THÈSE

présentée et publiquement soutenue devant

LA FACULTÉ D’AIX-MARSEILLE UNIVERSITÉ

Le Jeudi 29 Octobre 2015

Par

Dashiel MUNDING-MINIER

La dynamique spatio-temporelle de la production des mots: études par magnétoencéphalographie

Pour l’obtention du DOCTORAT d’AIX-MARSEILLE UNIVERSITÉ

Spécialité : PSYCHOLOGIE

Réalisée sous la direction de:

François-Xavier ALARIO

Membres du Jury de la Thèse :

Beth JEFFERIES, Professor of Psychology, University of York  Rapporteur
Christophe PALLIER, Directeur de Recherches, INSERM, CEA & CNRS  Rapporteur
Daniele SCHON, chargé de recherche, INS & INSERM-AMU F.-Xavier  Président du Jury
ALARIO, Directeur de Recherches, AMU & CNRS  Directeur de thèse
Acknowledgements

Writing a thesis is a life changing event. Getting off the plane for the first time in Marseille to be interviewed for the PhD position, I had no idea how much I would experience and learn, and how different things would be at the end of the job. It has been an education science and language, in compromise and perseverance, in friendship and in love, made possible by the support of many people from inside and outside the lab, who deserve my gratitude.

Thanks to Xavier, for believing in me in the first place, for supporting me through some frustrating and arduous times. Your insight, acumen, sense of humour and most of all, patience, made possible the successes we've achieved.

I must also thank my wife, Laure, for her support, care and love. It has been wonderful to have the voice of reason to share my ideas and problems with - now that we've written our theses in parallel, I know there's nothing we can't manage together!

Enormous thanks and recognition also to my colleagues at St. Charles and La Timone who were there for me and the project in various capacities. Jean-Michel, Sophie, and Bruno who made everything possible at the MEG lab, Marieke and Anne-Sophie who were so important in giving the project its shape.

Thanks to all my friends from the salle des doctorants, the LPC and the surrounding environs: Sebastiaan & Lotje, Delphine & Alban, Gabriel, Anaïs & Matthieu, Svetlana, Marina, Kim and the girls, Jasmin, and the rest of Team Alario, you were an unending source of friendship and fun, and a tonic for a frazzled mind. Thanks to my French family, especially Chantal, Alain, Claire, Damien, Theo and Lucas for making me so welcome and accepted. The same to my English family, who have been there for me even when they weren't here with me, I love you all!

My most respectful thanks to my jurors, for their benevolence in evaluating this work. I hope that I've managed to show the value of the collaborative effort made by all who collaborated on this thesis.

Finally, I would like to recognise the institutional support I've had. From the LPC, ED 356, AMU, CNRS and ERC, and most particularly Fred and Colette, who have been my spirit guides in the daunting world of the French bureaucracy.

DHM, 14/7/15

För min farfar.
# Contents

Preamble: ........................................................................................................................................... 3
Préambule: ........................................................................................................................................... 5

1. Introduction....................................................................................................................................... 8
   1.1. Psycholinguistic models of language processing and speech production: ......................... 8
   1.2. Simulated models – parameterisation of verbal models: ..................................................... 10
   1.3. Simulated models - complexity and the question of parallel/serial functionality: ............... 12
   1.4. Changing perspectives: the evolution of psycholinguistics as a consequence of developments in neuroimaging: ................................................................................. 16
   1.5. Time, space and neurons: instantiations of psycholinguistic models in the brain: .............. 19
   1.6. References: ............................................................................................................................... 28

2. Literature review ................................................................................................................................. 34
   2.1. Abstract: ................................................................................................................................. 35
   2.2. Introduction: ........................................................................................................................... 36
   2.3. Methods: ............................................................................................................................... 38
   2.4. Results: .................................................................................................................................... 41
   2.5. General Discussion: ............................................................................................................... 59
   2.6. Conclusions: ........................................................................................................................... 66
   2.7. References: ........................................................................................................................... 67
   2.8. Supplementary materials ......................................................................................................... 73

3. Addressing the problem: Experimental goals and development .................................................... 78
   3.1. The role of neuroimaging evidence in assessing psycholinguistic models: ......................... 78
   3.2. Requirements for adding to the current body of evidence: ................................................... 79
   3.3. Motivation for an empirical study of word production: ....................................................... 80
   3.4. Paradigm selection, experimental design and development: ............................................... 81
   3.5. The importance of a full exploration of the network versus the necessity of simplicity: ....... 88
   3.6. References: ........................................................................................................................... 89

4. Technical development of a MEG protocol to study language production ................................... 93
   4.1. Magnetoencephalography: scanning and analysis techniques: ........................................... 93
   4.3. Methods of MEG analysis: Beamforming: .......................................................................... 97
   4.4. Methods of MEG analysis: Minimum Norm Estimation ..................................................... 97
   4.5. Characteristics of MEG ........................................................................................................... 98
4.6. Developing the technical element of an empirical MEG investigation of language production ................................................................................................................................. 99
4.7. Adapting our experimental aims and procedures to MEG: .................................................. 101
4.8. Pilot MEG naming study ........................................................................................................... 102
4.9. Post-pilot refinements: ................................................................................................................. 111
4.10. References ..................................................................................................................................... 118
5. Early semantic and late visual differences in the cortical response during picture naming ...... 123
  5.1. Introduction: .................................................................................................................................. 123
  5.2. Methods: ....................................................................................................................................... 125
  5.3. Results: .......................................................................................................................................... 129
  5.4. Discussion: .................................................................................................................................... 134
  5.5. Limitations: ..................................................................................................................................... 139
  5.6. Conclusions: .................................................................................................................................. 139
  5.7. References: .................................................................................................................................... 141
  5.8. Supplementary materials: ........................................................................................................... 147
6. Summary and General Discussion ................................................................................................. 150
  6.1. Introduction and background: .................................................................................................... 150
  6.2. Literature review: ......................................................................................................................... 150
  6.3. Experimental design and technical development: ....................................................................... 151
  6.4. Empirical study: ........................................................................................................................... 152
  6.5. General interpretation: ................................................................................................................ 153
  6.6. Limitations and potential developments: .................................................................................... 154
  6.7. Summary and future directions/perspectives: ............................................................................ 155
  6.8. Conclusion: ..................................................................................................................................... 156
  6.9. References: ..................................................................................................................................... 159
7. Abstract: .......................................................................................................................................... 161
7b. Résumé: ....................................................................................................................................... 161
Preamble:

The efficient communication of ideas through speech underpins human endeavour, from the mundane to the transcendent. A parent persuading the children that it is now most assuredly and unavoidably their bedtime, relies upon the same basic, infinitely variable skill as an actor's portrayal of Shakespeare's fading Lear, or a cognitive psychologist co-ordinating the most complex and elegant of language studies with their team. It seems right then, that much psychological research is devoted to the subject of language production, and psycholinguistics has made an enormous amount of progress using reaction time and other behavioural experiments. In the last thirty years, neuroimaging techniques like functional magnetic resonance imaging [fMRI] (Price, 2010, 2012) and magnetoencephalography [MEG] (Hansen, Kringelbach, & Salmelin, 2010; Salmelin, 2007) have been developed to the point where predictions made by verbal or 'boxological' psycholinguistic models, originally constructed on the basis of behavioural experiments, can be tested in the living brain. Neuroimaging investigations of speech remain technically challenging, as many complex elements must be carefully gelled together to obtain trustworthy and informative evidence.

In order to achieve such an outcome, points of distinction between competing models must be clearly identified. A protocol for accessing this point of divergence must then be designed and programmed in a manner compatible with the constraints of the imaging technique in question, all this before the challenges of practical execution and intricate analytical techniques are applied. Errors at any stage can drastically compromise experimental results, and even with flawless execution, available scanning methods have limits of resolution and signal-to-noise ratio.

Speech requires combining multiple words: the more expert among us are capable of stringing several words together into sentences to express an idea we need to communicate. This process requires more planning than our habitual ease of expression suggests, especially when viewed from the perspective of a neuroimaging investigation. It is difficult to extract one word from another as their sounds run together, and impossible to extract the elements of a complex concept that is being maintained in the brain during the course of its articulation. For the moment then, fluent speech remains an extreme challenge in neuroimaging, and most research is conducted on the production of single words, normally using a word reading or picture naming paradigm.

The brain response during word production is widespread and highly dynamic. The rhythmicity of speech itself is around 3-5Hz, and the brain uses neural oscillations on a frequency range between ≈0.05Hz to ≈500Hz (Buzsáki & Draguhn, 2004), although responses at frequencies as high as 600Hz have been found using MEG (Hansen et al., 2010; p. 39). fMRI's temporal resolution is too poor to dissect the haemodynamic response in time windows much finer than a second (Huettel, Song, & McCarthy, 2004 p197-206), and Blood Oxygenation Level Dependent signal only provides a secondary measure of neural activity. Of the non-invasive neuroimaging techniques available for studying the dynamics of word production, MEG provides the best combination of temporal and spatial resolution. However, it demands a high price for this level of detail in terms of technical complexity; setup is complex, and analysing MEG data is extremely time consuming compared to behavioural methods, and these concerns are exacerbated when applying it to a system as complex as word production.
This thesis will attempt to dissect some of this complexity by exploring and discriminating between psycholinguistic theories of language and word production. It will assess how neuroimaging data, in particular MEG evidence, contributes to the debate. In this context, it will discuss the design and implementation of our own MEG investigation and how our findings add to existing knowledge.

**Overview of chapters:**

- **Chapter 1.** Introduction and consideration of psycholinguistic and computational models of speech.

In this introduction we will consider the development of psycholinguistic models of language processing from behavioural experiments, and the conversion of such verbally specified models into formal computational models. We will then discuss the necessity of relating such simulations to biological evidence of the cortical response to substantiate their claims, and the best way of providing such evidence in the context of the cortical language network.

- **Chapter 2.** Review of MEG neuroimaging literature on speech/word production: The cortical dynamics of speech: a review of MEG evidence from studies of language production.

This chapter assembles the evidence from the available literature on MEG studies of speech/word production, performing an analysis of the reported data. It reviews and assesses the reliability and interpretability of the evidence, discusses the relationship between cortical anatomy and cognitive function, and evaluates how to exploit the strengths of MEG to better investigate the language production network.

- **Chapter 3.** Addressing the problem: identification of experimental goals and development of procedures for adaptation into a MEG scanning protocol.

This chapter explores the theoretical questions and goals behind the adaptation of a behavioural psychological protocol to a neuroimaging context in general, and the specifics of designing a picture naming experiment compatible with MEG.

- **Chapter 4.** Technical development of a MEG protocol to study language production: a summary of MEG recording/analysis techniques, and the technical development of an experimental protocol.

Here we discuss the technical properties and capabilities of MEG in the context of designing a new empirical investigation, then outline the design steps, piloting process and subsequent refinements made in creating the experimental protocol and analysis pipeline.

- **Chapter 5.** Empirical investigation of picture naming. Early semantic and late visual differences in the cortical response to picture naming in Magnetoencephalography.

This chapter takes the form of a neuroimaging study of the cortical response during picture naming, performed at the University of Aix-Marseille and the MEG scanning facility at La Timone hospital.

- **Chapter 6.** Summary and discussion.
This sections consists of a contextual evaluation of the existing models, the literature review, and the empirical study, and an exploration of potential future avenues of investigation.

**Préambule:**


Les études en neuroimagerie portant sur la production du langage demeurent techniquement difficiles, car de nombreux éléments doivent être soigneusement combinés afin d’obtenir des données fiables et informatives. Pour atteindre cet objectif, des points de distinction entre les modèles théoriques en compétition doivent être précisément identifiés. Un protocole d’accès à ces points de divergence doit être conçu et programmé d’une manière qui soit compatible avec les contraintes de la technique d’imagerie utilisée. Ces difficultés ne font que précéder celles de la réalisation pratique de l’expérience et de l’exécution d’analyses complexes. Des faux pas commis à n’importe quelle étape peuvent gravement polluer les données récoltées, et même avec une exécution sans faute, les scanners ont leurs limites de résolution et de ratio signal/bruit.

Parler requiert la combinaison de multiples mots: les plus experts d’entre nous sont capables d’enchaîner plusieurs mots dans une phrase pour exprimer et communiquer des idées complexes. Du point de vu neurolinguistique, ce processus demande bien plus de planification que notre facilité d’expression le suggère. A l’heure actuelle, étudier la production du langage courant en neuroimagerie reste un extraordinaire défi. En effet, il est difficile de distinguer un mot du suivant car la fin du premier et le début du prochain sont coarticulés, et il est impossible de déterminer auquel des concepts complexes est associé une réponse corticale durant l’articulation d’une phrase complète. C’est pourquoi la plupart des recherches sont conduites sur la production de mots isolés, habituellement avec un protocole de dénomination d’images, ou de lecture de mots individuels.

La réponse cérébrale pendant la production des mots est répandue et très dynamique. La rythmicité de la parole elle même est d’environ de 3 à 5 Hz, et les neurones cérébraux utilisent une gamme de fréquences entre ≈0.05Hz et ≈500Hz (Buzsáki & Draguhn, 2004), bien que des réponses jusqu’à 600Hz aient été observées en MEG (Hansen et al., 2010). La résolution temporelle de l’IRMf est trop pauvre pour pouvoir analyser la réponse hémodynamique dans les fenêtres temporelles plus fines qu’une
seconde, et le signal dépendant du niveau d'oxygène sanguin n'est qu'une mesure secondaire de l'activité neuronale. Parmi les techniques de neuroimagerie, la MEG nous donne la meilleure combinaison de résolution temporelle et spatiale. Cependant, elle demande un haut niveau d'expertise technique. L'analyse des données MEG demande beaucoup de temps par rapport aux analyses de données comportementales, et la mise en place d'un protocole est complexe. Ces difficultés sont exacerbées lorsque la MEG est utilisée pour étudier un système aussi complexe que la production des mots.

Cette thèse essayera de disséquer cette complexité en explorant et discriminant entre des théories psycholinguistique et neurolinguistique de la production des mots. Elle évaluera comment les données de neuroimagerie, et en particulier de MEG, sont capables de contribuer au débat. Dans ce contexte, elle discutera de l'élaboration et implémentation d'une étude empirique en MEG, et de l'apport de nos résultats aux connaissances existantes.

Sommaire des chapitres:

- **Chapitre 1.** Introduction et considération des modèles psycholinguistiques et computationnels de la parole.

Dans cette première partie, on considère le développement des modèles psycholinguistiques de la parole basés sur des expériences comportementales, et la conversion de tels modèles verbaux en modèles computationnels formels. On discute ensuite de la nécessité de relier ces simulations aux réponse corticale pour justifier leurs propositions, ainsi que du meilleur moyen de munir des arguments dans le contexte du réseau cortical du langage.

- **Chapitre 2.** Revue de la littérature MEG sur la parole/la production des mots intitulé: "The cortical dynamics of speech: a review of MEG evidence from studies of language production."

Ce chapitre rassemble les arguments empiriques disponibles dans la littérature MEG concernant la production des mots, et réalise une analyse des données rapportées. Il aborde aussi la fiabilité et interprétabilité des données, discute la relation entre l'anatomie corticale et fonctionnement cognitif, et évalue comment exploiter les compétences de la MEG pour mieux décrire le système de production du langage.

- **Chapitre 3.** Aborder la problématique: identification des buts expérimentaux et du développement des procédures pour les adapter à une expérience en neuroimagerie.

Ce chapitre discute des questions théoriques et des buts de l'adaptation d'une étude comportementale au domaine de la neuroimagerie en général, et de la dénomination des images en MEG en particulier.

- **Chapitre 4.** Le développement technique d'un protocole pour étudier la production du langage en MEG: un sommaire des techniques d'enregistrement et analyses avec la MEG, ainsi que de la mise en place d'un protocole expérimental.

Ce chapitre discute des propriétés techniques de la MEG, dans la perspective de construire une nouvelle étude empirique. Il résume les étapes de conception de l'étude, l'exécution d'une étude pilote et les mises au point faites par la suite dans le protocole et le chemin d'analyse.
• **Chapitre 5.** Etude empirique de dénomination d’images intitulée: "*Early semantic and late visual differences in the cortical response to picture naming in Magnetoencephalography.*"

Ce chapitre rend compte d'une étude empirique de la réponse corticale pendant la dénomination d’images, réalisée à l'université d’Aix-Marseille, et à l'hôpital La Timone.

• **Chapitre 6.** Résumé et discussion

Ce chapitre consiste en une évaluation conjointe des modèles existants, de la revue de la littérature, de l'étude empirique, ainsi qu’une exploration des perspectives pour la poursuite potentielle de ces recherches dans l’avenir.

*References:*


1. **Introduction**

This chapter will discuss the recent history of psycholinguistic models, and their development alongside neuroimaging techniques and technology. It will go on to discuss modern models of language production, in particular the divide between the serial and parallel schools of thought as they pertain to brain function, and what neuroimaging and MEG specifically can contribute to this debate.

1.1. **Psycholinguistic models of language processing and speech production:**

Over the last half-century, thinking about language has progressed from verbal models of the network to fully implemented computational models with plausible neurological consequences. This section will use some representative (although not always archetypal) models from different periods and of different architectural types to illustrate the evolution from labelled boxes and arrows to closely defined cortical areas and neural activity at the scale of tens of milliseconds.

One of the first holistic models of speech production is proposed by Garrett (1975). In his 'Analysis of Sentence Production', he states that “somewhere in the recesses of the central nervous system” a message arises and is translated into a set of instructions that guide the articulatory apparatus. The dry wit belies frustration; Garret goes on to say that even though the vocabulary of the system is beginning to be understood, there is no formal description of the interaction between the system's components, nor the interactions between the structural types involved in the translation from cognition to expression. Despite this, his model uses elements that survive (in somewhat modified form) in many contemporary models. He describes a system of orthographic and auditory word recognition with distinct processing steps covering the processing in different receptive and expressive (or generative) situations, represented in Figure 1.
Figure 1. Adapted from Garrett (1975). A model of speech production from sensory input or internally generated intention to articulation. While the stages of processing are relatively analogous to that of more modern models, the descriptions of processes and interactions are not detailed enough to build a computational model that would permit testing empirical predictions.

Building on the basis of data collected from speech errors, Garrett's model provides an example of a localist network model, allowing parallel paths between input modalities and a generic 'memory store' which integrates these elements to call modality independent 'lemmas'. A proliferation of unidirectional arrows indicate a sequential, non-interactive flow of information through the network. While it does provide a point of departure for discussion and development, this version of Garrett's model is 'underspecified' in its description of component processes and their interactions, and does not include any attempt at mathematical or computational modelling of the system. A description at another level of detail is necessary if we are to fully understand the cognitive and cortical organisation of speech production in a precise and quantitative way.

To instantiate a verbal theoretical model, the modeller is compelled to make many specific theoretical choices, setting values and types of interaction between levels. Normally the goal is to reproduce behavioural outputs with the minimum number of operations. As Coltheart, Rastle, Perry, Langdon, & Ziegler, (2001) argue, all nodes and links in a model should each be functionally motivated, and the removal of any one should result in a failure of the model to perform a linguistic task that humans are capable of. A secondary consideration motivating simplicity of models is that as they become more expansive and inclusive, their complexity allows so many possible configurations of parameters that it is possible to arrive at similar outputs using different set-ups of the model. Without some evidence of
how these cognitive/conceptual structures are instantiated in the brain, which can be tested against the behaviour of the model, it is not possible to decide which set-up of the model describes the truth of the behaviour being modelled.

In the words of Rapp & Goldrick, (2000), “It clearly makes no sense to evaluate theoretical claims that make precise quantitative predictions with qualitative behavioural findings. It is equally senseless, however, to evaluate broad theoretical claims with detailed empirical results.” The gap between the broad theoretical claims it is possible to state using verbal models, and the level of detail and specificity necessary to replicate brain mechanisms is obviously huge, but the first step towards a true description of how language is instantiated in the brain is through computational modelling and simulation of behaviour.

1.2. Simulated models – parameterisation of verbal models:

In the speech domain, the first instances of such simulations were of single component elements of the network rather than multiple integrated steps. For example, following on from Garrett’s work, Dell, (1986) used data from speech errors to produce and evaluate a computer-simulated model of the phonological encoding process, assembling morphemes into phonological units. This work was later developed into the broader ‘interactive two-step’ model of lexical access (Gary S. Dell, Burger, & Svec, 1997; Gary S. Dell & O’Seaghdha, 1992), which accounts for the process of retrieving a word from its associated semantic properties, and subsequently identifying the phonological information necessary to produce it. Roelofs, (1992) proposed the WEAVER model of word production covering “…the activation of lexical concepts, the selection of lemmas, the morphological and phonological encoding of a word in its prosodic context, and the word’s phonetic encoding”, and producing a computational model of the same. (This was later developed by Levelt, Roelofs, & Meyer, (1999) into the LRM model, and further refined by Indefrey and Levelt (2004) into the I&L model.)

Other descriptions followed: (Coltheart et al., 2001) constructed a computational version of the Dual-Route Cascaded [DRC] model, the verbal version of which is graphically reproduced in Figure 2. This model comprises two routes to reading aloud from print, a non-lexical and a lexical. The first of these is a direct grapheme-phoneme transform, in which letters are identified according to the presence or absence of certain features, and then assembled into a graphemic form, from which appropriate phonemes are serially generated to produce the word’s (or pseudo-word’s) phonological form, which can then be articulated once a threshold of activity has been attained. This is referred to as the sub-lexical route, as no single, semantically meaningful form need be activated for articulation to occur. The other route is the lexical or direct pathway, in which a holistic orthographic representation belonging to a known word is recognised, and the phonology can be directly accessed from a stored trace or pattern in the phonological level. This pathway results in quicker pronunciation and implies access to semantic information about the spoken word in healthy speakers. The modelled system would also account for picture naming by allowing the semantic system to activate both the orthographic and phonological levels on the basis of the semantic properties of a pictured item.

To illustrate how quickly the complexity of a model can multiply, let us look at the sub-lexical (non-semantic) route of word selection. Each level of processing comprises multiple units of an appropriate kind, changes in a unit's activity at each time-point being governed by 3 factors: “the unit's previous
activation, how quickly this activation decays, and what the net input of activation into the unit from other units is”. In the 2001 version of this model, there are 14 feature-present and corresponding feature-absent units that identify a letter unit in each of 8 positions, which in turn pass activation to 7,981 orthographic lexical units, each one of which has a frequency weighting, high-frequency units being activated more quickly than low-frequency ones.

Figure 2. A schematic adapted from the Dual-Route Cascaded model of visual word recognition and reading aloud modelled in Coltheart et al. (2001). In this model, activity cascades through units in lexical and non-lexical routes between letter units and a phonemic processing function, the non- or sub-lexical orthography-to-phonology route being a direct connection that bypasses semantic processing.

For just one stage of one of the direct grapheme-phoneme route, the simpler of the two routes in this model, there are 28 feature present/absent units, 8 letter positions, 3 factors governing unit activation state from one time point to the next (previous state, decay rate, net input), and 7,981 graphemic/phonological lexical units, the whole model having 31 separate parameters controlling many thousands of units at different levels. It is easy to see how the potential for different setups to give similar results might is inherent in such a complex system. This is before considering the effect that other less straightforward variables like priming and hysteresis (“the time-based dependence of a
Coltheart and his co-authors are acutely aware of this problem, stating “it is true that a mathematical model with 31 parameters can fit perfectly an extremely complex data set - any empirical function with up to 31 points in it.” before going on to point out at that their parameters are constrained by a minimal functional architecture, itself motivated by empirical data.

