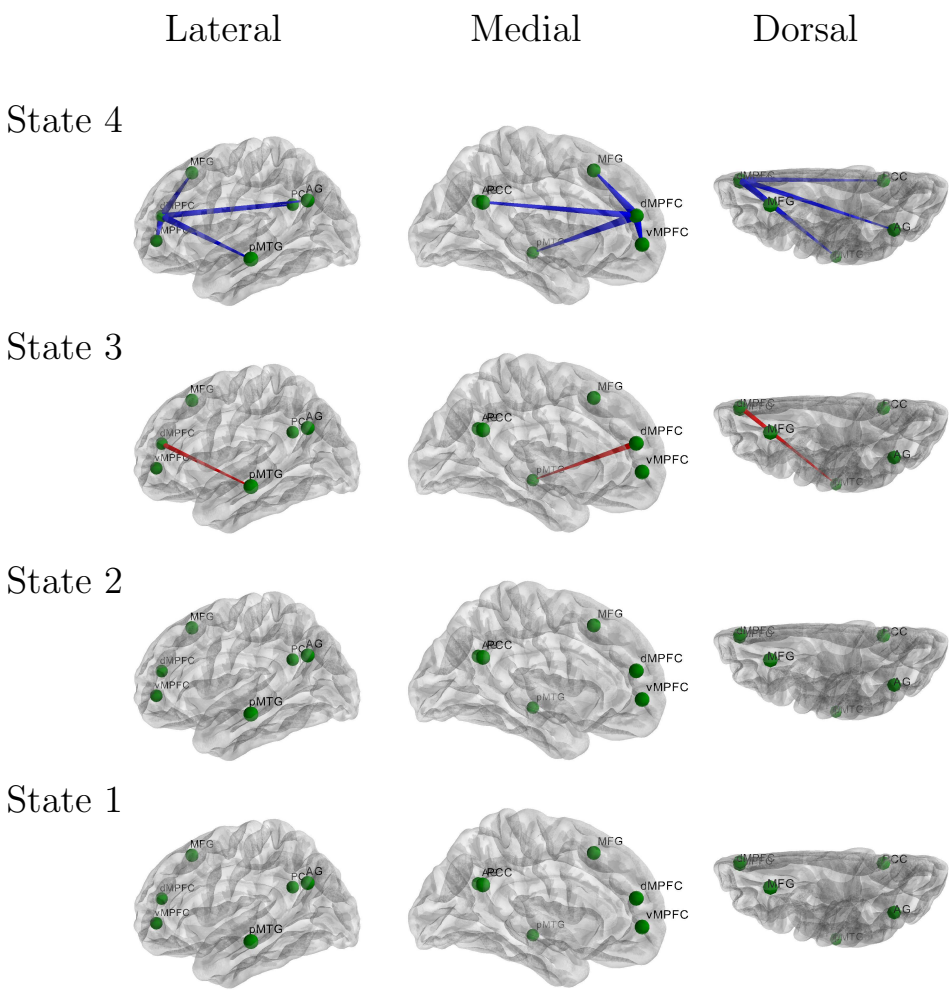

Figure 6.9: Representation of the state time course regression betas for the 55-65Hz frequency band within a glass brain. The edges indicate significant β estimates from the state changes indicated in the row. A red edge denotes a positive beta and a blue edge indicates a negative beta. The taper in the edge describes the direction of the connection in which the thick end of the edge is the source and the thin end the target, an untapered connection indicates that both directions were modulated by that state. The overall width of an edge indicates the magnitude of the beta estimate. Three views on the glass brain are shown in the columns, from left to right: looking at the lateral face of the left hemisphere, the medial wall of the left hemisphere and down on the dorsal part of the hemisphere.

ational Bayes fit computationally tractable. As such, the key analyses in this chapter arose from the state time series (u) computed by Viterbi decoding as seen in figure 6.3. The first analysis was applied the directionality metric from chapter 4 to the entire DMN for each state. This showed relatively subtle differences in directionality as a function of stats. Critically, there were no instances in which a strongly directional connection changes directionality between states. Though there are instances in which a weak directionality in one state becomes insignificant, or even weakly directional in another.

The objective of this analysis was to establish whether different functional states as defined by the HMM might be associated with different directionality within the DMN. The evidence supports a weak form of this statement in which directionality is modulated, but only reversed in cases with weakly directional connections. This supports suggestions from the computational literature which say that the brain is exploring a dynamic repertoire which is constrained by the anatomical skeleton provided by its biology (Deco et al., 2011).

The second analysis was the state-time course regression. This linked the HMM state observation probabilities (which were fitted to principle components of the PDC) back to the original PDC estimates. This was done by predicting the epoch-by-epoch PDC estimates from the the state time course, established by selecting the state with the highest *a posteriori* evidence for each time point in order to identify connections which are modulated by state transitions. In contrast to the directionality analysis, which does not directly test for whether the HMM actually explains any variance, this approach is able to validate the HMM by establishing whether is has captured any structure in the PDCs variance over time. The evidence showed that the HMM was able to capture structure in the evolution of the PDC over time. Figure 6.6 shows the variance explained by each state in each connection. These values are quite small (around $2 - 10\%$) however each point is the variance explained over the entire time course from five participants from a single state. Moreover, many potential sources of noise or variance in the PDC estimates are not modelled (for instance the time courses from all

five participants were concatenated and the joins between scanning sessions within or between individuals not accounted for) or potentially not possible to model (noise/distractions during the scan or movement). As such, these values, though small, are taken as evidence that the HMM has identified some real structure in the data.

The state-time course regression identified a very different range of modulated connections for different frequencies (see figures 6.7-6.9. In general the pattern on connections modulated by the state transitions is quite complex, even within a frequency band, however some specific results can be identified.

Firstly, the majority of the state modulations in the theta band involve the MPFC. State 2 being characterised by an increase in the influence from dMPFC and state 3 showing an increase in influence from vMPFC. This is in contrast to the alpha band which mainly modulates the connectivity from the AG and PCC. State 3 (which shows an increase in influence from vMPFC in the theta band) is associated with a significant decrease in connectivity from AG and PCC in the alpha band. In addition, state 1 (which shows no modulation in theta) shows large increases in PDC in alpha. Critically this suggests that increases in theta connectivity from vMPFC tend to co-occur with a decrease in PDC from PCC and AG in alpha. Moreover, when AG and PCC are strongly influential (state 1 in alpha) neither the dorsal or ventral MPFC show modulation with state transitions in either theta or alpha.

State 4 is associated with decreases in influence from MPFC in all three frequency bands, specifically the vMPFC $\rightarrow$ dMPFC connection in theta and output from dMPFC in alpha and gamma. It is important to note that this state dominates the state time course for participant 1, who spends the entire time (across five scanning sessions) in this state. It is unlikely that this participant showed no fluctuations in functional connectivity within the DMN during this time, rather the four states from the HMM may not be sufficient to unfold the dynamics during this time. As a result this state is dominated by epochs from one or two participants. Further recordings from a much larger cohort of participants will be needed to establish whether some

participants show less variance than other or if there is currently not enough evidence from a limited number of participants to describe the dynamics during this time.

6.4.2 Individual Differences

The HMM in this chapter was fitted on the concatenated PDC estimates for all five participants, however figure 6.3 clearly shows differences in both the fractional occupancy and transition probabilities between participants. Several approaches might allow us to characterise these differences. The first option would be to fit a separate HMM to each participant, however this would severely limit any group inferences that we could make as each participant would have a unique set of hidden states. An alternative might be to fit one HMM for the group but compute the descriptive statistics within each individual.

The critical question relating to individual variances is whether these differences reflect typical variance in the recording and model estimation process or if they reflect some cognitive/behavioural differences between the participants. Further investigation into the typical variance in these measures across participants and the behavioural/cognitive correlates of the observed brain states will be needed to definitively answer this question. Though there are quite large differences between participants, the transitions between states between individual scanning sessions within each participant appear remarkably consistent. We might take this as an indication that the differences between participants do reflect genuine individual differences in the dynamics of functional connectivity within the DMN rather than an confounding influence in the scanning/analysis.

6.4.3 Methodological Considerations

The PDC estimates for the five participants were concatenated in time prior to fitting the HMM, this results in one HMM with one set of states for all participants. This approach has been used in previous applications of HMMs to MEG state dynamics (Baker et al., 2014) however it makes several

critical assumptions. Firstly, that there is one consistent set of state that is representative for all participants. This point is challenged by participant one who remains in state 1 for almost the entire recording, moreover state 1 occurs very infrequently in three of the other four participants. It is unclear whether this represents typical individual differences in functional state dynamics or that the group states do not sufficiently represent the dynamics of that participant.

The Viterbi decoding assumes that only one state can be active at a given time. This is done by assessing the posterior evidence for each state at each time point and assigning the state with the most evidence to that time point. This may not accurately represent the dynamics of the system in which two or more state might be active at a given time. A "fuzzy" decoding in which the *a posteriori* evidence for each state at each time point is carried forward to the state-time course regression might allow for more realistic predictor variable rather than the binary series used in this chapter.

6.4.4 Conclusion

This chapter has applied a dynamic state allocation from Hidden Markov Model methodology (Woolrich et al., 2013; Baker et al., 2014) to explain the epoch-by-epoch evolution of directed functional connectivity within the DMN. This has been successful in estimating sets of epochs whose observations may have been generated from the same hidden functional state. state time course regression was then used to establish whether the transitions of these hidden states predict variance in any of the PDC estimates within the DMN. This established which connections might be modulated by the transitions of the hidden states.

The present work builds upon past methods by exposing the HMM directly to directional connectivity estimates based on TDE and MVAR modelling. This allows for characterisation of dynamics within the DMN on two time-scales. Firstly, time delay embedding is used to transform the data into a surrogate state space based on time delays of the data. This representation of the data provides a picture of the dynamics within the DMN on

a millisecond timescale. Time-delayed interactions which predict future observations consistently over a two second epoch are then parameterised with an MVAR model and expressed in the frequency domain with PDC. This first dynamic level describes the extent to which activity at one frequency in one node is able to predict activity in another node several milliseconds before it occurs. The slower modulations of this first level are then characterised on the time-scale of seconds using the HMM. This time-scale is rather slower than used in previous MEG-HMM studies (Woolrich et al., 2013; Baker et al., 2014), however this is necessary to ensure sufficient data points for the accurate characterisation of the rapid dynamics on the first level (see chapter 3 and 4 for a discussion).

This data limit prevents the direct application of this methodology to the very rapid dynamic changes seen in the brain in response to visual stimuli such as written words. In order to characterise both the rapid information flow at the first dynamic level and its slower modulation over time on the millisecond time-scale seen required for cognitive tasks, we need to compute the first level dynamics across many realisations of the cognitive process. The next two chapters explore this how this can be achieved in the context of visual word recognition

Chapter 7

Neural Dynamics Of Visual Word Recognition

7.1 Introduction

The last chapter explored the endogenous connectivity within a set of brain regions known as the Default Mode Network. Though recent years have seen an increase in interest in these self-generated brain networks, ultimately they are challenging to link to specific behaviours. As such, the following chapters will deal with brain dynamics related to specific cognitive and task demands.

7.1.1 Visual Word Recognition

Word reading is commonly thought to involve the decoding of three aspects of the word:

Orthography The visual structure of the word and its letters

Phonology The sounds associated with the word as it might be spoken

Semantics The meaning of the word

These "aspects" of a word have formed the basis of much of the theory on visual word recognition, appearing in many cognitive and computational models.

7.1.2 Models Of Visual Word Recognition

While the cognitively relevant features of a visual word form outlined above remain relatively uncontroversial in the literature, there is much debate about the processing framework in which they are resolved. Two approaches have dominated the behavioural and computational literature on this subject: the modular-hierarchical models which suggest that visual input produces and orthographic representation which feeds forward again to phonological and semantic processes, and secondly the interactive approach which allows for high level information to influence orthographic processing at an early stage. These contrasting frameworks form one of the oldest debates in visual word recognition (Carreiras et al., 2014).

A major argument from proponents of the feedforward framework is that phenomena such as the word superiority effect can be solved using

computational models with both feedforward and interactive architectures. As such, the simpler feedforward models should be preferred (Norris et al., 2000). A critical implication being that low level visual pattern recognition systems are sufficient for orthographic decoding and do not require linguistic experience or expectation (Grainger et al., 2012).

In contrast to this view, connectionist models such as the Interactive Activation Model (McClelland and Rumelhart, 1981) and Parallel Distributed Processing (PDP) readily include the potential for protracted interactions in between orthographic, phonological and semantic processes. In such models, visual input feeds into the orthographic representation, however the representation of the orthography, phonology and semantics converge together. Feedback in these models might allow for partially computed phonological representations to aid in the convergence on the appropriate orthographic representation. Interactions of this kind do and must occur in connectionist models (if all connections to all nodes are weighted), however the critical further question is the extent to which this feedback is influential or even necessary for accurate recognition of a word form.

Not all connectionist models contain a feedback component and many complex recognition tasks can be completed without feedback or backpropogation (Riesenhuber and Poggio, 1999). The importance of feedback increases greatly during unsupervised learning processes, whereby backpropogation of errors provides a very powerful method for learning weights and top-down modulation of visual processes such as attentional modulation. The question of top-down influence within a triangle model of word recognition is different from both of these processes. Rather than post-hoc modulation of weights or the influence of a previously known processes such as attention, the focus is on whether partially computed orthography, phonological and semantic processes are able to influence each other during the recognition process itself.

One type of computation model in which feedback is critical and clearly defined is the deep learning (Hinton, 2006, 2007). These generative models contain feedback connections from layers encoding abstract concepts with weights trained to reproduce the data observations associated with them.

Though such models have successfully been applied to visual letter (Hinton, 2007, 2006) and word recognition (Di Bono and Zorzi, 2013) the feedback connections are only engaged during the learning of feedforward connections rather than during perceptual inference itself (Lee and Mumford, 2003). Such an idea might be supported by the priming effects seen in MEG.

In contrast, an increasing number of computational models are including explicit top-down influence to visual processes during perception itself (Rao and Ballard, 1997). Moreover, feedback influences during the act of perception/recognition can be conceptualised as a hierarchical Bayesian system (Lee and Mumford, 2003). Such a system would allow for higher order areas to act directly or indirectly as priors influencing the inferences at lower levels. Low level sensory areas would then form *a posteriori* estimates of the stimulus causing novel sensory input by multiplying the probability of the sensory evidence by the contextual prior expectations and maximising the result by competition to arrive at a most probable external stimulus (Lee and Mumford, 2003). In such a model feedback may represent the prior expectation about a stimulus at a given time and would serve to shape the posterior probability of a given stimulus at lower levels in the perceptual hierarchy.

The studies and theories above do not directly interrogate the brain basis of visual word recognition, rather they focus on theoretical points relating to the cognitive and computational aspects of reading. The recent development of neuroimaging methodologies with the power to interrogate the brain basis of word recognition and potentially shed some light on whether orthographic recognition occurs without influence from semantic and phonological processes/expectations.

7.1.3 Brain Basis Of Visual Word Recognition

In the skilled reader, a distributed and dynamic cortical network sub-serves the extraction and integration of orthographic, phonological and semantic information from a visual word form. Application of magnetoencephalography (MEG) to the study of visual word recognition has demonstrated a

complex pattern of cortical activation propagating from primary sensory cortices towards anterior temporal and inferior frontal regions (Marinkovic et al., 2003; Pammer et al., 2004; Cornelissen et al., 2009). Moreover, these regions show a dense pattern of functional connections (Kujala et al., 2007) which are likely to recruit a range of white matter pathways (Catani et al., 2005; Ben-Shachar et al., 2007).

The initial stages of reading are visual and potentially only depend on the presence of orthographic information. Noise masking paradigms have demonstrated modulation of activity in posterior occipital cortex by noise intensity but have equivalent responses to visual presentation of words and consonant strings around 100ms. In contrast, a later response around 150ms in the ventral occipito-temporal cortex (vOT) preferentially responds words over syllables, consonant strings and noise. (Tarkiainen et al., 1999, 2002). This later response is more invariant to changes in latter position in the left hemisphere than the right (Cornelissen et al., 2003) suggesting a more abstracted, object level representation of a word form in the left hemisphere. There is clear evidence suggesting that the vOT is involved in the reading process, though its precise role remains unclear.

One account states that the vOT acts as a visual perceptual stage in a specialised hierarchical processing stream in which information is processed from low-level retinotopic feature extraction to feature invariant representation of bi-grams and whole words (Dehaene et al., 2005; Vinckier et al., 2007). These models emphasise the internal organisation of the vOT and how it as a region within a hierarchy may support reading. In contrast, an interactive account of reading places the emphasis on interactions with non-visual regions of cortex (Price and Devlin, 2011). In this sense the internal organisation of the vOT may not involve reading specific mechanisms, such as bigram detectors etc, rather the reading specific phenomena arise from top-down predictions, based on a probabilistic representation of the cause of sensory inputs (Friston, 2010).

Recent studies modelling neural responses to visual word recognition support an interactive model in which top-down predictions may arise from the inferior frontal gyrus (IFG). The left hemisphere IFG responds rapidly to

word forms (Cornelissen et al., 2009). This rapid response is enhanced when a word is phonologically primed (Wheat et al., 2010). This may indicate that the IFG responds preferentially to more predictable words, possibly using the prediction to streamline lower level processes. In a recent Dynamic Causal Modelling (DCM) study on reading Yvert et al. (2012) showed the winning family of models contained a direction connection between early visual areas and the IFG, as well as feedback from the IFG to either the vOT or superior temporal sulcus. In addition, this early influence on IFG from visual areas is stronger when reading real words than false fonts (Woodhead et al., 2012).

In summary, though the computational modelling literature may find that exclusively feedforward models are the most parsimonious, there is mounting evidence for both the theoretical (Lee and Mumford, 2003) and neural (Cornelissen et al., 2009; Wheat et al., 2010) basis for early feedback influences during visual word recognition. DCM analyses of MEG data have demonstrated that this influence is likely to arise from the LIFG and is stronger for real words than false font stimuli (Woodhead et al., 2012). This chapter looks to build on these findings firstly by establishing the dynamics of the connectivity within the reading network using MVAR models and PDC.

Real words, consonant string and false font strings were presented to participants during an MEG scan, power changes in the brain were characterised for each condition prior to connectivity analysis. The connections which are sensitive to recognition of the different features associated with reading are identified. Critically, which connections are crucial for recognising orthography and are visual regions sufficient to distinguish consonant strings from real words without early interactions with higher regions.

It is predicted that if vOT is sufficient for orthographic representation without top-down influences, then there will be no influence from non-visual brain regions to vOT prior to 200ms after stimulus onset during any condition. If vOT is able to discriminate arbitrary visual forms from a known orthography, the connectivity between higher areas and vOT will only be present for the consonant string and real word stimuli. Finally, if early top-down influences are present for all three conditions, this will indicate

that a full orthographic representation is only possible with interaction with higher level processes. That is to say that visual cortex can only rule out the presence of orthography in a word form with help from higher regions.

7.1.4 Overview

This chapter will address the question of whether the neural response and connectivity of the ventral occipito-temporal cortex reflects only the presence of orthographic information or is modulated by existence of semantics and phonology in addition. The answer to this question is likely to be different at different times during the recognition process. For instance it is possible that top-down influence on vOT occurs, but after the encoding of the orthographic information. As such the MVAR modelling in this chapter is modified to characterise the delayed interactions within a network on a much shorter time-scale (around 100ms). This time-scale is much more appropriate for task dynamics that the 2 second windows used in the resting state chapters, however to ensure that there are sufficient observations to get a good model fit, the MVAR fitted here will reflect the average delayed covariance across many epochs rather than providing one model per epoch.

A short window 'evoked' MVAR model is computed for three stimulus categories, real words, consonant strings and false fonts within short sliding windows across the experimental epoch. The temporal evolution of the connectivity within the epoch is then characterised before differences between the three conditions at different times statistically evaluated.

7.2 Methods

7.2.1 Participants

Data were recorded from five healthy participants (4 males) with an average age of 24.44 (SD:1.41) years. Participants were right handed and achieved a maximum hand dominance score on an Annett's Test (Annett, 1970). All participants had normal vision and no history of dyslexia. This study was approved by the York Neuroimaging Centre Research Ethics Committee

and conducted according to their guidelines. Each participant provided full written consent at each session and was fully debriefed at the end of the experiment.

7.2.2 Experimental Design

Stimuli

The stimuli consisted of 300 real words, 300 consonant strings and 300 false font words. The word stimuli were between four and six letters long (Mean Length= 4.8; Mean Concreteness= 428.59; Mean Familiarity= 560.11; Mean Imagability= 464.44; Mean Kucera and Francis Frequency= 146.24) and taken from the MRC Psycholinguistics Database (Coltheart, 1981).

The consonant strings matched the real words for length and randomly generated from a uniform distribution of consonants (vowels and the letter Y were excluded). The false font stimuli were selected from a set of ancient fonts retrieved from http://users.teilar.gr/~g1951d/. Characters from the Lydian, Carian and Lycean alphabets were used. The ancient characters were matched to consonants from the Latin alphabet and the pairings used to generate false font words from the consonant strings. The false font stimuli share some of the basic visual features of Latin characters such as superpositions of straight lines and curves. Despite sharing these elements of visual form, these letters are not familiar to the participants and as such will not be associated with an phonological information.

Participants were asked to identify animal names presented during the experiment and to press a button with their left index finger whenever one appeared. There were 30 animal name trials in total. These catch trials ensured that the participants were attending to the stimuli and that they had to semantically process the word if possible. This task was not possible from visual matching, orthographic or phonological processes alone.

Words, Consonant strings and animal names were presented in white on a black background in an upper case Arial Mono-spaced font. The stimuli were presented using a Dukane 8942 ImagePro 4500 lumens LCD projector at a virtual visual angle of 1°and horizontal angle of 4.5°. Vertical red lines

	Word	Consonant Strings	False Fonts
Examples			
Four Characters	AWAY	LFNH	⅁ᖵᑭ‡
Five Characters	BOARD	MHPTN	ᔭᙢᑭϒᑭ
Six Characters	ADVICE	NPHDSC	⅁ʌᖵ⅄ᔭᔭ
Features			
Visual Form	✓	✓	✓
Orthography	✓	✓	×
Phonology	✓	×	×
Semantics	✓	×	×

Figure 7.1: *Examples* Example stimuli from each condition and each word length. *Features* Indication of which features of a real word are present in each stimulus category.

above and below the centre of the screen indicated the fixation point and remained on the screen for the duration of the experiment.

Procedure

The stimuli were split into 6 blocks each containing 90 experimental trials from each of the three conditions and 30 animal name catch trials. Each block lasted around 15 minutes. Trials from each condition were presented in a random order and each trial was only presented once across all blocks. In addition the order of the blocks was counter balanced across participants.

Figure 7.2 contains a schematic of the trial sequence and timings used in the experiment.

7.2.3 Data Acquisition

Data were continuously recorded at 4069.017Hz using a 4D Neuroimaging Magnes 3600 whole head system. The data were passed through a low-pass filter set at 1500Hz. Participant head shape and reference coil location were recorded using a 3D digitiser (Polhemus Fastrak) allowing for co-registration with a high resolution anatomical T1 MRI image acquired using a GE 3.0T Signa Excite HDx (Kozinska et al., 2001).

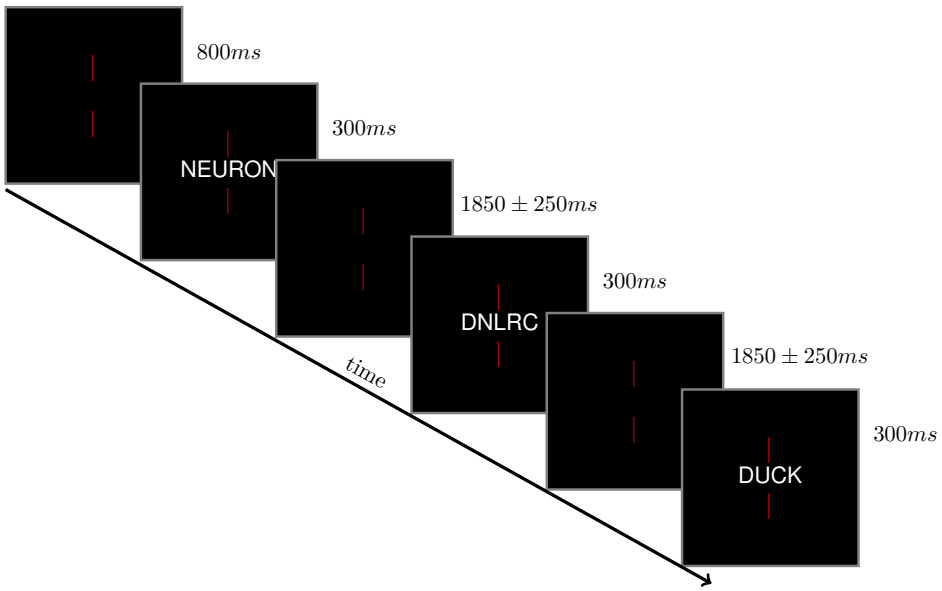

Figure 7.2: Schematic showing the stimuli presentation and ordering. The two red lines remain on the screen for the duration of the experiment and their center point indicated the fixation for the participant. The first two stimuli are from the Word and Consonant string conditions and the final stimulus is an animal name catch trial. The $1850 \pm 250ms$ inter stimulus interval includes an $800ms$ blink period.

Electrooculogram (EOG) and electrocardiogram (ECG) data were acquired alongside the MEG acquisition and used to aid the artefact rejection process.

7.2.4 Data Analysis

Artefact Rejection

Data from each participant were split into non-overlapping 2 second segments. These epochs were visually inspected for artefacts and trials containing muscle or eye movements were discarded from the analysis. The MEG data were then co-registered with the individual's structural T1 MRI scan and the MNI 152 standard brain to allow for source analysis to be conducted.

Beamformer Contrasts

Beamforming estimates the current at a source location in the brain from a weighted sum of the observed magnetic fields on the sensors. A type 1 vectorized, linearly constrained minimum variance (LCMV) beamformer (Van Veen et al., 1997; Huang et al., 2004) was used before the oscillatory power in a passive window (−200ms to 0ms) was compared to the power in four sequential active windows (0ms - 200ms, 100ms - 300ms, 200ms - 400ms & 300ms - 500ms). These contrasts were performed in an alpha (5 - 15Hz) and beta (15 - 25Hz) frequency bands. Non-parametric label permutations were used to compute statistical thresholds (Nichols and Holmes, 2002). Local maxima within these contrasts were used alongside a review of network nodes from the literature to establish the 6 node reading network used in this chapter.

Virtual Electrodes

Connectivity analysis was based on six nodes identified from the beamformer contrast results (details are in table 7.2). Current was estimated for each trial in these locations using a vectorized LCMV beamformer as defined by

Van Veen et al. (1997) which has been shown to generate the most accurate source reconstructions (Johnson et al., 2011). VEs were estimated at the full sampling rate (4069.017Hz) across the entire epoch for each location.