1.3. Simulated models - complexity and the question of parallel/serial functionality:

Many different, equally valid models of language production and reading exist, some divided by only subtle differences in the flow of activity between conceptual boxes or stages, in how decision points are reached (or if they are present at all), directions of information flow, the number and type of boxes, stages etc. Several models have been refined to the point that they are capable of simulating human behaviour in word selection, error patterns, and the consequences of lesions. While there are many shared ideas between such current models, an important divide exists between the serial/cascading and parallel/interactive schools of thought. Taking the task of picture naming as an example here, logic would seem to dictate a serial order of function; visual information is processed to isolate recognisable semantic features, which identify the concept of a particular object. This concept is related to a word that has a phonological form, which must be retrieved before it can be encoded into an articulatory gesture. As neat and appealing as this sequence may be, a single picture naming trial (as with all events in our lives) does not exist in isolation. Experimental participants are aware of the required task, have biases from previous experiences (like the preceding trial) and all areas in the densely connected brain may have access to and even influence upon perceptual information at the same time. This conceptual difference motivates many empirical neuroimaging studies, and is a key topic in this thesis. How is it possible then, to use behavioural and neuroimaging data to distinguish between the plausible and the true explanation of human language processing? How do functional models from the different schools of thought describe the system, and what are the critical differences between them? Again, we will illustrate these questions with representative examples in an attempt to identify their distinguishing features.

The Parallel Distributed Processing [PDP] model of language processing and speech production presented by Harm and Seidenberg (2004) is based on Seidenberg and McClelland’s “triangle” model (Seidenberg & McClelland, 1989), and does not feature separate routes for lexical and non-lexical utterances. Indeed, there are no entries corresponding to individual words, nor fixed transforms for converting graphemes to phonemic equivalents. Instead, the phonology and pronunciation or reading from print is generated using a network of weighted connections between orthographic, phonological, semantic and contextual levels of representation, where weights are set using back-propagation of error (Rumelhart, Hinton, & Williams, 1986), intended to parallel human experience with the spelling/sound/meaning correspondences observed in the language being learned. Rather than having a lexical or lemma unit that indexes the location or ‘code’ of a particular word’s phonological, orthographical, grammatical and other information, this type of model has 'hidden units' that mediate the relationships between the codes in different modalities of processing.
Harm and Seidenberg's (2004) reading model operates on a similar level of complexity as the DRC model previously outlined; 10 letter slots could house 'features' corresponding to letters of the alphabet, giving 111 possible orthographic units derived from the training set of 6,103 words, with 500 hidden units mapping the relationship between these and 1,989 semantic features, and another 100 hidden units mediating between the orthographic level and 200 phonological units. These level specific and between-level hidden units are all activate simultaneously, settling into an 'attractor state' of activity that embodies a network-wide gestalt representation of the target word. The weights of connections between units are not directly specified, as in the DRC model, but 'learned' by the system through hundreds of thousands of training trials on the test set of stimulus words. The authors took the fundamental design decision to make all the connections and units function under the same set of behavioural rules, and not specify different parameters for units at different levels or the connections between them, stating: “There is no independent evidence, for example, that the different brain structures that support orthography to phonology conversion and phonology to semantics conversion,
respectively, have intrinsically different computational properties (e.g., temporal dynamics).” (Harm and Seidenberg, 2004)

Rapp and Goldrick (2000) exhaustively explore the possibilities of interaction and cascading between semantic, lexical and phonological levels of a language model. They set out 5 increasingly interactive models using these levels of representation, with some additions at higher levels of interactivity. From least to most interactive these are:

- The discrete feedforward account [DFA]
  - feedforward only semantic → lemma → phonology
  - strictly serial processing in only one stage at a time
  - only selected items passed from one stage to the next
  - selection points at the end of each processing stage

- The cascading feedforward account [CFA]
  - feedforward only
  - activation cascades between stages before selection point
  - activation from all items active can cascade forward

- The restricted interaction account [RIA]
  - cascading and selection as in CFA
  - backward flow of activation from phonology to lemma level

- The high interaction account [HIA]
  - cascading, selection, backward phonological → lemma feedback as in RIA
  - feedback from lemma to semantics

- The further interaction, low seriality account [FILSA]
  - cascading, selection and feedback as in HIA
  - addition of visual feature input domain

The parameters of these different model types were extensively tuned with the intention of producing outputs that matched the normal rates of error in healthy speakers, and in which simulated 'lesions' interrupting transmission of information between levels could reproduce highly specific error patterns observed in four post-stroke patients. Of the 5 network types employed, only 2 (the RIA and HIA) were capable of fulfilling these criteria, the results of their simulations being laid out in Table 1. They show that a certain degree of interactivity and feedback from later levels to earlier is necessary to replicate human patterns of behaviour and post-lesion pathology, but that reducing seriality in the system destroys this ability. In this series of simulations then, it is again shown that multiple valid simulations are possible with different parameter settings.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DFA</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>CFA</td>
<td>✓ Only for errors arising at the phoneme level</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>RIA</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>HIA</td>
<td>✓</td>
<td>✓</td>
<td>✓ Need to limit L-level to concept-level feedback</td>
<td>✓ Need to limit phoneme to Lemma-level feedback</td>
<td>✓ Need to limit L-level to concept-level feedback</td>
<td>✓</td>
</tr>
<tr>
<td>FILSA</td>
<td>✓ Not tested but likely to be</td>
<td>✓ Not tested but likely to be</td>
<td>x</td>
<td>x</td>
<td>✓ Not tested but likely to be</td>
<td>✓ Not tested but likely to be</td>
</tr>
</tbody>
</table>

Table 1. A summary of the capabilities of differently interactive setups in a two-step model of lexical and phonological selection. Adapted from Rapp and Goldrick (2000). DFA = discrete feedforward account; CFA = cascading feedforward account; RIA = restricted interaction account; HIA = high interaction account; FILSA = further interaction, low seriality account.

While the DRC and PDP models as well as those variations explored by Rapp and Goldrick (2000) are based on fundamentally different architectures and principles, they arrive at the correct selection of a word. Each uses many mathematical parameters with finely tuned characteristics to be specified and modified. The models discussed are well motivated and employ sophisticated design based on empirical evidence in their attempts to reproduce behavioural results, but are purely mathematical in what they model. This is to say that none model physiological correlates of the cognitive processes (but see Pulvermüller, 2010, 2012 for an in-depth consideration of how syntax and semantics could be physiologically implemented). This is a crucial point: even if a model's output perfectly mirrors human behaviour, it is impossible to be sure that it directly reproduces the cortical/cognitive organisation of language production without neuro-biological evidence. If two fundamentally models can be tuned to reproduce the same results, how can one decide which is a better representation of the human model of language without some insight into how the human neural machinery works?
1.4. **Changing perspectives: the evolution of psycholinguistics as a consequence of developments in neuroimaging:**

Neuroimaging has added an impressive array of tools to the psychologist's arsenal. After decades of work fractionating response times, extrapolating from clinical evidence and building databases of error rates and type, EEG studies of language became more commonly used, and PET language studies began to appear. Information about the timing and anatomical location of different functions began to be incorporated into arguments between different psychological schools of thought. Until this era, theories of language needed primarily to be consistent with behavioural results to be valid, but neuroimaging introduced another layer of constraints on models. In order for models to be a true representation of the human language system, they must simultaneously be compatible with the behavioural and neurological evidence.

The work of correlating cognitive function with cortical activity in language took a leap forward in the 1800’s with the clinical study of patients whose lesions caused speech deficits. Another epoch began with a set of experiments by Petersen and his colleagues measuring cerebral blood flow (Petersen, Fox, Posner, Mintun, & Raichle, 1989; Petersen, S. E., Fox, P. T., Posner, M., Mintun, M., Raichle, 1988). Their word demonstrated different active areas for visual and auditory presentations of the same single words, supporting the contemporary idea that the different modalities have different processing streams (Seidenberg & McClelland, 1989) as opposed to a single representation activated in all cases.

With access to the data about brain dynamics provided by non-invasive neuroimaging techniques, researchers could find direct biological evidence to support or confound proposed models. On this basis, anatomical or temporal separation of elements of the brain response can be argued to imply the separation of cognitive functions.

This use of neuroimaging evidence as a way of corroborating models, and finding direct correlates for functions in brain uses the logic of forward inference (Henson, 2006). Researchers make the assumption that differing patterns of brain activity imply psychological dissociations between the stimuli tested. The logic of forward inference also implies that different tasks which activate similar or overlapping brain areas share some functional element managed by those areas. Once a cortical area has been associated with a function through forward inference (brain area A is active when function Z is required) it is possible on the common but perilous basis of reverse inference (Poldrack, 2006) to postulate that activity in this area means that function is being performed (area A was active, so function Z must be required). This can lead to the description of whole networks of cortical areas on the basis of allocations of functions from previous studies, without ever testing the functions concerned in an experiment. This does require that brain regions support only one function, which may not be the case for some regions (Binder, 2006), for example the Left Inferior Frontal Gyrus [LIFG] which has been associated with phonological processing (Costafreda et al., 2006; Klein et al., 2014; Wheat, Cornelissen, Frost, & Hansen, 2010), semantic processing (Clarke, Taylor, & Tyler, 2011; Hirshorn & Thompson-Schill, 2006; Wang et al., 2012) and grammatical processing (Sahin, Pinker, Cash, Schomer, & Halgren, 2009).

Allocating cortical activity to a certain function on the basis of prior findings is common practice, but even less logically sound attributions are also attempted. For example, in the speech production
domain attempts have been made to assign function on basis of post-stimulus peak latency instead of a manipulation of stimulus properties (Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998). The logic here is that the peak of neural activity coincides with the crucial involvement of the cognitive function, which has not been unequivocally established. Maintaining a local representation and communicating with both upstream and downstream areas may require more neural ‘power’ than local processing of a stimulus or selecting a response among alternatives.

An alternative approach to relying on the cycle of forward and reverse inference is to map the entire response to a task in the whole brain when participants perform a task. This approach has been widely employed; presenting participants with faces (Liu, Harris, & Kanwisher, 2002), words (Cornelissen, Tarkiainen, Helenius, & Salmelin, 2003; Tarkiainen, Cornelissen, & Salmelin, 2002), and more complex tasks like reading phrases (Gehrig, Wibral, Arnold, & Kell, 2012) or naming celebrities (Resnik, Bradbury, Barnes, & Leff, 2014) and charting the response in space and time for both evoked and induced power has provided network-wide descriptions for many tasks. The accumulation of data across many studies of this type can confirm the individual correlations found between area and function, and hypotheses formed on this basis can then be tested with specific manipulations.

Performing analyses on the scale of the whole brain brings into play technical challenges not encountered in region-of-interest studies: in comparing the between the ‘active’ period of interest and a ‘passive’ baseline, or between conditions at high spatial and temporal resolution one encounters the problem of false positives. MEG commonly uses models of the cortical surface made up of ~15,000 vertices, each sampled many hundreds of times across an experimental epoch. The standard rates for detecting Type I errors are clearly not appropriate, as beautifully illustrated by Bennett’s socially sympathetic dead salmon. In this light-hearted demonstration, a post-mortem fish (a North Atlantic Salmo Salar) was asked to determine what emotion was portrayed in pictures of humans. Between conditions t-tests uncorrected for multiple comparison showed significant differences in cerebral blood flow in 16 of 8064 voxels between conditions, which disappeared when a false discovery rate correction was applied (Bennett, Baird, Miller, & Wolford, 2011). If we accept that the ex-salmon had indeed gone to meet its maker, we allow that exploratory MEG studies must control for multiple comparisons even more rigorously, or risk presenting un-replicable findings.

When properly applied, however, exploratory MEG studies reveal changes in activity and connectivity across the brain with high temporal resolution. This is particularly important in the language domain, as the network of areas supporting language is large and widely distributed and the timing of activity in the network is necessary to relate activity to function, and discriminate between competing models. The interplay between exploratory and hypothesis-driven studies over the last 20 years has seen the birth of a network-level understanding of the brain. From Petersen et al’s identification in PET of separate cortical orthographical and phonological representations of the same words on a timescale of minutes (Petersen, Fox, Posner, Mintun, & Raichle, 1988), we have progressed to network connectivity analysis at the sub-second level (Clarke et al., 2011; Gehrig et al., 2012; Honey et al., 2009; Pollonini et al., 2010; Stam, 2004), to observing the resting state of various networks vs. their task-active state (de Pasquale et al., 2010; Greicius, Supekar, Menon, & Dougherty, 2009; Zhao et al., 2011) and cluster pattern analysis reflecting specific states of mind (Manning, Sperling, Sharan, Rosenberg, & Kahana, 2012). The vision of the brain as a variably interacting network of multi-functional areas with multiple potential states of connectivity has begun to dominate over the schematics of single-operation boxes and fixed-function arrows presented in the earlier verbal models.
Psycholinguistics has evolved alongside technical developments: models can now use cortical function as a guiding principle. Imaging studies are less frequently designed to evidence psycholinguistic models, so much as models are adapted to explain the results of imaging studies. Indefrey's and Salmelin's reviews (Indefrey, 2011, 2013; Salmelin, 2007) of the MEG literature on language show how models adapt to take the empirical evidence into account, and examples like Giraud’s work on the preference for left- (28–40 Hz) and right-hemisphere (3–6 Hz ) lateral Heschl’s gyri demonstrates another way in which biological evidence constrain the boundaries of psycholinguistic models, supporting as it does the physical separation of syllable and phoneme processing in the brain.

Neuroimaging provides a means for gathering evidence on this issue; beginning with positron emission tomography [PET] studies of single word processing in the 1980's, and growing with the popularity of fMRI in the 1990's (Binder et al., 1997; Buckner, Raichle, & Petersen, 1995; Fiez & Petersen, 1998; Mazoyer et al., 1993; McCarthy, Blamire, Rothman, Gruetter, & Shulman, 1993; Petersen et al., 1989; Petersen, S. E., Fox, P. T., Posner, M., Mintun, M., Raichle, 1988), a body of neuroimaging work has associated local changes in the brain's response to stimulus with manipulations of particular components of linguistic stimuli (see Price, (2010, 2012) for recent reviews of the haemodynamic literature).

In the production domain, attempts have been made to probe the relative timing of processing for component processes of language processing using behavioural measures such as picture-word interference (e.g. Schriefers, Meyer, & Levelt, 1990). In this paradigm, a semantically or phonologically related distractor word is presented (either read or heard) while a participant is preparing to name a picture. Semantic distractors have a stronger disruptive, delaying effect when presented at shorter stimulus onset asynchrony (SOA) than phonetically related distractors (Indefrey & Levelt, 2004), suggesting that semantic processing is taking place prior to phonological selection. Such ideas are attractive, but cannot in themselves displace the problem of equally valid parallel and cascading networks; perhaps motor programming of articulation (the final stage in both parallel and cascading models) is implicated in the perception of heard or read words and is vulnerable to the presentation of articulatory gestural neighbours.

In this way, the latency and characteristics of their function in various circumstances might shed light on the order and type of their interactions. Given that the normal response times [RT's] for picture naming are in the range of 600ms-1000ms, and RT's for single word reading are quicker by around 200ms (Bajo, 1988; Glaser & Glaser, 1989; Glaser, 1992), fMRI and even more so PET do not provide the temporal resolution necessary to investigate the dynamics of such short-lived processes. Fortunately, evidence from EEG, MEG and intracranial recordings with far superior temporal resolution can also be brought to bear on this problem. Of these three techniques, MEG is the best suited to exploring the dynamics of the language network as a whole-brain system, as it offers better spatial resolution than EEG (Purves et al., 2008, chp. 3), and greater coverage than can be offered by most intra-cortical or cortical surface electrode recordings, which usually probe only relatively small sections in the brains of pre-operative patients (Llorens, Trébuchon, Liégeois-Chauvel, & Alario, 2011).
1.5. *Time, space and neurons: instantiations of psycholinguistic models in the brain:*

We have considered the ongoing process of developing verbally described psycholinguistic models built on the basis of behavioural evidence into computational models, and the necessity for neuro-physiological markers as a decisive factor in deciding between potentially viable alternatives. In this context, the next chapter of this thesis will review and evaluate the available MEG evidence, assessing the extent to which the evidence supports one of the available viewpoints regarding the cascading or parallel nature of the involvement of the component processes. Before going on to do so, we will first explore some current attempts to integrate cognitive psychological models with neuroimaging data.

The WEAVER/LRM/I&L model, as previously discussed, was first outlined by Roelofs, (1992) as a psycholinguistic model based on reaction time data (as opposed to speech errors). It was later instantiated by Levelt et al. (1999) as a full computational simulation, and subsequently attempted to be integrated with evidence from a multi-modal meta-analysis of physiological (i.e. neuroimaging and patient) and behavioural data (P Indefrey & Levelt, 2000; Peter Indefrey & Levelt, 2004; Peter Indefrey, 2011). The model deals with word reading, word listening and word production, using picture naming as the example protocol for the latter. The stages incorporated in the model in the context of word production (i.e. picture naming) are: Conceptual preparation (from stimulus presentation to selection of target concept), lemma retrieval, phonological code retrieval, syllabification, and phonetic encoding up until the onset of articulation. These stages are laid out in Figure 4, leading from the presentation of a picture stimulus to the moment before the onset of articulation when the articulatory program has been elaborated and is ready to be spoken.
The model assumes that the word production and perception networks are shared for the lemma and conceptual strata, and states that phonological codes are linked, but different for perception and production. Because production requires the opposite spread of information from one level to another, (i.e. from concept to lemma) connections between these two strata must be bi-directional. A lemma activated by perceptual input spreads its activation to the corresponding lexical concept, and the reverse is true for production. Activation/information flow in the case of picture naming thus proceeds from stimulus presentation to articulation through the following stages:

**Figure 4**. The sequence of conceptual stages and transformative processes involved in picture naming, according to the WEAVER/Levelt, Roelofs, Meyer [LRM]/Indefrey and Levelt [I&L] model of language processing. Adapted from Indefrey and Levelt (2004).
1. Conceptual preparation: 200ms @ 0ms post-stimulus.
   - The visual recognition of a picture leads to the activation of multiple concepts at different levels of information. A picture of a sheep activates the idea of that specific animal [SHEEP], but may activate [GOAT] as a neighbour at the one level of information, and the super-ordinate concepts of [MAMMAL] and [ANIMAL]. Depending on the context, the correct concept is selected from competing alternatives. Based on behavioural studies such as Jescheniak and Levelt (1994), and Thorpe, Fize and Marlot (1996), who used a word/picture matching task, and Schmitt, Munte and Kutas (2000), who used a go/no go task dependent on a conceptual judgement, the I&L model estimates the time required for conceptual access as being around 175ms, but Indefrey (2011) updates this to 200ms, specifying that the process of accessing conceptual information is ongoing even after that which is relevant for accessing the appropriate lexical entry is available to the speaker.

2. Lemma retrieval & selection: 75ms @ 200ms post-stimulus.
   - Activation feeds directly from each the activated node at the conceptual level to a specific corresponding node at the lemma level. The selection of this lemma node may be faster or slower depending on the number of co-active nodes and their level of excitation, but for the purpose of their description an estimate of approximately 75ms is given, based on an ERP study by Schmitt, Schiltz, Zaake, Kutas, and Münte (2001). This study employed a go/no go task based on conceptual (an item’s weight) or lexical (grammatical gender) properties to elicit a button press. They showed an N200 response relating to the response decision, peaking at 477ms for the conceptually based decisions and 550ms for the lexically based decisions.

3. Phonological code retrieval: 20ms/phoneme @ 275ms post-stimulus.
   - Once the lemma has been selected, the segments of its phonological structure (in fact, all phonological codes for all morphemes of the word) are activated. Based on a lateralized readiness potential experiment by van Turennout, Hagoort, & Brown (1997) suggesting that the first segment is available 40ms after syntactic gender information, Indefrey and Levelt (2004) give an estimate of 80ms post lemma selection for the availability of the first phonological segment of a word, and Indefrey (2011) specifies an onset of 275ms post-stimulus for the process of phonological segment retrieval.

4. Syllabification/phonological encoding: 50ms to 55ms/syllable @ 355ms post-stimulus
   - After retrieving and selecting the appropriate phonological segments, the process of clustering the segments together into syllabic patterns begins. In the LRM version (Indefrey and Levelt 2004) of the model, the syllabification of a word is not fixed, as a lemma is a unit, but instead constructed 'on-line' and incrementally. A speed of about 25ms per segment is suggested for this process.

5. Phonetic encoding: 145ms @ 455ms
   - As syllables are created, they are converted into 'motor action instructions'. A lexicon of some few hundred high frequency syllables is stored in the 'mental-syllabary', but less usual constructs are 'programmed' or created on demand. There is no estimate given for the speed of this transformation process in the 2004 description, as the authors have no other chronometric measure of phonetic encoding, and only the first
syllable need be ready for articulation to begin.

6. Articulation: @ 600ms

- The onset of articulation marks to completion of the stream for the LRM model, which does not describe the processes involved in motor control.

Self-monitoring is also discussed in this model, but the specific processes involved are difficult to separate from the other elements of speech production. Indefrey and Levelt (2004) place a limit of 25ms after the retrieval of the phonological code in the response (i.e. 355ms post stimulus presentation in picture naming) for the beginning of self-monitoring, with no limit on the offset, as self-monitoring continues up to the end of articulation, and even after that, given that errors can be perceived and corrected post utterance. In the LRM model, there exist two feedback loops, the external and the internal. The external loop involves the activation of the corresponding phonological codes and segments, and thus the lemma of the spoken word via a speaker listening to their own output, while the internal loop feeds phonological (pre syllabification) words into the perceptual network. This estimate is questionable depending on what the definition of 'self-monitoring processes' includes. Oppenheim and Dell (2008), for example, show that inner (imagined) speech is susceptible to errors, so articulation and feedback via the external loop are not necessary. It may be the case that some error-monitoring processes are domain general, and thus active for recognition as well as production, or that self-monitoring processes are ongoing even during the 'rest' or pre-stimulus phase of a task, evaluating past performance. A strong definition is needed to be able to separate and identify self-monitoring components, which could then be quantified using experimental manipulation. Such evidence is not discussed in the LRM/I&L model.

This model was originally informed by Levelt et al's (1999) performance of an extensive review of data behavioural studies. Thus they developed alongside their model a cognitive chronometry of its component processes. These ideas are developed by Indefrey and Levelt (2000, 2004), and Indefrey (2011) who use evidence from neuroimaging studies to associate cognitive functions with cortical areas, in an attempt to correlate the dynamics of the cortical response in these functional areas with the temporal predictions of their model. The functional-anatomical associations made by Indefrey and Levelt (2000, 2004), and refined by Indefrey (2011) for what they term the “core areas” in word production are summarised in the map reproduced as Figure 5, where we see a sequence of areas moving generally forward from occipital cortex (visual recognition) through temporal and parietal areas (lemma retrieval and selection and phonological processing respectively) and into inferior frontal and motor areas (where syllabification and articulation are controlled). These areas, their functions and the time-window of their involvement according to Indefrey are also presented in Table 2. Interestingly, while Indefrey and Levelt (2004) discuss some right hemisphere areas, and some are also mentioned in Indefrey's 2011 paper, the right hemisphere is not represented in the graphical map included in the later update (Peter Indefrey, 2011).
Indefrey and Levelt’s account is a serial, cascading description of the cognitive architecture of word production, and the clear statements they make about the relationship between cognition and brain function make it possible to search the empirical data for supporting evidence, and to contrast it with other competing theories. Hickok and colleagues (Hickok, Houde, & Rong, 2011; Hickok & Poeppel, 2007; Hickok, 2012) present such an account with the Hierarchical State Feedback Control [HSFC] model, which presents a more parallel or connectionist view, more in the tradition of the PDP school of thought, or Rapp and Goldrick’s (2000) high interaction account and further interaction, low seriality accounts.

In the HSFC model, Hickok draws from the motor control and psycholinguistic perspectives, proposing a model in which a speech action is decided upon by the conjunction of conceptual, lexical, and phonological systems, while the motor system issues action commands to muscles and an internal forward model or 'effference copy' of the intended action and its expected somatosensory and auditory consequences. This efference copy is generated by the motor system after the motor plan has been decided upon, and is compared against proprioceptive and heard feedback as perceived by the various senses. The efference signal from the motor system is also transformed into an inhibitory signal going to the auditory target, so that auditory feedback matching the prediction 'cancels out', but
discrepancies between prediction and perception will cause strong activation, which is used to correct errors and guide unintended movements and utterances back to their planned course.

As shown in Figure 6, the system comprises distinct but interacting layers of function; Hickok (2012) treats the conceptual and word (or lemma) stages as being distinct from but interactive with a network of areas that control phonological and articulatory processing through parallel activity. Syllables and phonemes are articulatory goals managed at two distinct levels, syllables being auditory targets that form an open-close cycle of mandibular articulation, which decompose into articulatory feature clusters [AFCs]/phonemes forming half-cycles. The word level interposes between the phonological-motoric system and the conceptual level. The conceptual and word levels interact bi-directionally with one another, and the word level interacts with both the sensory and motor elements of the phonological-motoric articulation and feedback control network to select syllable level articulatory goals as previously mentioned. Once syllable level goals have been established, these are transformed into sequences of lower-level articulatory-feature-cluster elements. Hickok (2012) also allows that direct connections between the word level and the lower-level circuit may also exist, although these are not depicted in Figure 6.
Figure 6. Schematic of the Hierarchical State Feedback Control [HSFC] model, reproduced from Hickok's 2012 paper. The conceptual and lemma/word levels are conceived of as stages distinct from articulation and phonological processing, which is controlled by a network cortical areas managing high- (syllable) and low-level (articulatory feature cluster) motoric and sensory feedback goals.
Like Indefrey (2011), Hickok (2012) makes clear claims about the correspondence between cortical structures and specific functions in his model. Although Hickok’s focus is more on the phonological and articulatory elements than at the lemma/word level, there is still a good amount of functional and anatomical overlap in the claims made by the two authors, as illustrated in Table 2, where the areas mentioned in both models are laid out with their attributed functions.

The HSFC and I&L models are products of different approaches, the former being a parallel model from a connectionist, neuropsychological perspective, and the latter being a serial, cascading model based in psycholinguistics and behavioural psychology. They share some important similarities despite their

<table>
<thead>
<tr>
<th>Area</th>
<th>HSFC (Hickok) Function</th>
<th>I&amp;L Function</th>
<th>I&amp;L onset</th>
<th>Indefrey (2011) MEG peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left thalamus</td>
<td>N.A.</td>
<td>Self-monitoring</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Left mid/post. superior temporal gyrus</td>
<td>Syllable level auditory feedback</td>
<td>Self-monitoring</td>
<td>N.A.</td>
<td>371</td>
</tr>
<tr>
<td>Mid-middle temporal gyrus</td>
<td>N.A.</td>
<td>Conceptually driven lemma retrieval</td>
<td>200</td>
<td>190</td>
</tr>
<tr>
<td>Left posterior middle temporal gyrus</td>
<td>N.A.</td>
<td>Phonological code retrieval</td>
<td>275</td>
<td>360</td>
</tr>
<tr>
<td>Right mid superior temporal gyrus</td>
<td>N.A.</td>
<td>Self-monitoring</td>
<td>N.A.</td>
<td>400-600</td>
</tr>
<tr>
<td>Sylvian fissure at parieto-temporal boundary</td>
<td>Syllable level sensory-motor coordinate transform</td>
<td>Phonological code retrieval/[self-monitoring]</td>
<td>275/N.A.</td>
<td>300-320</td>
</tr>
<tr>
<td>Left inferior frontal gyrus</td>
<td>Motor syllable programs (BA44)</td>
<td>Syllabification</td>
<td>355</td>
<td>500</td>
</tr>
<tr>
<td>Left fusiform gyrus</td>
<td>N.A.</td>
<td>Phonetic encoding*</td>
<td>455</td>
<td>N.A.</td>
</tr>
<tr>
<td>Anterior supra-marginal gyrus</td>
<td>AFC-level somatophoneme targets</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Left anterior insula</td>
<td>N.A.</td>
<td>Articulation</td>
<td>600</td>
<td>N.A.</td>
</tr>
<tr>
<td>Post-central gyrus</td>
<td>AFC-level somatophoneme targets</td>
<td>Articulation</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>Left precentral gyrus</td>
<td>AFC-level motor-phoneme programs</td>
<td>Articulation</td>
<td>600</td>
<td>600</td>
</tr>
<tr>
<td>Supplementary motor area</td>
<td>N.A.</td>
<td>Articulation</td>
<td>600</td>
<td>400-600</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>AFC-level coordinate transform</td>
<td>Articulation</td>
<td>600</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

Table 2. Areas involved in the HSFC model and their functions according to Hickok (2011), and Indefrey and Levelt’s (2004) core areas for word production, with their theorised and median empirically observed times of activation according to Indefrey (2011).
different origins; both allow concurrent activity at the levels of self-monitoring, phonological processing, syllabification, and articulation control elements. Also, they both allow for information to flow in any direction between the conceptual, lemma and phonological levels. In the I&L model this is apparent in the reversal of information flow in reading and picture naming; in reading, phonological/lemma level information activates the conceptual representation of an item, whereas in picture naming the concept activated by the picture activates the lemma and phonology. In the HSFC model, the flow of information between phonological/articulatory levels and the lemma level is specified as being bi-directional, as it is between the lemma and conceptual levels. The key difference that could allow distinction between these two models on the basis of neuroimaging data are their different predictions about the relative onset of processing in phonology, syllabification, and articulation. In the HSFC the parallel management of the articulatory and phonological functions should mean that all sites implicated in these functions (at least within their own level of goal-management) should become active simultaneously, whereas in the I&L model, phonology is active first, then syllabification, then articulation.

Since both models are presented with a high level of specificity regarding which cortical area manages which cognitive function, comparison is possible between them on the basis of neurobiological evidence. Given that the involved sites are spread across the whole brain, MEG’s whole-brain coverage, good spatial resolution and excellent temporal resolution make studies of speech production using this technique an ideal source of such evidence. Following on from this thought, the next section of this thesis will attempt to evaluate the available MEG evidence, and interpret it in terms of existing neuro-psychological models of speech production, and existing functional-anatomical maps. By collating studies featuring compatible experimental tasks and gathering information about the onset of activity in the cortical areas implicated in these models, we will assess how well these theories fit with the observed data, how capable MEG is of contributing evidence to this investigation, and what ‘best practice’ might be in terms of relating types of MEG analysis and the decisive questions in theories of speech production. Such assessment will then guide the evaluation of the empirical part of the thesis.
1.6. References:


Experimental Psychology, Learning, Memory, and Cognition, 23(4), 787–806. doi:10.1037/0278-7393.23.4.787


2. Literature review

Title: On the cortical dynamics of word production: A review of the MEG evidence

Dashiel Munding¹, Anne-Sophie Dubarry¹,² and F.-Xavier Alario¹, *

dashiel.munding@gmail.com; asdubarry@gmail.com; Francois-Xavier.Alario@univ-amu.fr

+33 6 52 22 65 80; +33 4 13 55 09 70; +33 4 13 55 09 78

¹ Aix Marseille Université, CNRS, LPC UMR 7290
³ place Victor Hugo (Bâtiment 9, Case D)
13331 Marseille CEDEX 3 - France

2 INSERM, UMR 1106, Institut de Neurosciences des Systèmes, 13005 Marseille, France

* Corresponding author

This work was supported by the European Research Council under the European Community’s Seventh Framework Program (FP7/2007-2013 Grant agreement n°263575), and by the Brain and Language Research Institute (grants ANR-11-LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A*MIDEX) at Aix-Marseille Université. We thank the "Féderation de Recherche 3C" (Aix-Marseille Université) for institutional support.
2.1. **Abstract:**

Spoken word production has been investigated using magnetoencephalography [MEG] for 20 years. We perform a review and meta-analysis of MEG studies of speech published between the publication of picture naming study and Resnik et al’s MEG study of the tip-of-the-tongue phenomenon (Formatting Citation). We assess significant findings from source papers with the goals of gauging the spatial resolution, temporal resolution, and between-studies consistency of the cortical response during word production, and the reliability of cognitive inferences made from these findings. We use the author’s own attribution of activity to function as well as examples of serial and interactive models to scaffold our analysis. We find that despite the appearance of a generally serial trend in the dynamics of component functions of speech, no strong claims can be made about the serial or parallel nature of cognitive processes on the basis of current data. We suggest that future studies might employ standardised protocols and analyses, and report their results more comprehensively, in order to make inferring cognitive function and relating MEG data to psycholinguistic models even more efficient.

**Keywords:**

magnetoencephalography, speech, production, meta-analysis
2.2. **Introduction:**

2.2.1. **Dynamics of language production:**

A large proportion of the cognitive neuropsychology of language has been devoted to understanding the production of language (Rapp, 2001). Studies employing neuroimaging and neurophysiological techniques have tended to focus more on comprehension processes (Gernsbacher & Kaschak, 2003; Price, 2010, 2012). With these biases, or preferences, neurolinguistic research still has generated a rather clear view of the anatomical organization of language and speech production processes (e.g. Indefrey & Levelt, 2004; Price, 2010).

The availability of such anatomical-functional models fuels sustained interest in describing the underlying cortical and cognitive dynamics based on behavioural and neurophysiological data (see reviews published in 2011: Ganushchak, Christoffels, & Schiller, 2011; Indefrey, 2011; Llorens, Trébuchon, Liégeois-Chauvel, & Alario, 2011; Strijkers & Costa, 2011). Magnetoencephalography [MEG] and electroencephalography [EEG] play important roles in this debate (Indefrey, 2011; Salmelin, 2007) as they provide whole-head coverage, and of the two, MEG provides greater spatial resolution than EEG, in the order of a few millimetres rather than centimetres (Parkkonen and Salmelin, 2010; Troebinger, López, Lutti, Bestmann, & Barnes, 2014). Such spatial resolution is lower than in intracranial recordings of speech production, which also benefits from excellent temporal resolution and are free from muscular and movement noise, but cortical implantation often has sparse anatomical coverage (Llorens et al., 2011).

The intrinsic complexities of MEG signal processing notwithstanding, this technique could therefore be the method of choice for investigating the cortical and cognitive dynamics of word and speech production. As a primary example, Salmelin, Hari, Lounasmaa, and Sams (1994) published a pioneering article describing the MEG neural sources identified by contrasting naming and passive viewing of object pictures. The authors performed a stimulus locked analysis that found a set of sources including visual responses present independent of task, posterior temporal sources that were marginally weaker for passive viewing than picture naming, as well as naming specific responses in left sensorimotor and bilateral fronto-temporal areas. Further, a response locked analysis (centred on speech onset) accentuated sources in pericentral areas and at the vertex, implicating them in the actual production of speech. Twenty years from that study, we review this thread of research, collating evidence from MEG studies of word and speech production.

2.2.2. **Reviewing the evidence**

Constructing an integrative and dynamic view of cognitive and cortical processes implies integrating at least four dimensions within a single frame of reference: anatomical localization, temporal dynamics, cognitive function and signal processing assumptions must all be reconciled. A tractable strategy is to take one of these dimensions as the frame of reference (e.g. Indefrey, 2011, uses a functional cognitive hypothesis) on which to project the evidence. Here, our primary goal is to gauge the resolution and reliability of the reported data, and the reliability of the cognitive and cortical dynamics that can be inferred from it. We do not give priority to any of the dimensions mentioned, or make strong prior
assumptions about the spatial, cognitive or temporal organization of the language and speech production network. We thereby knowingly sacrifice some of the precision (e.g. fine grained analysis of experimental tasks; focus on specific regions of major previous interest) in the hope of providing a more inclusive view of the available evidence.

Our review will be broadly guided by the classic distinction between serial and parallel types of processing models (for a detailed discussion in the context of word production, see Rapp & Goldrick, 2000), and by at least three previous syntheses of the cortical-functional anatomy of language production: Hickok, 2012; Indefrey, 2011; and Price, 2012.

As an example of the serial model, Indefrey's most recent update to his model of word production (Indefrey & Levelt, 2004; Indefrey, 2011), proposes a sequence in which cortical areas concerned process component functions of picture naming. The temporal engagement of these components is described with the (realistic) assumption that behavioural latencies are in the range of 600ms.

Price's (2012) review of the PET and fMRI literature collates an enormous amount of research on speech comprehension and production, and she provides a map of the cortical structures in the left hemisphere [LH] associated with particular function from her own work. Apart from distinctions within visual and semantic functions (e.g. extra-striatal visual areas coming after primary visual vortex, see Fig. 4 of her compendious review for more details), Price (2012) does not assign a particular sequence of functional engagement. Accordingly, she presents a more parallel image of language processing (see Price, 2012; Fig 2.) but does not give a detailed description of the functional dynamics of this schema.

As an example of a more interactive model, we considered the hierarchical state feedback control [HSFC] model proposed by Hickok (2012; see also Tourville & Guenther, 2011). This model describes a network of cognitive processes and neural sites involved in speech production that act in a hierarchical interactive fashion. The HSFC has two levels of feedback control, with motor cortex feeding out to articulation, which initiates a feedback loop with acoustic and somatosensory inputs to anterior supramarginal gyrus [aSMG], primary sensory cortex [S1] and superior temporal gyrus/superior temporal sulcus [STS/STG]. Following stimulus, Hickok's model suggests simultaneous onset of activation in STG, TPJ and BA44, followed by activity in vBA6, sensory and motor cortex, cerebellum and aSMG, then by sustained activity in all these areas throughout articulation. Hickok also makes an allowance for direct connections between the lemma level and articulatory feature cluster loop, but does not show them in the graphical representation of the HSFC model in his 2012 paper. Input to the articulatory system is from the lemma level, fed by the 'conceptual system'.

The I&L model has a sequence of onsets running through a similar group of areas to the HSFC model, which proposes activity in four distinct but interacting layers, with lemma selection followed by sustained activity in the entire post-conceptual network. A third network of sites processing the component functions of speech is described by Price (2012), but she provides only an association of structure with function, not an appraisal of dynamics or sequence of onsets of activity in the areas concerned. The timing of activity in the areas described by these models is crucial to evaluating them, and as we have already discussed, MEG can provide such evidence.

2.2.3. The current meta-analysis:
We review articles from 20 years of speech production research (1994-2014) that have reported MEG activity associated with word and speech production tasks, using the specific criteria described below in the methods section. Our primary goals are to gauge the spatial resolution, temporal resolution and between-studies consistency of the reported cortical dynamics, and the reliability of the cognitive dynamics that can be inferred from the data. We summarise the evidence as interpreted in the original publications, but also use examples of serial and interactive models to guide our own treatment of the collated data. We also discuss the trends in how such studies are designed and reported, the comparability that their differences afford, and whether these practices might be improved (largely inspired by Gross et al’s guidelines for good practice in MEG studies (Gross et al., 2013)).

2.3. Methods:

2.3.1. Inclusion criteria for published articles:

This meta-analysis of studies gathers evidence from MEG studies of language production published between Salmelin et al’s (1994) picture naming study and Resnik et al’s MEG study of the tip-of-the-tongue phenomenon (2014). Published MEG studies were included in our selection if they featured source-level analysis of tasks requiring the production of speech or speech like actions, including picture naming, reading, and/or speech delayed to a secondary cue. Protocols involving delayed responses were included, while those involving only covert responses were excluded. In the latter, it is uncertain how much a participant engages with and processes the stimuli; even when asked to envisage speaking the stimulus words, participants are conscious that they will not actually perform the action. Additionally, with infrequent responses even the most highly motivated of participants can suffer from fatigue over the course of several runs and begin mind wandering. In contrast, participants performing a delayed-response task know that a response is required, and must prepare accordingly. A delayed response protocol is thus more comparable to a direct- or immediate-response protocol than a covert response, which more closely resembles silent reading.

Within these limits, we identified studies containing diverse cognitive tasks, experimental aims and techniques, methodologies and analytical procedures. The source articles all report MEG data from healthy speakers, producing one or more words in their native tongue, either in immediate response to a stimulus or after a delay and subsequent cue.

2.3.2. Inclusion and exclusion criteria for activity contrasts:

Articles matching the inclusion criteria were geared to answer their own purposes, and not all were designed to provide evidence directly relevant to our questions. To avoid distorting the others’ findings, we have tried to balance inclusivity and rigour in the way we report the included data, according to the process described below.

We relay reports of reconstructed source level activity contrasting task-relevant periods with an appropriate baseline, or contrasts between two experimental conditions. In both cases we include only differences specifically reported as being statistically significant in the authors’ tests. We also include reports of Equivalent Current Dipole [ECD] activity at the source level, based on goodness-of-fit to the
sensor level data, from some earlier studies (e.g. Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998, Sörös, Cornelissen, Laine, & Salmelin, 2003). In this synthesis we make no distinction between the methods in which these effects were found, although we later discuss their diversity. As such, significant differences arising from Beamforming, Minimum Norm Estimation [MNE] and Dipole Fitting are treated equally; Event Related Fields [ERF’s], source strength, time frequency or connectivity effects, whether versus baseline or between contrasting conditions are all treated as equally valid. We also ignore differences in the methods of statistical analysis, assuming equal validity of the (extremely) various descriptive data and inferential methods used to produce measures of significance, any activation or difference in activity described as statistically significant by the authors being treated as ‘fair game' for inclusion. A count of the identified activity contrasts is presented in later summary figures, sorted by time as well as anatomical location and cognitive function (see below for details on cognitive attribution of identified activities). If an article reports two substantially different kinds of evidence (for example ERF’s and time-frequency data) then these are considered as being “spatiotemporally complementary” (Kujala, Vartiainen, Laaksonen, & Salmelin, 2012; Liljeström, Hultén, Parkkonen, & Salmelin, 2009) rather than redundant and both activation counts are entered additively into our analysis.

From among our target papers, we exclude from graphical analysis those with an analysis window of 1 second, as they are unable to inform us about the dynamics of the cortical response during our period of interest.

2.3.3. Description of identified activity contrasts:

Direct comparability between studies is problematic for several reasons. There is variation between studies as to the nomenclature of cortical structures; one author’s temporoparietal junction is another’s inferior parietal lobule. Further, there are differences in the way authors choose points in cortical space to represent a given region of interest or broader active patch, and in the way information is reduced or compressed for reporting. For example, equivalent (spatially proximate) ECD’s in multiple participants may be averaged to a single timeseries (e.g. Saarinen, Laaksonen, Parviainen, & Salmelin, 2006), spreading areas of activity may be reported as the average timeseries across many vertices or voxels (e.g. Bar et al., 2006), or as a single timeseries from the point of peak activity within an active area (e.g. (Klein et al., 2014; Laaksonen, Kujala, Hultén, Liljeström, & Salmelin, 2012); any of these point, area, mean or max, being attributed to a specific cortical structure.

When authors apply an anatomical label to an activity, we report it in Supplementary Table S4. The variety of names applied by different authors makes it impractical to transfer this to our timecourse graphs, and thus we apply our own names according to the schematic shown in Figure 1. Where no anatomical structure is defined by the original authors, we make our own allocations under the same regime (see column B of Table S4 for the authors’ original descriptions of source location). To conform to these stated limits and what is reasonable given the spatial resolution of MEG, we restrict our own designations of activities to relatively large scale structures, rather than trying to over-specify. Where reported, standard space (MNI-152 or Talairach) co-ordinates for activities are relayed in Table S4.
2.3.4. **Timing of identified activity contrasts:**

For activities included in our analysis, we report the onset, peak and offset of significant effects provided by the original authors, treating all significant effects reported as being equally relevant. Where only a peak latency is reported, we give the activity an arbitrary duration of 100ms centred on the peak, solely for illustrative purposes in the figures.

Reported activations with an onset later than 1000ms have been excluded from the Figures, both to give a scale more consistent with readability, and because average word production latencies are around the 600-700ms mark for healthy proficient speakers. Reported activations with a duration equal to or longer than our window length of 1000ms (e.g. Gehrig, Wibral, Arnold, & Kell, 2012) have been excluded from the figures on the timings of reported activities. Following conventional logic, the timing of activities is critical to understanding how the speech network areas interact, and such long time-windows cannot contribute evidence regarding this issue. These longer latencies and windows are still present in the various tables.

2.3.5. **Attribution of identified activity contrasts to cognitive function**

In our analysis, we make two classifications of the reported activity:

1. A report of the location of each reported activity, homogenised into relatively large anatomical areas.

2. A report of the cognitive function the author(s) specifically attribute to the significant effects of their experimental manipulation(s).

We then present the same data using 3 other systems of classification, based on the anatomical-functional pairings put forward in:

1. Indefrey’s multi-modal review (Indefrey, 2011)

2. Price’s most recent review of the fMRI literature (Price, 2012)

3. Hickok’s description of the areas involved in the HSFC model (Hickok, 2012)

In building the arguments for their anatomical-functional schema, all 3 authors provide substantial justification from empirical studies. All authors consider the possibility that a cortical area may manage multiple cognitive functions, discussing diverse sources of evidence which call other functions into play and forming their arguments accordingly. However, the cognitive maps and models proposed by the three authors rely on their one allocated function being in some way represented in a given area. Such association of a cognitive function with a cortical area does not preclude the possibility that this area may be responsible for other functions in other contexts. While some studies use similar tasks, which limits the number of processes under consideration and makes it likely that an area is handling the same function, taking into account the myriad possibilities of multiple function per area would result in an unmanageable explosion of complexity, so in this review we maintain a narrow view, assuming single-function-per-area description.
2.4. **Results:**

*Original reports:*

The 17 studies listed in Table S1 conformed to our inclusion criteria. Many others furnish our discussion with complementary evidence but among those that meet our inclusion criteria, only these few studies answer the challenge of having participants articulate in an MEG scanner.

These 17 papers report 154 sources of activity. The experimental and analytical methods used in the target studies to arrive at these activations or differences are diverse. As previously explained, we disregard much of this variability; yet, not all results are readily usable in the context of the current discussion. We exclude the results from Gehrig et al., (2012), Resnik, Bradbury, Barnes, & Leff, (2014) and Ruspantini et al., (2012), as their window(s) of analysis were >=1s (see their methods) and Salmelin, R., Hari, R., Lounasmaa, O. V., and Sams., (1994) because the results from their early methods of source localisation are not easily placed in our ‘cartoon’ of cortical areas (see Fig. 1). These exclusions leave 142 activities in 14 papers as being eligible for inclusion in our meta-analysis.