Data Preprocessing

A 1Hz high pass filter was used to remove drift from the reconstructed signals and line noise removed using tight notch filters at 50Hz and it's first 4 harmonics. The samples within each epoch were normalised by demeaning each sample and dividing by the standard deviation. The epochs from each of the five runs were then concatenated.

7.2.5 Connectivity

The MVAR model fitting procedure outlined in chapter two and applied in chapters 3 and 4 makes use of the lagged covariance matrix (R) estimated from the system. The accurate estimation of R is critical to ensure that the model parameters (A) are accurate. Chapter 3 demonstrated the effect of data length on estimation of A and showed that fitting many parameters from few observations is likely to result in an unreliable model. In chapters 4 and 5, this limitation resulted in relatively long time windows being used in chapters 4 and 5. As a result, chapter 5 characterises dynamic modulations in connectivity on the time scale of seconds. Though this may be appropriate for resting state dynamics, it is orders of magnitude slower than the brain responses arising from seeing a visual word. A solution to this issue can be found in the estimation of R from many very short realisations of a neural response (Ding et al., 2000). If we have many realisations of a brain response to a single category of stimulus, we can compute the average R across all realisations in a manner analogous with an ERP. The resulting 'evoked' R matrix will contain an accurate representation of the consistent covariance across all observed epochs. The R can then be used to accurately estimate A. This approach has been successfully applied to EEG data many times (Ding et al., 2000; Supp et al., 2007), with time windows as short as 50ms allowing for characterisation of dynamics on a time-scale relevant to the

neural response to a word (Sun et al., 2009a).

Each epoch was split into 150ms sliding windows which advanced in 25ms steps from stimulus onset until stimulus offset 300ms later as described in figure 7.3. All the time series from each sliding window were used to compute an evoked R matrix which was used to estimate A through the Vieira-Morf algorithm. The PDC is then calculated from the A matrix for each sliding window.

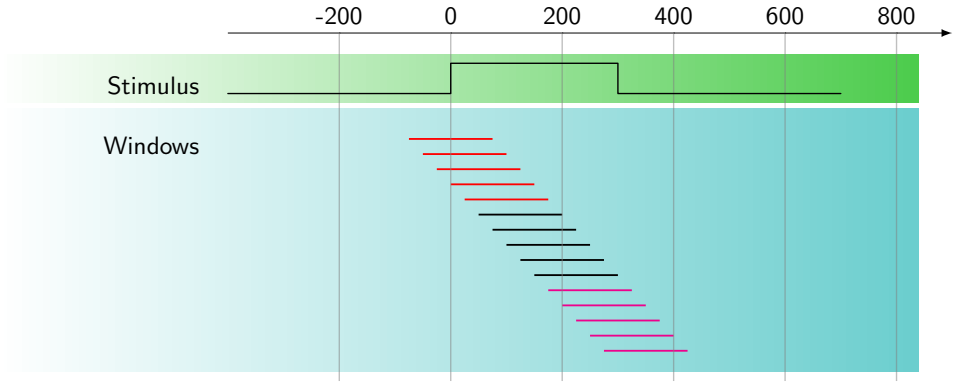

Figure 7.3: Schematic of the sliding analysis windows used in the connectivity analysis. The arrow at the top denotes time in milliseconds. The stimulus line shows when the stimulus appears within the epoch. The sliding windows are indicated with horizontal bars shown in the next section with the blue background. The PDC from within each window is presented in subsequent sections and labelled with the centre time of the window. The different colour of bar indicates the groups of windows which are combined to make the windows seen in figures 7.10, 7.11 and 7.12.

7.2.6 Condition Comparisons

A distribution of PDC estimates for each participant and sliding window was generated from jackknifing the trials used to fit the MVAR model. A total of 50 jackknives in which a random 80% of the trials were entered into the model. Once this distribution has been established, any differences between the conditions were characterised using unpaired t-tests. This produced three t-maps for the different contrasts (Word - Consonant, Word - False

Font and Consonant - False Font) for each direction of each connection.

The time-domain simulations in Chapter 4 indicated that the ratio between the number of observations and the number of parameters to be fitted in the MVAR model is of critical importance. A minimum acceptable ratio was suggested to be around 5 or 10 samples per parameter, but a ratio of 50 is more desirable. The visual word recognition data used in this experiment are likely to yield much more complex dynamics than the simulation used in Chapter 4, as such, a ratio of 100 to 150 data observations per parameter to be fitted was set as the minimum for this experiment. When sampling at $4069Hz$, a 150ms window will contain around 270 samples. Given this, 200 to 300 epochs are needed to make the required data ratio. 540 trials were collected from each participants for each condition to allow for sufficient epochs in each jackknife after manual artefact rejection. As a result, each participant committed around 4 hours of time in MEG (including set up). This large commitment that only 5 participants were recruited for this experiment, not a sufficient number to allow for a traditional group level analysis, though the analysis for each individual will have sufficient power to reliably model temporal dependencies.

A meta-analysis was performed across the available participants by z-transforming the t-maps for each participant and then combining them using the Stouffers method for combining the results from several independent tests of the same experimental hypothesis.

$$Z \approx \frac{\sum_{p=i}^{k} z_i}{\sqrt{k}} \tag{7.1}$$

in which z_i is a z-value from one participant and k is the total number of participants. The Z-map across participants from this meta-analysis was then arbitrarily thresholded at $Z = 2.3$.

7.3 Results

7.3.1 Behavioural

The participants were able to correctly identify 78.06% of the animal name catch trials with an average reaction time of 472.4ms. The results per subject can be seen in table 7.1. The reaction times are very consistent across participants, however there is a larger variance for the number of correctly identified catch trials.

Participant	Percent Correct	Reaction Time (ms)
1	86.6	480
2	83.3	474
3	55.5	476
4	71.6	467
5	93.3	465
Average	78	472

Table 7.1: Behavioural results per participant. Percent correct indicates the proportion of animal name catch trials which the participant was able to respond to across the six experimental blocks. Reaction time is the average time in milliseconds for a correct response from the participant. Only responses within 1 second of the catch trial were included in this analysis.

7.3.2 Node Identification

Beamformer Contrasts

The results of the beamformer contrasts were combined across participants using Stouffer's method for combining independent tests of the same hypothesis. The sliding window results for the three conditions in the alpha and beta bands can be seen in figures 7.4 and 7.5.

The beamformer power contrasts clearly replicate the posterior to anterior spread of activation over time previously reported in the literature (Marinkovic et al., 2003; Pammer et al., 2004). Critically, the word condition in the alpha band also shows considerable increases in power in the

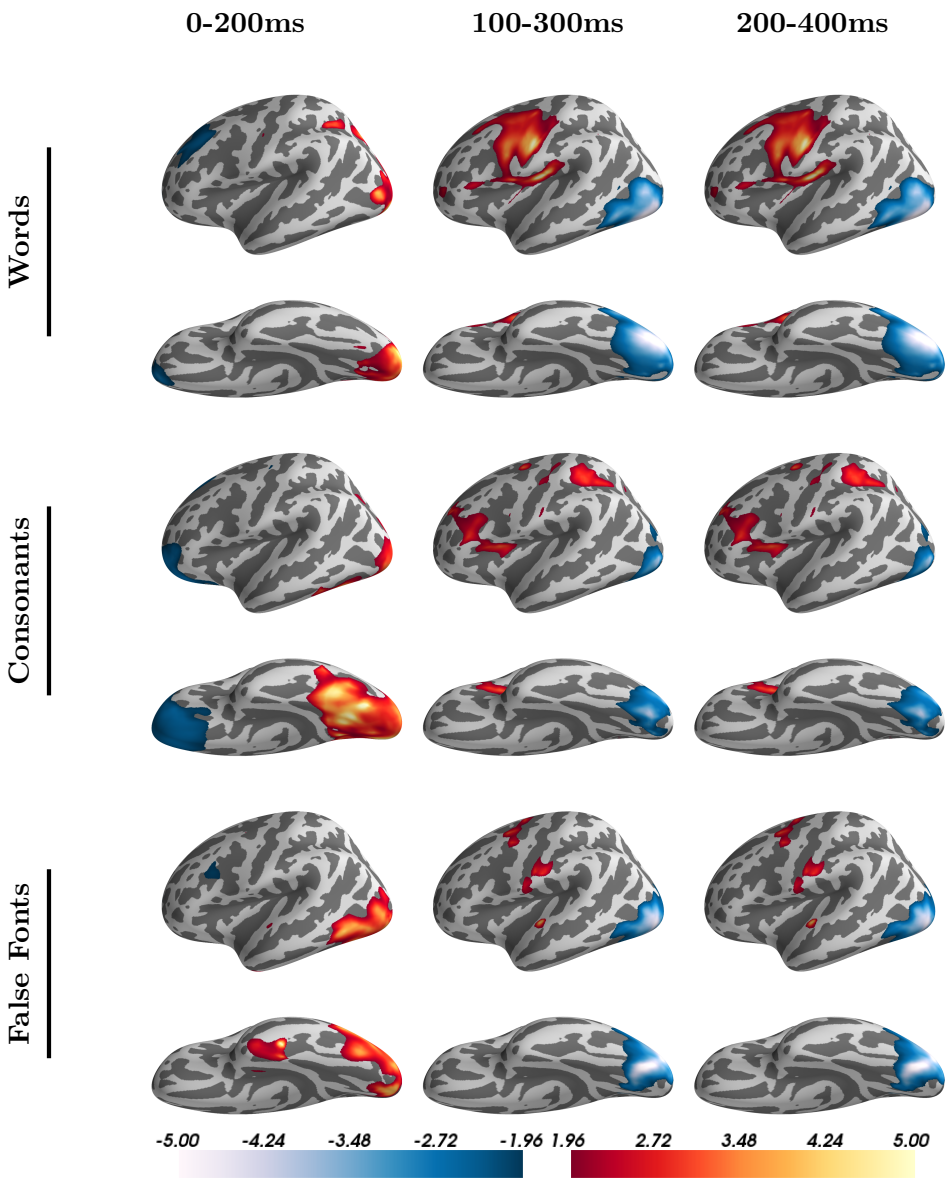

Figure 7.4: Z-stats from an active-passive beamformer power contrast in the alpha band ($5 - 15$ Hz), each surface contains the combined Z-stats across the five participants arbitrarily thresholded at $Z = 1.96$. The sliding window results from the three experimental conditions are shown. The results have been rendered onto the surface of a fsaverage brain with the lateral (top) and ventral (bottom) views of the left hemisphere shown.

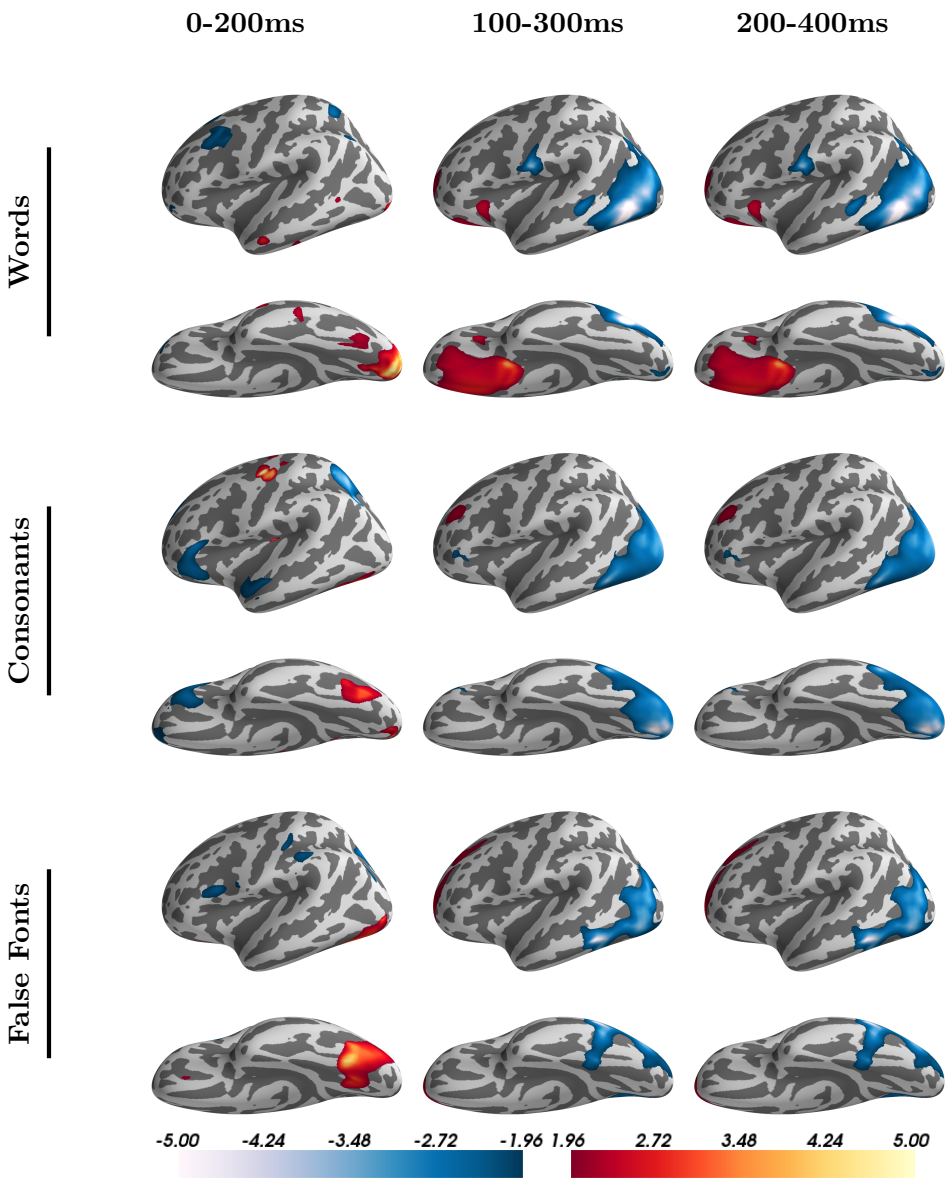

Figure 7.5: Z-stats from an active-passive beamformer power contrast in the beta band ($15 - 25$ Hz), each surface contains the combined Z-stats across the five participants arbitrarily thresholded at $Z = 1.96$. The sliding window results from the three experimental conditions are shown. The results have been rendered onto the surface of a fsaverage brain with the lateral (top) and ventral (bottom) views of the left hemisphere shown.

posterior inferior frontal gyrus in an early (100-300ms) time window. This finding has been identified several times in the literature and associated with a rapid access to phonological information in a word form (Cornelissen et al., 2009; Wheat et al., 2010). In addition, large increases in power can be seen in ventral visual areas, though this activation is stronger for the consonant strings and false font stimuli. This is in contrast to previous fMRI findings which showed that Word stimuli were more likely to activate anterior ventral occipital temporal regions (Vinckier et al., 2007). Several nodes in the reading network were identified from these beamformer analyses. The middle occipital gyrus, ventral occipital temporal cortex, inferior frontal gyrus *pars opercularis* and the superior temporal gyrus.

Though the results of the beamformer power analysis highlight the contributions of several brain regions critical to visual word recognition, an increase in power does not necessarily indicate any modulation in directed connectivity. As such, several additional brain regions were identified from a literature review of previous connectivity studies into the reading network. The locations of nodes within a reading network are shown in figure 7.6. Two locations which did not produce a power change in any contrast were included in the reading network after this literature search. Firstly the Angular gyrus which was identified in Jobard et al. (2003) has long been implicated in reading Brais (1993). Finally, the temporal pole included as seen in Yvert et al. (2012). This region is thought to be critical in semantic processing however it is rarely identified in fMRI studies of language potentially due to signal drop-out (Visser et al., 2010).

The final nodes included in the reading network are identified in table 7.2.

7.3.3 Model Validation

The model validation results are summarised in table 7.3. Though some participants have quite different number of epochs for the two conditions (as a result of the epoch rejection), the models are very consistent across participant and condition.All the stability indices are well below 1, indicating that the models are stable, which implies stationarity. Moreover the Durbin-

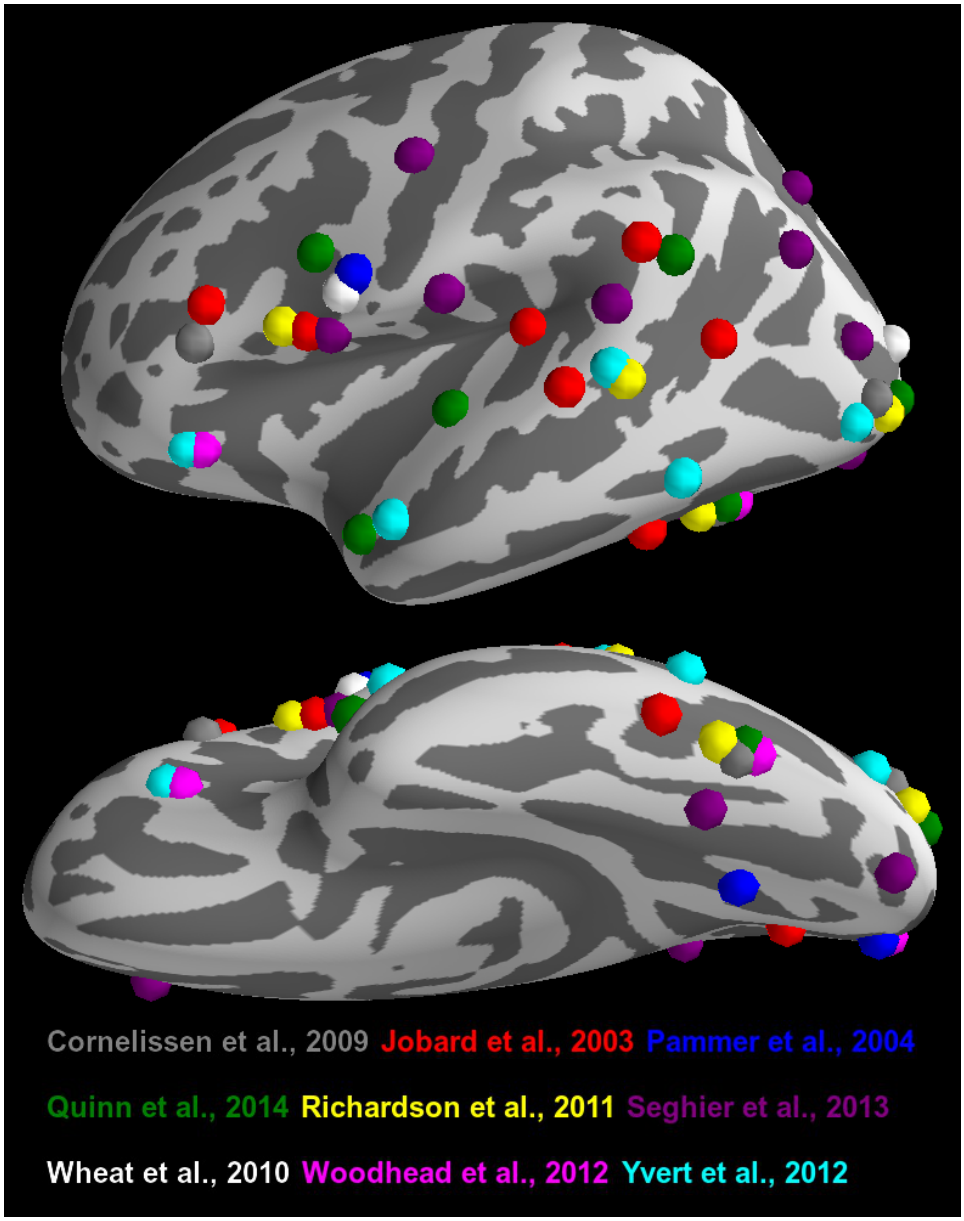

Figure 7.6: Critical nodes within the reading network identified from the literature. Nodes are taken from published studies localising locations in the reading network or investigating the dynamical interactions within a reading network. The node locations are projected onto an average cortical surface taken from the left hemisphere. The nodes used in the present chapter are identified in Green.

ROI	Abbreviation	MNI coord		
		x	y	z
Middle occipital gyrus	MOG	-30	-96	2
Ventral occipito-temporal cortex	vOT	-44	-58	-14
Angular gyrus	AG	-56	-52	22
Superior temporal gyrus/ Heschls gyrus	STG	-48	-14	2
Inferior frontal gyrus *pars opercularis*	IFG	-44	-2	24
Temporal pole	TP	-52	14	-20

Table 7.2: ROIs used in the connectivity analysis. These were generated from local maxima in the beamformer power contrasts and a review of the reading literature

Watson test shows that there is no indication of serial autocorrelation in the residuals of the model. Finally the percent consistency and R^2 shows that a good proportion of the variance in the data and its autocorrelational structure are being captured in four of the participants. The MVAR model fitted from participant four . Critically these values are very similar for both conditions, implying that any differences we see between the two are a result of the structure of the model itself rather than one condition being better described that the other.

7.3.4 Partial Directed Coherence

The average PDC across all participants for Words,Consonants and False Fonts can be seen in figures 7.7, 7.8 and 7.9 respectively. In contrast to the PDC estimated from the resting state data in chapter 4, there is little indication that the PDC here has a $\frac{1}{f}$ profile. This is because the task related model uses many realisations of the same process to allow for shorter analysis windows. As a result the samples are discontinuous which leads to a phase scrambling of any ongoing activity which would be picked up in a longer analysis. This is analogous to a traditional “evoked“ analysis which

Participant		Epochs	SI	DW(τ)	PC	R^2
1						
	Word	523	0.41	1.92	37.47	0.22
	Consonant	528	0.42	1.91	38.92	0.23
	False Font	534	0.42	1.92	38.2	0.23
2						
	Word	498	0.44	1.91	34.51	0.2
	Consonant	508	0.44	1.92	35.96	0.21
	False Font	506	0.44	1.92	33.62	0.2
3						
	Word	526	0.46	1.92	31.4	0.18
	Consonant	524	0.46	1.92	30.15	0.18
	False Font	524	0.46	1.92	31.24	0.18
4						
	Word	490	0.41	1.92	24.58	0.13
	Consonant	484	0.41	1.92	24.51	0.13
	False Font	495	0.4	1.92	26.66	0.13
5						
	Word	517	0.45	1.91	33.54	0.24
	Consonant	517	0.46	1.91	34.51	0.24
	False Font	513	0.46	1.91	34.94	0.24

Table 7.3: Model assessment and validation for the MVAR model fitted in this chapter for each condition and participant. *epochs*: the number of epochs included in the analysis after artefact rejection, *SI*: Stability Index, *DW(τ)*: Durbin-Watson with lag of τ. *PC*: Percent Consistency, R^2: Variance Explained.

would ongoing oscillations leaving only activity which is phase locked to the stimulus. In the same way, the model fitting across epochs will lead to the scrambling the phase of any cross correlations which are not consistent in all epochs.

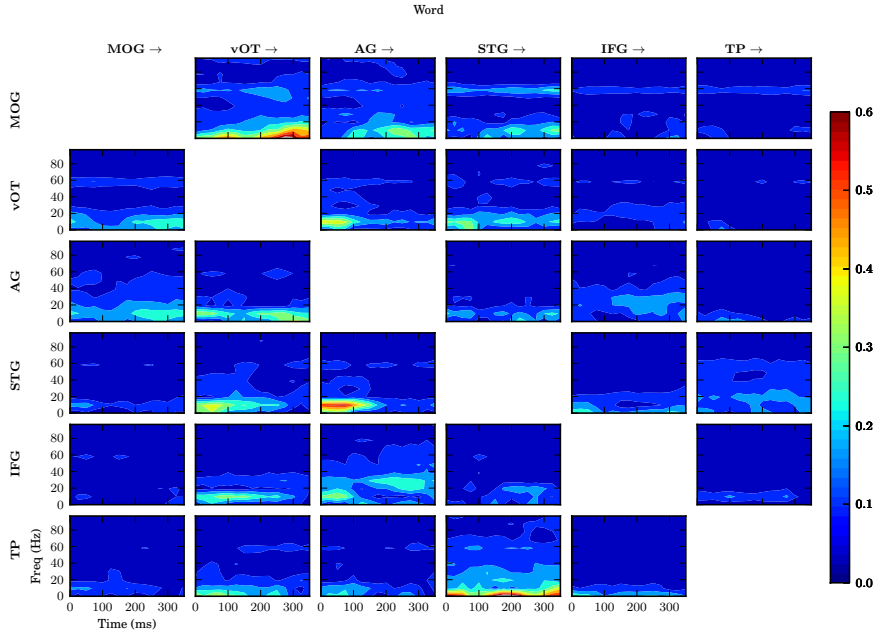

Figure 7.7: The average PDC estimates across all participants for the Word condition. In the matrix, a single subplot indicates the PDC for the connection in which the area in the column heading is driving the area in the row. The x axis indicates time and the y axis indicates frequency. Each time point represented a sliding window as described in figure 7.3. The black line indicates stimulus onset.

A clear result in all three conditions is that the PDC is not equal in all connections. This is simple but critical statement suggesting that some connections within the reading network may be more influential. For instance STG→TP, AG→IFG and STG→MOG all show relatively high PDC values for all conditions. Moreover, this is also true in the prestimulus period and as such is likely to reflect some ongoing interactions between these nodes rather than a specific response to the condition. In contrast, the

MOG→STG, MOG→IFG, MOG→TP and TP→AG connections are all relatively weak across conditions and frequencies. This may be an indication of which connections within the reading network are interacting even without a visual stimulus.

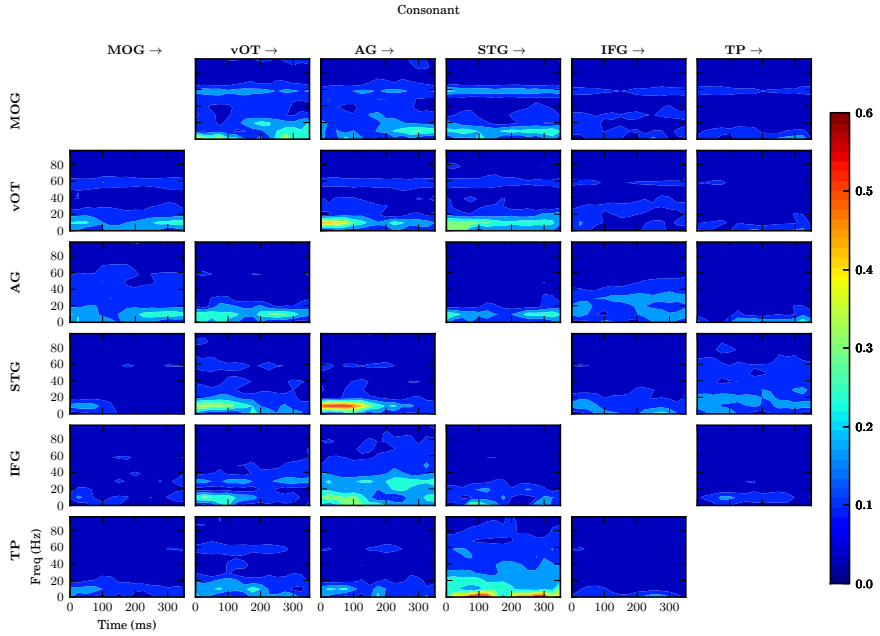

Figure 7.8: The average PDC estimates across all participants for the Consonant string condition. In the matrix, a single subplot indicates the PDC for the connection in which the area in the column heading is driving the area in the row. The x axis indicates time and the y axis indicates frequency. Each time point represented a sliding window as described in figure 7.3. The black line indicates stimulus onset.