*Anatomical distribution of the reported data:*

Figure 1 presents the schematic anatomical filter we have devised to classify reported activities. On the basis of this anatomical filter, we map in Figure 2 the timing of all differences (i.e. against relevant baseline or between conditions) identified in the target papers as significant, regardless of whether or not they were discussed by the original authors.
Figure 1. Regions of interest used to group activations into broad areas for homologation and graphing. Overlay applied to loci of activity described as large scale cortical structures/areas in source papers. The same layout is mirrored in the right hemisphere to account for the substantial number of activities reported there.
Figure 2: Timing of the activities identified in the target articles, sorted according to the cortical areas outlined in Fig. 1. Figure represents the activities in a stimulus that in our target is limited for open-ended. more than reported by a at a given some sources multiple types of ERF's, power in frequency baseline and conditions.

Colour number of given area at latency post-are reported articles, scale contrast, and There may be one effect single paper latency, as report of effect (e.g. oscillatory different bands, vs. between effects).
A broad bihemispheric network of areas is revealed in Fig. 2, with overlapping activities in many areas across both hemispheres. There are frequent reports of early activity in occipital cortex, largely bilateral. In certain reports, left occipital activity is sustained long enough to overlap with pericentral activity. Temporal lobe activity is typically reported from 100ms post-stimulus onwards. It is more lateralised to the left than occipital activity, and most frequently observed in left middle temporal gyrus around the 300 to 500ms time-window. In anterior temporal lobe [ATL], reports begin at 100ms (in LH), and frequency of report peaks at around 200ms (5 activations in 2 studies). The choice of time window in Klein et al., (100ms to 200ms post-stimulus) may be advancing this onset, but Clarke et al. (2011) provide very temporally precise data for activity in the ATL, sampling at 1000Hz and not integrating across time-windows, and showing active (vs. baseline) as early as 170ms for both source strength and phase locking between ATL and fusiform gyrus. Mid-temporal activity also begins at 100ms, and is sustained during most of the window of analysis, overlapping in time with activity in the pericentral areas, where sensorimotor activity is consistently reported as beginning at 300ms. Activity in temporo-parietal junction [TPJ] and superior parietal regions is often observed, and most often bilaterally. The reports point to a somewhat later onset than those of the preceding activities, in the 200ms range, although the mode of reported TPJ activity precedes, rather than trails, the mode of reported MTG activities. The inferior frontal gyrus [IFG] is active in both hemispheres, the LH very early and throughout the whole time window of analysis. The mode of IFG reported activation coincides more or less with the mode of reported activation in pericentral cortices, bilaterally. Insular activity is reported only occasionally.

The anatomical survey of reported active areas reveals a broad, bilateral network that involves visual cortex, ventral and dorsal streams and activity in sensorimotor areas. While the involvement of these systems is as might be expected, their temporal organization is not completely straightforward. There is substantial temporal overlap across regions, in particular between temporal, parietal and frontal activities. This description sets the stage for a more thorough cognitive interpretation of the activities, on the basis of the authors’ original attribution as well as through the lens of previous anatomical-functional meta-analyses.

2.4.1. Dynamics of function - the original author’s interpretation:

The source papers include a range of different experimental manipulations. Some are concerned with semantic processes (Clarke, Taylor, & Tyler, 2011; Maess, Friederici, Damian, Meyer, & Levelt, 2002), others grammatical properties (Liljeström et al., 2009; Sörös et al, 2003), phonological processing (Klein et al., 2014; Vihla, Laine, & Salmelin, 2006), or semantics (Liljeström et al., 2009; Sörös et al., 2003), while some attempt to describe word production in a more holistic manner (Maess et al., 2002). The largest group of studies (but not a majority) focuses on motor functions (e.g. Ruspantini et al., 2012; Salmelin, Schnitzler, Schmitz, & Freund, 2000). Despite this diversity, the processes identified by the authors can be grouped into broad categories corresponding to canonical stages in language processing. In Figure 3A we have compressed the diversity of authorial attribution into the following cognitive functions, conforming to the stages of processing that might be engaged by a word production task (we use the example of picture naming), other component processes of articulation is not practicable.

Visual processing → Semantics → Lexical selection → Phonology → Motor function
We exclude self-monitoring processes (internal, external or general) from this breakdown, primarily because no papers focusing on this aspect of word/speech production meet our inclusion criteria. Without any specific manipulation of self-monitoring processes, Attempting to fractionate them from other component processes of articulation is not practicable.

**Figure 3A** - Attribution of cortical activity to cognitive function on the basis of the original author’s analysis and discussion of their own results. Colour represents the number of significant activity contrasts (i.e. against relevant baseline or between conditions) reported for a particular processing stage (vertical axis), organized by latency (horizontal axis). The colour scale is capped at 7 for contrast, but is open-ended (i.e. values higher than 7 exist, but have the same colour). There may be multiple instances for a single source paper, particularly when different types of analysis are reported or when a given activity is attributed to distinct cognitive functions (e.g. semantics and phonology).

1 Tian & Poeppel (2010) present a series of MEG experiments on finger tapping and articulation (production of the syllable ‘da’) in which they interpret the results as pertaining to self-monitoring. Their analysis, however, is performed at the sensor level and thus was not included in our graphic synthesis and discussion. In this paper, they suggest that field changes recorded over parietal cortex before movement provide evidence of a kinesthetic/proprionceptive forward model. They also recorded activity over auditory cortex that they equate with motor estimation before articulation, interpreting it as anticipation of the auditory consequences of such, being a consequence of the creation of an ‘efference copy’ in a forward model of speech production.
Visual responses to stimuli appear in the earliest part of the chart, followed by many activations (the number of overlapping reports at the mode is 8) attributed to semantic processing in an early window (~180ms to ~250ms) after stimulus onset. In contrast to this, there is a relatively low frequency of reports attributing activity to lexical selection, in part due to the low proportion of papers that target this stage in particular. Lexical selection is most frequently reported as taking place just after 200ms (Levelt et al., 1998; Maess et al., 2002; Sörös et al., 2003). This is after the mode of visual processing attribution, but is at roughly the same point as the mode for semantic activity (circa 200ms). Notice however that the mode does not indicate a particular moment of transition in the process, and that spread of attribution over time to these three processing levels broadly overlaps (~180ms to ~250ms). Some of this overlap may be due to the time-windows reported. Fixed lengths and latencies of time window could pick up later visual activity and early lexical access and report them as significant for that period, although the functional responses might have different temporal characteristics. Even taking this and other data-processing caveats into account, the ensemble of activity is not indicative of serial processing stages but rather suggestive of concurrent or a strong temporally overlapping cascade of processes. In addition, there are 5 activities from 4 different articles attributed purely to semantic activity after 300ms post-stimulus, constituting a resurgence of semantic function lasting until as late as 875ms post-stimulus in one study. This is also consistent with overlapping stages of processing, perhaps interpretable in a connectionist context as a consequence of maintaining a holistic representation of the word in preparation for and during its utterance, or as a consequence of cascading reactivation of semantic/lemma information by the creation of a forward model for expected auditory input activating the phonological form of the word in a sequential model.
Attribution of activity to phonological processes is frequent, particularly in the period between 300ms and 600ms post-stimulus. The mode of attribution is around 350ms with 10 attributions from 6 different studies. Motor function, which comprises here both motor planning and articulation, is commonly reported in a window largely overlapping that of phonological activity, with two modes: the first essentially overlapping with that of phonology, and the second occurring rather late in the window, around 850ms. Again, these observations are consistent with the idea that the bulk of phonological and motor processes occur concurrently or have extensive temporal overlap, rather than having serial involvement. Additional early phonological attributions (2 observations around onset of stimulus) and late motor attributions (the second mode of motor function) might be linked to specifics of the involved studies. The early activity in the left inferior frontal gyrus [LIFG] comes from a study using a reading task (Klein et al., 2014) in which phonological access from orthography might differ in timing from access in other tasks. The later peak of motor activity relates mostly to the suppression of the 20Hz Rolandic rhythm in the hand and mouth motor areas reported by Salmelin et al (2000). This study involved delayed reading, stimulus words being followed by a question mark as a speech prompt 800ms later.

It is tempting to further interpret specific features of the counts we report, such as their onsets, offsets or the temporal sequence of modal attribution to a given function. For example, the periods of frequently reported effects (> 5 reports) follow the sequence of function mentioned earlier, flowing from early visual processing to late motor function in good time for the normal reaction times for single word production (around 500ms for single word reading and 600ms for picture naming). However, the offsets of these functions are not sequential and there are many reports outside the most frequently reported windows. For these reasons, a global description of the evidence, if less fine-grained, is preferable to a description that focuses on specific features of the time course.

The general trend visible in the pattern of authors’ attributions of activity to function is an overlapping cascade comprising two broad chunks of processes. Between 0 and 300ms, the graph is dominated by reports of visual, semantic and lexical activity. Only after that are phonological and motor processing reliably observed, in the period between 300ms and the end of our analysis window at 1000ms. These ideas have been mentioned in previous reports and models of the language production network using different experimental tasks and recording techniques (e.g. Hickok & Poeppel, 2007; Llorens et al., 2011; Strijkers & Costa, 2011, among others) and might perhaps be linked to earlier descriptions of word retrieval failures (Badecker, Miozzo, & Zanuttini, 1995; Henaff Gonon, Bruckert, & Michel, 1989). The current analysis gauges the reliability of such conclusions within the MEG evidence.

The sustained visual activity reported (note the resurgence of reports after 600ms) may represent an ongoing process of ‘rehearsal’, i.e. the maintenance of a visual image or semantic concept (Courtney, Ungerleider, Keil, & Haxby, 1997; Malecki, Stallforth, Heipertz, Lavie, & Duzel, 2009; Tallon-Baudry, Kreiter, & Bertrand, 1999). In our source papers, the LH is reportedly more active than the right, but this may be symptomatic of a priori assumptions affecting the search for, and reporting of, significant sources of activity.

In addition to this preliminary conclusion, it is worth noting that there are many activities reported by the authors but not discussed in detail, or ascribed by them to a particular cognitive function. In our count, as many as one third of reported activities (52 out of 143, i.e. 36%) did not receive a specific cognitive interpretation. As a reminder, these activities are present in the reports (tables, figures, or
text) as per the inclusion criteria described in the methods section, our contribution being only to compile them here. The time course of unattributed activities is visible in Fig. 3B running from 0-800ms post stimulus. The reasons for not interpreting these significant activities in the original reports may be diverse, including: focus of the study elsewhere, unexpected finding with no clear interpretation, lack of space and / or synthetic report, activity deemed less relevant, etc. The current review allows going beyond the original reports by interpreting these activities across studies and adding them to the “main” findings.

To interpret this previously undiscussed material, we attempted to allocate function to the reported areas on the basis of recent influential reviews of the neuroimaging literature: Indefrey’s review (Indefrey, 2011), which covers clinical, fMRI, EEG and MEG data; Price’s later review (Price, 2012) of the PET and fMRI literature, and Hickok’s HSFC model (Hickok, 2012). Our interpretation is guided by the relationship between anatomical structures and cognitive functions described in these reviews, irrespective of claims about timing in the case of Indefrey (2011). Thus, for example the early difference in IFG found by Klein et al (2014), which they attribute to phonological priming, would be to do with syllabification according to Indefrey’s account, and semantics according to Price.

In projecting MEG activities onto an fMRI anatomical-functional map, we run the recognized (and often taken) risk of reverse inference (Poldrack, 2006). The localization of a functional stage within a given study, on the basis of a specific experimental manipulation, does not exclude the possibility that a given area may be responsible for multiple functions in various contexts, similar or dissimilar as the case may be. The functional attribution of a similarly localized activity from different studies can therefore be disputed. While the logic behind such “reverse inference” argument cannot be disputed, it is also true that comparing studies involving quite similar cognitive tasks (thus quite similar cognitive processes) as we do here provides a prior that attenuates this risk. The similarity of tasks limits the number of processes under consideration and constrains the likelihood of their involvement. Interpreting MEG data based on previous functional attribution can thus provide a tentative bridge between different types of evidence.

2.4.2. Dynamics of function - interpretation according to Indefrey’s attribution of functional anatomy:

Table S2 lists all areas mentioned as being responsible for a function in Indefrey and Levelt’s model of speech production (Indefrey & Levelt, 2004; Indefrey, 2011). While Indefrey’s review does not specifically cover the early visual processes, we take the (moderate) liberty of attributing sources in posterior occipital cortex to visual processing. Figure 4 shows the timecourse of all reported activations, interpreted on the basis of their anatomical locus according to Indefrey’s allocation of ‘operation’ to cortical structure.
Figure 4: Timing of activity reported in source attributed to cognitive functions grouped by location according to Indefrey's (2011) anatomical-functional hypothesis. The functional attribution is based on Indefrey's spatial, rather than temporal, description. Visual processing is added to Indefrey's original list of which begins with conceptual preparation. Colour represents the number of studies reporting an activity related to a given function, limited for open-ended.
In general, we see a staggered sequence of the most frequently reported (>5) periods of reporting for the component functions of single word production. There are, however, many reports which disagree with a serial or sequential account when following this schema of functional anatomy.

Referring specifically to Indefrey's estimates of timing, several operations appear to begin earlier than estimated by the Indefrey and Levelt model, and others rather out of sequence. We examined visual processing in this context, finding that 12 activations from 4 studies included in the graph report the onset of visual activity in a window beginning at 0ms and lasting until at least 200ms. After this there is less frequent (but still substantial) reporting of sustained activity lasting until around 900ms post stimulus and overlapping with all Indefrey and Levelt's other functions. This may be explained as residual activity due to the continued presence of the stimuli on screen (for most studies), or of secondary stimuli (the question mark at 800ms post-stimulus in Salmelin et al, 2000, for example), but it is certainly not required in a cascading model. In this later model where syllabification is already underway at 355ms post-stimulus, as the visual processing and selection of the lemma to be spoken has necessarily been achieved before the onset of this later step (i.e. syllabification).

Indefrey suggests that lemma selection should begin at 200ms post-stimulus, and yet in our analysis of the data, 4 of our reviewed studies (including Levelt et al., (1998) report 10 significant activities in areas that he associates with lemma selection before that time. Additionally, Helenius et al. (2002), Kujala, Alho, Service, Ilmoniemi, & Connolly (2004), and Marinkovic et al. (2003), all provide evidence for lemma selection in this window. Again, activity in areas associated with lemma selection persists until long after cortical activity associated with articulation has begun, and even after activity for phonological code retrieval is reported to subside.

Reports of activity in areas Indefrey associates with phonological activity correspond well with his proposed onset, frequency of reported activation peaking between 260ms and 320ms post stimulus. Here, sustained activity is easily explainable in a speech production context as the consequence of producing (maintaining a phonological representation until articulated) and perceiving (reactivating by hearing) one’s own articulation.

Cortex associated with motor and articulatory function is active far in advance of Indefrey's suggested onset, beginning with Salmelin et al's (2000) report of activity in the mouth motor area, vertex and frontoparietal areas all within 120ms of stimulus onset. However, this early activity is once again unrepresentative: onset time in this study is keyed to a prompt for speech, the word to be spoken being presented for a 300ms period, 800ms before the presentation of this cue. Setting this first result to one side, Saarinen et al (2006) and Klein et al (2014) both report onset of significant differences in oscillatory activity vs. baseline in motor cortex with an onset of 200ms, followed by Hultén, Vihla, Laine, & Salmelin (2009), ‘ERF amplitude, “Right frontal (9)”’, at 300ms and Liljeström et al., (2009), ERF variance, insula) at 320ms post-stimulus. Reports of activity in cortex Indefrey relates to articulation become very frequent after 320ms, in advance of the estimated time of onset at 600ms in the model. Even taking into account the potential hastening effect of windowed analysis and allowing for the early spread of activation to articulatory areas from those concerned with feed-in functions like phonology and lemma representation, this begs for a reassessment of the latency of articulatory activity.

Self-monitoring processes as defined by Indefrey are not directly tested in the source papers we review and it is difficult to predict whether and when these processes might have a discrete onset. Participants must be aware throughout the experimental period that they are performing the task, and even
holding oneself in readiness to perceive the next stimulus requires a degree of self-control and auto-perception. If there is a general self-monitoring circuit serving speech production (Ries, Janssen, Dufau, Alario, & Burle, 2010; Nozari, Schwartz, Dell, 2011), it may be active to a greater or lesser degree at all times. Alternatively, local circuits monitoring e.g. lexical selection or articulation (Hickok & Poeppel, 2007; Hickok, 2012; Pickering & Garrod, 2004, 2013) could be presumed to become more active at the same time as those processes come online.

Under this functional mask, the data do not seem a good match for the dynamics suggested by Indefrey (2011), nor for a serial model of single word production. While there is a general trend towards sequentiality in the mode of reports of difference for functions, the latencies of onset and duration do not match Indefrey’s estimates. There is also a great deal of overlap in diverse functions’ active periods, and substantial numbers of reports of theoretically ‘late’ functions being implicated early in the post-stimulus period.

In the next section, we apply a finer grained parcellation of cortex and cognitive function, coming more from a cognitive neuroscience rather than a cognitive psychology perspective (Price, 2012).

2.4.3. **Dynamics of function - interpretation according to Price's attribution of functional anatomy:**

Price's breakdown of cognitive function by cortical area (Price, 2012) is much finer, compared to Indefrey’s, both in terms of spatial regions and in terms of processes of particular cognitive functions (See Table S2, and Price, 2012, p.6 for the source figure). In fact, this breakdown is based on the data gathered by Price and her own collaborators, and she gives an even finer fractionation of the elements of the broad cognitive processes consistently found in fMRI studies as part of the review. Even using the coarser level of functional distinction available in the review, the projection of MEG data onto Price’s anatomical-functional parcellation was challenging, and required a number of principled criteria. The MEG activity reported in our source studies tends to be more spatially extensive than the smaller regions considered in Price’s framework, particularly for group activities. Some activities from target papers spread across more than one of the designated regions, or the activities of different participants in a given condition may cross region boundaries. To establish our mapping, we made an allocation to a region based on either the peak or centre of group activity, or the centre of the cluster of individuals’ sources. Fig. 5 shows the timecourse of the activities in our source papers as attributed to the various functions and locations designated by Price (2012).
Figure 5. Timing of activity in source attributed different functions by cortical according review of and fMRI (Price, 2012).

A: Timing reported papers to cognitive grouped location to Price's the PET literature 2012.

Attribution of Activity to Function after Price 2012
represents the number of studies reporting an activity in a given area, scale is limited for contrast, and open-ended.
Figure 5B: Reported activations not fitting into a region defined by Price's cortical map.
Figure 5 reports a substantial cluster of reported activities attributed to visual and semantic processing but fewer activities attributed to word and auditory (i.e. phonological) processing. From ~300ms onwards, many reports include activities in areas associated with articulatory processes. Within the visuo-semantic cluster of attribution, modes of activities suggest a staggered sequence, with the mode of early visual processing (i.e. ‘visual 1’) occurring before the modes of simple semantic processing (i.e. ‘semantic 1’ & ‘semantic 2’) and that of more integrated semantic processing (‘Semantics (sentences)’). It is also interesting to note the substantial temporal overlap of activities attributed to different cognitive functions between 200ms and 500ms.

The second most populated cluster of attributions concerns articulatory processing. Very frequently activity associated with this function is reported as early as 300ms post stimulus. This overlaps to a large degree with the previously described visuo-semantic cluster and appears early with respect to typical average response times observed in naming and reading protocols (normally around 600ms or more).

This mapping displays a substantial temporal overlap of the different cognitive functions engaged. However, we must exercise caution in drawing conclusions. Attributing MEG activity to fine-grained anatomical distinctions derived from fMRI risks distortion by over-fitting large patches to inappropriately small areas. Accordingly, the functional labels used on Fig.5 might partly misrepresent the original evidence. The relative absence of word-level and phonological processing provides a good example; this is linked to the differences in spatial resolution between fMRI and MEG. Price's anatomical distinctions are based on functional brain imaging techniques that have higher spatial resolution than MEG, and afford tighter constraints on patches of BOLD difference, whereas sources in MEG tend to spread in space. Thus, Price maps many functions in an area that may be covered by the spreading field of a single peak in a beamforming/MNE study, or by a cluster of individual source locations in an ECD analysis. For example, in the posterior temporal lobe and angular gyrus of Prices functional map (Price, 2012; Fig. 4), 5 distinct functions (general action, integration, words, semantic and sentence processing) are assigned to the area covered by a single group of sources in e.g. Liljeström et al., (2009); Vihla et al., (2006), or by patches of activity in e.g. Klein et al., (2014); Laaksonen et al., (2012). Where this is the case, we have taken the peak or centre of the cluster or patch as the representative point, rather than count the same patch multiple times in multiple functions.

There are also functions included in Price's map that are not applicable to single word production, the outcome of all experimental tasks included. For example, a substantial number of attributions indicate semantic processing at the sentence level (“temporal pole” and “posterior MTG 1” on Price's map), which is unexpected in our target papers, given that none of them require participants to speak more than one word at a time. It is possible that there is shared processing between reading or naming temporally separate sequences of words or pictures and sentence processing that is differently visible in MEG and fMRI, which serves to highlight the danger of directly transferring regions of interest taken from one imaging modality or experimental task to another.

The current section highlights some limits on the conclusions that can be drawn in terms of the functional interpretation or cross-validation between MEG and fMRI studies (see Liljeström et al., (2009), for a comparison of the same protocol in the two modalities, and some limits identified in using such a close comparison). It is possible that regions identified in fMRI are not ideal, in the current state...
of the art, for direct translation or interpretation as regions of interest or functional boundaries in MEG; and yet such comparison is hardly avoidable when considering the cognitive or functional interpretation of neural activity across studies.

2.4.4. **Dynamics of function - interpretation according to Hickok's attribution of functional anatomy:**

Hickok's (2012) HSFC model (see also Tourville & Guenther, (2011)) makes specific, explicit claims about areas involved in the later processes of form encoding and response preparation and execution. Referring back to the areas implicated in this model (see introduction), Fig. 6 displays those for which evidence is present in our source papers. We exclude the cerebellum from this analysis as only Resnik et al. (2014) report cerebellar activity in a 1000ms analysis window, unhelpful for the purpose of diagnosing the time-course of its involvement.
Figure 6: Reported activities in areas
mentioned in Hickok's (2012) HSFC model of speech production. A serial or cascading model would predict a sequential order of onset running through the areas graphed from top to bottom. Colour represents the number of studies reporting an activity in a given area, scale is limited for contrast, and open-ended.
The different areas postulated in the HSFC model to drive phonological and subsequent speech production processes are commonly reported in the reviewed MEG studies, with the partial exception of the post-central region (S1). In this sense, Figure 6 is compatible with the anatomical hypothesis of the HSFC model. The status of “somato-phoneme” targets (physical/feedback target states for articulators corresponding to uttered phonemes, see Figure 4 in Hickok, 2012) remains open in the light of current evidence, as only one study reports activity in somato-sensory cortex. Lateralization of the activities reported is strongly lateralised to left hemisphere for STG as well as IFG. In contrast, activity in the critical hub at TPJ/Spt (see Hickok, Okada, & Serences, 2009 for a definition of area Spt’s extent and hypothesised function), is reported bilaterally (indeed, slightly more often in right hemisphere). Motor cortex is reported equally often in both hemispheres, as expected. In sum, the network revealed in this part of the meta-analysis is quite left lateralised with bilateral activity in TPJ and motor cortex.