7.3.5 Condition Contrasts

The distributions of PDC estimates between each pair of conditions was compared using a t-test, to combine these results across a small group of participants the resulting t-maps were Z-transformed and combined. This meta-analysis across participant was then thresholded at $Z = 2.3$. The

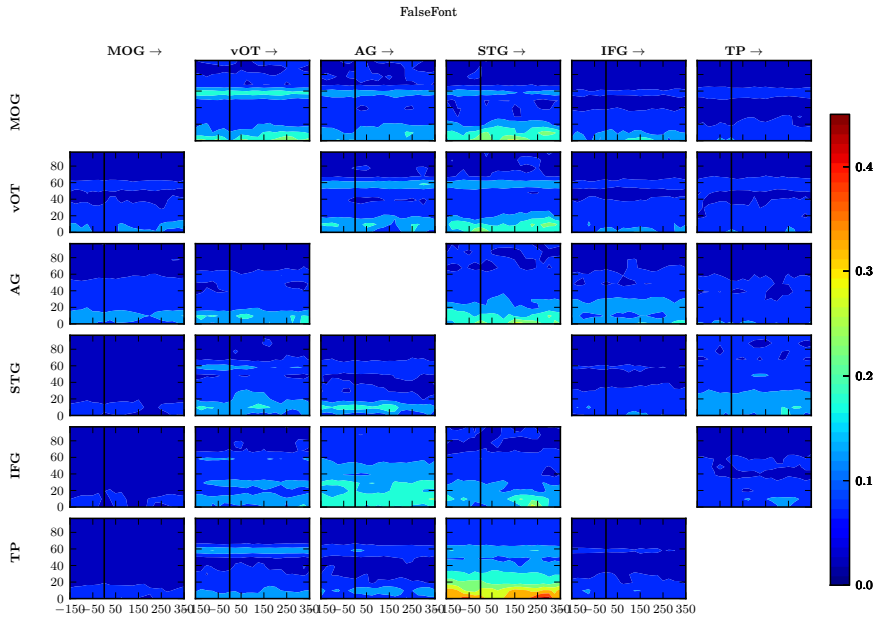

Figure 7.9: The average PDC estimates across all participants for the False Font condition. In the matrix, a single subplot indicates the PDC for the connection in which the area in the column heading is driving the area in the row. The x axis indicates time and the y axis indicates frequency. Each time point represented a sliding window as described in figure 7.3. The black line indicates stimulus onset.

results of this analysis for three comparisons. words - consonant strings, words - false fonts and consonant strings - false fonts can be seen in figures 7.10, 7.11 and 7.12 respectively. The Z-values which survived the combination across participants and the thresholding at $Z = 2.3$ were integrated across frequency within three time bins (Details of the time bins can be seen in figure 7.3). These integrated scores are presented in glass brains for easier visualisation in figures 7.13, 7.14 and 7.15.

Many of the condition differences within individuals were very strong, with t-values reaching as much as ± 50. The T-values were normally distributed around zero for each condition of each participants, as such, a Z-score which survives the group meta-analysis is likely to be associated with a very large t-value across all participants. This is a very conservative approach which should only preserve the largest differences which are consistent across participants.

All three plots show complex patterns of condition specific PDC differences which change rapidly in both time and frequency. The following sections discuss each contrast in more detail.

Words >Consonant Strings

The contrast between Words and Consonant Strings was designed to identify connections whose strength was modulated by the presence of phonological and semantic content. Words induce a larger PDC estimate that Consonant Strings in several outgoing connections from MOG. Within 100ms of stimulus onset Word stimuli produce larger PDC in both the MOG→AG and MOG→STG connections. The effect is more extensive in the connections from vOT in which words produce larger PDC estimates in vOT→AG,vOT→STG and vOT→TP. The vOT→MOG connection is more complex. Though this connection is initially larger for Words around 80Hz the largest difference within 100ms is a region a little later around 30Hz which is larger for consonant strings, though again this connection is larger for Words around 100-150Ms at 10Hz. Several connections originating from visual areas are more influential for Consonant Strings than Words between 100 and 300MS,

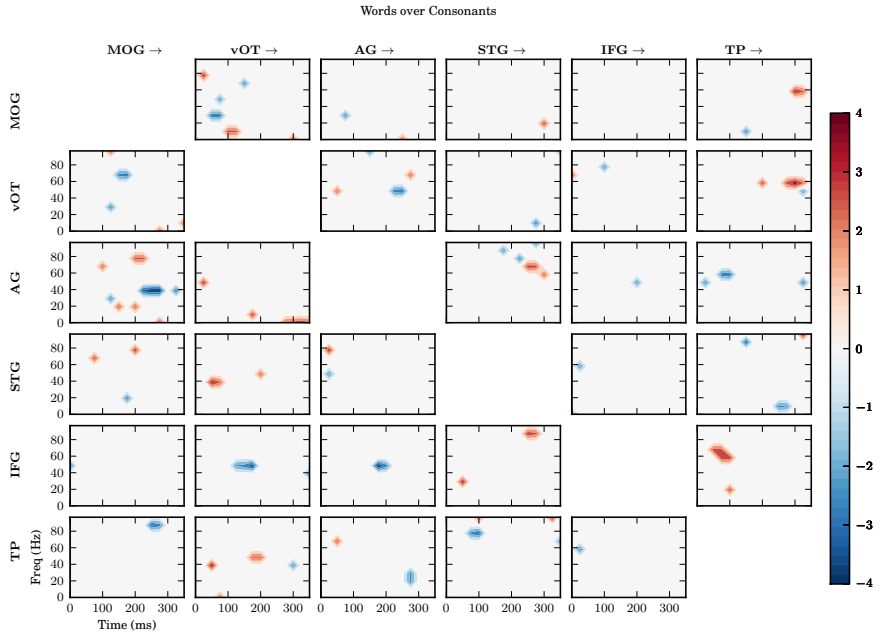

Figure 7.10: The thresholded meta Z map for the Words >Consonant Strings comparison. t-values from each individual were z-transformed and combined to make this Z-map which has been thresholded at Z=2.3. Each subplot indicates the PDC for the connection in which the region in column is driving the region in the row, the x-axis indicates time in milliseconds relative to stimulus onset and the y-axis indicates frequency. A positive value (red) indicates that the PDC for Words was greater than Consonant Strings across participants at that point, and a negative value (blue) indicates that the reverse was true.

most notably MOG→vOT, MOG→AG, MOT→TP and vOT→IFG.

There are no differences between Words and Consonant Strings in connectivity from visual areas to the IFG within 100ms, however both vOT and AG exert more influence on IFG between 100 and 300ms when viewing a consonant string. There are several small differences in top-down influence from IFG within 200ms of stimulus onset. The connection from IFG to vOT is slightly stronger for Words than Consonant strings in the very first window, however the remaining differences suggest that the IFG's influence is stronger for Consonant Strings. All of the differences in figure 7.10 from IFG to the rest of the network are small, however the meta-analysis method used here is very conservative.

Finally, the largest differences between Words and Consonant Strings arises between 50 and 150 ms after stimulus onset in the TP→IFG connection around 60Hz. At the same time, the TP→AG connection is stronger for Consonant Strings. A top-down feedback from TP to both MOG and vOT can be seen later in the experimental epoch around 300ms after stimulus onset.

Words >False Fonts

This was the broadest contrast which should identify differences in connectivity which arise from the presence of orthography, phonology or semantics. Within the first 200ms after stimulus onset, both MOG→vOT and MOG→AG show greater influence for the false font stimuli below 60Hz, though there is an indication of increased influence from MOG to vOT at 100Hz. Words elicited a stronger influence from vOT to both AG and STG within 200ms though only above 10Hz. At exactly 10Hz these same connections were stronger for False Fonts.

Two connections from AG showed a strong preference for False Fonts. A late top-down connection from AG to vOT at 60Hz and the AG→IFG connection around 40Hz around 200ms after stimulus onset. In contrast there is an indication for an early increase in PDC for Words in the AG→MOG, AG→IFG and AG→TP connections.

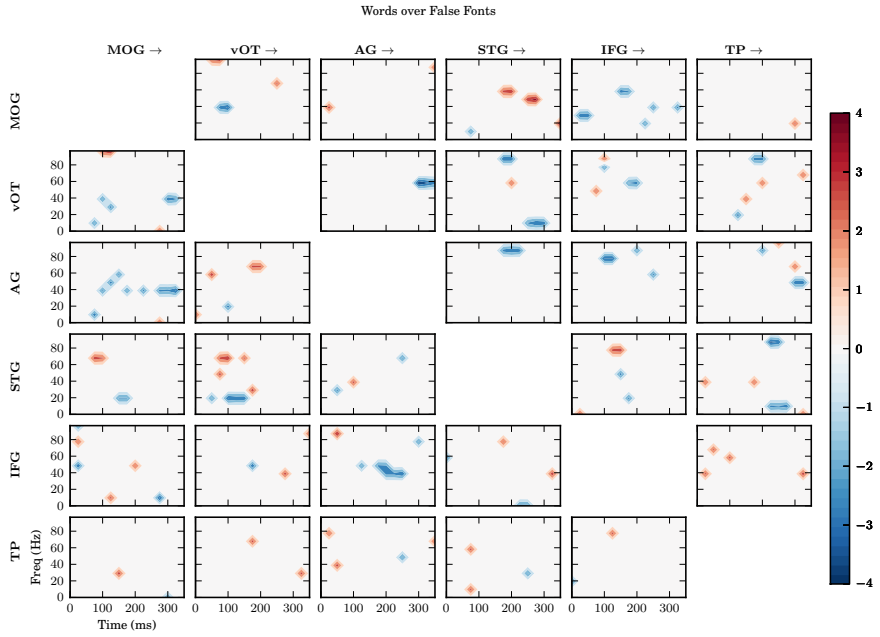

Figure 7.11: The thresholded meta Z map for the Words >False Fonts comparison. t-values from each individual were z-transformed and combined to make this Z-map which has been thresholded at Z=2.3. Each subplot indicates the PDC for the connection in which the region in column is driving the region in the row, the x-axis indicates time in milliseconds relative to stimulus onset and the y-axis indicates frequency. A positive value (red) indicates that the PDC for Words was greater than False Fonts across participants at that point, and a negative value (blue) indicates that the reverse was true.

The STG→vOT and STG→AG connections were larger for False Fonts than Words, both between 200 and 300ms after stimulus onset. In contrast STG→MOG was more strongly active for Words around the same time. Top-down influence from the IFG to MOG was stronger for False Fonts than Words within 200ms of stimulus onset, however the picture was more mixed in IFG→vOT and IFG→STG in which there are indications for greater PDC in both conditions. The IFG to AG connection showed a clear preference for False Font stimuli.

Finally there is a preference for Word stimuli within 100ms in both the TP→STG and TP→IFG connections, however this picture becomes more mixed later in the epoch.

Words >Consonants ∩ False Fonts

Several connections showed the same modulations in the Words >Consonant Strings and Words >False Fonts contrast. The connections which were more active to words were. MOG→STG around 100ms and TP→IFG before 100ms. As these connections were more active for Words than Consonant Strings and False Fonts, they are likely to reflect phonological or semantic processes.

In contrast, MOG→AG around 200-300ms, vOT→IFG around 150-200ms, AG→IFG around 150-250ms and TP→STG around 200-300ms were all more active to either False Fonts or Consonant Strings than Words. These connections are harder to interpret. They are all later in the epoch (generally around 200ms or after) these connections are recruited when a word is not easily or automatically recognised. As such, they may represent a second attempt by the system to extract information from the stimulus after the initial processing has failed to identify the word form.

Consonants >False Fonts

Only the longer range connections originating in the MOG show a difference between Consonant Strings and False Fonts within 100ms. These are a preference for Consonant Strings in MOG→STG and MOG→IFG and

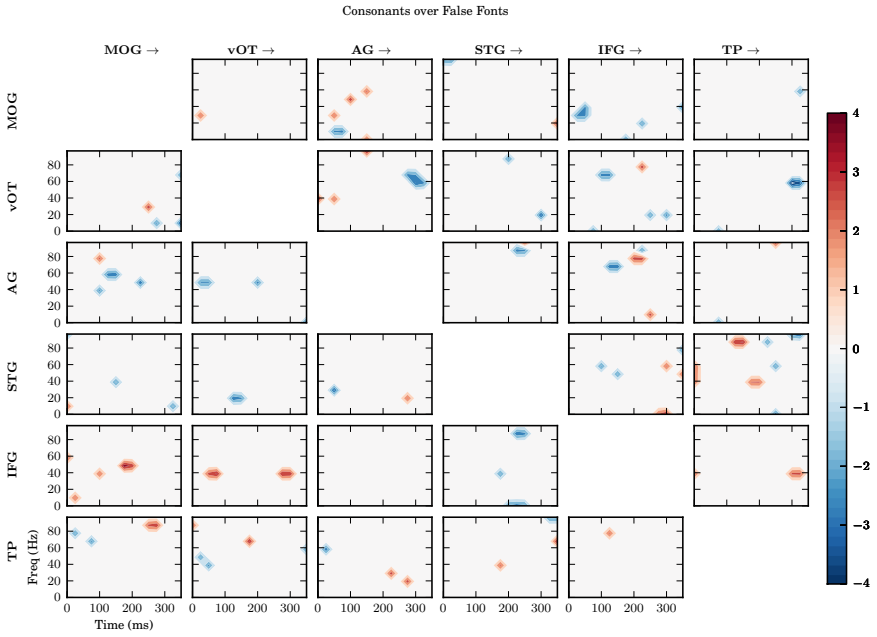

Figure 7.12: The thresholded meta Z map for the Consonant Strings >False Fonts comparison. t-values from each individual were z-transformed and combined to make this Z-map which has been thresholded at Z=2.3. Each subplot indicates the PDC for the connection in which the region in column is driving the region in the row, the x-axis indicates time in milliseconds relative to stimulus onset and the y-axis indicates frequency. A positive value (red) indicates that the PDC for Consonant Strings was greater than False Fonts across participants at that point, and a negative value (blue) indicates that the reverse was true.

a preference for False Fonts in MOG→TP. Later in the epoch different points in time and frequency show a preference for different conditions in the MOG→AG connection while Consonants induce a larger PDC estimate in the MOG→IFG connection and False Fonts produce more PDC in the MOG→STG connection.

Within 100ms of stimulus onset Consonants produce larger PDC in the vOT→MOG and vOT→IFG connections while False Fonts produce larger PDC in vOT→AG in the same time frame. The largest difference in this contrast is a preference for False Fonts late in the connections from AG to vOT.

There are larger PDC estimates in the IFG→MOG connection for False Fonts within 100ms of stimulus. False Fonts also produce larger PDC in the IFG→vOT and IFG→AG connections later in the epoch. Finally Consonant Strings produce larger PDC in early in the TP→STG connection and later in TP→IFG. In contrast, False Fonts produce a large amount of PDC in the TP→vOT connection around 300ms after stimulus onset.

Words ∩ Consonants >False Fonts

Several connections showed consistent behaviour in both Words and Consonant Strings relative to False Fonts. The MOG→IFG connection was larger in both Words and Consonant Strings around 200ms after stimulus onset at 40Hz. As such, this connection is likely to reflect an interaction associated with the existence of orthographic information in the stimulus.

In contrast, many connections were consistently larger for False Fonts than Words and Consonant Strings. These include MOG→AG between 100-200ms at 60Hz, vOT→STG at 100ms around 20Hz, AG→vOT at 300Hz at 60Hz and IFG→MOG around 50ms at 30Hz. These connections were consistently larger for False Fonts implying that their function might be related to visual processing of shapes without recognisable orthography. Another possibility is that these reflect processes which are only engaged when difficult or irregular orthographic challenges are present. In such cases when the first pass of orthographic decoding has not succeeded, a wider network

is engaged perhaps to produce a likely estimate.

7.4 Discussion

This chapter has applied the methodologies established in previous chapter to characterise the rapid dynamics within a network of brain regions as a person reads a word, consonant string or string of false font characters. The connectivity within the reading network showed widespread changes in response to all three stimulus categories. Moreover there was evidence of integration between visual and language regions within the network for all categories within 100ms.

7.4.1 Orthography

The critical contrasts to isolate connections which may carry orthographic information were Consonant Strings >False Fonts and Words >False Fonts. Both Words and Consonant Strings contain recognisable characters from the Latin alphabet which should engage processing related to orthographic decoding and representation. The contrast between Consonant Strings and False Fonts showed that the presence of such recognisable characters changes the pattern of connectivity within the reading network dramatically. Critically, when Latin characters are present both the MOG→IFG and vOT→IFG connections are strengthened within 100ms of the stimulus onset. In contrast, when the characters are not recognisable in the False Font condition, the strength of the IFG's influence on both MOG and vOT is increased. This shows that when recognisable visual characters are present the visual regions produce a greater influence over the IFG. In contrast, when the orthographic information cannot be decoded the IFG begins to influence the visual areas.

This suggests that the visual cortex nodes (MOG/vOT) are sufficient to produce a first pass orthographic decoding which feeds forward into the reading network, however when this first system does not reach a solid representation then influence from the IFG increases, perhaps providing top-down

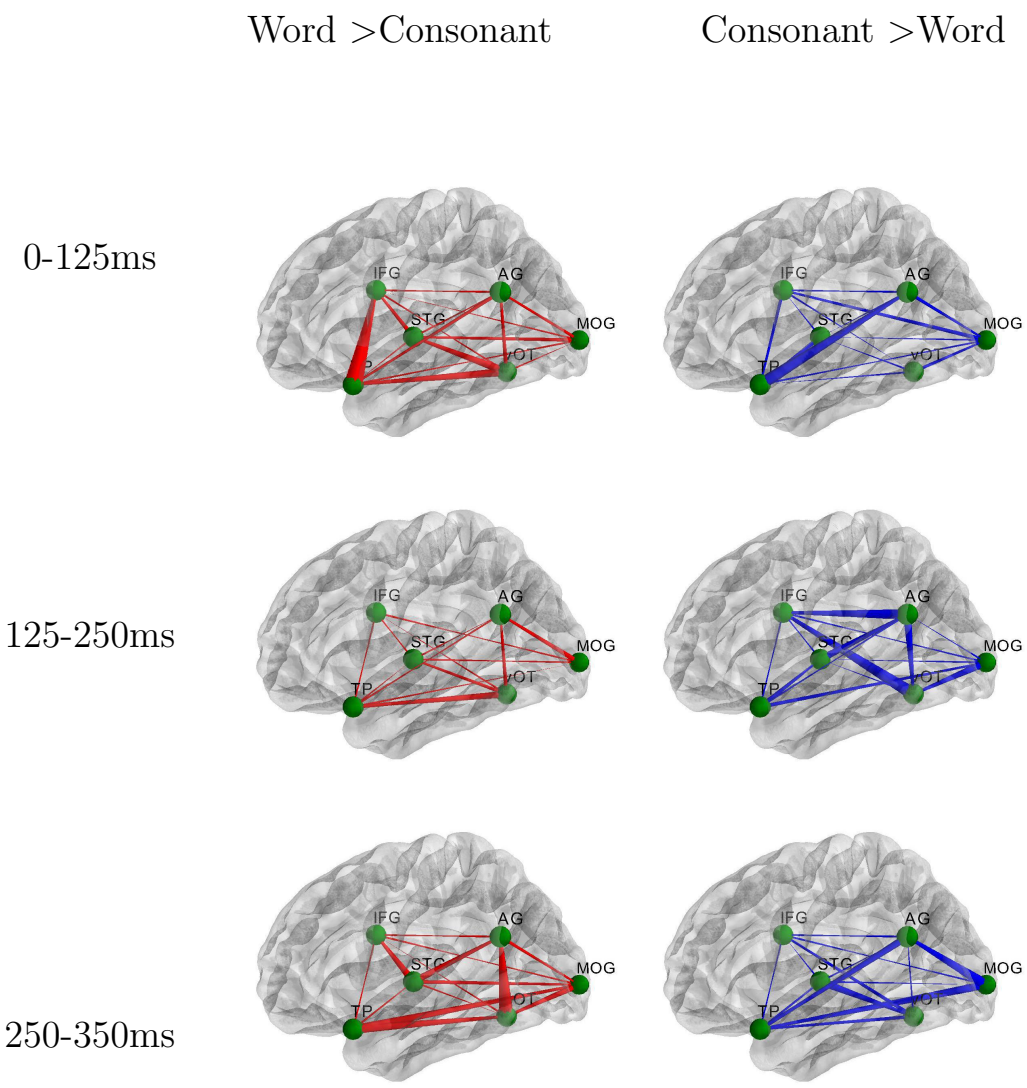

Figure 7.13: The connections within the reading network who show larger activation to Words relative to Consonant Strings. The size of the connections are established by integrating the results seen in figure 7.10 across frequency and within the three time bins described in figure 7.3. Each edge represents a connection whose source is the thick end and target the thin end. The overall size of the edge indicates the extent to which the connection is modulated by the condition contrast.

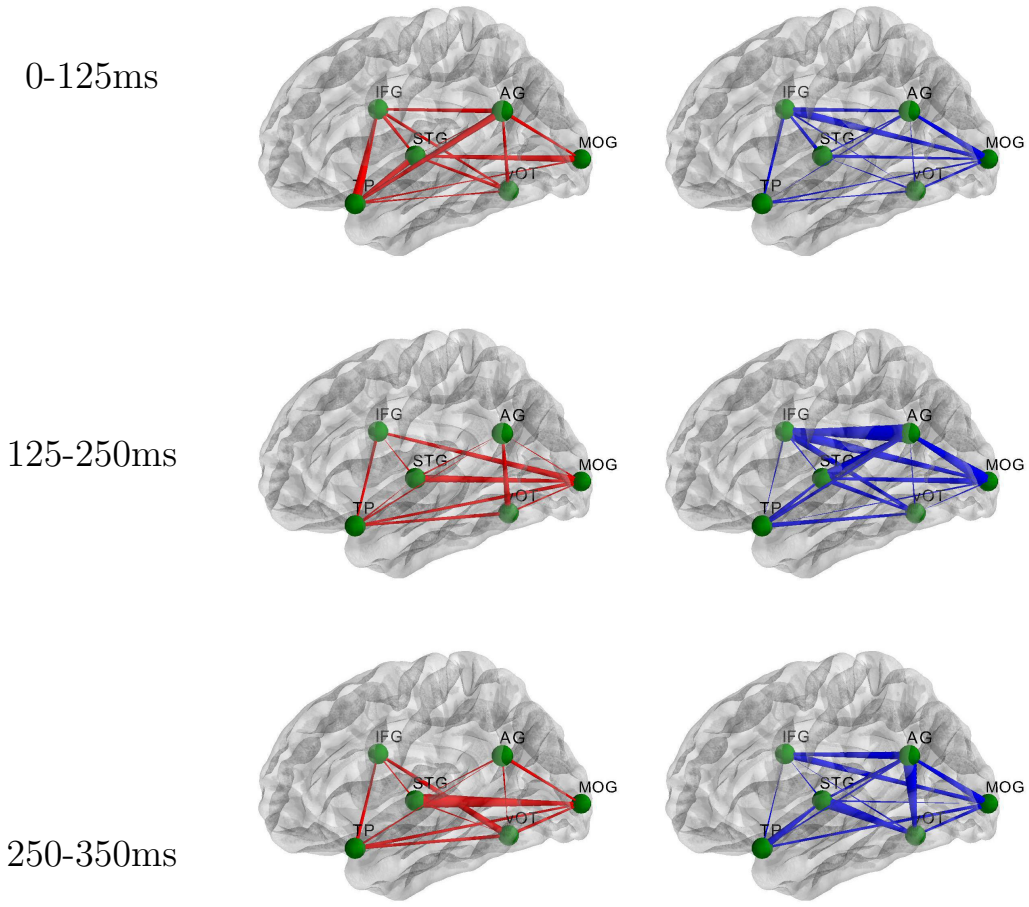

Figure 7.14: The connections within the reading network who show larger activation to Words relative to False Fonts. The size of the connections are established by integrating the results seen in figure 7.11 across frequency and within the three time bins described in figure 7.3. Each edge represents a connection whose source is the thick end and target the thin end. The overall size of the edge indicates the extent to which the connection is modulated by the condition contrast.