The original HSFC model does not make explicit claims about the timing of involvement in the different regions of interest. In the target papers, there is a tendency for repeated reports of STG and TPJ activity to precede IFG and pre-central activity (a finding at odds with the possible involvement of IFG in various other cognitive processes before motor syllable programming). As the HSFC model is more concerned with processes leading up to the overt response than those involved in stimulus processing and response selection, the timing of operations might have been better captured by response-locked rather than the stimulus-locked analysis. This remains an infrequent approach in MEG (but see Salmelin et al, 1994; 2000; and in EEG, see Laganaro & Perret, 2011; Riès, Janssen, Burle, & Alario, 2013).

2.5. General Discussion:

We have reviewed the methods used and data obtained from MEG studies of speech and word production published in the 20 years since Salmelin et al’s first single word production paper in Nature (1994), considered in the light of two prominent cognitive models (Indefrey & Levelt's (2004) and Hickok's (2012)) and an anatomical-functional framework built on a large body of empirical evidence (Price, 2012). Despite the technical challenge of recording speaking participants in the MEG scanner, a total of 154 separate significant activities were identified, across various analytical methods, in 17 articles that meet our inclusion criteria. From among those, 142 from 14 articles were compatible with our re-analysis methods. We believe these included articles represent a substantial body of evidence for secondary assessment.

2.5.1. The reviewed evidence can be interpreted to support both cascading and parallel models, and is not sufficient to exclude either.

Fig. 4 presents the data from our source papers viewed through the anatomy-to-function mask described in Indefrey’s 2011 review paper. In this interpretation of the data, there is a cascading sequence of overlapping activity moving through the brain. The bulk of evidence agrees with the order and timing of functional involvement suggested by Indefrey and Levelt (Indefrey & Levelt, 2004; Indefrey, 2011), and the few out-of-sequence activities and temporal spread can be attributed to task differences or other idiosyncrasies of analysis. However, data not included in his review (e.g. Klein et
al., 2014; Miozzo et al., 2014, etc.) are less compatible with a serial explanation, particularly if alternative functional parcellations are applied (Fig. 4).

Using the spatial filter provided by Indefrey's anatomical analysis to assign cortical structure to cognitive function, including those activities uninterpreted by the original authors, the data do not fully agree with a cascading model. There are effects reported in areas concerned with most 'higher' or post-perceptual functions from 100ms onwards (from 0ms onwards for syllabification), and looking at the times when frequency of reported activity in the discussed areas becomes common we see much earlier onsets than predicted. Articulatory activity shows above 10 reported activities before 400ms, where Indefrey (2011) proposes an onset of 600ms post stimulus. Likewise, phonological code retrieval is frequently reported 75ms before predicted, and onsets for lemma selection are spread to both sides of the predicted onset of 200ms.

This last bears further explanation, as it depends on the conjunction of Indefrey's (2011) attribution of the LIFG to syllabification (a proposal tempered by alternatives in his review) and the contribution of Klein et al's 2014 study of phonological priming.

When a spatial mask based on Price's attribution of structure to function is applied, the reviewed data shows a dramatically different sequence of functional latencies. Figure 5 shows that the appearance of a cascading flow is destroyed, with only visual and articulatory activities being distinct in their latencies of frequent reporting, although they still show overlap in a small number of activities. This map of structure to function also gives surprisingly early onsets for some functions that ought to come late in a cascading sequence (e.g. the 100ms onset of activity in cortex processing articulation). Areas associated with semantics, sentence processing and articulation in Price's (2012) interpretation are too early for a purely cascading model to be a plausible explanation of the speech network. The earliest onsets in “Semantics 2”, are those from Klein et al. (2014), which appeared as the early onset of syllabification under Indefrey's scheme. Perhaps the clearest thing apparent from this analysis is that combining fMRI and MEG evidence is difficult, at least in language tasks, and does not lead to a simple, single interpretation of the findings (n.b. Liljeström et al., 2008).

Group level MEG studies are less well suited to separating effects in closely neighbouring cortical areas, as tendency of patches of activity in MEG to have a diffuse rather than sharply defined extent, and differences of both cortical anatomy and differing location of active sites between individuals conspire to blur the picture, presumably more so than in fMRI individual differences in anatomy blur spatial resolution. In Price's parcellation (Fig. 5) of the fMRI map of cognitive function, there are many small areas in close proximity that subserve subtly different functions. Notice that the grain-size of this parcellation does stem from a single empirical study, but from the integration of many studies in the meta-analysis. It is difficult to obtain the spatial resolution in an MEG study to separate regions on this scale, and even more difficult to integrate information from multiple studies to address this issue of timing.

Focusing on the areas referenced by Hickok's HSFC model as graphed in Fig. 6, it is worth noting that some of the key areas in this model are observed largely bilaterally (precentral gyri, TPJ and IFG, for example). At the level of spatial acuity afforded by our analysis, there is no clear sequential pattern of activation predicted by a cascading or serial model, and the IFG seems to be active before the STG and TPJ rather than in parallel, as the HSFC model would seem to suggest.
In addition to the issues of early and non-sequential activation of component functions, there is extensive overlap in the active periods of the different functions. In some cases this makes sense; since articulation takes place over time, later elements (phonemes, syllables) of a word must be held in memory as earlier ones are spoken, thus the lemma, phonology, syllabification and articulation could be concurrently active. For other functions this sustained activity makes less sense in the context of a cascading model. Sustained activity of areas involved in lemma selection is superfluous, for example. Once a lemma has been selected, all following processes can continue cascading towards articulation without the need for maintaining or reprocessing the same decision. In parallel models however, this protracted activity makes more sense, as all elements of the network conspire to maintain a holistic representation of the spoken word and its concept throughout articulation.

2.5.2. All task-involved areas are in a state of relative excitation and task engagement, priming and habituation are manifestations of network states existing before stimulus delivery.

We see early activity contrasts in motor and late stage (phonology, syllabification, etc.) functional areas, reported by several studies. Recent neuroimaging research suggests two explanations for this; either rapid parallel activation of all network areas (Klein et al., 2014; Miozzo, Pulvermüller, & Hauk, 2014), or task engagement having the effect of preparing top-down areas to modulate bottom up information processing (Engel & Fries, 2010; Gehrig et al., 2012; Strijkers & Costa, 2011). Both of these ideas stand against a strongly sequential, bottom-up model of speech production, whilst allowing for sequentiality in the form of stimulus information percolating into an expectant network.

There is clear evidence that the brain is prepared to respond to a stimulus, and has certain biases or expectations about what that might be, particularly in the context of a psychological experiment. Oddball and priming experiments (Klein et al., 2014; J. Kujala et al., 2012; Shtyrov & Pulvermüller, 2007) show that short term hysteresis (“the dependence of the output of a system not only on its current input, but also on its history of past inputs” (http://en.wikipedia.org/wiki/Hysteresis)) is present in the brain and behaviour, and effects of word frequency are a good example of long term biases in learning and expectation. These biases, both short and long term, can be conceived of as ‘preferred states’, for the network of areas processing the task, familiar patterns of network attractor states (Rueckl, 2002) or expected representations that the brain may be guided into by top-down control, or some other local mechanism of neural co-ordination. This expectant pre-stimulus network state and its consequent modulating effect on the processing of stimulus information present an explanation for why high level areas might be active so early in otherwise stimulus driven tasks.

The early involvement of such anterior areas as the IFG and temporal pole suggest top-down control and selection processes are active very soon after primary sensory cortex. For reasons both of expectation (the system is engaged in a particular task, and expecting to respond to a stimulus) and hysteresis (the system has just seen one or many stimuli related to this same task) the initial state of these areas when the stimulus is presented will have a strong effect on how it is perceived. If task-relevant information passes through the brain sequentially, as in a cascading model, the IFG ought to be active subsequent to visual or auditory cortical areas, depending on input modality. For IFG as for other areas governing theoretically ‘later’ functions, this is not always reported to be the case, and several researchers present widespread, parallel onsets of vs. baseline activity or between conditions difference in inferior frontal, anterior temporal and articulatory areas. These areas have been
uncontroversially associated with post-perceptual or output processes (Clarke et al., 2011; Klein et al., 2014; Laaksonen et al., 2012; Miozzo et al., 2014; Saarinen et al., 2006; Salmelin et al., 2000; Sörös et al., 2003), and do not support a purely bottom-up sequence of information flow.

When performing an experimental task like picture naming, the ‘holding pattern’ of an expectant network is perturbed by the perception of a new stimulus. Beyond the very lowest levels of input to cortex, the timing of changes in activity due to diffusion of this new information may be less predictable than generally assumed. The perceptual system is capable of sending information to different functional circuits depending on stimulus. For example, semantic and orthographic systems are differentially recruited for picture naming and word reading. Additionally, task-irrelevant information can be activated: orthography is automatically accessed even for repeating spoken words (Pattamadilok, Perre, & Ziegler, 2011). In this context, expectation of a particular stimulus type may alter connectivity in the network so as to optimise transmission of information between particular systems, reducing noise or transmitting information first or faster. Alternatively, all systems may receive information (about the arrival of a stimulus) in parallel, but there may be mechanisms in place to prevent, inhibit or otherwise alter processing in task irrelevant systems.

2.5.3. The nature of ERF’s analysis may be biased towards a serial interpretation of activity patterns, and perhaps serial cognitive models.

The majority of papers we review report event related fields [ERF's], limiting for our understanding of the word production network as evoked responses report “the power of the average; while induced responses are the average power that cannot be explained by the power of the average.” (David, Kilner and Friston, 2006). Top down elements, particularly early ones, may not be phase locked to appearance of stimulus, indeed, inter-stimulus intervals are often jittered expressly to interrupt this cycle of expectation. Induced power may give more insight into this situation – the early activities reported by Clarke et al. (2011), Klein et al. (2014), lend some support to this idea. Further evidence from an MEG study of reading is provided by Wheat, Cornelissen, Frost, & Hansen, (2010), where T/F analysis of different phonologically related (or unrelated) primes for a target word give rise to differences in IFG within 100ms of stimulus presentation. For those areas graphed in Fig. 6, the majority of early activities observed are from T/F analyses, which may indicate a bias for such changes to have earlier onset for activity than ERF analysis in a given area.

In making the distinction between interactive and cascading models, observing the latency for ERF’s and TF power changes is key, as it can tell us about the order of processing steps and how different mechanisms (i.e. indexed by evoked and induced oscillatory power) of the network might interact. Studies that do manage to combine analysis techniques (Laaksonen et al., 2012) and examine the response in the whole brain (e.g. Gehrig et al., (2012); Klein et al., (2014)) are to be lauded in this context, as they provide internally consistent accounts of network activity as a whole.

2.5.4. fMRI shows that closely related functions are cortical neighbours, but MEG lacks the spatial resolution to differentiate. Consistent tasks and individual-level analyses can improve matters.
In projecting MEG activities onto an fMRI anatomical-functional map, we have run the recognized risk of reverse inference (Poldrack, 2006). While one study sees a localized significant difference for a specific experimental manipulation, the same area may be responsible for multiple similar functions. Our approach attempted to attenuate such risk by comparing studies with relatively similar cognitive tasks (thus relatively similar underlying cognitive processes). The similarity of tasks limits the processes under consideration and constrains the likelihood of their involvement.

In addition, MEG has poorer spatial resolution than fMRI (Vartiainen, Liljeström, Koskinen, Renvall, & Salmelin, 2011). It is thus difficult to transfer a distinct functional area as a region of interest from fMRI to MEG, as activity originating in an area that would be distinct in fMRI will blur across to other nearby structures in MEG. Differences in individual structural and functional anatomy only serves to compound this uncertainty at the group level, or where standard templates rather than individual anatomies are used.

We note that advances in participant immobilisation and co-registration techniques (Troebinger et al., 2014) allow a spatial resolution of < 2mm using anatomical models from individual MRI scans, meaning that characterisation of the word production network on same scale as Price's analysis might be possible in individuals. Identifying the network using functional localisers for each component function in individual participants, investigating the dynamics of their interaction and confirming these patterns across a group is a time-consuming but potentially viable method for distinguishing between the serial and parallel models.

2.5.5. Differences in task, protocol and analysis blur the information present in the reviewed material.

One of the issues revealed by this review comes from the variety of tasks applied by researchers (see Table S2 for details). This is unsurprising given that the neurobiology of language is a dynamic topic comprising diverse investigations with neuroimaging data. Important as this diversity might be, attempts at generalization are partly frustrated because potentially informative inconsistencies between comparable studies may be due to differences in the stimuli, protocol or task rather than differences in the true cortical response, and because it is difficult to assign consistent functions to areas active in inconsistent tasks. Differences in the protocols used in the source papers for this review surely have an effect on the cortical activity recorded (some use reading, some picture naming, some require articulatory gestures in response to symbols, etc.) in terms of both the anatomy and timing reported.

Because of the technically demanding nature of performing MEG studies, (particularly speech studies) it seems reasonable to advocate a more consistent approach in the cognitive and task dimensions of this research. Even a complex task like picture naming does not have a standard version against which variations may be compared. Designing empirical MEG studies of language and speech production with the inclusion of standard tasks or functional localiser might be an avenue to be considered in future research:

---

2 Ideally, it would be possible to capture participants’ cortical responses to a standard battery of tests, and use these results to localise specific functions before subjecting these participants to an experiment intended to manipulate some part of the functional network covered. This would provide a resource-intensive but

63
Moving beyond the variability in experimental task, even those studies where the task is superficially the same are not directly comparable due to differences in the way that the data is processed. The different analysis methods (MNE, Beamforming, ECD’s etc.) will show different configurations of the underlying cortical activity, such that even identical input data may give rise to different timings, source distributions and interpretations. MEG sensor data is projected onto the cortical surface using one of several available methods to resolve the ‘inverse problem’ of source reconstruction. Of the 17 MEG speech production papers we reviewed, 6 use beamforming (2 * DICS, 3 * LCMV, 1 * SAM), 1 uses current source density modelling, 9 use equivalent current dipole modelling, and 3 use minimum norm estimation (with some overlap where studies report more than one method). Not only do these techniques arrive at different solutions to the same (ill-posed) inverse problem, but the format of the data produced lends itself to different types of descriptive and statistical analysis. Beamforming, for example is a spatial filtering method generally thought to give better results for time/frequency analysis of distributed sources (Hansen et al., 2010, Ch. 7 & 8). In contrast, MNE is a technique of minimisation that can offer both focal and diffuse solutions to source localisation In the papers currently under review, MNE is more commonly analysed as the ERF timeseries of given sources. Among papers where the source modelling technique is shared, there is still plenty of room for variability; for example, sources in the projection can be constrained to the orientation of individual or standard cortical anatomy or be freely oriented without constraint, used to compare different post-stimulus time windows to various different pre-stimulus periods, data can be filtered in different frequency bands, can be analysed as either ERF’s or in the time/frequency domain (very seldom both together). Just as there is no unique solution to resolving the inverse problem in source projection, the number of equally valid ways to analyse MEG data can seem inexhaustible.

In addition to the normal array of options for MEG analysis, protocols demanding speech from participants must also confront the technical challenge of the artefactual noise arising from muscle movement during articulation, and the displacement of the head that can blur localisation of sources, the authors in our source papers employ various solutions, each with their own particular biasing effect on the reported data. For example, Klein et al, (2014) use a bite bar to maintain participant head position, but bite bars interfere with natural articulation, the creation of normal motor programs, and perturb proprioception and feedback from articulators. Bite bars can also cause discomfort to participants, particularly in long recording sessions, which has behavioural consequences. Continuous head tracking has been implemented in some MEG scanning systems, which can ameliorate the localisation of sources, but this does nothing to remove noise arising from muscle. Techniques such as tSSS (as employed by Clarke, Taylor and Tyler, 2011) substantially reduce such artefactual noise in the MEG signal, but again, this algorithm is implemented in a tool only compatible with Elekta MEG systems. It is also potentially possible to use other component separation techniques to separate noise from signal, and spatial filtering techniques are implemented in current freely available electrophysiological analysis packages such as Brainstorm (Tadel, Baillet, Mosher, Pantazis, and Leahy. 2011) and Fieldtrip (Oostenveld, R., Fries, P., Maris, E., Schoffelen, JM 2011) to enable trained users to identify and remove EKG, blink and other sources of artefactual noise. In EEG, the use of blind source separation based on canonical correlation analysis and independent component analysis have been employed and evaluated with interesting results (Porcaro, Medaglia and Krott, 2015; DeVos et al, 2010; McMenamin et al 2010); to our knowledge, no such evaluation has yet been conducted using MEG data as a comprehensive alternative to group averaging that would ameliorate the problems of spatial smoothing inherent in group averages.
data. There is also the consideration that the nature of MEG signal allows for good separation of muscular and cortical sources on spatial grounds when compared to EEG: beamforming analysis can allow muscular noise to be separated from cortical signal if allowed to project sources in the entire head space instead of constraining them to be inside the pial surface (see Laaksonen et al., 2012, supplementary figure 2B and 2C for a demonstration of this). Some authors among our sources attempt to avoid the problem of articulatory artefacts polluting cortical signal by using only very short analysis windows, ending before the onset of the response, but this solution is not ideal for two reasons; firstly, it does not allow the investigation of cortical signal during the critical moment of speech (i.e. actually speaking). This is highly regrettable in the context of the our current review, as areas involved in pre-articulatory processing may still be active or exhibit between conditions differences in the later period, but are excluded from reporting. Moreover, such conservative measures may not avoid the problem; as EMG activity has been observed to start as early as 250ms after stimulus onset (Van der Linden, 2014).

2.5.6. Incomplete reporting masks potentially useful information, and replication is rare.

Though highly complex and resource intensive, MEG is a very powerful technique, and carefully applied, provides the most complete picture of cortical dynamics available from non-invasive neuroimaging techniques. Our comment here is not to set out technical guidelines, or to advocate particular analysis techniques, such guidelines have already been excellently well described by others (Gross et al., 2013). Rather we urge comprehensive reporting of as many aspects of recorded MEG data as possible to facilitate a full understanding of the interactive network involved in speaking. For example, an identified challenge in understanding the word production network is that researchers often focus on specific (cognitive) contrasts of interest, while significance levels estimated in whole brain analysis are given much less attention (although see Klein et al., (2014); Gehrig et al., (2012); Wheat et al., (2010) etc.). Indeed, the majority of studies focus on a set of ROI, the method by which these are chosen being variable. These may be chosen a priori (e.g. Clarke et al., 2011), or sometimes via significance levels vs. a baseline level of activity. More general reporting of effects in the whole brain and not just in specific time windows or regions of interest would make the field of MEG speech research more tractable to review and comparison. Comparisons between the oscillatory and connective state of the brain between the resting state and the task-engaged period might give further insights into how the brain prepares itself to receive and manage stimulus information. This would lead to a better picture of whole brain functionality, which is necessary to discern the truth of models of speech production.

Our advocacy for more general reporting requires attention to the variability in the statistical methods used to assess significance levels in the data, particularly in how the multiple comparison problem can be addressed. The volume of data delivered by an MEG scan with high sample rates requires robust correction for multiple comparisons. This is an essential step, particularly when analysing and reporting whole-brain data without a priori restrictions, given the number of statistical tests necessary to compare against a baseline period or between conditions for thousands of voxels or vertices across many time-samples. The crucial point is to find a balance between overly conservative correction that suppresses potentially interesting sub-threshold effects, and liberal thresholds whose observations may turn out to be unreplicable. Table 1 reveals that the attention given to issues of multiple
comparisons has substantially increased over time, and an empirical reconsideration of earlier findings in that light might be in order.

Compounding the problem of limited analysis is that the same tasks are very seldom (if ever) replicated between studies (Drew, 2014). Different tasks will give rise to different types of activity in the brain, making direct comparison or interpretation of the cortical response precarious. This is especially true since there is no robustly or commonly observed baseline or standard established for the cortical response to canonical tasks that is seen across studies or different samples of MEG participants.

2.6. **Conclusions:**

The timing of the onset of activity, the relationships between time frequency and evoked potentials, the differences in connectivity between the resting and task-engaged brain, and the changes in connectivity across the whole period of task performance are all crucial to producing a full model of the brain's activity during word production. While existing MEG studies provide a good deal of such evidence regarding speech and word production, it is not yet possible to conclusively discard or accept any one current model on the basis of the currently assessed evidence. This is partly because consistency of experimental design and reporting is lacking. Greater care in preserving the temporal resolution of MEG would help resolve the issue of serial or parallel activation of functions, and a broader perspective on reporting ERF's, T/F data and resting state/task-engaged comparisons of effects in the whole brain and not just in those time windows or regions of interest would make comparison between studies easier and provide richer data for the conception and falsification of models. In general, more comprehensive reporting of as many aspects of the recorded and analysed MEG data as possible to facilitate a full understanding of the interactive network involved in speaking.
2.7. **References:**


Parkonnen and Salmelin, (2010). This is a chapter. In Hansen, Kringelbach and Salmelin (Eds.), 2010, p.74