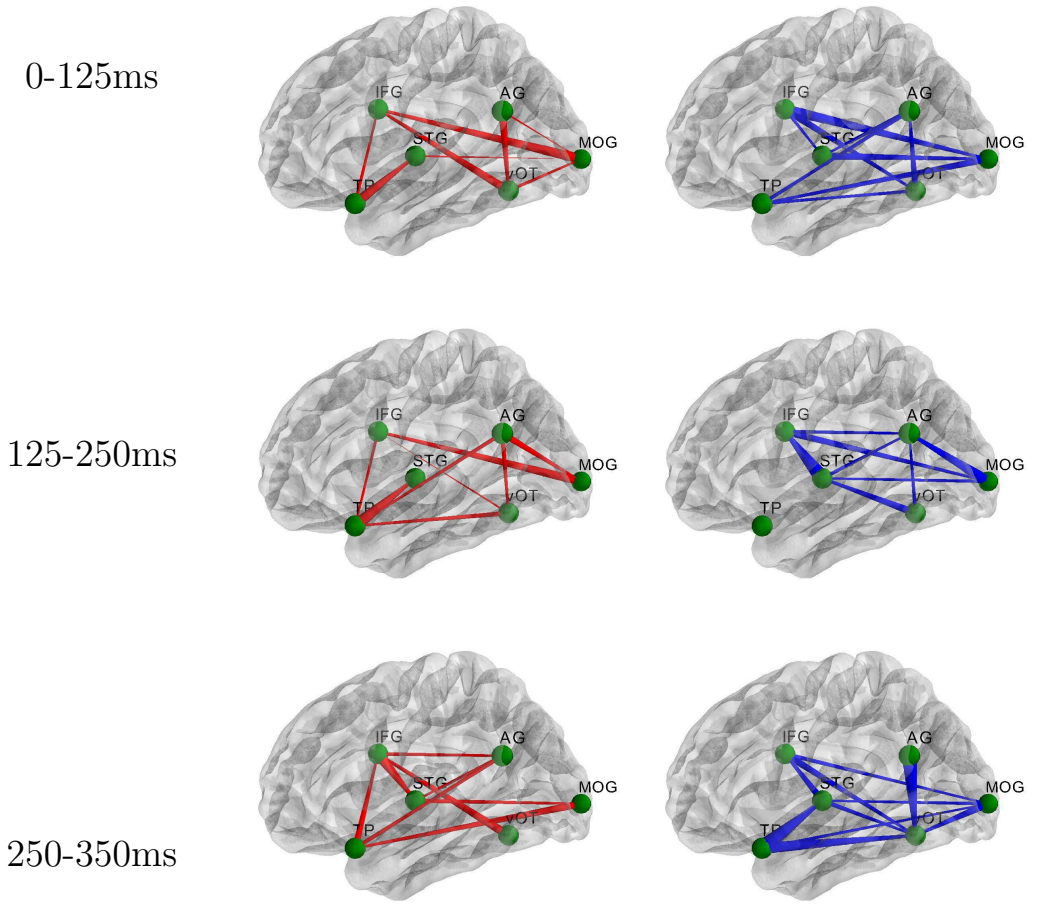

Figure 7.15: The connections within the reading network who show larger activation to Consonant Strings relative to False Fonts. The size of the connections are established by integrating the results seen in figure 7.12 across frequency and within the three time bins described in figure 7.3. Each edge represents a connection whose source is the thick end and target the thin end. The overall size of the edge indicates the extent to which the connection is modulated by the condition contrast.

influence which helps to converge on a representation of the orthography if there is one to be found. In the context of the literature on word recognition, this account sits between the feedforward (Dehaene et al., 2005; Vinckier et al., 2007) and interactive (Price and Devlin, 2011) models of word recognition. The present results suggest that the visual system may be sufficient for orthographic decoding when it is clearly present, however top-down influence from the IFG does come into play when no solution is reached. The precise effect of the influence from the IFG is not clear from this study alone. It may be the case that the IFG is not providing any influence relevant to the recognition of the word, rather the connections we see in this chapter reflect some other function which occurs *after* the visual system has ruled out linguistic content. In contrast, this may be an indication that the influence of the IFG on visual areas during word recognition is only necessary in challenging cases, and in fact the visual system may need input from IFG to rule out orthographic information.

Corroboration for this second claim comes from Wheat et al. (2010) who showed that rapid masked prime words containing orthographic or phonological information about a forthcoming target word modulates the induced activity seen in the LIFG within 100ms of stimulus onset. This suggests that relevant information about potential phonological content of a word is able to reach the IFG in time for the top-down influences seen in this chapter.

A study by Woodhead et al. (2012) used DCM to characterise connectivity within the reading network in response to real words and false font stimuli. They showed that the winning model included a top-down connection from left IFG to left vOT which was more active for the word rather than false font stimuli. The present results provide a partial replication of this word in that the Word >False Font contrast indeed shows an increase in PDC in the IFG→vOT connection within 200ms of stimulus onset, however this is a relatively small difference and occurs within the context of many other changes occurring at the same time.

7.4.2 Phonology And Semantics

The Word >Consonant String contrast is designed to isolate connections which may be carrying phonological or semantic information. Similarly to the previous contrasts the presence or absence of this information can be seen to produce large changes in the configuration of the reading network. Very quickly after stimulus onset, Words induce a stronger influence in both the vOT→STG and vOT→TP connections, in addition real words greatly increases the influence from TP to the IFG.

Later in the epoch, around 300ms after stimulus onset Words produces a larger connection from TP to both MOG and vOT relative to both Consonant Strings and False Fonts. As this difference in present in both contrasts it is likely that these top-down connections are relevant to the phonological or semantic information in the stimulus rather than its orthography or visual form. Similarly Words increase the influence from AG to TP within 100ms of stimulus onset in both the Word >Consonant String and Word >False Font contrasts, again this likely to reflect phonological or semantic processes. The involvement of the temporal pole in these differences might suggest that these modulations are more related to the semantics than the phonology.

Finally, the connection most likely to reflect phonological or sound based information in these results is from MOG to STG which shows larger PDC estimates for words than either Consonant String or False Fonts around 60Hz just prior to 100ms after stimulus onset.

7.4.3 Models Of Visual Word Recognition

The present results do not clearly support or reject either of the main theories of visual word recognition as discussed above. There is clear evidence of top-down influence from IFG to visual areas during all three conditions, however it's role remains unclear. The early changes in connectivity from IFG in the Consonant String and False Font conditions (within 100ms) suggests that this connectivity is related to the recognition process itself. However, the clearest evidence of strong feedforward influence from visual areas is in

the real word condition, suggesting that the top-down connection might be least important in this case.

In a Bayesian sense, the IFG might be providing a further information about prior expectations about the stimulus which the visual system can used to converge on the *a posteriori* most likely visual stimulus which gave rise to the incoming information. However this top-down expectation might be redundant when the orthographic information is trivially decoded from a clearly presented word or consonant string. The brain may only use the priors from the IFG when the visual evidence is not sufficient to provide an orthographic representation of an incoming word. This approach to cognition has been widely influential in recent year and the present results may partly support it (Lee and Mumford, 2003). This idea can be further refined by explicitly defining a "prior expectation" to include the current evidence in a brain region rather than the *a priori* expectation irrespective of the information currently in the system. There is increasing evidence that the low frequency content of a visual stimulus is able to rapidly ascend the processing hierarchy and reach frontal brain regions very quickly (Bar et al., 2006). The top-down influence from the IFG may then include a very rough representation of the linguistic content of the stimulus gleaned from the very rapidly processed low frequency visual information. Such an idea might be supported by the priming effects seen in MEG (Cornelissen et al., 2009; Wheat et al., 2010) which suggests that phonological information is very quickly available in IFG.

7.4.4 Blurring Orthography and Phonology

This chapter makes a simplifying assumption that orthographic and phonological information can be cleanly separated (see table 7.1. However the truth may not be this simple. Vinckier et al. (2007) used a similar false font to word stimulus set in an fMRI study to identify which regions in visual cortex are recruited with increasing amounts of psycholinguistic information. They conclude that the more anterior parts of the ventral visual stream are only activated with the most word-like stimuli. This is mainly interpreted

in terms of orthography. However, a similar response bias to more word-like stimuli can also be seen in the LIFG. This may arise as the more orthographically word-like stimuli will inevitably become more pronounceable as well. For example a consonant string will be associated with the sounds of it's constituent letters whereas a false font stimulus will not.

In terms of this present results, this may suggest that the top-down influence from the LIFG might be working the other way around. It might be the case that the LIFG's role is to interrogate words which are more challenging to pronounce rather than contributing to orthographic decoding.

7.4.5 Methodological Statements

A critical limitation of this study is the number of participants. A large number of samples are needed to fit a reliable MVAR model, and in order to use a short-analysis window in this chapter several hundred experimental trial were collected per condition per participant. This allowed for a detailed and robust MVAR model to be estimated for each person. Practical limitations meant data collection was restricted to 5 participants, severely limiting any group inferences that can be made to the meta-analysis described in the methods section. This is a conservative and principled approach for combining the results of independent tests of the same experiment, however the significance of the outcome was determined by an arbitrary thresholding of the final Z-map. This chapter has demonstrated the application of a TDE-MVAR-PDC approach for characterising the rapid dynamics associated with visual word recognition, however further work should endeavour to collect data from larger numbers of participants, allowing the application of group-level non-parametric permutation tests to establish the statistical significance of the final results. Moreover, the thresholding of the Z-maps was done without any cluster correction. Such correction may eliminate some of the smaller differences seen in the results above, particularly weaker Z-values with no visible neighbourhood.

More participants and cluster corrected group statistics may also have the benefit of cleaning up the highly complex results arising from this anal-

ysis. The theoretical discussion above has focuses on the larger connections surviving the group meta-analysis and in particular connections which show consistent modulations in two of the three contrasts. As such, many smaller or less consistent connections are not discussed. With more robust group statistics the functional relevance of these other connections may be interrogated as well.

Two of the participants who completed this study had a much lower accuracy rate in the in-scanner behavioural task than the rest of the cohort. Table 7.1 shows that participants 3 and 4 were only able to correctly identify around 55 and 73 percent of the catch trials respectively, though their reaction time was equivalent to the group. It is unclear whether this reflected a practical or attentional which may have impaired performance, however table 7.3 clearly shows that these participants also showed the lowest percent consistency and R^2 measures from the fitted MVAR models. This may imply that these participants were relatively disengaged with the experiment leading to noisier data in which the MVAR model under-performs, even when several hundred epochs are available for modelling.

7.4.6 Future Directions

The present results raise intriguing questions about the role of the IFG in relation to the recognition of visual words. There are, however, several outstanding questions. Firstly, does the influence from the IFG arise as the brain is trying "harder" to recognise the Consonant String and False Font stimuli? Visual input to the brain is noisy, and as such, when a Consonant String or False Font stimulus is perceived when a real word is anticipated the brain may seek more evidence from linguistic processes in addition to the visual system to rule out the possibility that the stimulus is a noisy presentation of a real word and only then stop looking for phonological and semantic associations with the stimulus.

Finally, the theoretical accounts discussed in this chapter do not extend far beyond the visual system. However there are many models of reading which include the entire reading network up to the point of reading out

loud. These models raise new questions not considered here, for instance, there may be more than one way to recognise a word. Different pathways within the reading network may be tuned towards different features and depending on the content of the word itself, one or other of these routes may be engaged.

7.5 Conclusions

This chapter has sought to characterise the dynamic changes in connectivity within the reading network and how these are modulated by the presence of different categories of linguistic information. Through the application of the TDE-MVAR-PDC approach outlined in previous chapters a wide network of connections has been established which is highly sensitive to the linguistic content of a word. Critically, the relationship between visual nodes and the IFG seen in this chapter suggests that there exists very rapid connectivity between these regions in all conditions. The direction of the connectivity seems to be related to the presence of orthographic information in the stimulus. The visual areas exert a driving influence when clear Latin characters are presented, while the False Fonts lead to a larger top-down influence.

Chapter 8

Multiple Routes To Recognition

8.1 Introduction

The previous chapter outlined the connectivity within the reading network in response to stimuli containing different levels of orthographic and semantic information. Critically two of the three conditions in this previous experiment were not recognisable as real words, as such the analysis was tuned to look at the response of the system to the presence or absence of specific types of information. This is a very blunt manipulation as many factors within the domain of real words modulate the speed and/or ease of their recognition. This chapter builds on the previous results characterising the connectivity associated with recognisable or unrecognisable stimuli by establishing the extent the recognition process is modulated by word frequency. This subtler contrast allows us to identify whether visual word recognition of any word is founded on a consistent set of brain state or is it the case that high and low frequency word recruit distinct functional states during the recognition process?

The faster recognition of words that appear more frequently in written language is one of the most robust effects in cognitive psychology and has influenced many cognitive and computational models of visual word recognition (Grainger, 1990).A major argument in the word recognition literature has been whether there are distinct routes to recognition or if phenomena such as the word frequency effect can arise from a densely interconnected network.

8.1.1 Multiple Routes To Recognition

Many models of reading have been based on the idea that there are two or more routes from a seeing a word to recognising it (Davelaar et al., 1978; Coltheart et al., 1993; Ellis and Young, 1996). Of these a dual route architecture with two routes to recognition has been most influential (Coltheart et al., 1993). Dual-route models suggest that there are two pathways from visual input to lexical access: a direct mapping from graphemes to the lexicon and an indirect process in which graphemes are mapped to the lexicon via phonology. The mapping from orthography to phonology is rule gov-

erned, however the mapping for a given word may either be in line with these rules or not. All words are processed by both streams and according to dual-route theory, for regularly spelt words the routes should converge on the same answer, however in the case of irregular words the indirect route would process the word according to regular spelling rules and interfere with the direct route, thus slowing the correct identification. The rapid recognition of highly frequent words is though to arise from the direct route which is so efficient in processing high frequency words that an identification is achieved before the indirect route has a chance to produce any interference.

This explanation has some support from neuroimaging results which suggest the direct pathway activates bilateral visual cortex and the left middle temporal gyrus (MTG) whereas the indirect pathway leads to greater activation in the left inferior frontal gyrus (LIFG) *pars opercularis*, insula, thalamus and caudate nucleus (Fiebach et al., 2002). Moreover, the graphophonological and lexico-semantic processes relating to the indirect and direct pathways respectively are consistently associated with distinct clusters of activity within the left hemisphere (Jobard et al., 2003). Connectivity analysis of fMRI data within the left hemisphere further supports a dual route model Levy et al. (2009) in which the occipito-temporal cortex is involved in sublexical orthographic processing however familiar words can be read with a direct connection from early occipital cortex to parietal regions. In this model the occipitotemporal cortex would only be necessary for the processing of infrequent words.

While dual-route models remain influential they have given rise to a lot of debate, critically there it is not the case that the dual-route architecture is the only mechanism that is able to account for a large amounts of behavioural effects (Humphreys and Evett, 1985). Moreover, some neuroimaging data suggests that the connectivity within the reading network in the brain are rather more complex than we might expect from a simple dual-route account. Kujala et al. (2007) mapped coherences within the left hemisphere during reading and showed that many of the nodes within the network were densely interconnected with little evidence for two separate routes. In addition, there is evidence for three separate processing routes

from occipital cortex to regions within the temporal cortex during reading (Richardson et al., 2011).

8.1.2 Connectionist Account Of Recognition

Connectionist accounts of recognition take a different approach. In Connectionist models such as the triangle model, information enters the system, before all cognitive aspects of the word are resolved simultaneously. Though there may be some separation between the orthographic, phonological and semantic nodes within this network, they converge on a final representation simultaneously and information in any one point in the network may influence any other point. As such, these models may be better represented by a series of states within the network rather than an ordered set of discrete processes.

Connectionist/interactive models of reading such as the Parallel Distributed Processing (PDP) models (McClelland et al., 1986; Seidenberg and McClelland, 1989) would suggest the structure of the network processing a word should depend on the networks previous experience with that word. These networks do not comprise a hierarchical processing stream, rather they distribute incoming information to all parts of the network at once. The 'triangle' models include three clusters of nodes representing orthography, phonology and semantics. Visual information enters the model through the orthography cluster, however this cluster interacts with the phonology and semantics clusters immediately, even before the orthographic representation has completely resolved.

The connections between the nodes in are weighted and may be trained or adapted to optimise the recognition of a set of words. In this framework, more frequent exposure of a word to a Connectionist model would lead to the weights having more opportunity to optimise. This would suggest that the word frequency effect arises as the links between spelling and sound are learned statistically over many presentations of many different words. This task will be simpler for regular spelling-to-sound relations however with sufficient presentations the weights within the network could optimise to

accommodate irregular words.

Similarly a Bayesian model might adapt its priors to account for the greater or lesser likelihood of a given word appearing. This would imply that high frequency words would be more efficiently processed, or in the Bayesian framework, require less evidence for them to be confidently recognised (Norris, 2006). Though Bayesian models of word recognition and perception in general (Knill and Pouget, 2004) are gaining in popularity, there are few specific hypotheses about the origin of prior expectations within the brain. In contrast, there are specific hypotheses about the location of the aspects of the triangle model in the brain (Carreiras et al., 2014). Moreover, there is considerable overlap with the brain regions thought to play a role in dual-route theories (Jobard et al., 2003).

According to arguments such as these, the PDP approach would predict that the whole reading network should be modulated by word frequency. A skilled reader would have a system optimised for frequent words and as such the network would converge on a correct identification faster than a low frequency word. In the case of low frequency words, a regularity effect arises as the network has been (as yet) unable to compensate for the statistically unlikely spelling of irregular words and as such identification becomes more difficult. Neuroimaging results looking to characterise the performance of a whole network have found interconnected functional networks with the strength of many connections being modulated by task demands such as reading sentences in which each word is presented individually at a manageable or unmanageable rate (Kujala et al., 2007). The success of this statistical learning mechanisms used in many PDP models implies greater global efficiency when processing more frequent words rather than a greater efficiency in one distinct route.

8.1.3 Recognition And Functional State

The section above outlines the evidence for several routes to word recognition each with different functional roles, though there is very strong evidence for this the computational literature on Connectionist modelling might sug-

gest that this is an oversimplification. Perhaps difference in the recognition process for different word are the result of globally different functional brain states rather than a greater or lesser recruitment of one or more relatively discrete pathways. This chapter will characterise the directional influences within the reading network over time in using the same methodology as the previous chapter. The dynamic changes in connectivity are then characterised as a series of functional states.

8.1.4 Connectivity

The previous chapter showed that the psycholinguistic content of a word-form can modualate the directional connectivity within the reading network, critically the top-down influence from the LIFG onto visual areas. The design from Chapter 7 interrogated differences when psycholinguistic content (orthography, phonology or semantics) are present or absent, the present chapter builds on these findings by looking for similar modulations in connectivity depending on the ease with which psycholinguistic content can be accessed. This contrast is performed with the word frequency effect as described above.

8.2 Methods

8.2.1 Participants

Six right handed participants (4 male, mean age 29) took part in this experiment. All participants had normal vision and no history of dyslexia. This study was approved by the York Neuroimaging Centre Research Ethics Committee and conducted according to their guidelines. Each participant provided full written consent at each session and was fully debriefed at the end of the experiment.

8.2.2 Experimental Design

Stimuli

The stimuli were 1500 words evenly split between three word frequency groups: high, medium and low. The words were taken from the CELEX2 database (Baayan et al., 1996) and details on the average word frequency for each group can be seen in table 8.1. The words were uniformly distributed between four to six characters in length.

	Word Frequency		
Condition	mean	max	min
High	2.2	3.54	0.33
Medium	0.63	0.86	0.37
Low	-0.21	0	-0.47

Table 8.1: Word frequency statistics for the three conditions used in this study. All values are in log_{10} frequency per million.

Procedure

Participants silently read the words which were projected at 60Hz in a white font onto a black screen with a viewing distance of around 75cm. Vertical nonius lines acted as a fixation reference.

Recording was done over ten sessions across multiple days. Each trial consisted of a single word presentation of $300ms$ followed by a $3000 \pm 500ms$ uniformly jittered interstimulus interval (ISI). 150 catch trials were evenly dispersed throughout scanning in which the word from the preceding experimental trial was repeated and the participant pressed a button with their left index finger, an additional $2000ms$ was added to the ISI after catch trials to ensure motor artefacts did not affect subsequent trials. Stimuli were presented using Presentation $V14$ (http://www.neurobs.com).

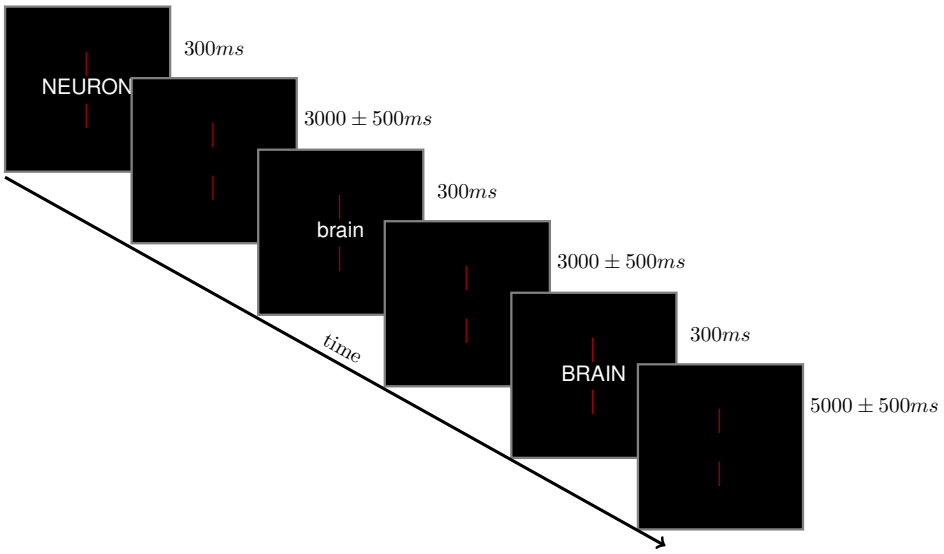

Figure 8.1: Schematic showing the stimuli presentation and ordering. The two red lines remain on the screen for the duration of the experiment and their center point indicated the fixation for the participant. The first two stimuli are experimental conditions and the final stimulus is an example of a catch trial. The case of the font was alternating for the duration of the experiment. The $3000ms$ interstimulus interval included a 'blink period' during which participants were encouraged to blink in order to minimise ocular artefacts contaminating the experimental epochs.

8.2.3 Data Acquisition

Data were continuously recorded at 4069.017Hz using a 4D Neuroimaging Magnes 3600 whole head system. The data were passed through a low-pass filter set at 1500Hz. Participant head shape and reference coil location were recorded using a 3D digitiser (Polhemus Fastrak) allowing for co-registration with a high resolution anatomical T1 MRI image acquired using a GE 3.0T Signa Excite HDx (Kozinska et al., 2001).

8.2.4 Data Analysis

Artefact Rejection

Data were visually inspected for artefacts and trials containing muscle or eye movements were discarded from the analysis. The MEG data were then co-registered with the individual's structural T1 MRI scan and the MNI 152 standard brain to allow for source analysis to be conducted.

Beamformer Contrasts

Beamforming estimates the current at a source location in the brain from a weighted sum of the observed magnetic fields on the sensors. A type 1 vectorized, linearly constrained minimum variance (LCMV) beamformer (Van Veen et al., 1997; Huang et al., 2004) was used before the oscillatory power in a passive window (-200ms to 0ms) was compared to the power in four sequential active windows (0ms - 200ms, 100ms - 300ms, 200ms - 400ms & 300ms - 500ms). These contrasts were performed in an alpha (5 - 15Hz) and beta (15 - 25Hz) frequency bands. Non-parametric label permutations were used to compute statistical thresholds (Nichols and Holmes, 2002)

Virtual Electrodes

Connectivity analysis was based on six nodes identified from the beamformer contrast results. Current was estimated for each trial in these locations using a vectorized LCMV beamformer as defined by Van Veen et al. (1997) which

has been shown to generate the most accurate source reconstructions (Johnson et al., 2011). VEs were estimated at the full sampling rate (4069.017Hz) across the entire epoch for each location.

8.2.5 Connectivity

The MVAR model fitting procedure was identical to the one used in the previous chapter. Each epoch was split into 150ms sliding windows which advanced in 25ms steps from 75ms prior to stimulus onset to 350ms after (see "Windows" in figure 8.2). All the time series within each sliding window were used to compute an evoked R matrix which was used to estimate A through the Vieira-Morf algorithm. The PDC is then calculated from the A matrix for each sliding window.

The PDC within each functional window was compared to a control or passive window defined at -225ms to -75ms relative to stimulus onset. To allow for statistical comparison the trials within each window were jackknifed 50 times. In other words the PDC within each window was calculated using a random 80% of the available epochs each time. The distributions of the PDC estimates across jackknives were then compared to the control window using an independent samples t-test. p values were Bonferroni corrected to correct for multiple comparisons.

8.2.6 Condition Contrasts

Finally the connectivity within the reading network for the high and low frequency words are directly compared using the same Z-score based meta-analysis seen in the previous chapter. In brief, any differences within each connection are characterised in each individual using a t-test, the results of which are then Z-transformed. These Z-scores are then summed across participants and divided by the square root of the total number of participants. This is Stouffer's method for combining Z-score from several independent iterations of the same test. This group Z-map is then thresholded at Z=2.3.

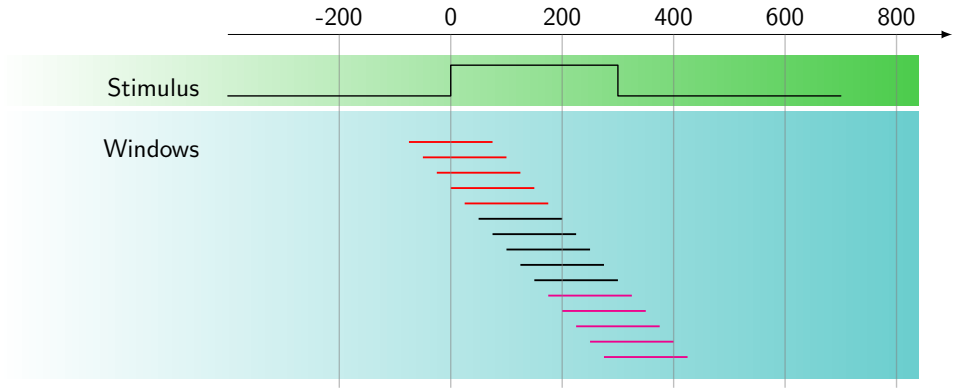

Figure 8.2: Schematic of the sliding analysis windows used in the connectivity analysis. The arrow at the top denotes time in milliseconds. The stimulus line shows when the stimulus appears within the epoch. The active sliding windows are shown in the section with the blue background. The PDC from within each window is presented in subsequent sections and labelled with the centre time of the window. The different colour bars in this section denote the three time bins in which the condition differences in PDC are integrated over for the glass brain plots.

8.3 Results

8.3.1 Behavioural Data

The six participants correctly responded to 94.7% of the catch trials with an average reaction time of 740 milliseconds.

8.3.2 Node Selection

The reading network in the previous chapter was identified from a combination of local maxima and minima in a set of sliding window beamformer contrasts and the literature. This chapter uses the same approach.

8.3.3 Model Validation

The beamformer contrast results again show the general trend for activation to spread from posterior to anterior brain regions as reported in the previous chapter and the literature Marinkovic et al. (2003); Pammer et al. (2004).

Participant	Percent Correct	Reaction Time (ms)
1	98	325.4
2	91.3	318.7
3	88.7	325.0
4	86	336.0
5	93.3	342.3
6	98	326.8
Average	92.55	329.03

Table 8.2: Behavioural results per participant. Percent correct indicates the proportion of animal name catch trials which the participant was able to respond to across the six experimental blocks. Reaction time is the average time in milliseconds for a correct response from the participant. Only responses within 1 second of the catch trial were included in this analysis.