2.8. Supplementary materials

This page intentionally left blank
<table>
<thead>
<tr>
<th>Authors (Year)</th>
<th>Cognitive task</th>
<th>Type of evidence</th>
<th>Analysis reported</th>
<th>Whole brain reporting</th>
<th>Multiple comparisons correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmelin, Hari, Lounasmaa and Sams (1994)</td>
<td>Picture naming</td>
<td>ECD’s</td>
<td>Goodness-of-fit, source strength; no inferential statistic reported</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Levelt, Praamstra, Meyer, Helenius and Salmelin (1998)</td>
<td>Picture naming</td>
<td>ECD’s</td>
<td>Goodness-of-fit, source strength; no inferential statistic reported (no effect of word freq. manipulation)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Salmelin, Schnitzler, Schmitz and Freund (2000)</td>
<td>Single word reading [delayed to cue]</td>
<td>ECD’s</td>
<td>Goodness of fit, source strength; ANOVA of T/F in 8-14Hz and 16-28Hz bands</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Maess, Friederici, Damian, Meyer and Levelt (2002)</td>
<td>Picture naming</td>
<td>ECD’s/Current source density estimates, principal component analysis</td>
<td>Principal component analysis of 251 source dipoles, source strength; ANOVA of principal components for semantic manipulation</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Sörös, Cornelissen, Laine and Salmelin (2003)</td>
<td>Noun/verb naming from pictures</td>
<td>ECD’s</td>
<td>Goodness of fit, 8-11 dipoles/participant, source strength, no inferential statistic reported</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Saarinen, Laaksonen, Parvainen and Salmelin (2006)</td>
<td>Single word reading</td>
<td>ECD’s</td>
<td>Goodness of fit, 4 dipoles/participant (hand + mouth motor), source strength; T/F analysis of 16-24Hz band; ANOVA of T/F power suppression onsets and latencies in levels of word-like articulations,</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Vihla, Laine and Salmelin (2006)</td>
<td>Picture naming</td>
<td>ECD’s</td>
<td>Goodness-of-fit, 9-11 dipoles/participant, source strength; paired t-tests of dipole source strength between naming and visual, phonological, and semantic judgement tasks</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Hultén, Vihla, Laine and Salmelin (2009)</td>
<td>Picture naming</td>
<td>ECD’s</td>
<td>Goodness-of-fit, =&lt;10 dipoles per participant, source strength; ANOVA of source strength between naming familiar and unfamiliar items, and defining items</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Liljestrom, Hultén Parkkonen and Salmelin (2009)</td>
<td>Noun/verb naming from pictures</td>
<td>ECD’s/MNE</td>
<td>Goodness-of-fit, source strength (ECD’s), dynamic Statistical Parametric Maps [dSPM] (MNE),</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Clarke, Taylor and Tyler (2011)</td>
<td>Picture naming (domain and basic level naming)</td>
<td>MNE</td>
<td>DSPM; a priori ROI’s; ANOVA between levels of semantic manipulation</td>
<td>No</td>
<td>Cluster mass permutation correction</td>
</tr>
<tr>
<td>Gehrig, Wibral, Arnold and Kell (2012)</td>
<td>Sentence reading (overt and covert)</td>
<td>Beamforming</td>
<td>Linearly constrained minimum variance source model; T/F analysis of 2-6Hz, 7-13Hz, 14-30Hz and 31-120Hz bands; Group level t-tests between overt and covert conditions</td>
<td>Yes</td>
<td>Cluster mass permutation correction (sensor level), Family Wise Error correction (source level)</td>
</tr>
<tr>
<td>Laaksonen, Kujala, Hultén, Liljestrom and Salmelin (2012)</td>
<td>Picture naming</td>
<td>ECD’s / Beamforming (erDICS)</td>
<td>Meta-analysis of 3 experiments; Goodness-of-fit, source strength; T/F analysis of 7-12Hz and 17-22Hz bands</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Authors</td>
<td>Repetition of /pa/</td>
<td>Beamforming (DICs)</td>
<td>Corticomuscular coherence at varying rates of articulation</td>
<td>Yes/No</td>
<td>Bonferroni (t-tests of coherence)</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------</td>
<td>--------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Ruspantini, Saarinen, Belardinelli, Jalava, Parviainen, Kujala and Salmelin (2012)</td>
<td>Repetition of /pa/</td>
<td>Beamforming (DICs)</td>
<td>Corticomuscular coherence at varying rates of articulation</td>
<td>Yes</td>
<td>Bonferroni (t-tests of coherence)</td>
</tr>
<tr>
<td>Klein, Grainger, Wheat, Millman, Simpson, Hansen and Cornelissen (2014)</td>
<td>Single word reading</td>
<td>Beamforming (LCMV)</td>
<td>T/F analysis in 5-15Hz, 15-25Hz, 25-35Hz and 35-50Hz bands, oscillatory power vs. baseline (whole brain); T/F 5-50Hz full spectra for a priori ROI's</td>
<td>Yes</td>
<td>No (but permutation distribution testing (Nichols and Holmes 2004))</td>
</tr>
<tr>
<td>Miozzo, Pulvermüller and Hauk (2014)</td>
<td>Picture naming</td>
<td>MNE</td>
<td>Signal to noise ratio over time of dSPM MNE; event related regression of source strength with 4 predictor factors, semantic, visual, number of action features and word form (no. of phonemes and neighbourhood density)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Resnik, Bradbury, Barnes and Leff (2014)</td>
<td>Picture naming (faces)</td>
<td>Beamforming (LCMV)</td>
<td>T/F analysis in 5-15Hz, 15-30Hz and 30-70Hz freq. Bands; t-tests between know, don’t know and tip-of-the-tongue face recognition conditions</td>
<td>Yes</td>
<td>Family wise error (whole-brain)</td>
</tr>
<tr>
<td>Piai, Roelofs, Jensen, Schoffelen and Bonnefond (2014)</td>
<td>Picture naming (word distractors)</td>
<td>MNE (evoked activity), Beamforming (Time frequency analysis)</td>
<td>Graphical T/F analysis in 4-20 Hz band, statistical tests on 4-12Hz and 4-8Hz bands. Cluster based permutation tests on source level between Related and Unrelated distractor word conditions.</td>
<td>Yes</td>
<td>Yes (Maris and Oostenveldt, 2007)</td>
</tr>
<tr>
<td>Pyllkänen, Bemis, and Elorietta (2014)</td>
<td>Two-word phrase or list reading; adjectives and nouns</td>
<td>MNE</td>
<td>Evoked activity in ROI's (1-40Hz bandpass filtered), t-tests between list and phrase conditions (expt. 1) and t-tests between phrase and noun only/adjunctive only conditions (expt. 2)</td>
<td>Yes</td>
<td>Yes (Maris and Oostenveldt, 2007)</td>
</tr>
</tbody>
</table>

Table S1. Published articles identified with the inclusion criteria as providing neural sources of direct evidence: i.e. comprising overt reading, picture naming, delayed speech or other articulation. Note that even where the basic task is the same, different manipulations were used in almost all cases (please refer to original papers for details). Those papers that were not included in graphical analysis are highlighted in **bold italics**.
<table>
<thead>
<tr>
<th>Operation</th>
<th>Onset of Operation post-stimulus (ms)</th>
<th>Cortical Area(s) Responsible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual processing (After Binder et al., 2009, cited by Indefrey)</td>
<td>0</td>
<td>Posterior inferior parietal lobe Middle Temporal Gyrus Fusiform Gyrus Parahippocampal Gyri Dorsomedial prefrontal cortex</td>
</tr>
<tr>
<td>Lemma selection</td>
<td>200</td>
<td>Middle Temporal Gyrus Anterior Temporal Lobe Superior Temporal Gyrus</td>
</tr>
<tr>
<td>Phonological code retrieval</td>
<td>275</td>
<td>Temporo-Parietal Junction (Wernicke) Superior Temporal Gyrus Middle Temporal Gyrus Inferior Parietal Lobe (attribution by timing, see schematic in Fig. 4)</td>
</tr>
<tr>
<td>Syllabification</td>
<td>355</td>
<td>Area Spt (Sylvian fissure at TPJ) Inferior Frontal Gyrus (Broca's area) [from schematic, not text]</td>
</tr>
<tr>
<td>Phonetic encoding</td>
<td>455</td>
<td>Ventral premotor cortex Inferior Frontal Gyrus (Broca)</td>
</tr>
<tr>
<td>Articulation</td>
<td>600</td>
<td>Cerebellum Precentral Gyrus Thalamus Insula</td>
</tr>
<tr>
<td>Self-monitoring (in italics = after Christoffels et al., 2007, cited by Indefrey)</td>
<td>-</td>
<td>Superior Temporal Gyrus (bilateral) Cingulate cortex Insula SMA Motor cortex Cerebellum Thalamus Basal Ganglia</td>
</tr>
</tbody>
</table>

Table S2 — Cortical areas and their operational roles in speech production after Indefrey's 2011 review and update to the model of speech he previously proposed in 2004 (Indefrey 2011, Indefrey and Levelt, 2004).
|-------------------------|---------------------|
| Visual 1                | Inferior Occipital Gyrus (IOG)  
|                         | Occipital (OCC)  
|                         | Occipito-temporal (OT)  |
| Visual 2                | Globus Pallidus (GP)  
|                         | Occipital dorsal (dOCC)  
|                         | Occipito-temporal middle ventral (mvOT)  
|                         | Occipito-temporal posterior ventral (pvOT)  
|                         | Putamen (PUT)  
|                         | Thalamus (Th)  |
| Auditory 1              | Auditory cortex (A)  |
| Auditory 2              | Planum Temporale (PT)  
|                         | Superior Temporal Gyrus (STG)  
|                         | Superior Temporal Gyrus anterior (aSTG)  |
| Semantic 1              | Middle Temporal Gyrus (MTG)  
|                         | Inferior Temporal Gyrus posterior (pITG)  
|                         | Occipito-temporal anterior ventral (avOT)  |
| Semantic 2              | Angular Gyrus dorsal (AGd)  
|                         | Angular Gyrus mid (AGm)  
|                         | Angular Gyrus ventral (AGv)  
|                         | Middle Temporal Gyrus anterior (aMTG)  
|                         | Pars Opercularis ventral  
|                         | Pars Orbitalis  
|                         | Pars Triangularis  
|                         | Superior Frontal Gyrus  |
| Sentences               | Middle Temporal Gyrus posterior, anterior section  
|                         | Temporal Pole  |
| Words                   | Superior Temporal Sulcus anterior  
|                         | Temporo-parietal junction  
|                         | Superior Temporal Sulcus posterior  
|                         | Superior Gyrus posterior  |
| Word retrieval          | Middle Frontal Gyrus  
|                         | IFG pars Opercularis dorsal  
|                         | Middle Frontal Gyrus  
|                         | Inferior Frontal Sulcus pars triangularis  |
| Integration             | Pars Opercularis  
|                         | Superior Temporal Sulcus posterior  |
| Premotor                | Cerebellum VI left  
|                         | Cerebellum VI right (medial anterior)  
|                         | Cerebellum VII right (lateral posterior)  
|                         | Parietal Operculum  
|                         | Premotor ventral + Precentral (tongue)  
|                         | Supplementary Motor Area  |
| Sensorimotor            | Postcentral  |
| General action          | Caudate  
|                         | Middle Temporal Gyrus posterior, posterior section  
|                         | Precentral Supplementary Motor Area  |
| Articulatory            | Precentral (larynx)  
|                         | Precentral dorsal  
|                         | Supramarginal Gyrus ventral  |

Table S3 — Cortical areas and their operational roles in speech production according to Price's 2012 review of the fMRI and PET literature. This allocation to function based on Price (and her collaborators) own data. (Price, 2012)
3. Addressing the problem: Experimental goals and development

3.1. The role of neuroimaging evidence in assessing psycholinguistic models:

The development of formal psycholinguistic models based on behaviour and error patterns has informed sophisticated computational models, featuring multiple levels of processing and many thousands of parameters and permutations. With careful construction, such computational models provide robust replication of the original word selection and error patterns, and models with different architectures and directions of interactivity between functional layers (Rapp & Goldrick, 2000), or using serial and parallel architectures as in the Indefrey and Levelt [I&L] (Peter Indefrey & Levelt, 2004) and Dual Route Cascading [DRC] (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Zevin & Seidenberg, 2006) models can be equally valid (see chapter 1 for a fuller discussion of these models). Since the results of competing models from different theoretical perspectives are plausible, we must turn to new types of evidence to decide between them.

Fortunately for psycholinguistic researchers, alongside the development of psycholinguistic and computational models, electrophysiological neuroimaging techniques have also developed since the turn of the last century. Magnetoencephalography [MEG] has now reached the point where we are able to observe neural function at the centimetre (even the millimetre) scale (Troebinger, López, Lutti, Bestmann, & Barnes, 2014; Troebinger, López, Lutti, Bradbury, et al., 2014), at a temporal resolution of tens of milliseconds (Gross et al., 2013; Hansen, Kringelbach, & Salmelin, 2010; Salmelin, 2007; Troebinger, López, Lutti, Bestmann, et al., 2014). As far as speech production is concerned, this sophistication is the complement of the increasingly specific and detailed psycholinguistic models that have been refined alongside the technology to explain the sequence of cognitive function(s) that underlie language. In order to provide evidence capable of supporting one or other of the competing models, experiments identifying associations between manipulations of stimulus properties and changes in the cortical response to these stimuli can be considered, along with predictions based on behavioural data about the time-window in which certain types of information become accessible in the system (as in e.g. Indefrey & Levelt, 2000; Levelt, Roelofs, & Meyer, 1999). It is thus potentially possible to distinguish between competing models based on their differing predictions about the attribution of different functions to the cortex, and/or the temporal dynamics of the cortical response in the areas processing the involved functions.

Such research is both theoretically and technically demanding, requiring many elements to be in place before hypotheses can be made and an appropriate experiment to test them can be designed. In order to achieve such an outcome, a full set of functional elements must be decided upon, and cast into the cortical anatomy (see Indefrey, 2011 and Price, 2012 for
examples of this using evidence from MEG and fMRI respectively)\(^3\). A clear point of distinction between the predictions of the competing models must be clearly identified. A protocol for accessing this point of divergence must then be designed and implemented in a manner compatible with the technical constraints of MEG, all this before perfect practical execution and intricate analytical techniques are applied. Given such constraints, the scope of experiments must be limited, and a focus on single word production is justified.

### 3.2. Requirements for adding to the current body of evidence:

While existing MEG studies provide a good deal of such evidence regarding speech and word production, it is not yet possible to conclusively discard or accept any one current model on the basis of the currently assessed evidence. This is partly because consistency of experimental design and reporting is lacking. In addition, greater care in preserving the temporal resolution of MEG would help resolve the issue of serial or parallel activation of functions, and a broader perspective on reporting ERF’s, T/F data and resting state/task-engaged comparisons of effects in the whole brain and not just in those time windows or regions of interest would make comparison between studies easier and provide richer data for the conception and falsification of models. In general, more comprehensive reporting of as many aspects of the recorded and analysed MEG data as possible to facilitate a full understanding of the interactive network involved in speaking.

What is involved in a full, source-level description of the cortical response during word production? In both EEG and MEG, a broad range of event related or evoked potentials have been identified relating to sub-components of speech production in reading and picture naming; the visual evoked potential relating to stimulus presentation (Liljeström, Hultén, Parkkonen, & Salmelin, 2009), a lateralised readiness potential before response onset (Buján, Galdo-Álvarez, Lindín, & Díaz, 2012; Peter Indefrey, 2011; Schmitt, Kutas, & Münte, 2000; van Turenout, Hagoort, & Brown, 1997), other components associated with audition (Hickok, Houde, & Rong, 2011; Houde, Nagarajan, Sekihara, & Merzenich, 2002) and speech errors (Acheson & Hagoort, 2014; Ries, Janssen, Dufau, Alario, & Burle, 2010). The rhythmicity of speech itself is around 3-5Hz, a rhythm which coheres with cortical neuronal activity in motor areas during speech (Ruspantini et al., 2012), and the brain uses neural oscillations on this and other frequencies in a range between ~0.05Hz to ~500Hz (Buzsáki & Draguhn, 2004), although responses at frequencies as high as 600Hz have been found using MEG (Hansen, Kringelbach, & Salmelin, 2010, p. 39). The timing of the onset of activity, the relationships between time/frequency [T/F] changes and evoked/event related potentials [ERP’s], the differences in connectivity between the resting and task-engaged brain, and the changes in connectivity across the whole period of task performance are all complementary sources of information.

\(^3\)fMRI’s temporal resolution is too poor to dissect the haemodynamic response in time windows much finer than a second (Huettel, Song, & McCarthy, 2004, p197-206; Purves et al., 2008, p80-81), p197-206), and BOLD signal only provides a secondary measure of neural activity.
regarding the brain's activity during word production (Clarke, Taylor, & Tyler, 2011; Gehrig, Wibral, Arnold, & Kell, 2012; Laaksonen, Kujala, Hultén, Liljeström, & Salmelin, 2012; Papoutsi, Stamatakis, Griffiths, Marslen-Wilson, & Tyler, 2011).

Reporting on every single aspect mentioned above throughout the brain represents more than is feasible for reporting in a single study, and most investigations feature only a single method of analysis. Moreover, most authors of word production studies elect to focus on manipulations of a single aspect of their stimuli and a single facet of the speech network, and often restrict themselves to reporting on fixed regions or time windows of interest. Such an approach can be informative when discussing a particular issue or point in the debate about the speech/word production network, but the number of different approaches can make generalising from one study or integrating findings from many studies difficult. The complementary nature of the different approaches might seem advantageous, but when the relationship between ERF's and T/F changes in an area, or different types of connectivity between areas is unclear, interpreting the whole becomes more difficult, particularly when only one type of information is available from each study, and differences in task, participants, and data capture/analysis come into play.

In Munding, Dubarry, & Alario, (2015) [Chapter 2] we performed a review of MEG studies of speech production in the context of prominent functional maps of the cortex, and two current anatomically specified models of the word production network, namely the I&L (Peter Indefrey, 2011) and hierarchical state feedback control [HSFC] (Hickok, 2012) models.

### 3.3. Motivation for an empirical study of word production:

The existing MEG literature is insufficient to make a clear case for either the serial or parallel schools of thought concerning word production. Associations between function and cortical area have already been made in many neuroimaging studies. While there is general 'grosso modo' consensus about the network involved, studies often provide different data concerning location, extent and timing of response and effects (see Chapter 2 for more discussion). While cognisant of the valuable contributions of preceding authors, we wished to conduct our own investigation into the cortical dynamics of speaking a word. Since we conclude that restricting an analysis on the basis of expectations from previous studies might bias experimental findings, or prevent the discovery of interesting effects outside expected times and cortical structures, we decided to employ a data-driven approach making no a priori assumptions about the locus, timing or type of response expected in any element of the network. To do this, we reasoned that our study must comprise certain elements in its theoretical and technical design to contribute usefully to the debate.

1. Full, ecologically valid engagement of the word production network.
2. Contrasts between experimental conditions that allow independent association of function and cortical structure.
3. Whole brain analysis and reporting.
4. High temporal resolution.
5. No a priori assumptions of when and where activity and differences might be found (an exploratory, data driven approach).

3.4. **Paradigm selection, experimental design and development:**

It was desirable that the process of speaking be as natural as possible for participants and maximally ecologically valid, and that the task be well-established in the existing literature, to facilitate both comparison with the existing literature and potential replication or variations upon the study. Multiple word protocols add complexity to an already difficult task, so single-word production is favourable. Of the two most common single-word production tasks, word reading and picture naming, picture naming is task that provides the fullest engagement of the production system (Glaser, 1992), since it is arguably not necessary to engage the semantics of a read word (a direct route from orthography to phonology exists), while a picture must be semantically evaluated in order to be identified and named. Additionally, picture naming has a longer response time than single word reading (Glaser, 1992), and having a longer window before verbal response onset may make it easier to observe and separate the distinct processing stages in the brain, as they will be less compressed in time.

Having selected a task that fully engages the system, it is then important to observe the response throughout the brain, and across the whole time period of the response. This is desirable for several reasons: to allow confirmation across studies that signature elements of the cortical response in picture naming are present, and to compare the responses across different instances of the same task, and across variations of the same type of task. An exploratory approach also allows one to avoid missing potentially meaningful or informative changes in the brain response between conditions due to a focus on regions or time windows of interest selected *a priori*. The technical limitation of noise at the MEG sensor level arising from muscular activity still remains after the onset of response, and avoiding, removing or otherwise compensating for this noise is necessary.

In order to fully explore the dynamics of the response then, we wished to use a manipulation of the stimuli that would allow us to observe differences in several of the theorised sub-components of the network. A manipulation of visual complexity that should elicit the largest differences in response in brain areas concerned with visual processing, and a semantic manipulation that should do the same for the areas concerned with conceptual processing. We hoped to manipulate these factors in a single protocol, in a single population, in an effort to avoid the usual problems of blurring between studies, participant groups, scanners, analysis techniques etc.
Sadly, adapting a paradigm to MEG is not as simple as copying the files from a behavioural experiment across to the stimulus control computer in the scanner room. MEG requires around 100 valid trials to have a signal-to-noise ratio adequate for acceptable source localisation, although more trials can enhance spatial acuity, particularly for deep structures (Attal & Schwartz, 2013; Attal et al., 2007; Gross et al., 2013; Hansen et al., 2010; Troebinger, López, Lutti, Bradbury, et al., 2014). Ideally, one would run a surplus of trials and have enough to spare, but the length of scanning runs and participant sessions is an issue. Fatigue becomes a factor in task performance after around 15 minutes, and even motivated, experienced subjects can become tired after 15 minutes continuous scanning, leading to excessive blinking, mind-wandering, inattention to the task etc. Hansen et al., (2010) recommend that each scanning block be no more than 15 minutes in length. There is an expectation that approximately 25% to 30% of trials will be rejected, either due to behavioural error, or because they are contaminated by artefactual noise arising from swallowing, blinks, head movements, SQUID jumps etc.

These guidelines in mind, in order to be relatively confident of obtaining 100 clean trials from each participant, we programmed 256 normal experimental trials, with the expectation of rejecting a certain proportion, finding number of trials that marked a limit of participant’s comfort and performance, and reducing the number of trials to that limit after piloting.

Again referring to Hansen et al. (2010) we find the recommendation that the stimulus onset asynchrony [SOA] be around 1-3s for time/frequency analysis, and that stimulus onset and offset both give rise to evoked potentials, but little sustained activity. Since we expect response times to be in the order of 600ms-1000ms (Glaser, 1992), and the act of speaking a word is likely to take around a second, we set the length of time for which the stimulus is present on screen to be 2s, so that there was no chance of overlapping a second visual evoked potential with other potentially interesting components in the experimental epoch. We wished to leave ample time for articulation to stop, and any residual muscular activity to subside before the onset of the next trial, so we built a jittered pause between trials with an average of 2s, stepped in 200ms intervals. This gives a trial SOA with an average of 4s.

The number of blinks can be minimised by putting proper procedures in place, such as instructing participants to avoid blinking during critical periods of the epoch. Refraining from blinking for extended periods is obviously difficult and tiring, so we combined the instruction to not blink during trials with blink pauses inserted after every third trial, allowing participants to rest, hoping that the time lost to these rest periods would save trials overall. Adding these blink trials, being 3s after every 3rd trial gives us a 15s cycle over 3 experimental trials, and a duration of 21 minutes and 20 seconds for our 256 trial blocks.

The factors in our protocol were conceived of as 4 different conditions:
I. [COLOUR] Colour naming - a semantically minimal condition, which activates the full production network but requires only a perceptual judgement with limited semantic processing.

II. [SINGLE] Naming line drawings of items from a single semantic category, requiring more semantic engagement than the colour naming condition in the recognition of a picture, but using a small picture set with many repetitions to make the task relatively easy and predictable.

III. [MULTI] Naming a larger set of line drawings of items from several different semantic categories. While names will be repeated, each type of item will have multiple pictures, meaning that the semantic recognition element must be done anew every trial, rather than being a repetition of an already seen item.

IV. [COVERT] A re-randomisation of the stimuli previously seen in the MULTI condition, which participants would name internally to avoid any noise from articulation polluting the MEG signal and provide a foil for the analysis of the 3 expressive conditions.

These conditions parameters established, our manipulations had the following format:

1. **Visual manipulation**

The visual manipulation consisted of three levels: in the COLOUR condition, participants would be required to name colours from patches of visual noise with no coherent shape. Colour naming is possible based on purely perceptual information, and requires minimal semantic engagement.

The colours to be used were decided upon by choosing high frequency colour names\(^4\) for unambiguous colours that had good name agreement between participants when exposed to the colour. The French names used, their frequency in language, and their English translations are detailed in Table 1.

---

\(^4\) The word frequency measures used throughout our experimental design are those for frequency of occurrence occurrences per million in books (freqlivres), as listed in the LEXIQUE database (v3.55) (New, 2006)
Several approaches to the design of the stimuli were tried. At first we attempted to use solid squares of colour, but reasoned that such shapes might inadvertently activate the name of the shape in participants' thoughts, polluting the semantically minimal, visual only purpose of this condition. Knowing that we wished to use line drawings for other conditions, we created 8 pseudo-object stimuli using abstract shapes, somewhat in the style of the Rey-Osterreith complex figure (Rey, 1941), to match the physical properties of these line drawings, but without having any determinate or nameable shape to introduce any semantically valent element. A selection of these is displayed in Figure 1(A). The colour of these pseudo-objects would have been varied randomly across trials to prevent any association between shape and colour. 8 noise patterns would be presented 4 times in each of the 8 colours to make up our 256 trials.
Figure 1: (A) Abstract pseudo-objects intended as stimuli for the colour naming condition, to match the physical properties of line drawings in other more semantically complex conditions. (B) Scrambled, coloured visual noise patterns created by randomly shuffling 20x20 pixel squares from line drawing stimuli.

After testing, these pseudo-objects were decided to have too many recognisable geometric shapes composing them, and to be too likely to induce saccades as participants examined the different parts of the picture. By this point, we had selected potential stimuli for the condition in which participants were to name objects from a single semantic category, and another style of colour stimuli was tested, created by randomly scrambling small tiles of the line drawings from that condition, randomly rotating them and shuffling their position in the image. While less globally similar to the line-drawings than the pseudo-objects, these had several advantages. They have exactly the same amount of coloured or 'inked' part of the picture as the line drawings, they possess no distinct nameable component forms, and they are highly disordered, noisy images, intended to be stimulating to lower-level visual areas (Field, 2003; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999). In this format, we hoped the colour stimuli would produce a maximal response in visual cortex, with a minimal response in the other component elements of the word-production network.

2. Semantic manipulation

Having designed the stimuli for our colour condition to be visually distinct from the black-and-white line drawings we intended to use for the picture naming conditions, we then chose pictures in the style of Snodgrass and Vanderwart’s famous stimulus set for the latter (Snodgrass & Vanderwart, 1980). Wishing to use multiple semantic categories and match some lexical properties of the colour words, we were unable to use many of the pictures from the Snodgrass set, but made efforts to use their drawings as a template for selecting our own on the basis of complexity, style and size. In so doing we hoped to minimise the low-level visual differences between the stimuli in the SINGLE and MULTI conditions. We chose to use
common (i.e. often seen), distinctive and easily recognisable objects for the picture naming conditions; vehicles, furniture, household tools and musical instruments.

While holding the low-level visual components steady between the picture naming conditions, we wished to ensure that the semantic recognition/lemma selection element of the task was different between the two picture naming conditions. To this end, we chose to have a set of only 8 pictures, one for each of 8 vehicles in the SINGLE condition, repeating each one 32 times. In this way we hoped that the stimuli would be extremely easy to recognise, and represent a middle level of semantic engagement. The MULTI condition would then comprise line-drawings of a similar style, but a much wider semantic 'target area', featuring 32 different names from all 4 semantic categories, with each trial featuring a new picture, forcing full semantic exploration of the picture each time.

A range of names was selected in each semantic category according to the starting letter and lexical frequency of each colour name. Some hundreds of pictures were sourced and tested on a group of 8 native French speakers for naming agreement. Those that achieved less than 87.5% agreement were rejected outright, and final stimuli were selected from the remainder, according to the goodness of the lexical frequency match, the similarity of the onset sound in French, and finally the characteristics and style of the picture.

Response times are variable. This will mean that if the same cortical areas are active in the same sequence, the later elements of the response will be more spread out in time and thus their signal-to-noise ratio will benefit less than early elements from the effect of stimulus locked averaging across trials. If our assumption about the timing of cognitive functions and cortical processing holds, then both evoked and induced changes in source level power associated with post-lemma and articulatory processes will be more evident in a response-locked than in a stimulus-locked analysis of the data. Figure 2 represents a simplistic illustration of this idea.
Figure 2: A simple schematic illustration of how the distribution of response times has consequences for the latency of the cognitive processes and the associated cortical activity in a serial or cascading model of word production. Response onsets are illustrated with a vertical red bar, stimulus onsets a vertical black bar. Coloured curves represent the activity associated with the stages of cognitive processing.

To take advantage of these ideas, a clear locking point for the onset of the response is necessary. The use of voice recordings to identify the moment at which speech begins is common practice in picture naming protocols (Kempen & Huijbers, 1983; W. J. M. Levelt & Maassen, 1981; Willem J. Levelt, Schriefers, Vorberg, Meyer, & Et Al, 1991; Maess, Friederici, Damian, Meyer, & Levelt, 2002; Pulvermuller, Shtyrov, & Ilmoniemi, 2003; Ries, Janssen, Burle, & Alario, 2013). Following suit, we set up a system using an optical microphone in the MEG scanner, and planned to use a Python script implementing the Teager-Kaiser Energy Operator (see Chapter 5, and Khoubrouy & Panahi, 2013; van der Linden et al., 2014 for details) to automatically identify voice onset times in the voice recordings, and find their latency relative to stimulus onset in each trial. After initial testing, it was found that the different onset sounds for the stimulus names had very different acoustic characteristics, which made it extremely difficult to identify equivalent onset points either automatically or manually. To solve this problem, we decided that participants should begin each response with either “du” in the case of colour naming, or “des” in the case of object naming, in order that the onset syllable of each

---

5 See Chapter 4 of this thesis for details of the equipment setup and testing procedures.
This is completely natural in French for colour naming (du bleu, du rouge etc.), but implies the plural form of the noun. We thus duplicated the line drawings in each stimulus picture to have two of the same items, one behind the other, to make the pictures appropriate for the desired response.

Having thus fixed a clear response onset point to which the epochs can be locked for averaging in analysis, we hoped to be able to identify the elements more strongly associated with articulatory processes by their relatively greater strength in a response locked rather than a stimulus locked analysis. The greater articulatory variability in the semantically more complex MULTI condition should show a greater proportional increase than the semantic elements in a response locked vs. stimulus locked comparison, due to the blurring of the average over time for the semantic elements of the response caused by the variable response latency. Such elements having been spatially located according to the dynamics of the response, the cortical regions involved could then be compared between the simple and complex conditions of articulation. Any significant differences could then provide confirmatory evidence for the function of these loci, and provide information about the characteristics of the cortical response.

3.5. **The importance of a full exploration of the network versus the necessity of simplicity:**

Our preparations led us to a very complex set of between-conditions manipulations, which runs the risk of combining with the technical complexity of implementing a MEG experiment to confuse the data produced to a dangerous degree. This valid criticism must be balanced against the value of having a dataset in which many elements have been manipulated while the technical setup, sample population, equipment and analysis methods will all be consistent. These factors have a strong effect on the data, and are very seldom shared between studies, making it difficult to draw comparisons and make inferences across different MEG studies. Combining this advantage with a data-driven, exploratory approach would mean that any significant effects found that agree with existing findings have the distinction of being unbiased and natural, and opens up the possibility of finding new and unique associations between psycholinguistic models, cognitive function and the cortical anatomy and dynamics or the response to naming.
3.6. **References:**


Rey, A. (1941). L’examen psychologique dans les cas d’encéphalopathie traumatique.(Les problems.). *Archives de Psychologie.*


4. Technical development of a MEG protocol to study language production

In the previous chapter, we established some desirable experimental goals and procedures and outcomes for a magnetoencephalography [MEG] neuroimaging study of the dynamics of the cortical network underpinning picture naming. We will now discuss the technical capabilities and requirements of MEG in general, and how our procedures were adapted into a full scanning protocol and analysis pathway.

4.1. Magnetoencephalography: scanning and analysis techniques:

Modern MEG neuroimaging uses a helmet-shaped array of super-cooled quantum interference devices [SQUIDs] inside a dewar of liquid helium to measure fluctuations in the magnetic field around the head, providing a direct measure of neural activity in the brain (Hansen, Kringelbach, & Salmelin, 2010). Essentially a small loop of metal with a gap across which current can flow, at super low temperatures (around 4°Kelvin, or -279°C) changes in the magnetic field induce changes in the resistance of the metal in the loop, changing the amount of current that flows across the gap (Kleiner, Koelle, Ludwig, & Clarke, 2004). These fluctuations in current are amplified for the hundreds of sensors in the array, and these measurements analysed to provide a reconstructed measure of the magnetic field change in the scanner (Ermer, Mosher, Huang, & Leahy, 2000; Hillebrand, Singh, Holliday, Furlong, & Barnes, 2005; Vrba & Robinson, 2001).

With a participant’s head inside the helmet-like aperture in the Dewar, the electrical activity in the neurons making up the brain perturbs the magnetic field around the sensor array. Layers II, III and V of the neocortex contain populations of pyramidal nerve cells, which are aligned with one another and perpendicular to the cortical surface (Hansen et al., 2010; Murakami & Okada, 2006). This arrangement means that synchronisation of post-synaptic potentials across approximately $10^5$ of these cells can cause magnetic fluctuations large enough to be measured by the scanner. To put this in context, the cortex has an surface area of approximately 750cm$^2$ (Attal & Schwartz, 2013; Attal et al., 2007), with $10^{11}$ neurons connected by approximately $10^{14}$ synapses (M. S. Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993).

Performing any cognitive task (and even the brain’s resting state) involves co-ordinated, synchronised oscillatory communication between neurons, using electrical impulses across synapses. The post-synaptic potentials between these neurons and the corresponding magnetic field change can be summed over time (Baillet, Mosher, & Leahy, 2001), and where the population is large enough and properly oriented with respect to the MEG sensor array, it can be sensed and recorded. MEG signal then, is generated by certain populations of the aforementioned pyramidal neurons where their arrangement perpendicular to the cortical surface means the magnetic field generated by the aligned cells passes through the skull to reach the sensors. This is generally along the walls of sulci, whereas the cells along the crests of gyri and in the troughs of the sulci generate radial field changes parallel
with the skull’s surface, that are not externally readable (Molins, Stufflebeam, Brown, & Hämäläinen, 2008).

As shown in Figure 1, MEG recordings take place in a magnetically shielded room, the participant sitting or lying with their head inside the Dewar. The shielded room minimises disturbances from external magnetic sources (Hansen et al., 2010), but there is still the problem of noise generated inside the room. The heart creates electro-magnetic fields 100–1000 times larger than the cortical signals of interest (Vrba, 2002). These are attenuated by the greater relative distance from the sensors (magnetic fields drop in power exponentially over distance), but the heartbeat, as well as eye blinks and swallows all contribute noise to the sensor readings. Measurement and rejection or correction of such artefacts, are used to increase the signal-to-noise ratio further (Gross et al., 2013; Hansen et al., 2010).

Figure 1A: A participant in an experiment seated in the MEG scanner, head inside the Dewar helmet. Figure 1B: The Dewar can be rotated through 90° to the horizontal position, and the couch placed flat to accommodate the participant in the supine position.

Before scanning, the participant’s head shape is digitised using a 3D modelling technique, and once installed in the scanner the participants head position is measured within the Dewar using coils that emit signals of a particular frequency, which are placed on fiducial points on the head. This
measurement is then used in combination with a model of the participant’s anatomy (or a standard template anatomy) to fix the position of sources relative to the sensors in source-space analysis. (Gross et al., 2013; Hansen et al., 2010).

MEG data is recovered from the scanner as a large file containing the continuous trace from all the sensor channels. Depending on the analysis intended and the software used, various pre-processing and cleaning steps are necessary. The continuous data are high-pass filtered with a cut-off between 0.3Hz to 1Hz (this high latter limit is useful if breathing related artefacts are contaminating the data) (Gross et al., 2013) before they are segmented into epochs containing the individual experimental trials, their length defined a priori and normally locked to the presentation of the stimulus. A low-pass filter can also be applied, either for filtering out muscular artefacts, which appear as high-frequency oscillatory bursts, or for smoothing the data in preparation for ERP analysis. In the example of the experiment proposed in the previous chapter, our trial epochs are unusually long, containing 2s of stimulus presentation. For a stimulus locked analysis concerned with only the early part of the response in picture naming and wishing to avoid artefactual noise from articulation, the cut-off point for the epoch could be defined by the onset of the response, at around 600ms post-stimulus (Glaser, 1992). An epoch would then span some quiet or resting period that would provide a baseline level of activity prior to the stimulus onset, say 400ms pre-stimulus, until 600ms post stimulus onset, giving a full second of data in each epoch.

Epochs containing magnetic artefacts, such as blinks, swallows etc. are then identified, which can be done by manually examining the raw sensor traces, using automated methods to detect abnormal or strong fluctuations in the power or variance of the sensor trace, by or using electro-myographic [EMG] or electro-oculographic [EOG] data recorded from mouth or eye muscles during the scan. Artefacted epochs must then be removed or subjected to some form of correction before the clean trials are subjected to further analysis. For the analysis of event related potentials (evoked power) epochs are averaged within conditions and participants in order to increase the signal-to-noise ratio, while for time/frequency analysis (of induced power) the signal is split into frequency bands according to the needs of the researcher, and the power over time in each band calculated before being averaged across trials. In both cases, the averaging is performed under the assumption that the cortical response to stimulus remains consistent across trials, and that other fluctuations present in the epochs are uncorrelated noise (Gross et al., 2013; Hansen et al., 2010).

The simplest type of MEG analysis plots signal recorded at the sensor level directly (see Figure 2). The size and latency of the changes in the magnetic field can then be compared across conditions. Contour plots are another way to represent sensor level information. Adding back in a spatial element, contour plots showing isometric lines of field strength can be plotted, showing the field strength across sensors as they are laid out in the helmet as a 2D plot at a given time-point in the epoch (see Figure 3). Plotting the field at the peaks of signal strength in time is common practice, and experienced researchers can learn to recognise characteristic patterns such as visual, sensory or auditory evoked responses in the contour plots. Sensor space analyses do not ordinarily account for the size or position of the participant’s head in the dewar, and thus it is difficult to draw conclusions about the relative field strength between participants at this level. For this reason, group analysis is usually performed at the source level, after a model of the activity in the participants’ brains has been estimated.
Figure 2: A ‘butterfly’ plot of the response to picture naming, reporting the signal recorded at the sensor level. Each coloured line represents the fluctuation of the magnetic field measured by one sensor across the average experimental epoch. Time runs from -400ms to 600ms either side of stimulus on the X axis, and sensor signal strength is in femto-tesla on the Y axis.

Figure 3: Contour plot of the magnetic field across the helmet at 100ms post-stimulus. Sensors are displayed as white dots, and lines represent contours of equal magnetic field strength.
For the purpose of relating cortical function to experimental manipulation, sensor space analysis does not provide an adequate description of where in the brain the neural activity is taking place. Techniques such as equivalent current dipole [ECD] modelling (Scherg & von Cramon, 1985), minimum norm estimation [MNE] (M. S. Hämäläinen & Ilmoniemi, 1994) and beamforming (Van Veen, van Drongelen, Yuchtman, & Suzuki, 1997) can be applied to project the sensor data back into a model of the sources which generated the measurements. These approaches all attempt to overcome the ‘inverse problem’ of transforming the data from a 2D to a 3D representation, which has no unique solution (M. S. Hämäläinen et al., 1993). As Helmholtz (1853, as cited by Hämäläinen, et al., 1993) demonstrated, knowledge about the electromagnetic fields outside a spherical conductor is not sufficient to provide a unique description of the current distribution inside that conductor. That is to say, multiple different satisfactory solutions involving different numbers, strengths and positions of electrical sources could give rise to the shape of the magnetic field as measured at the sensor level. In order to model the electrical sources, structured assumptions must be made to constrain the number of possibilities and explain the variance of the sensor signal over space and time (Hillebrand et al., 2005).

### 4.2. Methods of MEG analysis: Equivalent current dipole [ECD] modelling

ECD modelling uses estimates of the location, strength and orientation of neural sources from the magnetic fields measured at the sensors (M. Hämäläinen & Hari, 2002). In the simplest case, a single equivalent current dipole is placed to best explain the observed magnetic fields at a given moment, but ECD models in functional neuroimaging studies usually contain multiple ECD’s, based on the assumption that a minimal number of neural sources contribute to the field and either using a priori estimates of the number of sources, or setting a threshold for the minimum amount of goodness-of-fit that placing an extra dipole must add to the model (M. Hämäläinen & Hari, 2002; Hansen et al., 2010).

### 4.3. Methods of MEG analysis: Beamforming

Beamforming is a spatial filtering technique for MEG source space analysis, which uses narrowband filters to maximise the sensitivity of each sensor to signals from a location of interest. A grid of points (or virtual electrodes) is cast in the volume of the modelled brain, and the filtered contributions of the sensors are mixed in different proportions to estimate the current source at each of these points. As with MNE, this method is usually constrained by anatomical information from MRI, ensuring the grid is only cast inside the head, and not in other tissue or empty air.

### 4.4. Methods of MEG analysis: Minimum Norm Estimation
MNE assumes neural sources are distributed across the cortical surface or within the brain volume, taking a surface (usually the pial, or grey matter surface, but sometimes the grey/white matter boundary) from anatomical MRI scans to constrain the orientation of sources (Hansen et al., 2010; Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). MNE estimates a distribution of sources that has the least energy necessary for minimising the difference between the magnetic fields estimated from the model and the measured magnetic fields (Hillebrand et al., 2005).

**4.5. Characteristics of MEG**

Employing these source-level analysis techniques, MEG has excellent temporal resolution, but relatively poor spatial resolution compared to fMRI and intracranial recordings. Because it requires relatively large populations of well-aligned neurons in a particular orientation relative to the skull, it is less able to resolve sources on the crests and troughs of gyri, where the magnetic fields are radial to the skull surface, or to get clear signal from deeper structures that do not have the same neural organisation as the cortex. While it is still possible to resolve such structures as the hippocampus and putamen, more trials are required to obtain the same quality of spatial resolution as for cortical structures (Attal & Schwartz, 2013; Attal et al., 2007; Hillebrand & Barnes, 2002; Troebinger, López, Lutti, Bestmann, & Barnes, 2014). Additionally, due to the aforementioned exponential drop-off in magnetic field strength with distance, deeper sources are more difficult to resolve than superficial ones. Figure 4 shows the probability of detecting a source of equal strength at different locations in the brain, and demonstrates clearly the advantage of orientation and distance from the sensors. Hillebrand & Barnes (2002) point out that while radial sources are difficult to detect, they represent only around 5% of the cortical surface.

**Figure 4.** Adapted from Hillebrand and Barnes (2002). The relative probability of detecting a source of a given strength with MEG at different locations and depths in the brain.
Hillebrand & Barnes, (2002) estimated that using the average current density value in cortex of 50pAm/mm², a source can be localised as a cube with a volume of 2mm³. This has been developed to the point where MEG can discriminate between the different cortical laminae using MEG in an auditory evoked response paradigm, with a spatial resolution of about 2mm, using only 145 trials (Troebinger et al., 2014).

Although this is still a lower spatial resolution than is available using fMRI (which can achieve approx. 1 mm (Huettel, Song, & McCarthy, 2004)), this is compensated for by the gain in temporal resolution, which we would argue is critical to the issue of parallel or serial function in the speech production network. The millisecond temporal resolution of MEG is essential for discerning the order of cognitive processing in picture naming.

4.6. Developing the technical element of an empirical MEG investigation of language production.

In chapter 2 we discussed the current state of MEG research on word production, exploring some technical aspects of research, and how particular types of evidence might contribute differently to the debate on the serial or parallel involvement of different cognitive processes. As mentioned, much debate centres around the idea that distinct functions are localised in specific cortical areas, and many studies focus on reporting the timing of changes in activity and/or connectivity in and between areas, or the latency at which a given area responds differently in reaction to a manipulation of stimulus properties, thereby inferring onset times for different processes. These latencies of onset for different functions are key to the debate as to whether the component functions of word production are processed serially or in parallel by the brain.

Let us take the example of the Left Inferior Frontal Gyrus [LIFG], an area central to neurolinguistics for more than 150 years, whose place in the language system is a subject of ongoing investigation (Costafreda et al., 2006; Katzev, Tuscher, Hennig, Weiller, & Kaller, 2013; Price, 2012; Sahin, Pinker, Cash, Schomer, & Halgren, 2009; Wright, Randall, Marslen-Wilson, & Tyler, 2011). It is of particular interest to us as an example of a site where both function and timing of activation in various language processes are debated. It is uncontroversial to assert that LIFG plays some role in single word reading and picture naming, and many studies use these tasks to probe its function (Picture naming: Cornelissen et al., 2004; Liljeström, Hultén, Parkkonen, & Salmelin, 2009; Salmelin, R., Hari, R., Lounasmaa, O. V., and Sams., 1994; Sörös, Cornelissen, Laine, & Salmelin, 2003; Vihla, Laine, & Salmelin, 2006 Reading: Klein et al., 2014; Salmelin, Schnitzler, Schmitz, & Freund, 2000; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999; Wheat, Cornelissen, Frost, & Hansen, 2010). Using these tasks, the LIFG has been attributed a role in both semantic and phonological function, and recent studies (Clarke, Taylor, & Tyler, 2011; Klein et al., 2014; Wheat et al., 2010) show some strikingly early effects in manipulations of both these elements. Wheat et al. (2010) describe a difference in IFG
response corresponding to a phonological manipulation in Rapid Serial Visual Presentation [RSVP] of words within 100ms of stimulus presentation. Klein et al (2014) find effects of phonological priming in the IFG for a single-word reading task in the first 50ms after stimulus, and Clarke, Taylor & Tyler et al. (2011) found that orbitofrontal cortex adjacent to IFG was differently active for basic (CAMEL) and domain-level (ANIMAL) picture naming within 170ms of stimulus presentation. These latencies disagree with estimates provided in serial accounts of spoken word production (Indefrey, 2011), based upon reports of other MEG and EEG studies.

What gives rise to these conflicting accounts of the onset of activity in LIFG, when experimenters are ostensibly manipulating similar properties of their stimuli? Clearly, the effects reported by the studies above are a consequence not only of the experimental manipulation, but of the methods of analysis applied. Common practice is to use a priori regions or time windows of interest to limit the scope of the investigation and ameliorate the multiple-comparisons problem often encountered in modern neuroimaging research. Levelt, Praamstra, Meyer, Helenius, & Salmelin (1998) provide an example of this approach in the time dimension, associating observed activities with particular function on the basis of the latency of peak activation, whatever their cortical location, Clarke, Taylor and Tyler (2011) provide an example of limitation in space. Often these parameters are based on previous studies (Clarke, Taylor and Tyler (2011) take their ROI's from Bar et al., (2006) for example). While the results thus arrived at may be perfectly solid and justifiable, in the mass they present a form of experimenter bias that skews the image of the word production network in meta-analyses and reviews.

Price's 2012 review does an excellent job of summarising the commonly studied areas in the fMRI field, and her exhaustive meta-analysis demonstrates the relatively greater number of effects reported in the left hemisphere over the right in fMRI studies of language (see Table 2, p. 823). This left-handed weight of evidence is similarly apparent in both Salmelin (2007) and Indefrey, (2011), who provide reviews of the MEG literature. This bias is surprising, as in many MEG papers where the whole brain response in language tasks is reported, there is substantial right-hemisphere activity for both receptive (P. Cornelissen, Tarkiainen, Helenius, & Salmelin, 2003; Vartiainen, Liljeström, Koskinen, Renvall, & Salmelin, 2011; Katherine Louise Wheat, 2012) and productive (Gehrig, Wibral, Arnold, & Kell, 2012; Klein et al., 2014; Levelt et al., 1998; R. Salmelin et al., 1994) tasks.

In this context, we wished to carry out an MEG experiment that would allow further exploration of the whole brain response, perhaps give confirmation of early semantic effects in frontal cortices. We also hoped to gather evidence, regarding the onset of activity and difference in areas previously associated with semantic processing in the hope of contributing to the debate on the serial or parallel nature of the cortical network supporting picture naming. We wished to achieve this without using regions or times of interest, in a data-driven and exploratory manner.
Picture naming is a task that fully activates the word production network, enabling a full exploration of the cortical response, but imposing some technical difficulties. We anticipated problems in that articulation in the MEG scanner gives rise to substantial artefactual noise from movement and muscular activity. In order to approximate the contribution of this noise to the measured signal, we envisaged two procedures. Firstly we thought to include a covert naming condition, in which participants would imagine the response without overtly articulating it. This would provide a possible path for the isolation of artefactual effects, as the same stimuli could be repeated in an overt and covert condition, the covert condition providing an approximate model for the cortical response that might be expected in the overt, albeit missing the motor/articulatory elements. Secondly we created a suite of tests in which a volunteer would make movements of their articulators, and have examples of the signals emitted, in the hope of better identifying these as components of the experimental data. In addition to these practical elements, we were also aware that muscular activity gives rise to noise focussed in the gamma frequency band (Kujala, Vartiainen, Laaksonen, & Salmelin, 2012; Muthukumaraswamy, 2013), and planned to adapt our analysis by low-pass filtering the data in an initial ERF analysis of the data.