		MNI coord		
ROI	**Abbreviation**	x	y	z
Middle occipital gyrus	MOG	-30	-96	2
Vetral occipito-temporal cortex	vOT	-44	-58	-14
Angular gyrus	AG	-56	-52	22
Superior temporal gyrus/ Heschls gyrus	STG	-48	-14	2
Inferior frontal gyrus *pars opercularis*	IFG	-44	-2	24
Temporal pole	TP	-52	14	-20

Table 8.3: ROIs used in the connectivity analysis. These were generated from local maxima in the beamformer power contrasts

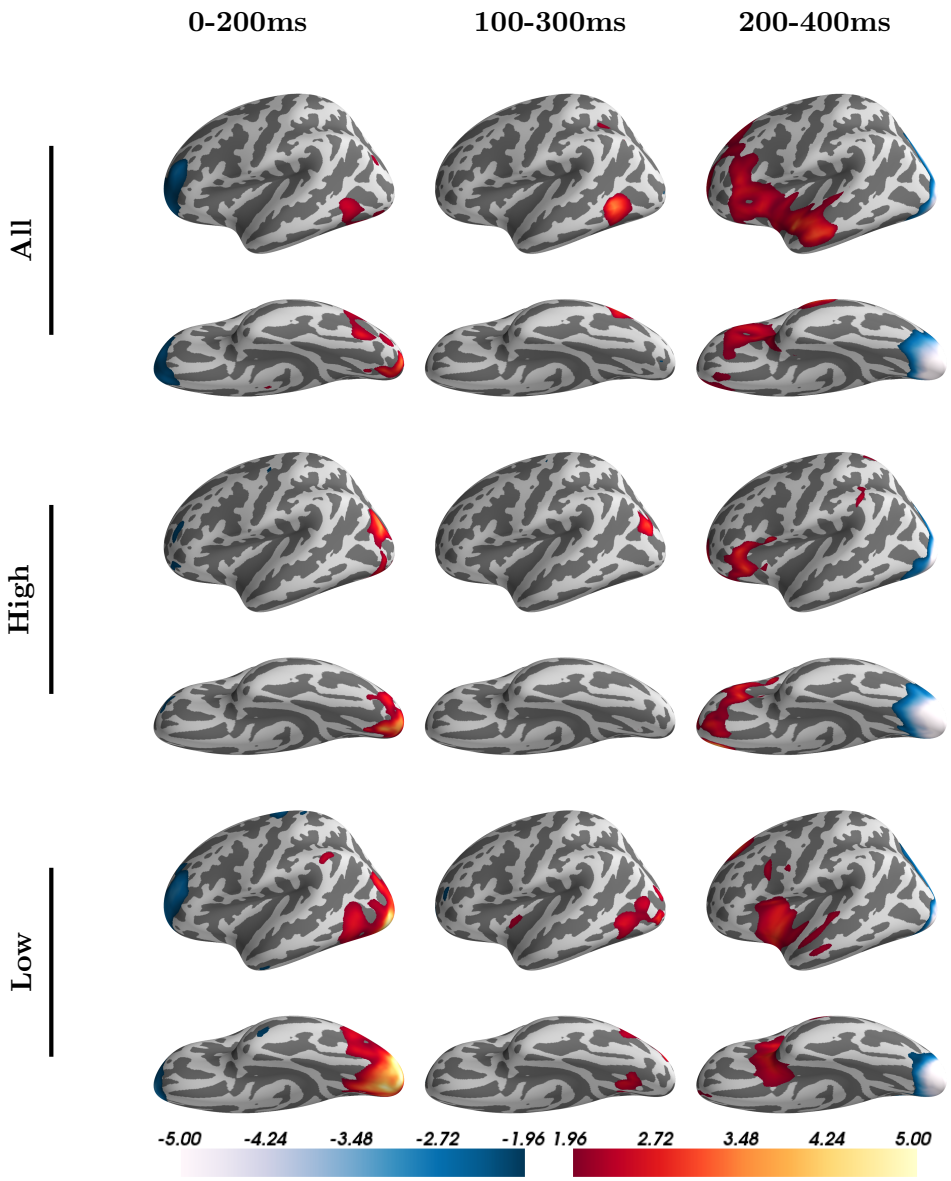

Figure 8.3: Z-stats from an active-passive beamformer power contrast in the alpha band ($5 - 15$ Hz), each surface contains the combined Z-stats across the five participants arbitrarily thresholded at $Z = 1.96$. The sliding window results from the sliding window contrast across all words is shown followed by the same contrast for just the high frequency word and finally just the low frequency words. The results have been rendered onto the surface of a fsaverage brain with the lateral (top) and ventral (bottom) views of the left hemisphere shown.

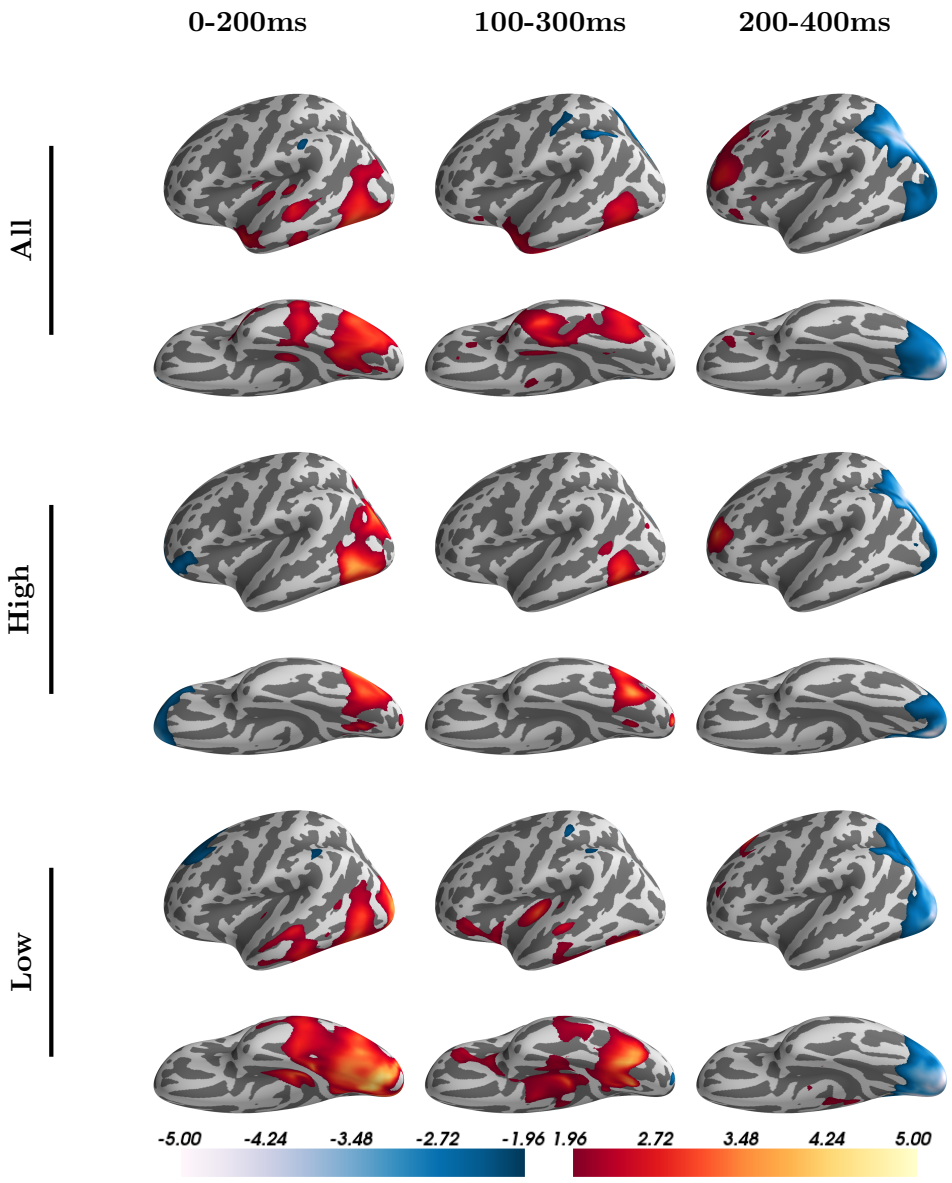

Figure 8.4: Z-stats from an active-passive beamformer power contrast in the beta band ($15 - 25$ Hz), each surface contains the combined Z-stats across the five participants arbitrarily thresholded at $Z = 1.96$. The sliding window results from the sliding window contrast across all words is shown followed by the same contrast for just the high frequency word and finally just the low frequency words. The results have been rendered onto the surface of a fsaverage brain with the lateral (top) and ventral (bottom) views of the left hemisphere shown.

The large early activations in the LIFG seen in chapter 6 are not seen in these results, however as this finding has been replicated a number of times Pammer et al. (2004); Cornelissen et al. (2009); Wheat et al. (2010) this is likely to be due to the small number of participants included in the study.

The beamformer contrasts show a strong justification for inclusion of the middle occipital gyrus, ventral occipito-temporal cortex, temporal pole and superior temporal gyrus as seen in the previous chapter. Moreover there is an indication that the angular gyrus is activated in response to low frequency words in the alpha band. Guided by these results and the literature, the same reading network is used as seen in chapter 6. This is justified in part by the present beamformer contrasts and comes with the large advantage that the connectivity results in chapters 6 and 7 will be easily comparable. The nodes included in the reading network are summarised in table 8.3.

The model validation results are summarised in table 8.4. Though some participants have quite different number of epochs for the two conditions (as a result of the epoch rejection), the models are very consistent across participant and condition.All the stability indices are well below 1, indicating that the models are stable, which implies stationarity. Moreover the Durbin-Watson test shows that there is no indication of serial autocorrelation in the residuals of the model. Finally the percent consistency and R^2 stats show that a good proportion of the variance in the data and its autocorrelational structure are being captured. Critically these values are very similar for both conditions, implying that any differences we see between the two are a result of the structure of the model itself rather than one condition being better described that the other.

8.3.4 Partial Directed Coherence

PDC was estimated from the MVAR parameters for each sliding window for each participant and each condition. The average PDC across participants for the high and low word frequency conditions can be seen in figure 8.5. The first critical result is that, similarly to the previous chapter, the distinctive $\frac{1}{f}$ frequency profile seen in the resting state chapter has been strongly

Participant	Condition	Epochs	SI	DW(τ)	PC	R^2
1						
	Low	467	0.45	1.91	35.38	0.21
	Medium	459	0.44	1.91	35.13	0.2
	High	456	0.45	1.91	33.86	0.2
2						
	Low	349	0.5	1.92	28.99	0.19
	Medium	358	0.5	1.92	28.95	0.19
	High	350	0.49	1.92	28.38	0.18
3						
	Low	414	0.41	1.91	36.85	0.22
	Medium	423	0.41	1.91	35.63	0.21
	High	423	0.41	1.91	35.95	0.22
4						
	Low	388	0.45	1.91	31.38	0.22
	Medium	398	0.46	1.91	31.31	0.22
	High	404	0.46	1.91	32.21	0.22
5						
	Low	420	0.44	1.91	37.83	0.21
	Medium	421	0.44	1.91	37.93	0.21
	High	427	0.43	1.91	37.58	0.21
6						
	Low	479	0.41	1.9	41.97	0.24
	Medium	477	0.41	1.9	42.16	0.24
	High	472	0.42	1.91	41.56	0.24

Table 8.4: Table containing the model validation results for the six participants in this chapter. SI: stability index, DW(τ): Durbin-Watson test with lag of τ, PC: Percent consistency (Ding et al., 2000) and R^2: The amount of variance explained by the model.

attenuated. This is as the MVAR models in this chapter are calculated from the average lagged covariance matrix from a short window across many trials. This allows for the estimation of directed influences that are consistent within a time window across many realisations. As these windows are discontinuous any ongoing oscillations that are not task related may not be modelled. In contrast, the longer continuous windows used in the resting state analysis will be able to pick up on such oscillations. A deeper discussion into how this may lead to a $\frac{1}{f}$ frequency profile can be found in chapter 4.

Similarly to the directionality analysis in chapter 4, several clear asymmetries can be seen in figure 8.5. This directionality is often consistent over the two conditions (for instance vOT→MOG and STG→AG) though may be modulated in strength over the course of the epoch.

Many of these connections are similar to the Words condition seen in the previous chapter. For instance the influence from MOG is stronger over vOT and AG than the rest of the network. STG has a stronger influence on MOG then in the reciprocal connection and finally the Temporal Poles influence is stoniest on the STG than the rest of the network.

8.3.5 Low vs High Frequency Contrast

In order to more directly characterise the differences between the high and low frequency words in time, a direct comparison between the conditions was carried out within each individual and combined using Stouffer's method. The results of this test can be seen in figure 8.6.

As seen in the previous chapter, the condition differences have a widespread and complex impact on the connectivity of the reading network. There are many rapidly differences in almost all connections

High >Low Frequency

The blue values in figure 8.6 indication connections which were stronger for the high frequency words than the low frequency words. The earliest differences occur within 100ms in which the MOG→vOT connections is

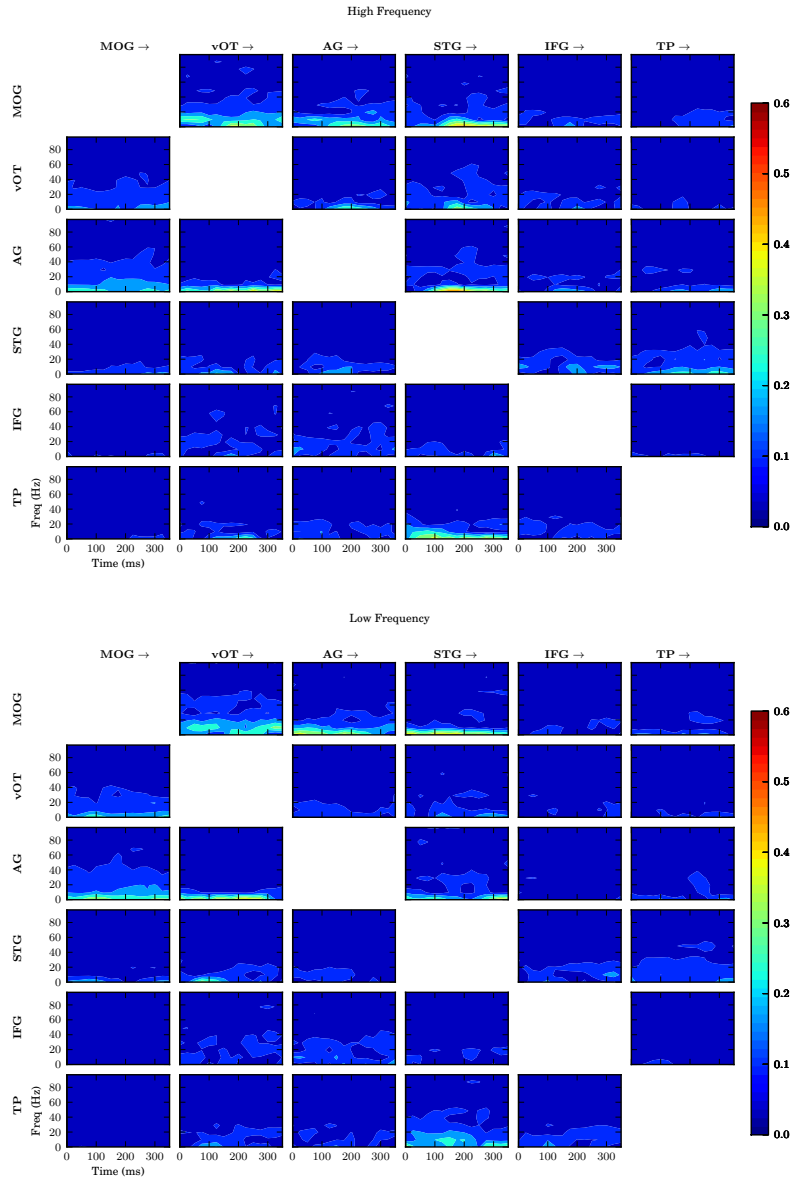

Figure 8.5: *top*: The average PDC estimates across all participants for the high word frequency condition. *bottom*: The average PDC estimates across all participants for the low word frequency condition. In each matrix, a single subplot indicates the PDC for the connection in which the area in the column heading is driving the area in the row. The x axis indicates time and the y axis indicates frequency. Each time point represented a sliding window as described in figure 8.2.

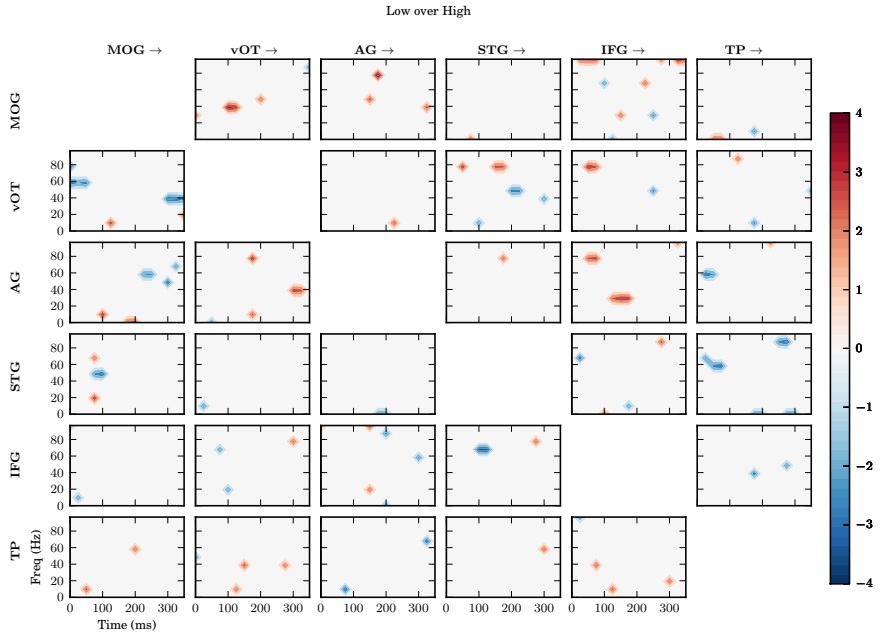

Figure 8.6: The difference in PDC between the Low and High word frequency conditions. The difference within an individual was characterised using a t-test before the t-maps were converted to Z-scores and combined using Stoffer's method for combining the results of independent repetitions of the same test. This group level Z-map was arbitrarily thresholded at $Z = 2.3$. Each subplot indicates the connectivity for the connection in which the region in the column is driving the region in the row. The x-axis indicates time in ms relative to stimulus onset and the y-axis denotes frequency in Hz. A red value indicates that that connection was more active for low frequency words, whilst the blue values indicate that the PDC was higher for high frequency words.

more influential for high than low frequency words. There is a much smaller preference for high frequency words in MOG→IFG around 10Hz at the same time.

The next strongest early preferences for high frequency words occur within 100ms in the TP→STG and TP→IFG connections. Around the same time point there is a small indication at the AG→TP connection shows the same bias for high frequency words. Around 100ms there is a larger PDC estimate in the STG→IFG connection before the STG→vOT connection shows the same effect around 200ms.

In general, the high frequency words tend to increase the early influence from MOG to vOT and IFG while also increasing the influence from TP onto AG and STG. STG has a more driving influence on vOT and IFG later in the epoch.

Low >High Frequency

The red values in figure 8.6 indicate connections which were stronger for low than high frequency words. The largest differences within 100ms which are larger for low frequency words are in connections originating in the IFG. the IFG→MOG/vOT/AG and TP connections are all stronger for the low frequency words within 100ms, though the IFG→TP difference is small. This influence continues in IFG→AG in between 100 and 200ms after stimulus onset.

There is a small indication of larger PDC in the MOG→TP connection around 10Hz very early in the epoch. Similarly, there are small increases in MOG→vOT/AG/STG from around 80ms to 200ms. The top-down connection vOT→MOG is also larger for infrequent words during this time.

Both AG and STG appear to influence visual areas more for infrequent words, though this occurs later than the top-down influence from the IFG.

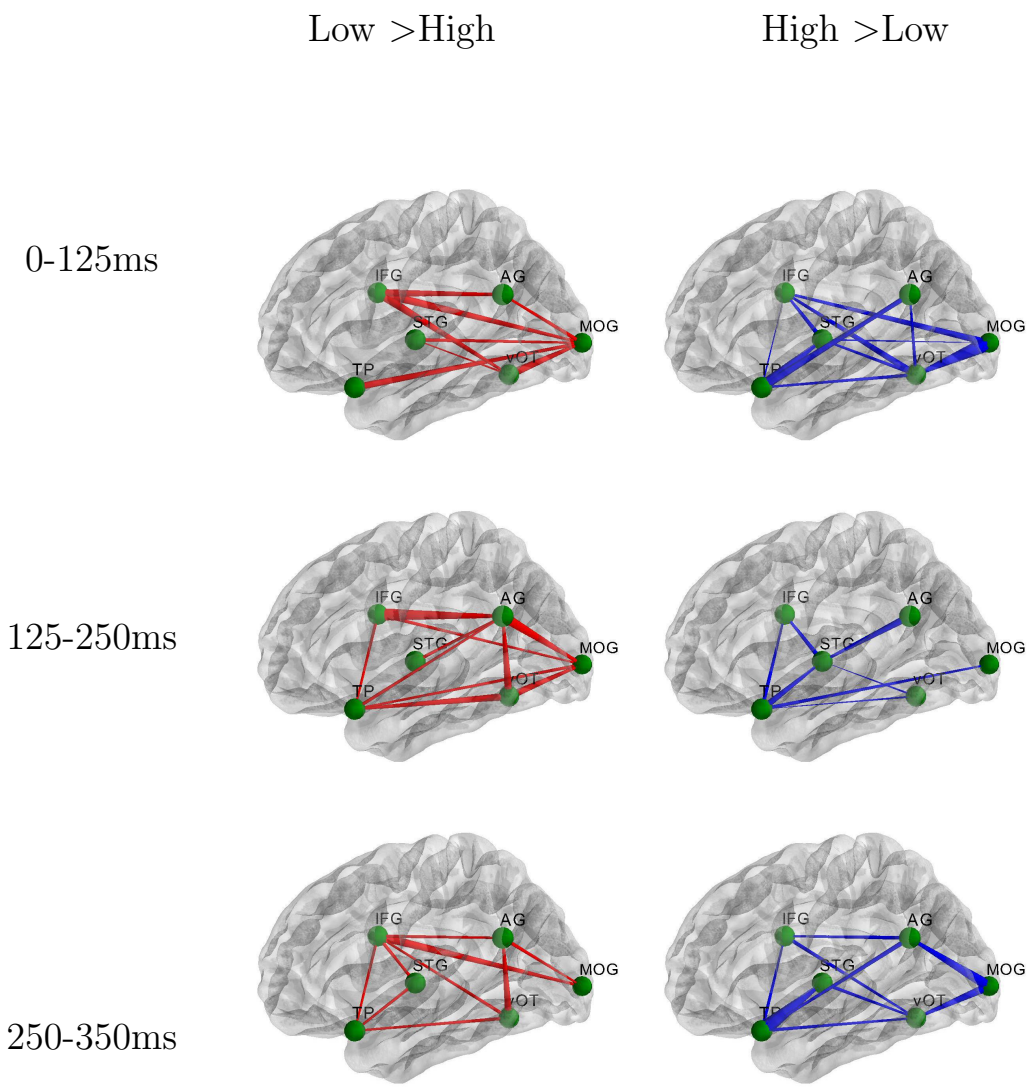

Figure 8.7: The connections within the reading network who show larger activation to Words relative to False Fonts. The size of the connections are established by integrating the results seen in figure 8.6 across frequency and within the three time bins described in figure 8.2. Each edge represents a connection whose source is the thick end and target the thin end. The overall size of the edge indicates the extent to which the connection is modulated by the condition contrast.

8.4 Discussion

This chapter seeks to characterise the connectivity within the reading network using the TDE-MVAR-PDC approach outlined in previous chapter. The critical question arose from the results of chapter 6, which showed that top-down influence from the IFG to MOG was strongest for False Font stimuli. This chapter aimed to ask whether the role of the IFG was specific to False Font stimuli or if it played a role in recognition of challenging words in general.

The neural substrate which may underlie the rapid recognition of high frequency words can be seen in these results. Critically, high frequency word induce stronger feedforward connections from visual areas to IFG within 100ms. At the same time, the influence from the TP on STG and AG is stronger for high frequency words. This earlier engagement of the temporal pole might indicate an earlier access to semantic content which can begin to feedback into the system within 100ms. In contrast the low frequency words show little indication of large modulations of the connectivity from the temporal pole. Rather driving influences from AG,STG and IFG are more prominent.

8.4.1 The IFG

The connectivity from the IFG to the rest of the reading network was considerably stronger for the more challenging low frequency words. This modulation occurs very rapidly after stimulus onset, with some of the long range top-down connections increasing within 100ms of stimulus onset. In relation to the results in the previous chapter, the present results strongly support the argument that the IFG's early role in visual word recognition might be much more prominent with challenging/ difficult to recognise stimuli. The present results build on the previous chapter by demonstrating that this early top-down influence does occur with challenging linguistic stimuli rather than just word-like stimuli with no linguistic content.

These findings are further supported by MEG evidence that linguistic aspects of a visual word such as its phonology can modulate IFG activation

on a very rapid time-scale (Cornelissen et al., 2009; Wheat et al., 2010). A TMS investigation into whether this fast phonological access is *necessary* for visual word recognition has suggested that disruption to the IFG prior to 225ms after stimulus onset does not affect reading speeds (Wheat et al., 2013). However, this study used high frequency words, which the present results show to induce the least top-down influence from the IFG. Moreover, the study was carried out in Dutch which has a very transparent orthography which, unlike English, will not contain any irregular spelling-to-sound conversions.

Overall, there is clear evidence that the linguistic content of words and word-like stimuli can modulate inferior frontal and temporal brain regions within 200ms of stimulus onset (Pammer et al., 2004; Cornelissen et al., 2009; Wheat et al., 2010) though the function of this fast response remains unclear. Previous DCM studies have suggested that the IFG may provide rapid feedback to visual an temporal areas during word recognition (Yvert et al., 2012; Woodhead et al., 2012). The results of the previous two chapters does support the existence of such early feedback, however it is most prominent in words and word-like stimuli which are not easily recognised, such as low frequency words and non-words.

A Bayesian approach may provide the best interpretation as it allows for a fully probabilistic account. As visual information feeds into the brain it combines with *a priori* expectations (perhaps such as word frequency) to produce an *a posteriori* probability distribution across the set of possible stimuli which could have produced the sensory input. The rapid access to phonology discussed above suggests that this first process may happen simultaneously on several levels, resulting in a probability distribution based on a very fast visual and phonological analysis of the sensory input. If the *a posteriori* evidence in the visual system is sufficient, that suggests the brain has recognised the word and the phonological evidence not further considered in the context of recognition. In contrast, if the *a posteriori* visual probabilities are ambiguous (as would be the case for infrequent words or non-words) in that several or many words are similarly likely to have produced the incoming bottom-up sensory information, then the rapid phonological decoding

may provide further top-down evidence. The combination of these two may then produce a match, or be sufficient to conclude that the stimulus was not a word.

8.4.2 Multiple Routes To Recognition?

The results show that the IFG is preferentially influential for low rather than high frequency words, and that the semantic regions are influential earlier for high frequency words. This may be taken as support for a dual-route account in which the indirect phonemic route is only influential for difficult words (Coltheart et al., 2001). However the interpretation may not be that simple. The wider results from the last two chapters have shown that False Font, Consonant String, Frequent Words and Infrequent words all induce modulations in activity across most of the reading network. Even the False Font stimuli induce large increases in directed functional connectivity between higher "linguistic" brain regions which would not be predicted by a hierarchical feedforward model of word recognition.

It would appear that the entire network responds to a word-like visual stimulus as would be predicted by a distributed Connectionist account. However the psycholinguistic and orthographic content of the word does induce differences in specific parts of the reading network as might be predicted by a more modular approach. The answer is likely to lie in between these two accounts in which a widely distributed and highly interactive network of specialised brain regions co-operate to identify an incoming stimulus. The connectivity within the reading network does change rapidly over time, further suggesting that there may be several stages of processing, though each stage may be associated with the coordinated activity of a number of nodes in the network. Moreover, the same node may not play the same role in different stages of processing with different network contexts (Price and Friston, 2005). It may be that the functional "state" of whole network is better able to describe how psycholinguistic aspects of a word are processed than a modular or route based account.

8.4.3 Methodological Statements

The results in this chapter are subject to the same statements about group level statistics as the previous chapter.

8.4.4 Future Directions

This chapter has provided direct evidence that top-down influence from the IFG occurs during the recognition (or attempted recognition) or challenging word-like stimuli with or without linguistic content. Further work looking to characterise this effect should first look to build a larger cohort of participants to provide more robust statistical confidence to the presented findings.

The Hidden Markov Model approach applied to the resting state data in chapter 5 could potentially be applied to task data such as this. By fitting a set of hidden states and their transitions, the very complex dynamics seen in the previous two chapter could be simplified considerably.

Spelling Regularity

Finally, more information about the precise function of the top-down influence from the IFG could come from a design which manipulates both word frequency and spelling regularity. The relationship between these two variables has been the source of considerable debate in the computational modelling literature and dual-route and Connectionist model provide different accounts of how the reading network might cope with irregular and infrequent words. Critically, the dual route account would suggest that irregular words such as *yacht* cannot be easily identified by the indirect grapheme-to-phoneme conversion route and this would interfere with the direct recognition process which identifies a word form directly from its visual form. In contrast, a Connectionist account would suggest that the entire network is able to compensate for irregular spelling, but only through repeated exposure to the stimulus. As such, the whole network should be able to deal with a frequent but irregular word but converge slower for an infrequent irregular word.

As such, an early modulation of the connectivity of brain regions associated with phonology for any irregular word might support a dual-route account whereas if any frequent word is processed similarly irrespective of regularity then this would support the Connectionist models.

8.4.5 Conclusion

Overall, this chapter aimed to identify the neural substrate which gives rise to the word frequency effect. High frequency words were found to induce larger feedforward connections from MOG and top-down influence from the temporal pole. In contrast, the high frequency word were associated with much stronger top-down influence from the IFG on several other regions. This builds on the results of the previous chapter by confirming the linguistic relevance of the top-down influence from the IFG. Moreover, in the context of previous literature, there is increasing evidence for the role of the IFG in processing difficult word and word-like stimuli is becoming clearer.

Chapter 9

Discussion

9.1 Chapter summaries

9.1.1 Maths

An analysis pipeline which reconstructs local brain dynamics from observed multivariate data before statistically assessing the direction of information transfer was introduced in this chapter. The theoretical justification of the first stages of this pipeline are drawn from non-linear dynamics, specifically, the work of Takens (1981) who showed that the dynamical structure of a system may be reconstructed from a vector of time lagged observations of its behaviour. The representation of the system in its embedding space is then used as the predictors in a multivariate autoregressive model which characterises which delays contain information about any future observations. This information transfer was quantified in the frequency domain using partial directed coherence, which expresses the information flow from node A to B as a proportion of the total influence from A to the whole network.

9.1.2 Simulations

While the theoretical justification of the analysis pipeline is outlined in chapter 2, this chapter validated several important practical issues. Firstly, the accuracy with which a model can be fitted from a set of observed data is critically dependent on the ratio of observations to parameters. A simple simulation was used to demonstrate that while more observations are always better, a ratio of around 5:1 should lead to an accurate model. Secondly, several different information transfer metrics can be computed from a fitted MVAR parameter matrix. The use of PDC in this thesis is validated by demonstrating that it can establish the direction of information transfer within a system (unlike coherence) and disambiguate a direct from an indirect connection (in contrast to coherence and Directed Transfer Function). Finally the issue of frequency resolution is discussed. The relationship between the sampling rate, window length and frequency resolution is outlined before a novel set of simulations introduced. These simulations produce multivariate systems with a known spectral content in the cross

correlations. In other words, the cross correlation between nodes oscillated at a known frequency. In contrast to claims from the literature, this simulation demonstrated that the frequency resolution of a connectivity metric estimated from an MVAR model is completely determined by the time delay between coefficients and the total number of predictors.

9.1.3 Directionality In The Resting State

The first application of the analysis pipeline outlined in chapters 2 and 3 sought to identify directionality within the Default Mode Network. This is a very well known endogenous brain network which has been robustly identified across many modalities (fMRI and MEG) and methods (ICA, Seed based correlation) however the causal structure between its component nodes is not well known. As this was the first application of the analysis pipeline a thorough search through the time delay embedding parameter space was conducted to identify which combination of model order and delay produced a model most able to predict future observations consistently and without over fitting. This identified a delay of around 5ms and an order of 20. The final model order was determined with the additional constraint that it should provide a reasonable frequency resolution.

This approach identified many asymmetries in information flow within the DMN. Critically, the ventral and dorsal medial prefrontal cortex,posterior cingulate and angular gyrus were identified as predominantly driving nodes. In contrast the middle frontal gyrus and posterior middle temporal gyrus were predominantly driven. Moreover the connection between the anterior and posterior driving nodes was show to vary as a function of frequency, with influence from vMPFC to PCC decreasing with increasing frequency.

9.1.4 Dynamics In RSN Directionality

The directional asymmetries within the DMN were significant across the entire recording time, but recent results suggest that the strength of connections within endogenous brain networks might fluctuate over time. This temporal variance was characterised using a hidden markov model. This

produces a number of hidden states which are associated with an particular distribution of observations and provides a probabilistic description of the transitions between these states over time. This chapter first reduced the dimensionality of the PDC estimates from the 2 second windows generated in the previous chapter with a PCA before the HMM was fitted to the weights from each component. Four states were generated as this is the smallest number of states which produce HMMS with the lowest expected Free Energy (an approximation of the model evidence). Finally the predictive power of the state transitions on specific connections within the DMN was characterised by regressing the original PDC estimates against the state time course. This identified which connections at each frequency increase or decrease in strength with the presence of each HMM state. Many dynamic changes were identified, critically the influence of the posterior driving nodes across the network in the alpha band increased in one state and decreased in another while remaining constant in the other. The anterior driving nodes were not modulated by state transitions at this frequency.

9.1.5 Visual Word Recognition

Dynamics in directionality within brain networks occur on multiple timescales. The previous chapter demonstrated variance in directed functional connectivity on a time-scale of seconds. While this is informative over the course of many minutes, this rate of change is an order of magnitude slower than the brains response to a stimulus such as a word. This chapter modified the MVAR parameter matrix fitting to parameterise temporal dependencies which are consistent within very short time windows over many experimental trials rather than within one large time window. This is applied to a visual word recognition task in which False Fonts, Consonant Strings and Words are presented to a participant.

Several theories describe how the brain might recognise a word or word like stimulus. Feed forward hierarchical models would suggest that increasingly complex features of a visual input are represented by layers in a processing hierarchy. Features such as simple shapes and angles are represented

by early visual cortex and single letters, bigrams and eventually words are represented by more anterior regions along the ventral occipital temporal cortex. In contrast, an interactive account would say that complex features such as a word are not explicitly represented by visual cortex. Rather, the visual input is compared to a top-down prediction of which stimulus may have caused the sensory input. Through interaction between the visual representation and top-down prediction the word is identified. Critically this does not occur without interactions with "higher" areas.

It was found that the false font and consonant string stimuli produced a larger top-down PDC from LIFG to visual brain regions within 150ms of stimulus onset. In contrast the words produced more early feed forward connections from the vOT and early influences from the temporal pole. These connections show different magnitudes in different conditions in very short time scales after stimulus onset. Moreover the magnitude of the differences between conditions changes rapidly over the course of the first 300ms after stimulus onset.

These results generally support a mixture of the two accounts outlined above. Firstly, a word may most simply be recognised through a feed forward process potentially similar to the hierarchical models in the literature. However these models may not account for cases in which the stimulus is challenging or impossible to recognise. False Font and Consonant Strings are impossible to recognise but word like-stimuli. These stimuli support a more interactive account of recognition, in which visual cortex receives influence from IFG during the early processing of the stimulus. This top-down influence is interpreted as a source of extra information to aid recognition when the visual cortex is unable to identify the stimulus itself.

9.1.6 Multiple Routes To Recognition

The previous chapter suggests that the reading network features more top-down influence on visual cortex when the stimulus is impossible to recognise as a word. However the interpretation that this top-down influence is an aid to recognition may be spurious if the top-down interactions were unre-

lated to word reading and only occurred with non-word stimuli. Chapter 7 investigates this confound by comparing the PDC within a reading network as a participant reads either high or low frequency words. Both of these categories are real words, however the high frequency words are very easily recognised by a skilled reader. In contrast the low frequency words present more of a challenge to the reading system.

It was found that the low frequency words did increase the top-down influence from LIFG to both visual regions in the the reading network within 100ms of stimulus onset. In contrast the high frequency words elicited stronger feed forward connections from visual cortex.

9.2 Concept Summaries

Several of the major themes within this thesis are discussed with reference to the relevant chapters below.

9.2.1 Dynamics of Dynamics

Time delay embedding is used to reconstruct the evolution of a systems state over time (Abarbanel et al., 1994). A TDE is a vector containing observed data at a given time and its delays. This representation of the local dynamics of a brain network may have a very fine temporal resolution and is likely to be very high dimensional. Though complex, this representation is a some unknown transform of the internal state dynamics of the hidden system which is generating the observations. Due to the fine temporal resolution and inclusion of delayed observations in the TDE vector, estimates of Granger causality and temporal precedence may naturally be estimated.

The first state in this causality estimation is a parameterisation of the relationship between an observation at time t and its past. This is performed by fitting a multivariate autoregressive model to the TDE. This fits a parameter to each dimension in the TDE vector, however to ensure that the parameters represent an ongoing dynamic relationship between an observation an its time delay, the parameters are fitted across many TDE vectors

within a wider time window. In chapters 4 and 5, this window was 2 seconds long and the TDE based on a delay of 5ms and an order of 20. At the 4 kilohertz sampling rate this resulted in around 7700 TDE vectors each containing 720 dimensions. The MVAR parameter matrix contained 720 parameters which describe the extent to which each dimension in the TDE is able to predict future observations across all 7700 vectors within a window. Critically, though the time-window is two seconds long in this case, the MVAR model does describe the very rapid dynamics of the system on a millisecond time-scale. However it will only model the rapid dynamics which were consistent across the two second period.

Chapters 6 and 7 used a different approach designed to allow good MVAR fits across much shorter time windows. Instead of one continuous time window from which to extract the TDE vectors, this approach uses many realisations of much shorter time windows. This as the MVAR parameters may be directly estimated from a lagged covariance matrix R, the covariance between signals at different time delays (delays $= \tau, 2 * \tau. \ldots, p * \tau$). Chapters 4 and 5 estimate this across all TDE vectors within a window, however in chapters 6 and 7 the time-windows being considered are too short to ensure a good estimate of R. Ding et al. (2000) show that a good estimate of R can be obtained from a very short time window if many comparable realisations of the same dynamics are available. This short-window approach uses the average R estimated from within many short time windows, picking up on any delayed covariances which are consistent across all realisations. In this way the time-window can be shorted from 2 seconds to tens or hundreds of milliseconds. This approach was not applied in chapters 4 and 5 as we cannot assume that consecutive time windows in a resting state scan could be considered different realisations of the same dynamics. In contrast, the controlled stimulus presentation used in chapters 6 and 7 should evoked a similar dynamic process associated with word recognition and as such the average covariance over many realisations should pick up on consistencies in the responses across experimental epochs.

After the estimation of the lagged covariance matrix by either method, the MVAR model fit parameterises the extent to which the covariance struc-

ture is predictive of future observations. As mentioned above, the fitted MVAR parameters may be further transformed to provide estimates of Granger causality in the frequency domain. The Partial Directed Coherence expresses the amount of influence the past history of one node has on future observations in the rest of the network. This is performed in the frequency domain after a Fourier transform of the MVAR coefficients for each pair of nodes across all delays. This representation provides a further simplification of the rapid dynamics originally reconstructed by the TDE vectors. The rapidly changing and very high dimensional TDE representation is transformed and simplified until the PDC expresses the extent to which specific delayed interactions are predictive of future observations in the frequency domain. Critically the PDC is a representation of the causal influences in the rapid dynamics which are consistent over a time-window (or many windows over the same local dynamics.

Chapter 5 looked to extend this representation of local dynamics within a single window to describe how these local and rapid dynamics might change over longer time-scales. A hidden Markov model was used to characterise how the PDC description of brain dynamics change over the course of minutes. The dimensionality of the PDC was reduced using PCA before the HMM used to provide a probabilistic description of the change in PDC through a discrete space comprising several PDC states. It is important to note that the HMM states and state transitions are entirely different from the TDE vector state-space. The HMM space is discrete over a predetermined set of states each associated with a PDC description of the data. In contrast, the TDE state dynamics are continuous and constructed from the observed data itself.

The HMM approach was not applied to the dynamics of the task data for several practical reasons. Firstly, the resting state analysis allowed for several thousand epochs to enter the HMM estimation with a ratio of around 125:1 observations to dimensions. In contrast the task data would only provide several hundred epochs with a much smaller ratio between observations and dimensions in the data. Secondly, the overlapping windows in the task chapters make the concatenations seen in chapter 5 problematic as succes-

sive windows may be interdependent. Despite these practical challenges, this approach remains a highly promising approach for future research. Identification of functional connectivity states on a millisecond timescale during the brain's response to a range of stimuli would provide a powerful approach for characterising the similarities and differences in the processing of different visual word forms.

9.2.2 Task Dynamics

The previous section outlined the theory behind modelling the rapid task dynamics seen in chapters 6 and 7 in relation the modelling seen in previous chapters. Across both chapters 6 and 7 this approach has lead to several insights which are summarised in this section.

A critical advantage of estimating the covariance across many experimental epochs is that we can work with shorter time windows. This allows for a detailed model of the structure of the reading network on a time-scale of tens or hundreds of milliseconds. This method does not require the definition of a set or family of model structures *a priori* nor presumptions about when a task effect might occur. A set of nodes comprising the reading network were required *a priori* as well as the time delay embedding parameters (a deeper discussion into the TDE parameter requirements can be found in the methodological statements).

Both chapters 6 and 7 showed considerable variance in the connection strength between different nodes in the reading network. Critically, the average PDC for the connections from MOG to both vOT and AG were higher than the other connections below 20Hz. In addition the vOT→MOG connection was larger than the MOG→vOT. These biases in the PDC estimates were generally present in the passive period prior to stimulus onset, though modulation over time is seen, it appears within the context of these apparently ongoing directional biases in connection strength.

The task contrasts in both chapters highlighted several interesting outcomes. Firstly, almost all connections within the reading network showed some significant modulation in strength by different stimulus categories at

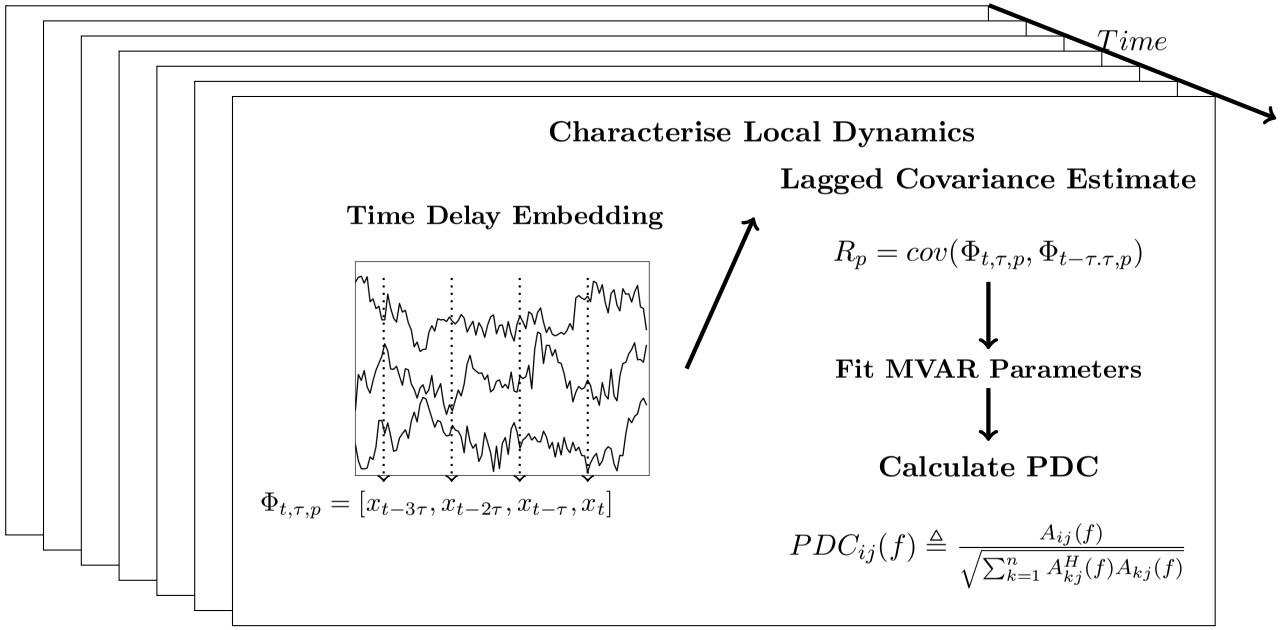

Figure 9.1: A schematic illustrating the relationship between the local dynamics reconstructed by the Time Delay Embedding (contained within each rectangle) and the larger scale change in these local dynamics across multiple sliding windows. The local dynamics within each window in chapters 4 and 5 were reconstructed from a single 2 second window whereas the more rapid dynamics in the visual word recognition tasks in chapters 6 and 7 arose from the average of several hundred experimental epochs within one condition. The change over time in chapter 5 was characterised using a hidden markov model. In contrast the change over time in chapters 6 and 7 came from condition contrasts thresholded across all windows.

some point in time or frequency. This result is most consistent with a parallelised and distributed account of visual word recognition in which identity of a recognised word is encoded in the state of the network as a whole. However, these PDP models doe not include any encoding of the temporal differences in connectivity. For instance, the Words >False Font contrast in chapter 6 shows early modulations in connectivity in the PDC from LIFG within 150ms, in contrast there are more late modulations in the connections from Temporal Pole. Though the widespread modulations do follow predictions made from PDP models, the same models do not incorporate any temporal predictions such as the findings highlighted above. This is perhaps as many sets of temporal constraints might produce models which accurately predict behavioural outcomes. As such this may provide an avenue in which modelling of neural dynamics may provide a set of biological constraints on the temporal evolution of computational models.

Secondly, these results highlight the need for information about the dynamics in functional connectivity on a millisecond time-scale. Many highly influential accounts of the connectivity subserving visual word recognition provide a relatively static picture of the network interactions (Kujala et al., 2007). Moreover, many models of neural interactions during recognition derived directly from experimental data show equally static images. The present results build upon previous demonstrations of rapidly spreading activation across the cortex by showing that this representation of the connectivity is highly simplistic. This dynamic representation of brain connectivity may provide more detailed hypotheses for future experimental work, for instance the example in the paragraph above might suggest a double dissociation between early and late TMS to LIFG and Temporal Pole whilst reading word a False Fonts. This avenues of future work might be critical in attributing cognitive and behavioural outcomes to connections in specific time windows.

Finally, the rapid modulations seen in almost all contrasts across chapters 6 and 7 support the increasingly popular view that highly complex information may be extracted from a visual stimulus very rapidly. Results from EEG and MEG have suggested that high level linguistic information

such as phonology and semantic representations may arise within 100ms of stimulus onset (Wheat et al., 2010; Cornelissen et al., 2009), challenging earlier hierarchical models of processing. This distinction between sequential and parallel decoding of linguistic stimuli has been made previously primarily based on ERP evidence (Pulvermüller et al., 2009). The present results build on this by showing that the connectivity within the reading network also shows rapid modulations in connectivity between brain regions thought to encode/process high level linguistic information.

9.2.3 Dynamics In The Resting State

Previous work on dynamics during the resting state in humans has typically focused on transitions between entire networks defined by either seed based correlation or ICA decomposition (Brookes et al., 2011a; Baker et al., 2014; de Pasquale et al., 2010). The present work presents a different approach in which the connectivity within an endogenous brain network is investigated. This distinction is important as the present approach allows more easily for investigations into directional influences between brain regions, particularly if the analysis can be constrained to a relatively small number of brain regions which can be expected to interact *a priori*. There has been an immense amount of work in this area using fMRI to identify brain regions which are functionally related during the "resting state" (Yeo et al., 2011; Raichle et al., 2001; Buckner et al., 2008). Chapter 4 used the large analysis in Yeo et al. (2011) to identify 5 brain regions which comprised the default mode network and several related brain regions. An additional anterior prefrontal hub was included from the literature to ensure that a ventral MPFC region was included in the analysis as this area is considered to be critical in the core of the DMN (Andrews-Hanna et al., 2010).

The present work builds upon this existing literature not by questioning whether the regions within this network are connected but whether there are asymmetries in their connectivity. That is, whether one region may exert more influence on a second than vice versa. The results showed that many of the connections within the DMN and associated areas were strong asym-

metrical. Intriguingly, these directional asymmetries were often consistent across frequency. One of the only two connections which showed differences across frequency were between the two areas within the "midline core" of the DMN (Andrews-Hanna et al., 2010) which showed that the PCC→vMPFC was most consistently directional across participants at frequencies higher than 20Hz. In addition, the vMPFC→PCC connection was dominant for frequencies below 20Hz.

Though the functional relevance of these directional connections is not known at this point, the present result raises intriguing directions for future research. These avenues take advantage of individual differences in the strength of connections and the extent to which connections are directional within the DMN. Previous work with fMRI has shown that individual differences in coupling strengths between nodes in resting state networks may be correlated with explicit behavioural outcomes. Specifically the extent to which the two regions in the mid-line core of the DMN were connected was predictive of comprehension in a reading task (Smallwood et al., 2013). This link between DMN connectivity and cognitive performance arises from differences between individual participants. This approach could readily be applied to the data from chapter 4, but perhaps the most intriguing possibility for future research arises from the Hidden Markov Model approach used in chapter 5. Critically chapter 5 shows considerable individual differences in the fractional occupancy and the transition matrix for across time both within and between participants. This highlights the possibility that behavioural outcomes such as comprehension in reading tasks might be related to the occupancy of a given state prior to stimulus onset. This may provide a promising avenue for future work in this area, explicitly relating the function states identified in chapter 5 to specific cognitive and behavioural outcomes.

9.2.4 Multiple Routes to recognition

Several sections above have outlined the theoretical justification of the dynamical modelling used in this thesis and general outcomes from its appli-

cation to task related neural responses measures from MEG. The specific theoretical contributions of this work in relation to visual word recognition is outlined in this section.

Two questions are tackled in chapters 6 and 7, firstly chapter 6 asks whether interactions between visual and higher linguistic brain regions are an integral part of the recognition process or are epiphenomenal. The relationship between visual nodes and the LIFG seen in this chapter suggests that there exists very rapid connectivity between these regions in all conditions. The direction of the connectivity seems to be related to the presence of orthographic information in the stimulus. The visual areas exert a driving influence when clear Latin characters are presented, while the False Fonts lead to a larger top-down influence. These results implicate the LIFG.

Chapter 7 asks whether the top-down influence from the LIFG is related to a linguistic process by comparing high and low frequency words. Critically both these categories The neural substrate which may underlie the rapid recognition of high frequency words can be seen in these results. Critically, high frequency word induce stronger feed forward connections from visual areas to IFG within 100ms. At the same time, the influence from the TP on STG and AG is stronger for high frequency words. This earlier engagement of the temporal pole might indicate an earlier access to semantic content which can begin to feedback into the system within 100ms. In contrast the low frequency words show little indication of large modulations of the connectivity from the temporal pole. Rather driving influences from AG,STG and IFG are more prominent.

A Bayesian approach may provide the best interpretation as it allows for a fully probabilistic account. As visual information feeds into the brain it combines with *a priori* expectations (perhaps such as word frequency) to produce an *a posteriori* probability distribution across the set of possible stimuli which could have produced the sensory input. The rapid access to phonology discussed above suggests that this first process may happen simultaneously on several levels, resulting in a probability distribution based on a very fast visual and phonological analysis of the sensory input. If the *a posteriori* evidence in the visual system is sufficient, that suggests the brain has

recognised the word and the phonological evidence not further considered in the context of recognition. In contrast, if the *a posteriori* visual probabilities are ambiguous (as would be the case for infrequent words or non-words) in that several or many words are similarly likely to have produced the incoming bottom-up sensory information, then the rapid phonological decoding may provide further top-down evidence. The combination of these two may then produce a match, or be sufficient to conclude that the stimulus was not a word.

It would appear that the entire network responds to a word-like visual stimulus as would be predicted by a distributed connectionist account. However the psycholinguistic and orthographic content of the word does induce differences in specific parts of the reading network as might be predicted by a more modular approach. The answer is likely to lie in between these two accounts in which a widely distributed and highly interactive network of specialised brain regions co-operate to identify an incoming stimulus. The connectivity within the reading network does change rapidly over time, further suggesting that there may be several stages of processing, though each stage may be associated with the coordinated activity of a number of nodes in the network. Moreover, the same node may not play the same role in different stages of processing with different network contexts (Price and Friston, 2005). It may be that the functional "state" of whole network is better able to describe how psycholinguistic aspects of a word are processed than a modular or route based account.

9.3 Methodological Statements

9.3.1 Delay Embedding Parameters

The MVAR methodology used in this thesis has often been criticised due to difficulties in selecting an appropriate model order. Several studies using MVAR methods have reported that Akaike's Information Criterion monotonically decreases with increasing model order, making the choice of p seemingly arbitrary (Schlögl and Supp, 2006; Supp et al., 2007). Moreover, there

are large inconsistencies in the model orders used in the literature. These inconsistencies are compounded by the fact that many studies using MVAR methodologies do not explicitly parameterise delay in the model. The table in appendix A of chapter 4 summarises the variance in model order and delay selection in a selection of the literature.

Though there is evidence that an incorrect selection of model order can distort estimates of functional connectivity (Porcaro et al., 2013) others suggest that the choice of model order may make little difference to the performance of the PDC estimate if the ratio of parameters to observations remains high (Florin et al., 2011; Supp et al., 2007). The results of the parameter space search performed in chapter 4 can provide some context to these results. The change in AIC as a function of model order is different at different delays, critically there is a middle range of delay values in which the AIC does decrease almost monotonically as previously reported in the literature. In contrast for very short or very long values the AIC does plateau or reach a local minima, suggesting that an overestimate of the order might result in overfitting the model. As such, both the expected minima in the AIC and reported monotonic decreases can exist within the same dataset at different delays.

In the context of the time delay embedding the model order parameter denotes the dimensionality of the embedding space. The embedding theorem of Takens states that the embedding space must be of at least dimension $2d + 1$ in which d is the dimensionality of the hidden state space. In the context of MEG data, the underlying dimensionality of the brain is unknown, as is the dimensionality of the observed signal. Methods such as false nearest neighbours are able to estimate the dimensionality of a system however the estimation process is extremely computationally intensive particularly for large datasets of richly sampled observations. As a consequence the model order used in this thesis was selected from two key criterion. Firstly, the upper bound for the range of possible model orders is set by a local minima in the AIC estimates seen in chapter 4. Secondly, the largest model order within this bound that still allowed for a good ratio between the number of parameters to be estimated and observed data points. This allows both

for a good frequency resolution and by taking as large a model order as possible we increase the likelihood that our embedding space is sufficient to reconstruct the dimensionality of the hidden state space.

9.3.2 Participants

Each of the experimental chapters have been restricted by a practical trade off between collecting sufficient data from a single participant to allow for complex models to be accurately fitted and obtaining enough complete datasets to allow for robust group statistics. As this work represents a novel application of TDE-MVAR-PDC methodology to the resting state and reading networks, it was decided that the first criterion was most important. As such, each chapter contains several hour of MEG recordings from each of a handful of participants. The results individual analysis are therefore very detailed and robust, however this is at the expense of a second level group analysis. Future work along these lines building to a larger group dataset would allow for stronger generalisations of the results presented in this thesis. However, as highlighted in chapter 3, this should not be completed at the expense of a ratio of observations to model parameters of at least 5:1.

9.4 Conclusions

Overall, this thesis presents a rigourus application of Time Delay Embedding and Multivariate Autoregressive modelling to identify directed functional connectivity during the resting state and visual word recognition. This modelling approach was firstly placed in a detailed theoretical context before more practical issues such as the effect of sample size and frequency resolution were probed with simulations. Next several experimental chapters apply this method to build on contemporary work in two areas of cognitive neuroscience. Firstly, whether resting state networks exhibit directional connections and how these might change over time and secondly, how brain regions within the reading network communicate during the process of recognising a visual word form.

The methodology is applied to a resting state network featuring a core of the Default Mode and several associated areas revealing many asymmetrical connections. A Hidden Markov Model then characterised the dynamics changes in this directed functional connectivity and identified four functional states across participants. Critically, each state is predictive of modulations in difference connections within the network. Task dynamics are then probed using a visual word recognition paradigm comparing network connectivity in the brain in response to Word, Consonant String and False Font stimuli. Many rapid and directional connections were modulated by each stimulus category. Intriguingly a top-down connections from the Inferior Frontal Gyrus was found to be stronger for False Font stimuli than Words, raising the possibility that this connection is associated with processing a stimulus which is challenging, if not impossible to recognise. The final chapter demonstrates that this increase in top-down influence from the IFG exists for challenging real-word stimuli.

Appendix A

Vieira-Morf Algorithm

The Vieira-Morf algorithm is an extension of the multichannel Levinson-Durbin-Whittle recursion for estimating multivariate autoregressive coefficients from a set of time series. The VM algorithm uses estimates of the partial correlation matrix based on the residual variance and covariance matrices.

Initialisation The covariance of the raw data is estimated and the error variance is initialised as the raw time series.

$$\mathbf{P}_0^f = \mathbf{P}_0^b \quad = \quad \frac{1}{N}\sum_{n=1}^{N} x[n]x^H[n] \tag{A.1}$$

The error variance is initialised as the raw time series for $1 \leq n \leq N$.

$$\mathbf{e}_0^f[n] = \mathbf{e}_0^b[n] \quad = \quad x[n] \tag{A.2}$$

The algorithm iterates over all model orders from 1 to the predefined maximum M. The current model order for a given iteration is denoted by p.

Main Loop

- In the main loop, estimates of the covariance at the next lag are ob-

tained from the error variance left from the previous step. These are only estimates, the real values come from the error variance remaining after the new parameters are estimated.

$$\hat{\mathbf{P}}_p^f = \frac{1}{N}\sum_{n=p+2}^{N} \mathbf{e}_p^f[n]\mathbf{e}_p^{fH}[n] \tag{A.3}$$

$$\hat{\mathbf{P}}_p^b = \frac{1}{N}\sum_{n=p+2}^{N} \mathbf{e}_p^b[n-1]\mathbf{e}_p^{bH}[n-1] \tag{A.4}$$

$$\hat{\mathbf{P}}_p^{fb} = \frac{1}{N}\sum_{n=p+2}^{N} \mathbf{e}_p^f[n]\mathbf{e}_p^{bH}[n-1] \tag{A.5}$$

- the current model order is incremented

$$p = p + 1 \tag{A.6}$$

- An estimate of the partial correlation matrix is generated from these estimated covariances.

$$\hat{\Lambda}_{p+1} = (\hat{\mathbf{P}}_p^{f1/2})^{-1}(\hat{\mathbf{P}}_p^{fb})(\hat{\mathbf{P}}_p^{b1/2})^{-H} \tag{A.7}$$

- This estimate of the partial correlation matrix is then used with the actual error covariances from the previous lag to compute the forward and backward reflection co-efficients.

Appendix B

Multivariate Likelihood

In most brain connectivity estimation questions we will want to assess the fit of a multivariate rather than univariate model. The univariate likelihood above can be extended to a multivariate case by using a multivariate normal distribution. The non-degenerate case the multivariate normal distribution is

$$\mathbf{y}_i = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} e^{(-\frac{1}{2}(\mathbf{y}_i - \mathbf{x}_i\mathbf{b})'\Sigma^{-1}(\mathbf{y}_i - \mathbf{x}_i\mathbf{b}))} \tag{B.1}$$

B.1 Mahalanobis Distance

The exponential term in equation B.1 is intimately related to the Mahalanobis distance. This is a measure of the distance between two data points, in this case the observation and the model's prediction. If we take:

$$\hat{\mathbf{y}}_\mathbf{i} = (\mathbf{y}_i - \mathbf{x}_i\mathbf{b}) \tag{B.2}$$

The Mahalanobis distance is then:

$$D = \sqrt{\hat{\mathbf{y}}'\Sigma^{-1}\hat{\mathbf{y}}} \tag{B.3}$$

This provides a measure of how far the observed data point is from the models prediction, normalised by the covariance matrix of the residuals. The

likelihood for the whole multivariate sample is the product of the likelihood of each individual sample

$$L(\mathbf{y}|\mathbf{x},\mathbf{b},\Sigma) = \prod_{i=1}^{N} \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} e^{(-\frac{1}{2}Tr(\Sigma^{-1}\hat{\mathbf{y}}'\hat{\mathbf{y}})} \tag{B.4}$$

the log likelihood is therefore

$$\begin{aligned}
\ln L &= \sum_{i=1}^{N} \ln\left(\frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} e^{(-\frac{1}{2}Tr(\Sigma^{-1}\hat{\mathbf{y}}_\mathbf{i}'\hat{\mathbf{y}}_\mathbf{i})}\right) \\
&= \sum_{i=1}^{N} \left(\ln(1) - \ln((2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}) - \frac{1}{2}Tr(\Sigma^{-1}\hat{\mathbf{y}}_\mathbf{i}'\hat{\mathbf{y}}_\mathbf{i})\right) \\
&= N\ln(1) - N\ln((2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}) - \sum_{i=1}^{N}(\frac{1}{2}Tr(\Sigma^{-1}\hat{\mathbf{y}}_\mathbf{i}'\hat{\mathbf{y}}_\mathbf{i})) \\
&= N\ln(1) - N(\ln((2\pi)^{\frac{N}{2}} + \ln(|\Sigma|^{\frac{1}{2}})) - \sum_{i=1}^{N}(\frac{1}{2}Tr(\Sigma^{-1}\hat{\mathbf{y}}_\mathbf{i}'\hat{\mathbf{y}}_\mathbf{i})) \\
&= -N\ln((2\pi)^{\frac{N}{2}}) - N\ln(|\Sigma|^{\frac{1}{2}}) - \sum_{i=1}^{N}(\frac{1}{2}Tr(\Sigma^{-1}\hat{\mathbf{y}}_\mathbf{i}'\hat{\mathbf{y}}_\mathbf{i})) \\
&= -\frac{N^2}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \sum_{i=1}^{N}(\frac{1}{2}Tr(\Sigma^{-1}\hat{\mathbf{y}}_\mathbf{i}'\hat{\mathbf{y}}_\mathbf{i})) \qquad \text{(B.5)}
\end{aligned}$$

Appendix C

Analytic Confidence Limits

In contrast to the permutation method, several analytical methods are available to compute confidence intervals based on the experimental estimate alone. One such technique developed by Schelter et al. (2006) first estimates the covariance matrix from the data observations

$$\mathbf{Q}(k,l) = \begin{pmatrix} Q_{11}(kl) & \dots & Q_{1m}(kl) \\ \vdots & \ddots & \vdots \\ Q_{m1}(kl) & \dots & Q_{mm}(kl) \end{pmatrix} \tag{C.1}$$

where m is the number of channels in the system. The inverse of this matrix is then a critical term in the calculation of C

$$C_{ij}(f) = \Sigma_{ii}\Bigg[\sum_{k,l=1}^{p} \mathbf{Q^{-1}}_{jj}(k,l)\Big(\cos(kf)\cos(lf) + \sin(kf)\sin(lf)\Big)\Bigg] \tag{C.2}$$

Which can be used to approximate the significance level with

$$CL_{i,j} = \left(\frac{\hat{C}_{ij}(f)\chi^2_{1,1-alpha}}{N\sum_{k=1}^{n}\bar{A}^H_{kj}(f)\bar{A}_{kj}(f)}\right) \tag{C.3}$$

In which $\chi^2_{1,1-\alpha}$ is the $1 - alpha$ percentile of the χ^2-distribution with 1 degree of freedom and N the total number of observations in the model. Note that the denominator of the confidence limit is a scaled version of the denominator in the PDC calculation seen in equation 3.36.

This approach is far more computationally efficient than the permutation method and may be computed alongside the experimental PDC estimation due to shared terms in the denominator. As such, this method will be used to establish confidence limits from PDC estimates in this thesis.

Appendix D

Previous MVAR Literature

Paper	Method	Model Order	Sample Rate	Resample Rate
Ding et al. (2000)	TBE	5	200Hz	-
Brovelli et al. (2004)	LFP	10	200Hz	-
Kus et al. (2006)	EEG	5	250Hz	125Hz
Supp et al. (2007)	EEG	15	512Hz	-
Gómez-Herrero et al. (2008)	Simulated EEG	5	?	-
	EEG	?	256Hz	-
Astolfi et al. (2008)	Simulated EEG	3	256Hz	-
	EEG	16/17	?	-
Gow et al. (2008)	M/EEG	10	601Hz	1000Hz
Korzeniewska et al. (2008)	ECoG	10	1000Hz	333.33Hz
Gow and Segawa (2009)	M/EEG	6	601Hz	1000Hz
Gow et al. (2009)	iEEG	2	500Hz	-
Sun et al. (2009b)	EEG	?	1000Hz	-
Boatman-Reich et al. (2010)	ECoG	?	1000Hz	-
Astolfi et al. (2010)	EEG	16/17	200Hz	-
Korzeniewska et al. (2011)	ECoG	?	1000Hz	333.33Hz
Porcaro et al. (2013)	EEG	?	4069Hz	-
(Fasoula et al., 2013)	MEG	5/6	1250Hz	62.5Hz

Table D.1: Review of papers using MVAR models to describe electrophysiological data, the papers highlighted in bold contribute points to the past literature seen in figure 5.3. The modality and task employed in each paper are stated - note several papers employed both simulated and real EEG data. Finally the model parameters used in each case are stated, '-' means the parameter was not applicable to that study and '?' indicates that the parameter was not reported. Abbreviations are as follows: *TBE*- Transcortical bipolar electrode, *LFP* - Local field potential, *EEG* - Electroencephalogram, *MEG* - Magnetoencephalography, *M/EEG* - MEG and EEG, *ECoG* - Electro corticography

Bibliography

Abarbanel, H. and Parlitz, U. (2006). Nonlinear analysis of time series data. In Schelter, B., Winterhalder, M., and Timmer, J., editors, *Handbook of Times Series Analysis: Recent Theoretical Developments and Applications*. Wiley-VCH, Weinheim.

Abarbanel, H. D., Carroll, T., Pecora, L., Sidorowich, J., and Tsimring, L. S. (1994). Predicting physical variables in time-delay embedding. *Physical review E*, 49(3):1840.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.

Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, 27(3/4):217.

Albo, Z., Di Prisco, G., Chen, Y., Rangarajan, G., Truccolo, W., Feng, J., Vertes, R., and Ding, M. (2004). Is partial coherence a viable technique for identifying generators of neural oscillations? *Biological Cybernetics*, 90(5):318–326.

Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H., and Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *Journal Of Comparative Neurology*, 412(2):319–341.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., and Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4):550–562.

Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology*, 61(3):303–321.

Anwander, a., Tittgemeyer, M., von Cramon, D. Y., Friederici, a. D., and Knösche, T. R. (2007). Connectivity-Based Parcellation of Broca's Area. *Cerebral Cortex*, 17(4):816–25.

Astolfi, L., Cincotti, F., Mattia, D., De Vico Fallani, F., Tocci, A., Colosimo, A., Salinari, S., Marciani, M. G., Hesse, W., Witte, H., Ursino, M., Zavaglia, M., and Babiloni, F. (2008). Tracking the time-varying cortical connectivity patterns by adaptive multivariate estimators. *IEEE Transactions on Bio-medical Engineering*, 55(3):902–13.

Astolfi, L., Cincotti, F., Mattia, D., De Vico Fallani, F., Vecchiato, G., Salinari, S., Vecchiato, G., Witte, H., and Babiloni, F. (2010). Time-Varying Cortical Connectivity Estimation from Noninvasive, High-Resolution EEG Recordings. *Journal of Psychophysiology*, 24(2):83–90.

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., de Vico Fallani, F., Salinari, S., Ursino, M., Zavaglia, M., Ding, L., and et al. (2007). Comparison of different cortical connectivity estimators for high-resolution EEG recordings. *Human Brain Mapping*, 28(2):143–157.

Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Baccala, L. A., Fallani, F. D. V., Salinari, S., Ursino, M., Zavaglia, M., and Babiloni, F. (2006). Assessing cortical functional connectivity by partial directed coherence: simulations and application to real data. *Biomedical Engineering, IEEE Transactions on*, 53(9):1802–1812.

Baccala, L. and Sameshima, K. (1998). Studying the interaction between brain structures via directed coherence and Granger causality. *Applied Signal ....*

Baccalá, L. a. and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–74.

Baccala, L. A. and Sameshima, K. (2006). Comments on 'Is partial coherence a viable technique for identifying generators of neural oscillations?' - Why the term 'Gersch Causality' is inappropriate: Common neural structure inference pitfalls. *Biological Cybernetics*, 95(2):135–141.

Baillet, S., Mosher, J., and Leahy, R. (2001). Electromagnetic brain mapping. *Signal Processing Magazine, IEEE*, 18(6):14–30.

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Probert Smith, P. J., and Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. *eLife*, 3(0):e01867–e01867.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M., Marinkovic, K., Schacter, D., Rosen, B., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2):449–454.

Barnett, L. and Seth, A. K. (2011). Behaviour of Granger causality under filtering: theoretical invariance and practical application. *Journal of Neuroscience Methods*, 201(2):404–19.

Ben-Shachar, M., Dougherty, R. F., and Wandell, B. a. (2007). White matter pathways in reading. *Current Opinion in Neurobiology*, 17(2):258–270.

Berker, E. A. (1986). Translation of broca's 1865 report. *Arch Neurol*, 43(10):1065.

Besner, D. (1987). Phonology, Lexical Access In Reading, And Articulatory Suppression - A Critical-Review. *Quarterly Journal Of Experimental Psychology Section A-Human Experimental Psychology*, 39(3):467–478.

Binder, J. R., Desai, R. H., Graves, W. W., and Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12):2767–96.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., and Prieto, T. (1997). Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. 17(1):353–362.

Biswal, B., Yetkin, F., Haughton, V., and Hyde, J. (1995). Functional Connectivity In The Motor Cortex Of Resting Human Brain Using Echo-Planar MRI. *Magnetice Resonance In Medicine*, 34(4):537–541.

Bitan, T., Booth, J. R., Choy, J., Burman, D. D., Gitelman, D. R., and Mesulam, M.-M. (2005). Shifts of effective connectivity within a language network during rhyming and spelling. *The Journal of Neuroscience*, 25(22):5397–403.

Blinowska, K., Kuś, R., and Kamiński, M. (2004). Granger causality and information flow in multivariate processes. *Physical Review E*, 70(5):1–4.

Blinowska, K. J. (2011). Review of the methods of determination of directed connectivity from multichannel data. *Med. Biol. Engineering and Computing*, 49(5):521–529.

Boatman-Reich, D., Franaszczuk, P. J., Korzeniewska, A., Caffo, B., Ritzl, E. K., Colwell, S., and Crone, N. E. (2010). Quantifying auditory event-related responses in multichannel human intracranial recordings. *Frontiers In Computational Neuroscience*, 4.

Bolger, D. and Perfetti, C. (2005). Cross cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping*.

Brais, B. (1993). Jean Martin Charcot and aphasia: Treading the line between experimental physiology and pathological anatomy. *Brain and language*.

Britz, J., HernÃndez, L., Ro, T., and Michel, C. M. (2014). Eeg-microstate dependent emergence of perceptual awareness. *Front. Behav. Neurosci.*, 8.

Britz, J., Van De Ville, D., and Michel, C. M. (2010). Bold correlates of eeg topography reveal rapid resting-state network dynamics. *NeuroImage*, 52(4):1162–1170.

Brookes, M. J., Hale, J. R., Zumer, J. M., Stevenson, C. M., Francis, S. T., Barnes, G. R., Owen, J. P., Morris, P. G., and Nagarajan, S. S. (2011a). Measuring functional connectivity using MEG: methodology and comparison with fcMRI. *NeuroImage*, 56(3):1082–104.

Brookes, M. J., Stevenson, C. M., Barnes, G. R., Hillebrand, A., Simpson, M. I., Francis, S. T., and Morris, P. G. (2007). Beamformer reconstruction of correlated sources using a modified source model. *NeuroImage*, 34(4):1454–1465.

Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., Barnes, G. R., Smith, S. M., and Morris, P. G. (2011b). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proceedings of the National Academy of Sciences of the United States of America*.

Broomhead, D. and King, G. P. (1986). Extracting qualitative dynamics from experimental data. *Physica D: Nonlinear Phenomena*, 20(2):217–236.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., and Bressler, S. L. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26):9849–9854.

Bub, D., Arguin, M., and Lecours, A. (1993). Jules Dejerine and his interpretation of pure alexia. *Brain and language*.

Büchel, C., Price, C., and Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature*, 394(July):14–17.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124(1):1–38.

Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., Andrews-Hanna, J. R., Sperling, R. A., and Johnson, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to alzheimer's disease. *Journal of Neuroscience*, 29(6):1860–1873.

Bullmore, E. T. and Sporns, O. (2012). The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349.

Burnham, K. P. and Anderson, D. R. (2010). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer.

Cabral, J., Kringelbach, M. L., and Deco, G. (2014a). Exploring the network dynamics underlying brain activity during rest. *Progress in Neurobiology*, 114:102–131.

Cabral, J., Luckhoo, H., Woolrich, M., Joensson, M., Mohseni, H., Baker, A., Kringelbach, M. L., and Deco, G. (2014b). Exploring mechanisms of spontaneous functional connectivity in meg: How delayed network interactions lead to structured amplitude envelopes of band-pass filtered oscillations. *NeuroImage*, 90:423–435.

Carreiras, M., Armstrong, B. C., Perea, M., and Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences*, 18(2):90–98.

Carreiras, M., Seghier, M. L., Baquero, S., Estévez, A., Lozano, A., Devlin, J. T., and Price, C. J. (2009). An anatomical signature for literacy. *Nature*, 461(7266):983–6.

Catani, M. and Ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain*, 128(Pt 10):2224–39.

Catani, M., Jones, D. K., and Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of neurology*, 57(1):8–16.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, 11(41):63–65.

Cellucci, C., Albano, A., and Rapp, P. (2003). Comparative study of embedding methods. *Physical Review E*, 67(6, 2).

Chang, C. and Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fmri. *NeuroImage*, 50(1):81 – 98.

Cohen, D. (1972). Magnetoencephalography: Detection of the brains electrical activity with a superconducting magnetometer. *Science*, 175(4022):664–666.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. a., and Michel, F. (2000). The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123 (Pt 2:291–307.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., and Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125(Pt 5):1054–69.

Coltheart, M. (1981). The mrc psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4):497–505.

Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review*, 100(4):589.

Coltheart, M., Rastle, K., and Perry, C. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological ....*

Cornelissen, P., Tarkiainen, A., Helenius, P., and Salmelin, R. (2003). Cortical effects of shifting letter position in letter strings of varying length. *Journal of Cognitive Neuroscience*, 15(5):731–46.

Cornelissen, P. L., Kringelbach, M. L., Ellis, A. W., Whitney, C., Holliday, I. E., and Hansen, P. C. (2009). Activation of the left inferior frontal gyrus in the first 200 ms of reading: evidence from magnetoencephalography (MEG). *PLoS One*, 4(4):e5359.

Daunizeau, J., David, O., and Stephan, K. E. (2011). Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *NeuroImage*, 58(2):312–22.

Davelaar, E., Coltheart, M., Besner, D., and Jonasson, J. T. (1978). Phonological recoding and lexical access. *Memory & Cognition*, 6(4):391–402.

David, O., Cosmelli, D., and Friston, K. J. (2004). Evaluation of different measures of functional connectivity using a neural mass model. *NeuroImage*, 21(2):659–73.

de Pasquale, F., Della Penna, S., Snyder, A. Z., Lewis, C., Mantini, D., Marzetti, L., Belardinelli, P., Ciancetta, L., Pizzella, V., Romani, G. L., and Corbetta, M. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):6040–5.

de Pasquale, F., DellaPenna, S., Snyder, A., Marzetti, L., Pizzella, V., Romani, G., and Corbetta, M. (2012). A cortical core for dynamic integration of functional networks in the resting human brain. *Neuron*, 74(4):753–764.

Deco, G., Jirsa, V. K., and McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, 12(1):43–56.

Deco, G., Jirsa, V. K., and McIntosh, A. R. (2013). Resting brains never rest: computational insights into potential cognitive architectures. *Trends in Neurosciences*, 36(5):268–274.

Dehaene, S. and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2):384–98.

Dehaene, S. and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6):254–62.

Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in Cognitive Sciences*, 9(7):335–41.

Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J., Bihan, D. L., and Cohen, L. (2004). Letter binding and invariant recognition of masked words: Behavioral and neuroimaging evidence. *Psychological Science*, 15:307–313.

Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D., and Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13(3):321–5.

Di, X. and Biswal, B. B. (2014). Identifying the default mode network structure using dynamic causal modeling on resting-state functional magnetic resonance imaging. *NeuroImage*, 86:53–59.

Di Bono, M. G. and Zorzi, M. (2013). Deep generative learning of location-invariant visual word recognition. *Front. Psychol.*, 4.

Dien, J. (2009). The neurocognitive basis of reading single words as seen through early latency ERPs: a model of converging pathways. *Biological psychology*, 80(1):10–22.

Ding, M., Bressler, S. L., Yang, W., and Liang, H. (2000). Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biological cybernetics*, 83(1):35–45.

Donner, T. H. and Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends in Cognitive Sciences*, 15(5):191–9.

Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, 28:233–244.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4):409.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. ii. *Biometrika*, 38(1/2):159.

Eggert, G. (1977). *Wernicke's works on aphasia : a sourcebook and review.* Mouton, The Hague.

Eichler, M. (2006). On the evaluation of information flow in multivariate systems by the directed transfer function. *Biological Cybernetics*, 94(6):469–82.

Ellis, A. and Young, A. (1996). *Human Cognitive Neuropsychology: A Textbook with Readings.* Psychology Press.

Engel, A. K., Gerloff, C., Hilgetag, C. C., and Nolte, G. (2013). Intrinsic Coupling Modes: Multiscale Interactions in Ongoing Brain Activity. *Neuron*, 80(4):867–886.

Faes, L., Erla, S., and Nollo, G. (2012). Measuring connectivity in linear multivariate processes: definitions, interpretation, and practical analysis. *Computational and mathematical methods in medicine*, 2012:140513.

Fair, D. A., Cohen, A. L., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Barch, D. M., Raichle, M. E., Petersen, S. E., and Schlaggar, B. L. (2008). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences*, 105(10):4028–4032.

Fasoula, A., Attal, Y., and Schwartz, D. (2013). Comparative performance evaluation of data-driven causality measures applied to brain networks. *Journal Of Neuroscience Methods*, 215(2):170–189.

Fiebach, C., Friederici, A., Muller, K., and von Cramon, D. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal Of Cognitive Neuroscience*, 14(1):11–23.

Fiez, J. a. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. *Human Brain Mapping*, 5(2):79–83.

Florin, E., Gross, J., Pfeifer, J., Fink, G. R., and Timmermann, L. (2011). Reliability of multivariate causality measures for neural data. *Journal of Neuroscience Methods*, 198(2):344 – 358.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. V., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. USA*, 102:9673–9678.

Fraser, A. M. and Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Phys. Rev. A*, 33.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews. Neuroscience*, 11(2):127–38.

Friston, K., Daunizeau, J., and Stephan, K. E. (2013). Model selection and gobbledygook: Response to lohmann et al. *NeuroImage*, 75:275–278.

Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, 19(4):1273–1302.

Friston, K., Moran, R., and Seth, A. K. (2012). Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology*, pages 1–7.

Friston, K. J. and Price, C. J. (2001). Dynamic representations and generative models of brain function. *Brain research bulletin*, 54(3):275–285.

Frobenius, G. (1912). Uber Matrizen aus nicht negativen Elementen. pages 456–477.

Gardner, H. (2008). *The Mind's New Science: A History of the Cognitive Revolution*. Basic Books.

Gersch, W. (1972). Causality or driving in electrophysiological signal analysis. *Mathematical Biosciences*, 14(1–2):177 – 196.

Gersch, W. and Goddard, G. (1970). Epileptic Focus Location - Spectral Analysis Method. *Science*, 169(3946):701–&.

Geschwind, N. (1965a). Disconnexion syndromes in animals and man. I. *Brain*, 88(2):237–94.

Geschwind, N. (1965b). Disconnexion syndromes in animals and man. II. *Brain*, 88(3):585–644.

Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77(378):304–313.

Ghosh, A., Rho, Y., McIntosh, A. R., Kötter, R., and Jirsa, V. K. (2008). Noise during rest enables the exploration of the brain's dynamic repertoire. *PLoS Comput Biol*, 4(10):e1000196.

Glezer, L. S., Jiang, X., and Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the "visual word form area". *Neuron*, 62(2):199–204.

Gómez-Herrero, G., Atienza, M., Egiazarian, K., and Cantero, J. L. (2008). Measuring directional coupling between EEG sources. *NeuroImage*, 43(3):497–508.

Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–5.

Gow, D. W., Segawa, J. a., Ahlfors, S. P., and Lin, F.-H. (2008). Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. *NeuroImage*, 43(3):614–23.

Gow, Jr., D. W., Keller, C. J., Eskandar, E., Meng, N., and Cash, S. S. (2009). Parallel versus serial processing dependencies in the perisylvian speech network: A Granger analysis of intracranial EEG data. *BRAIN AND LANGUAGE*, 110(1):43–48.

Gow, Jr., D. W. and Segawa, J. A. (2009). Articulatory mediation of speech perception: A causal analysis of multi-modal imaging data. *COGNITION*, 110(2):222–236.

Grabowski, T. J., Damasio, H., and Damasio, a. R. (1998). Premotor and prefrontal correlates of category-related lexical retrieval. *NeuroImage*, 7(3):232–43.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29(2):228–244.

Grainger, J., Dufau, S., Montant, M., Ziegler, J. C., and Fagot, J. (2012). Orthographic processing in baboons (papio papio). *Science*, 336(6078):245–248.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 37(3):424–438.

Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., and Binder, J. R. (2010). Neural systems for reading aloud: a multiparametric approach. *Cerebral Cortex*, 20(8):1799–815.

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, 100(1):253–258.

Grill-Spector, K. and Malach, R. (2004). The human visual cortex. *Annual review of Neuroscience*, 27:649–77.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, a., and Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 98(2):694–9.

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., and Lounasmaa, O. V. (1993). Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of modern Physics*, 65(2):413.

Heim, S., Eickhoff, S. B., Ischebeck, A. K., Friederici, A. D., Stephan, K. E., and Amunts, K. (2009). Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. *Human Brain Mapping*, 30(2):392–402.

Hillebrand, A., Barnes, G. R., et al. (2005). Beamformer analysis of meg data. *International Review of Neurobiology*, 68:149.

Hinton, G., Dayan, P., Frey, B., and Neal, R. (1995). The "wake-sleep" algorithm for unsupervised neural networks. *Science*, 268(5214):1158–1161.

Hinton, G. E. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507.

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, 11(10):428–434.

Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., and Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience*, 15(6):884–U110.

Hochstein, S. and Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804.

Hodges, J. R., Patterson, K., Oxbury, S., and Funnell, E. (1992). Semantic dementia. progressive fluent aphasia with temporal lobe atrophy. *Brain*, 115(6):1783–1806.

Horwitz, B., Rumsey, J. M., and Donohue, B. C. (1998). Functional connectivity of the angular gyrus in normal reading and dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8939–44.

Hotton, S. and Yoshimi, J. (2011). Extending Dynamical Systems Theory to Model Embodied Cognition. *Cognitive Science*, 35(3):444–479.

Huang, M. X., Shih, J. J., Lee, R. R., Harrington, D. L., Thoma, R. J., Weisend, M. P., Hanlon, F., Paulson, K. M., Li, T., Martin, K., Millers, G. A., and Canive, J. M. (2004). Commonalities and differences among vectorized beamformers in electromagnetic source imaging. *Brain Topogr*, 16(3):139–158.

Humphreys, G. W. and Evett, L. J. (1985). Are there independent lexical and nonlexical routes in word processing? an evaluation of the dual-route theory of reading. *Behav Brain Sci*, 8(04):689.

Humphreys, G. W., Evett, L. J., and Quinlan, P. T. (1990). Orthographic processing in visual word identification. *Cognitive Psychology*, 22(4):517 – 560.

Hämäläinen, M. and Ilmoniemi, R. (1994). Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1):35–42.

Jacobs, A. M. and Grainger, J. (1994). Models of visual word recognition: Sampling the state of the art. *Journal of Experimental Psychology: Human perception and performance*, 20(6):1311.

Jiao, Q., Lu, G., Zhang, Z., Zhong, Y., Wang, Z., Guo, Y., Li, K., Ding, M., and Liu, Y. (2011). Granger causal influence predicts bold activity levels in the default mode network. *Human Brain Mapping*, 32(1):154–161.

Jobard, G., Crivello, F., and Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. *NeuroImage*, 20(2):693–712.

Johnson, S., Prendergast, G., Hymers, M., and Green, G. (2011). Examining the Effects of One- and Three-Dimensional Spatial Filtering Analyses in Magnetoencephalography. *PLoS One*, 6(8).

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, 87(4):329.

Kaminski, M. and Blinowska, K. (1991). A new method of the description of the information flow in the brain structures. *Biological cybernetics*, 65(3):203–210.

Kaminski, M. and Blinowska, K. (2014). Directed transfer function is not influenced by volume conduction—inexpedient pre-processing should be avoided. *Frontiers in Computational Neuroscience*, 8(61).

Kaminski, M., Ding, M., Truccolo, W. A., and Bressler, S. L. (2001). Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological Cybernetics*, 85(2):145–157.

Keller, S. S., Crow, T., Foundas, A., Amunts, K., and Roberts, N. (2009). Broca's area: Nomenclature, anatomy, typology and asymmetry. *BRAIN AND LANGUAGE*, 109(1):29–48.

Kennel, M., Brown, R., and Abarbanel, H. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A*, 45(6):3403–3411.

Kiebel, S. J., Garrido, M. I., Moran, R. J., and Friston, K. J. (2008). Dynamic causal modelling for EEG and MEG. *Cognitive neurodynamics*, 2(2):121–36.

Kleiman, G. (1975). Speech Recoding In Reading. *Journal Of Verbal Learning And Verbal Behavior*, 14(4):323–339.

Knill, D. C. and Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12):712 – 719.

Koenig, T., Studer, D., Hubl, D., Melie, L., and Strik, W. K. (2005). Brain connectivity at different time-scales measured with eeg. *Philo-*

sophical Transactions of the Royal Society B: Biological Sciences, 360(1457):1015–1024.

Korzeniewska, A., Crainiceanu, C. M., Kuś, R., Franaszczuk, P. J., and Crone, N. E. (2008). Dynamics of event-related causality in brain electrical activity. *Human Brain Mapping*, 29(10):1170–92.

Korzeniewska, A., Franaszczuk, P. J., Crainiceanu, C. M., Kus, R., and Crone, N. E. (2011). Dynamics of large-scale cortical interactions at high gamma frequencies during word production: Event related causality (ERC) analysis of human electrocorticography (ECoG). *Neuroimage*, 56(4):2218–2237.

Koyama, M. S., Kelly, C., Shehzad, Z., Penesetti, D., Castellanos, F. X., and Milham, M. P. (2010). Reading networks at rest. *Cerebral Cortex*, 20(11):2549–59.

Kozinska, D., Carducci, F., and Nowinski, K. (2001). Automatic alignment of EEG/MEG and MRI data sets. *Clinical Neurophysiology*, 112(8):1553–61.

Kugiumtzis, D. (1996). State space reconstruction parameters in the analysis of chaotic time series — the role of the time window length. *Physica D: Nonlinear Phenomena*, 95(1):13–28.

Kujala, J., Gross, J., and Salmelin, R. (2008). Localization of correlated network activity at the cortical level with MEG. *NeuroImage*, 39(4):1706–20.

Kujala, J., Pammer, K., Cornelissen, P., Roebroeck, A., Formisano, E., and Salmelin, R. (2007). Phase coupling in a cerebro-cerebellar network at 8-13 Hz during reading. *Cerebral Cortex*, 17(6):1476–85.

Kujala, J., Vartiainen, J., Laaksonen, H., and Salmelin, R. (2012). Neural interactions at the core of phonological and semantic priming of written words. *Cerebral Cortex*, 22(10):2305–12.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Kus, R., Ginter, J. S., and Blinowska, K. J. (2006). Propagation of EEG activity during finger movement and its imagination. *Acta Neurobiologiae Experimentalis*, 66(3):195–206.

LaBerge, D. and Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive psychology*, 6(2):293–323.

Lamme, V. A. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences*, 23(11):571–579.

Lee, T. S. and Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 20(7):1434.

Lefèvre, J. and Baillet, S. (2009). Optical flow approaches to the identification of brain dynamics. *Human Brain Mapping*, 30(6):1887–1897.

Lehmann, D. (1984). Eeg assessment of brain activity: Spatial aspects, segmentation and imaging. *International Journal of Psychophysiology*, 1(3):267–276.

Lehmann, D., Strik, W., Henggeler, B., Koenig, T., and Koukkou, M. (1998). Brain electric microstates and momentary conscious mind states as building blocks of spontaneous thinking: I. visual imagery and abstract thoughts. *International Journal of Psychophysiology*, 29(1):1–11.

Levinson, N. (1946). The wiener rms (root mean square) error criterion in filter design and prediction. *Journal of Mathmatical Physics*, 25:261–278.

Levy, J., Pernet, C., Treserras, S., Boulanouar, K., Aubry, F., Demonet, J.-F., and Celsis, P. (2009). Testing for the Dual-Route Cascade Reading Model in the Brain: An fMRI Effective Connectivity Account of an Efficient Reading Style. *PLoS One*, 4(8).

Liu, J., Li, J., Rieth, C. a., Huber, D. E., Tian, J., and Lee, K. (2011). A dynamic causal modeling analysis of the effective connectivities underlying top-down letter processing. *Neuropsychologia*, 49(5):1177–86.

Liu, Z., Fukunaga, M., de Zwart, J. A., and Duyn, J. H. (2010). Large-scale spontaneous fluctuations and correlations in brain electrical activity observed with magnetoencephalography. *NeuroImage*, 51(1):102–111.

Lohmann, G., Erfurth, K., Müller, K., and Turner, R. (2012). Critical comments on dynamic causal modelling. *NeuroImage*, 59(3):2322–2329.

Lohmann, G., Müller, K., and Turner, R. (2013). Response to commentaries on our paper: Critical comments on dynamic causal modelling. *NeuroImage*, 75:279–281.

Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., and Yang, Y. (2012). Rat brains also have a default mode network. *Proceedings of the National Academy of Sciences*, 109(10):3979–3984.

Lütkepohl, H. (1991). *Introduction to multiple time series analysis*. Springer-Verlag.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., and Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38(3):487–97.

Marple, S. L. (1987). *Digital Spectral Analysis: With Applications (Prentice-Hall Series in Signal Processing)*. Prentice Hall.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 275(942):pp. 483–519.

Maus, A. and Sprott, J. (2011). Neural network method for determining embedding dimension of a time series. *Communications in Nonlinear Science and Numerical Simulation*, 16(8):3294–3302.

McClelland, J. and Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5):375.

McClelland, J. L., Rumelhart, D. E., Group, P. R., et al. (1986). Parallel distributed processing. *Explorations in the microstructure of cognition*, 2.

McIntosh, a. R. (2000). Towards a network theory of cognition. *Neural networks : the official journal of the International Neural Network Society*, 13(8-9):861–70.

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Lambon Ralph, M. a., Patterson, K., McClelland, J. L., and Price, C. J. (2005). Dissociating reading processes on the basis of neuronal interactions. *Journal of Cognitive Neuroscience*, 17(11):1753–65.

Meiss, J. (2007). Dynamical systems. *Scholarpedia*, 2(2):1629.

Mesulam, M. M. (2000). *Principles of behavioral and cognitive neurology*. Oxford University Press, Oxford New York.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76(2):165–178.

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., and Tyler, L. K. (2005). Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*, 15(11):1723–35.

Musso, F., Brinkmeyer, J., Mobascher, A., Warbrick, T., and Winterer, G. (2010). Spontaneous brain activity and eeg microstates. a novel eeg/fmri analysis approach to explore resting-state networks. *NeuroImage*, 52(4):1149–1161.

Nakagawa, T. T., Woolrich, M., Luckhoo, H., Joensson, M., Mohseni, H., Kringelbach, M. L., Jirsa, V., and Deco, G. (2014). How delays matter in an oscillatory whole-brain spiking-neuron network model for {MEG} alpha-rhythms at rest. *NeuroImage*, 87(0):383 – 394.

Nalatore, H. and Rangarajan, G. (2009). Short-window spectral analysis using AMvar and multitaper methods: a comparison. *Biological Cybernetics*, 101(1):71–80.

Nichols, T. and Holmes, A. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, 15(1):1–25.

Nobre, A., Allison, T., and McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10):2292–307.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, 113(2):327–357.

Norris, D. (2009). Putting it all together: A unified account of word recognition and reaction-time distributions. *Psychological Review*, 116(1):207–219.

Norris, D. and Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review*, 119(3):517–545.

Norris, D., McQueen, J. M., and Cutler (2000). Merging information in speech recognition: Feedback is never necessary. *Behav. Brain Sci.*, 23(3):299–325.

Norris, D., Mcqueen, J. M., Norris, D., Mcqueen, J. M., Cutler, A., and Butterfield, S. (2008). Shortlist b: A bayesian model of continuous speech recognition. *Psychological Review*, pages 357–395.

Packard, N., Crutchfield, J., Farmer, J., and Shaw, R. (1980). Geometry from a time series. *Physical Review Letters*, 45(9):712–716.

Pammer, K., Hansen, P. C., Kringelbach, M. L., Holliday, I., Barnes, G., Hillebrand, A., Singh, K. D., and Cornelissen, P. L. (2004). Visual word recognition: the first half second. *NeuroImage*, 22(4):1819–25.

Pascual-Marqui, R. D., Biscay, R. J., Bosch-Bayard, J., Lehmann, D., Kochi, K., Kinoshita, T., Yamada, N., and Sadato, N. (2014). Assessing direct paths of intracortical causal information flow of oscillatory activity with the isolated effective coherence (icoh). *Frontiers in Human Neuroscience*, 8(448).

Pelli, D. G., Farell, B., and Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, 423(6941):752–756.

Peressotti, F. and Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception & Psychophysics*, 61(4):691–706.

Perron, O. (1907). Zur theorie der matrices. *Mathematische Annalen*, 64(2):248–263.

Petersen, S. E., Fox, P. T., Snyder, a. Z., and Raichle, M. E. (1990). Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science (New York, N.Y.)*, 249(4972):1041–4.

Pinneo, L. R. (1966). On noise in the nervous system. *Psychological Review*, 73(3):242–247.

Porcaro, C., Coppola, G., Pierelli, F., Seri, S., Di Lorenzo, G., Tomasevic, L., Salustri, C., and Tecchio, F. (2013). Multiple frequency functional connectivity in the hand somatosensory network: An EEG study. *Clinical Neurophysiology*, 124(6):1216–1224.

Price, C. J. (2012). A review and synthesis of the first 20 Years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2):816–47.

Price, C. J. and Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, 19(3):473–481.

Price, C. J. and Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15(6):246–53.

Price, C. J. and Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive neuropsychology*, 22(3):262–75.

Priestley, M. (1981). *Spectral Analysis and Time Series: Multivariate series, prediction and control*. Probability and Mathematical Statistics. Academic Press.

Pugh, K. R., Shaywitz, B. a., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., Bronen, R. a., Shankweiler, D. P., Katz, L., Fletcher, J. M., and Gore, J. C. (1996). Cerebral organization of component processes in reading. *Brain*, 119 (Pt 4:1221–38.

Pulvermüller, F., Shtyrov, Y., and Hauk, O. (2009). Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain and language*, 110(2):81–94.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Ragwitz, M. and Kantz, H. (2002). Markov models from data by simple nonlinear time series predictors in delay embedding spaces. *Physical Review E*, 65(5, 2).

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, 98(2):676–682.

Rao, R. P. and Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, 9:721–763.

Rezek, I. and Roberts, S. (2005). Ensemble hidden markov models with extended observation densities for biosignal analysis. *Advanced Information and Knowledge Processing*, page 419–450.

Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S. C., and Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *The Journal of Neuroscience*, 31(22):8239–47.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience.*, 2(11):1019–1025.

Salmelin, R. and Kujala, J. (2006). Neural representation of language: activation versus long-range connectivity. *Trends In Cognitive Sciences*, 10(11):519–525.

Sameshima, K. and Baccalá, L. A. (1999). Using partial directed coherence to describe neuronal ensemble interactions. *Journal of Neuroscience Methods*, 94(1):93–103.

Sameshima, K. and Baccalá, L. a. (1999). Using partial directed coherence to describe neuronal ensemble interactions. *Journal of Neuroscience Methods*, 94(1):93–103.

Schelter, B., Winterhalder, M., Eichler, M., Peifer, M., Hellwig, B., Guschlbauer, B., Lücking, C. H., Dahlhaus, R., and Timmer, J. (2006). Testing for directed influences among neural signals using partial directed coherence. *Journal of Neuroscience Methods*, 152(1-2):210–9.

Schlögl, A. (2006). A comparison of multivariate autoregressive estimators. *Signal Processing*, 86(9):2426–2429.

Schlögl, A. and Supp, G. (2006). Analyzing event-related eeg data with multivariate autoregressive parameters. *Progress in Brain Research*, 159:135–147.

Schlögl, A. and Supp, G. (2006). Analyzing event-related EEG data with multivariate autoregressive parameters. *Progress in Brain Research*, 159:135–47.

Schoffelen, J.-M. and Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Human Brain Mapping*, 30(6):1857–65.

Schoonbaert, S. and Grainger, J. (2004). Letter position coding in printed word perception: Effects of repeated and transposed letters. *Language and Cognitive Processes*, 19(3):333–367.

Seghier, M. L. and Friston, K. J. (2013). Network discovery with large DCMs. *NeuroImage*, 68:181–91.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science.*, 14(5):238–242.

Seidenberg, M. S. (2006). *Connectionist Models of Reading*, chapter 14, pages 235–25. Oxford University Press, first edition. As advertised: a pretty simple overview of connectionist models of reading. Might dispel some misconceptions about methodology, general approach.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, 96(4):523–68.

Simos, P. G., Breier, J. I., Fletcher, J. M., Foorman, B. R., Castillo, E. M., and Papanicolaou, A. C. (2002). Brain mechanisms for reading words and pseudowords: an integrated approach. *Cerebral cortex (New York, N.Y. : 1991)*, 12(3):297–305.

Smallwood, J., Gorgolewski, K. J., Golchert, J., Ruby, F. J. M., Engen, H., Baird, B., Vinski, M. T., Schooler, J. W., and Margulies, D. S. (2013). The default modes of reading: modulation of posterior cingulate and medial prefrontal cortex connectivity associated with comprehension and task focus while reading. *FRONTIERS IN HUMAN NEUROSCIENCE*, 7.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., et al. (2009). Correspondence of the brain's functional architecture during activation

and rest. *Proceedings of the National Academy of Sciences*, 106(31):13040–13045.

Stam, C. J. (2005). Nonlinear dynamical analysis of eeg and meg: review of an emerging field. *Clinical Neurophysiology*, 116(10):2266–2301.

Sun, Y., Zhang, H., Feng, T., Qiu, Y., Zhu, Y., and Tong, S. (2009a). Early cortical connective network relating to audiovisual stimulation by partial directed coherence analysis. *IEEE transactions on bio-medical engineering*, 56(11 Pt 2):2721–4.

Sun, Y., Zhang, H., Feng, T., Qiu, Y., Zhu, Y., and Tong, S. (2009b). Early Cortical Connective Network Relating to Audiovisual Stimulation by Partial Directed Coherence Analysis. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, 56(11):2721–2724.

Supp, G. G., Schlögl, A., Trujillo-Barreto, N., Müller, M. M., and Gruber, T. (2007). Directed cortical information flow during human object recognition: Analyzing induced eeg gamma-band responses in brain's source space. *PLoS ONE*, 2(8):e684.

Takens, F. (1981). Detecting strange attractors in turbulence. *Lecture Notes in Mathematics*, page 366–381.

Tarkiainen, A., Cornelissen, P. L., and Salmelin, R. (2002). Dynamics of visual feature analysis and object level processing in face versus letter string perception. *Brain*, 125(5):1125–1136.

Tarkiainen, a., Helenius, P., Hansen, P. C., Cornelissen, P. L., and Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain*, 122 (Pt 1:2119–32.

Taylor, J. S. H., Rastle, K., and Davis, M. H. (2012). Can Cognitive Models Explain Brain Activation During Word and Pseudoword Reading? A Meta-Analysis of 36 Neuroimaging Studies. *Psychological Bulletin*.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. a. (2002). Meta-Analysis of the Functional Neuroanatomy of Single-Word Reading: Method and Validation. *NeuroImage*, 16(3):765–780.

Twomey, T., Kawabata Duncan, K. J., Price, C. J., and Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *NeuroImage*, 55(3):1242–51.

Uddin, L. Q., Clare Kelly, A., Biswal, B. B., Xavier Castellanos, F., and Milham, M. P. (2009). Functional connectivity of default mode network components: Correlation, anticorrelation, and causality. *Human Brain Mapping*, 30(2):625–637.

Uddin, L. Q., Supekar, K., Amin, H., Rykhlevskaia, E., Nguyen, D. A., Greicius, M. D., and Menon, V. (2010). Dissociable connectivity within human angular gyrus and intraparietal sulcus: Evidence from functional and structural connectivity. *Cerebral Cortex*, 20(11):2636–2646.

Ungerleider, L. G. and Mishkin, M. (1982). Two cortical visual systems. In Ingle, D. J., Goodale, M. A., and Mansfield, R. J. W., editors, *Analysis of Visual Behavior*, volume 549, chapter 18, pages 549–586. MIT Press.

van den Heuvel, M., Mandl, R., Luigjes, J., and Hulshoff Pol, H. (2008). Microstructural organization of the cingulum tract and the level of default mode functional connectivity. *Journal of Neuroscience*, 28(43):10844–10851.

van den Heuvel, M. P. and Sporns, O. (2011). Rich-Club Organization of the Human Connectome. *Journal Of Neuroscience*, 31(44):15775–15786.

van Oort, E., van Cappellen van Walsum, A., and Norris, D. (2014a). An investigation into the functional and structural connectivity of the default mode network. *NeuroImage*, 90:381–389.

van Oort, E., van Cappellen van Walsum, A., and Norris, D. (2014b). An investigation into the functional and structural connectivity of the default mode network. *NeuroImage*, 90:381–389.

Van Veen, B., Van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9):867–880.

Vartiainen, J., Liljeström, M., Koskinen, M., Renvall, H., and Salmelin, R. (2011). Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography evoked responses yield different neural functionality in reading. *The Journal of Neuroscience*, 31(3):1048–58.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., Zempel, J. M., Snyder, L. H., Corbetta, M., and Raichle, M. E. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, 447(7140):83–86.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., and Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron*, 55(1):143–56.

Visser, M., Jefferies, E., Embleton, K. V., and Ralph, M. A. L. (2012). Both the Middle Temporal Gyrus and the Ventral Anterior Temporal Area Are Crucial for Multimodal Semantic Processing: Distortion-corrected fMRI Evidence for a Double Gradient of Information Convergence in the Temporal Lobes. *Journal Of Cognitive Neuroscience*, 24(8):1766–1778.

Visser, M., Jefferies, E., and Ralph, M. A. L. (2010). Semantic Processing in the Anterior Temporal Lobes: A Meta-analysis of the Functional Neuroimaging Literature. *Journal Of Cognitive Neuroscience*, 22(6):1083–1094.

von Stein, A. and Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *Int J Psychophysiol*, 38(3):301–313.

Vrba, J. and Robinson, S. E. (2001). Signal processing in magnetoencephalography. *Methods (San Diego, Calif.)*, 25(2):249–71.

Wandell, B. a., Dumoulin, S. O., and Brewer, A. a. (2007). Visual field maps in human cortex. *Neuron*, 56(2):366–83.

Wheat, K. L., Cornelissen, P. L., Frost, S. J., and Hansen, P. C. (2010). During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: evidence from magnetoencephalography. *The Journal of Neuroscience*, 30(15):5229–33.

Wheat, K. L., Cornelissen, P. L., Sack, A. T., Schuhmann, T., Goebel, R., and Blomert, L. (2013). Charting the functional relevance of Broca's area for visual word recognition and picture naming in Dutch using fMRI-guided TMS. *Brain And Language*, 125(2, SI):223–230.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: the SERIOL model and selective literature review. *Psychonomic bulletin & review*, 8(2):221–43.

Whitney, H. (1936). Differentiable manifolds. *The Annals of Mathematics*, 37(3):645.

Woodhead, Z. V. J., Barnes, G. R., Penny, W., Moran, R., Teki, S., Price, C. J., and Leff, a. P. (2012). Reading Front to Back: MEG Evidence for Early Feedback Effects During Word Recognition. *Cerebral Cortex*, pages 1–9.

Woolrich, M. W., Baker, A., Luckhoo, H., Mohseni, H., Barnes, G., Brookes, M., and Rezek, I. (2013). Dynamic state allocation for meg source reconstruction. *NeuroImage*, 77:77–92.

Wright, P., Randall, B., Marslen-Wilson, W. D., and Tyler, L. K. (2011). Dissociating linguistic and task-related activity in the left inferior frontal gyrus. *Journal of Cognitive Neuroscience*, 23(2):404–13.

Yan, C. and He, Y. (2011). Driving and driven architectures of directed small-world human brain functional networks. *PLoS ONE*, 6(8):e23460.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zoeller, L., Polimeni, J. R., Fischl, B., Liu, H., and Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal Of Neurophysiology*, 106(3):1125–1165.

Yvert, G., Perrone-Bertolotti, M., Baciu, M., and David, O. (2012). Dynamic causal modeling of spatiotemporal integration of phonological and semantic processes: an electroencephalographic study. *The Journal of Neuroscience*, 32(12):4297–306.

Zeki, S. (2001). Localization and globalization in conscious vision. *Annual Review of Neuroscience*, 24(1):57–86. PMID: 11283305.

Zevin, J. D. and Seidenberg, M. S. (2006). Consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, 54(2):145–160.

Zhou, Z., Wang, X., Klahr, N. J., Liu, W., Arias, D., Liu, H., von Deneen, K. M., Wen, Y., Lu, Z., Xu, D., and et al. (2011). A conditional granger causality model approach for group analysis in functional magnetic resonance imaging. *Magnetic Resonance Imaging*, 29(3):418–433.