4.7. **Adapting our experimental aims and procedures to MEG:**

We planned to use an experimental protocol and task that activates the whole of the word-production network, and manipulations of the stimuli (colour and semantic engagement). To avoid limiting our investigation in time or space in the ways mentioned earlier, it was decided that at least the first round of analysis should be exploratory and data-driven in order not to exclude potentially interesting effects by focusing on particular regions of interest. Between conditions contrasts in an exploratory, whole-brain analysis allow us to see differences caused by our manipulations even in unexpected areas, and make inferences about the possible functional role of those areas implicated without relying on reverse inferences from other studies (Poldrack, 2006). This comprehensiveness comes at the expense of potentially confounding factors, exacerbating the problem of multiple comparisons inherent in analysing high-resolution neuroimaging data, and of requiring clear predictions and objectives to avoid becoming lost in the volume of data output by the analysis.

With these elements in mind, we designed a pilot study to test the feasibility of our aims and methodology, to assess the level of articulatory noise present in a picture naming protocol using MEG, and determine how much of this articulatory noise could be separated from cortical signal.
4.8. **Pilot MEG naming study**

4.8.1. **Methods:**

4.8.1.1. **Participants:**

Participants were 4 volunteers from among our team in the psychology department at Aix-Marseille University. Three were females, all were right handed, with mean age 29.75 years. Participants had normal vision, no speech deficits and self-reported as having no other cognitive abnormalities. Participants wore no metallic items, but one had a metal dental implants (a metallic crown on a tooth).

4.8.1.2. **Experimental materials:**

Three experimental conditions were used, in which participants named black line drawings of common objects, or distorted colour images derived from those drawings by dividing them into 4 sections, distorting the lines, rotating, overlaying and colouring them. There were 256 trials in each condition.

The three pilot conditions were:

1. **COLOUR** – Naming a set of 8 colours from distorted coloured lines derived from the 8 drawings of vehicles in the SINGLE condition. Each distorted image was presented in each colour 4 times.

2. **SINGLE** – Black line drawings of 8 vehicles. Each drawing was presented 32 times. The vehicles used here were recycled as part of the MULTI condition.

3. **MULTI** – Black line drawings of vehicles, musical instruments, furniture and tools. There were 8 items in each of the 4 categories, and 8 different examples of each item. Each of these was presented once only, so that each presentation would require a new search and identification of the stimulus picture, even if there was repetition of item names.

4.8.1.3. **Design:**

Participants were verbally briefed on the task, then seated in the MEG scanner and shown an instruction slide repeating the briefing in detail. They were then shown the COLOR, SINGLE and MULTI conditions in an order counterbalanced across participants. Each condition comprised 256 trials, the first 8 of which were to be discarded as ‘warm-up’ trials. These trials were presented in pseudo-random orders created in Mix (van Casteren & Davis, 2006), with the constraints that items names should not be repeated, and that no two consecutive items should have the same onset letter.

4.8.1.4. **Procedure:**
Participants fixated a central cross for 1300ms plus a jitter of up to 1400ms, which was then replaced by a stimulus picture for 2000ms. Participants named the stimulus picture in the format “Des [items]” for the picture conditions and “Du [colour]” for the colour naming condition. Participants were instructed not to blink before or during their response, and to move as little as possible apart from speaking normally.

4.8.2. Data Acquisition: Apparatus

Sound capture system: In order to properly identify the onset of vocal response, it was necessary to record participants' speech during the experiment. Tests using the intercom built into the magnetically shielded room showed insufficient signal to noise, so an AKG C520 microphone fixed to a positionable non-metallic support was placed next to the scanner, connected to an Alesis i/o2 external power source and amplifier outside the scanning room via a metallic XLR cable.

4.8.3. Data acquisition: Sound/stimulus synchronisation procedure

The output from the microphone was recorded on one channel of a stereo track, and the signal from a photo-diode placed in a concealed position on the projection screen was recorded on the other channel. The presence of a stimulus on screen was accompanied by a white patch that shone on the photo-diode, sending either an 'on' or 'off' pulse that was used to signal the onset of the stimulus for the dual purpose of synchronising MEG and sound data, and measuring the latency of response. Sound was recorded at 16 kHz using the open-source Audacity v2.0.0 software (http://www.audacity.sourceforge.net).

4.8.4. MEG scanning procedure

We used the 4D NeuroImaging Magnes 3600 whole head 248 channel scanner at La Timone, sampling at 2034.15Hz. Head shape and position coil location were recorded using a Polhemus Fastrak 3-D digitising stylus. Typically more than 2000 points were captured. Head position was measured at the beginning and end of each run. Participants were tested in a seated position, and padding was placed between the rim of the helmet and participants' heads to minimise head movement.

4.8.5. Analysis: Pre-processing

4.8.5.1. Behavioural data: response scoring:

Behavioural data were scored offline by listening to the recording and cross checking responses with a sheet of the stimulus order. Trials containing incorrect responses, hesitations longer than 1000ms, non-speech vocalisations (“umm”, “err”, etc.) and double utterances
4.8.5.2. **Onset detection:**

A development version of the 'onset detective' software package (https://github.com/lvanderlinden/OnsetDetective) was used to automatically number and separate the individual trials using the photo-diode's 'on' pulse from the sound recording, and the onset of speech in that trial identified using the Teager-Kaiser Energy Operator, a measure of the change in variance in a waveform signal over time. In each segmented trial a baseline measure of the ambient noise in the recording was taken during a 500ms period of silence. The variance of power [K] in this baseline sample was compared with that in a 100ms sliding time window that was advanced from the stimulus latency onwards. If 5% of the samples within this sliding window had energy greater than 12*K, the first sample surpassing this threshold was automatically marked as the onset of response. These latencies were then manually inspected to check for anomalies and corrected by hand to the appropriate time as necessary.

4.8.5.3. **AnyWave: Filtration of continuous data, trigger code conversion**

The continuous MEG data were imported into the AnyWave software package (La Timone MEG laboratory, INS, Marseille, http://meg.univ-amu.fr) and high-pass filtered at 0.1Hz. Trigger codes for stimulus onsets were manually exported for later transfer into other analysis packages. Due to a feature of the MEG/EEG trigger code synchronisation in place at La Timone, we were obliged to manually check and correct a number of triggers that overlapped with an EEG synchronisation pulse on the same channel as our protocol's triggers.

4.8.5.4. **Anatomy and head/brain models: production of surface texture from individual anatomies**

We recorded 4 pilot participants that had had structural MRI scans taken within the previous 5 years, and obtained copies of their anatomical data directly from them. This data was imported first into Freesurfer and BrainVISA, where attempts were made to produce white matter, pial and scalp surface textures for coregistration with the MEG data. This proved unusually problematic due to anomalies in the MRI scans, potentially arising from an abnormally large bias field in the MRI scanner at La Timone. A software plugin for the BrainVISA suite designed to compensate for this is available here: http://marspack.free.fr, but was not successful in automating the process of surface generation for all participants, and time-consuming manual refinement of the surfaces was necessary. Head shape and fiducial points were sampled at the time of MEG scanning using a Polhemus 'Fastrak' stylus. The individual MEG data and anatomical surfaces were imported into the Brainstorm software package and coregistered with each other. All analysis from this point onwards was undertaken using the Brainstorm software suite.
4.8.6.  **Sensor level analysis: Channel rejection**

Flat and noisy channels were manually rejected based on their amplitude, the number of SQUID jumps (artefacts of the electrical control system in the MEG scanner) and expert assessment of each channel's signal, resulting in the rejection of 3 channels across all participants.

4.8.7.  **Sensor level analysis: Segmentation & baseline correction of trials**

Continuous MEG data was segmented into trials from -2000ms to +3000ms around stimulus, DC offset corrected between -2000ms and 0ms, then re-sampled to 300Hz. The period between -1300ms to -800ms was used for baseline correction, as being a period of minimal noise (both cognitive and articulatory) in the trial cycle. The baseline period for response locked analysis was the same period as for the stimulus locked analysis.

4.8.8.  **Sensor level analysis: Artefact/blink rejection**

Segmented data were inspected manually for artefacts using a sub-selection of MEG sensor channels near the rim of the helmet. Trials containing amplitude spikes in the frontal region indicative of blinks, or sudden peaks in amplitude after the average articulation were rejected.

4.8.9.  **Sensor level analysis: Time/Frequency analysis**

After a preliminary visual check of the average MEG sensor level data for each participant in each condition, an exploratory Morelet wavelet T/F analysis was performed in clusters of sensors in the pericentral, occipital, frontal and temporal regions, in the canonical frequency bands (Delta, 0.5-2Hz; Theta, 4-7Hz; Alpha, 8-12Hz; Beta, 15-29Hz; Gamma, 30-59Hz; High Gamma, 60-90Hz)

In the time/frequency analysis there was a strikingly large amount of power in the very low frequency range around 0.3Hz, lasting through the whole trial window, extending back before stimulus onset. There is also a very strong peak visible at the sensor level around the onset of response, as visible in Figure 5.
Figure 5. Pilot participant S02: avg. stimulus locked activity. Time window between -2000ms and +3000ms around stimulus in the MULTI condition. A drift away from the average in the sensor level data was present in the whole trial window, extending long before stimulus onset, and corresponding to a low frequency oscillation.

While it is possible that oscillatory power at this frequency is of cortical origin (Buzsáki & Draguhn, 2004) give an approximate lower bound of 0.05Hz for oscillations of cortical neuronal origin), we thought that it was more likely to be of muscular origin, as it matched better with the cadence of respiration, a cycle necessarily locked to response onset. As participants were aware that they were responding to an ongoing sequence of trials with a fixed onset, we thought they may have settled into a cycle of breathing and articulation that meant slight movements of the body were summed over many trials to produce this slow, powerful oscillation. This supposition is compatible with the fact that this oscillatory cycle is present in the epoch at least 2 seconds before the stimulus is presented, in the period of rest and fixation between trials.

We also suspected that the high gamma band power change around the time of articulation may be of muscular rather than cortical origin, as has previously been found to be the case by other researchers using MEG to examine speech production (Laaksonen, Kujala, Hultén, Liljeström, & Salmelin, 2012 supplementary materials). We therefore decided to focus on an analysis of source level Event Related Fields [ERFs] filtering out the high and very low frequency ranges.

4.8.10. Source level analysis: Potential artefacts in the gamma band:

We performed an exploratory source level T/F analysis using an unconstrained whitened Minimum Norm Estimation [wMNE] of the sources. This was produced using an overlapping
spheres source model, the individual's cortical anatomy and a noise covariance matrix produced from a period of rest before the beginning of the experimental condition.

We cast a group of scouts (groups of vertices in manually selected regions of interest) in regions reported in the literature to subserve key linguistic functions. Using a Morelet wavelet T/F analysis in the canonical frequency bands (see Figure 6 for details), and computing the Z-scores of the power in these bands with respect to the baseline period between -200ms and 0ms pre-stimulus, we observed a strikingly large amount of change in the Gamma frequency band at the temporal poles around the time of articulation (from approximately 500ms post-stimulus onwards). This area is associated with semantic knowledge and categorisation, but it is also proximate to the tongue and jaw muscles, large sources (relative to cortex) of electromagnetic activity during articulation.
Fig. 6a. Pilot participant S02: Z scores of vs. baseline (-200ms to 0ms pre-stimulus) oscillatory power for a cluster of 10 sensors over Left Hemisphere [LH] pericentral cortex. Time window between -200ms and +600ms around stimulus in the MULTI condition. Note the strong delta band change between 200 and 400ms post-stimulus, and the peak of gamma change around 550ms to 600ms.
Figure 6b. Locations of left and right hemisphere scouts in the temporal poles of the COLIN27 template anatomy used for estimation of frequency band power.
Figure. 6c. Z-scores vs. a baseline period of -200ms to 0ms pre-stimulus of the oscillatory power in the temporal poles. Morelet wavelet analysis was used in the Alpha, 8-12Hz, Beta, 15-29Hz, Gamma 30-59Hz (gamma1) and High Gamma 60-90Hz (gamma2) frequency bands.
4.9. *Post-pilot refinements:*

4.9.1. *Training condition:*

In order to familiarise participants with the procedure before they began the experimental trials, a new mini-condition featuring 16 pictures of 8 different animals was created. These practice trials were run while the participant was being installed in the MEG scanner to check their understanding and responses, and to test the function of visual presentation and sound capture equipment.

4.9.2. *Blink controls:*

Despite the generally low level of trials contaminated by blinks, we found that one of the pilot participants showed a systematic pattern of blinks after the response, and another had a tendency to blink around the time of stimulus onset. To reduce the risk of contaminating data in the full protocol with blinks, it was decided to instruct participants to refrain from blinking in the course of normal trials, and implement a pause after every 3rd trial to allow them to rest their eyes.

4.9.3. *Replacement of condenser microphone with optical microphone:*

The microphone used for the pilot experiment was a condenser-type, which operates on the principle of amplifying changes in electrical current through a resistive medium, whose resistivity varies with the compression and expansion due to changes in air pressure on an external membrane. This microphone thus incorporated metallic and electromagnetically active parts, and it was considered desirable to replace it for the experiment proper with a magnetically neutral alternative.

We ran comparative tests of both MEG and audio recording quality with the AKG C520 microphone and a Sennheiser MO2000 microphone, which uses the scattering of light reflected down an optical fibre onto the receiving membrane to detect its movement. This allows all metallic and electrical components to be kept at a distance from the head of the microphone, outside the magnetically shielded room.

Sound of fixed volume was played over the intercom system, and the microphones were placed one after the other in two recording situations. First on the chest of a volunteer, so that the microphone would be displaced by the movement of breathing, and then on the rim of the Dewar helmet with no participant present.

In both situations, the optical microphone produced audio recordings with an audio signal-to-noise ratio approximately 3 times greater than the condenser microphone (see Figure 7), and when fixed in an identical position in close proximity to the MEG sensors, magnetic noise levels were more than twice as high for the condenser microphone (>2000 femtoTesla [fT] as opposed to <1000 fT, see Figure 8).
Figure 8a. AKG C520 condenser microphone: sound recording of pulsed alarm played over MEG shielded room intercom at consistent volume.

Figure 8b. Sennheiser MO2000 optical microphone, recording of identical sound as in Fig 2a. Signal-to-noise ratio was superior to AKG by a ratio of 1.62:1.

Figure 8c. MEG signal recorded using the 2 different microphones in the scanner room. Left: AKG microphone, Right: Sennheiser optical microphone. Scales are identical.
Given the advantages in both MEG and audio signal quality, we employed the optical microphone for the experiment proper. The better audio recordings had the benefit of making response onset detection more accurate, although all automatically assigned onset times were still manually checked.

4.9.4. Fewer trials:

For the pilot study we employed experimental conditions comprising 256 trials each. On average a trial lasted 4 seconds, so each of our 3 conditions took 17 minutes to record, 51 minutes of scanning time total. Even for our experienced and motivated volunteers, immobility and unbroken concentration was difficult to maintain for this length of time, and fatigue affected performance. For the experiment proper, we reduced the number of trials in each condition by half to 128 to minimise fatigue, mind wandering and the risk that participants would move to reduce discomfort.

4.9.5. Dealing with signal artefacts from speech:

We saw high levels of gamma power around the time of articulation both at the sensor and source levels. At the source level, this power was strongest around the temporal poles and basal frontal area, closest to the tongue and larynx. Changes in gamma (30-59Hz) and high gamma (60-90Hz) power were also highly visible at the sensor level in clusters of sensors over the pericentral (sensorimotor) cortex. This region is coincidentally where the mastoid muscles which close the jaw are anchored to the skull.

Muscular activity (particularly during speech) is well known to be a source of noise in the gamma power band for EEG and MEG recordings (de Vos et al., 2010; Muthukumaraswamy, 2013). Thinking that the source level gamma power change in the temporal poles, as well as the sensor level gamma power fluctuations around the time of articulation may be related to this rather than of cortical origin, we decided to exclude these frequencies from our analysis in the experiment proper. We thus focussed on an ERP analysis, using a bandpass filter between 0.5 and 35Hz.

4.9.6. Breathing noise/slow drift: Anticipation of response/removing the prefixed determiner ‘des’/’du’ from naming

As mentioned in the chapter on developing the experimental procedure, we used picture stimuli featuring duplicated items and instructed participants to respond with “des [items]”/”du [couleur]”. This was done in order to have a uniform point (the initial D sound) at which to mark the onset of vocal response. As this onset word was fixed within conditions, participants could thus inspire regardless of the duration of the jitter, and begin their response as soon as the stimulus appeared on screen, finishing the identification of the item as they spoke a more or less elongated determiner. Adding this concern to the problems of dual rather than single utterances, it was decided to eliminate the requirement to pronounce the determiner, and ask only that participants speak the object/colour name. In this way,
participants were less able to predict and plan their response.

4.9.7. Artefact suite and covert condition:

Having found several anomalies in our pilot data that we suspected came from extra-cortical sources, we thought it useful to better understand the separate elements of articulatory movement. We ran a test recording with a volunteer in the MEG scanner, trying to intentionally generate artefacts of muscular noise and articulator movement, and find 'template' versions of these which could be used to identify articulatory noise in, and perhaps separate it from, the task-related MEG signal of cortical origin.

The 'artefact suite' consisted of 4 task conditions, each to be performed on the appearance of the same cue. A jaw movement opening and closing the mouth; a tongue movement, running the tongue back across the roof of the mouth from the teeth to the tonsils, producing a brief humming sound, and repeating the word “caramel”.

Instead, implemented COVERT condition, as in the average, trials should have the same cortical response, and removing that element should leave us with just the articulatory part. We could then isolate this from the signal, and remove it to see what differences and similarities there were between the overt and covert conditions.

4.9.8. Distance to projector screen:

The screen on which stimuli were displayed is free-standing, and must be positioned each time a new participant is installed in the scanner. The distance between the projector and the screen was not controlled in our pilot, meaning that the visual angle subtended by the stimuli was not necessarily consistent between participants. To ensure this was constant in the experiment proper, we measured the distance between projector and screen, and screen and scanner, adjusting as necessary to have the same visual angle and stimulus size for all participants.

4.9.9. Analysis:

4.9.9.1. Behavioural scoring online

For the pilot we performed the response scoring offline, listening to the audio recordings after the event. This required the expenditure of a considerable amount of time, and we switched to assessing the accuracy of responses while participants were being scanned, using a detailed labelling system to note error types. This had the additional benefit of being able to give the participant feedback about their performance and further instruction if necessary.

4.9.9.2. Trigger code identification:
The inter-system synchronisation process at La Timone was altered as a consequence of the trigger code problems encountered during the pilot process. In consequence, it was possible to automate the removal of the EEG synchronisation trigger pulses from the data, and script automatic importation of the stimulus and response onsets into Brainstorm.

4.9.9.3. **Choice of cortical surface template**

Recruitment of participants with existing MRI’s proved problematic. Given the technical problems that we had encountered in constructing satisfactory cortical surfaces and the additional cost of performing individual MRI scans, it was decided that we would use the Montreal Neurological Institute's COLIN-27 template anatomy to model activity at the source level. Since our intention was to use unconstrained source orientation, and average our sources at the group level after modelling, it was thought that the loss of anatomical specificity would not too adversely affect the spatial resolution at the group level.

To achieve this group level average, we collaborated with Brainstorm's developers to develop a point-for-point warping process where the template anatomy was warped to the individual's space on the basis of the head-shape recorded during the MEG scanning session. This final warping process is described in the next chapter, as a part of the empirical report.

4.9.9.4. **Change of baseline period:**

Since we were focussing our first analysis on the stimulus locked data only, we decided to move the baseline period for our epoch to the period of minimal difference with the post stimulus activity, -200ms to -0ms before stimulus onset.

4.9.9.5. **Choice of descriptive and inferential statistics:**

Minimum norm source estimation in Brainstorm provides a measure of current source density across a mesh of vertices modelling the cortical surface. This comprises a measure of current in 3 orthogonal directions, and we chose to use the norm of the power in the 3 directions to inform a Z-score for each time-sample post stimulus, the baseline measure being the average variance in the 200ms prior to stimulus onset.

As an inferential statistic for the between conditions comparisons, we used t-tests to compare the populations of group average Z-scores at each vertex on the template anatomy across all time samples.

4.9.9.6. **Choice of statistical thresholds:**

Our analysis of the pilot ERF data displayed the expected pattern of activity (ergo, an early peak of vs. baseline activity in visual cortex, followed by a forward spreading wave of activity in the dorsal and ventral language streams. This assessment was made solely on the base of a descriptive statistics (Z-scores vs. a pre-stimulus baseline period). To improve upon this in a
disciplined way, we wished to implement a between conditions comparison using a statistical threshold to calculate inferential statistics. We also wanted to have a minimum of anatomical and functional constraints on our analysis both to avoid any bias, and to retain the opportunity of discovering unexpected effects of our manipulation with an exploratory analysis.

Given the high temporal (scanners can sample at rates of up to ≈2kHz) and relatively high spatial resolution of MEG (standard template models of cortical anatomy used for analysis are made up of >15,000 vertices), such an exploratory approach encounters the multiple comparisons problem in calculating inferential statistics between conditions, but this can be countered using various correction methods, such as permutation testing (Maris & Oostenveld, 2007; Nichols & Holmes, 2001), false discovery rate correction (Efron, 2005) and the random field theory method (Pantazis, Nichols, Baillet, & Leahy, 2005).

Our template cortical surface was rendered as 15002 vertices, with an MEG sampling rate of 2034.15Hz, meaning that to test a statistic on all vertices between 2 conditions in an experimental epoch of 600ms would require 18,309,790 comparisons. Under these circumstances, the classic Bonferroni and False Discovery Rate corrections for multiple comparisons would impose an unrealistically high threshold on any potentially significant between conditions differences. We therefore opted to use the local False Discovery Rate (IFDR) correction method described by Efron (Efron, 2005). This statistical technique compares the observed frequencies of p values in the tails of the distribution of observed values to that expected for a model distribution, and reports values that differ from that model distribution beyond a certain threshold. Only a proportion of the most extreme values reported as being significant in the original test are admitted as being truly significant, the others are ignored as being likely to contain erroneous or false 'discoveries' of a non-null effect of the manipulation. In Efron's paper he recommends that the 'q' threshold for this proportion of values be set at 0.2, and we follow his recommendation.

4.9.9.7. Minimal values for significant differences in time and space:

The IFDR method of error rate correction is more permissive than family-wise error correction (it runs less risk of Type II error), so with the purpose of further reducing the risk of reporting false positives and limiting the number of between conditions differences we report and discuss, we decided to take only those differences present in 6 or more adjacent vertices that lasted for at least 20 consecutive milliseconds (around 40 samples).

4.9.10. Summary of developments:

As a result of the piloting process, substantial modifications and improvements were made to our experimental protocol, technical execution and analysis. With regards to the procedure, a training condition was implemented to familiarise participants with the stimuli and their task during their induction and installation. We also included the requirement for participants to control their blinks during normal trails, restraining themselves until an allocated blink period. We reduced the number of trials per condition, hoping to reduce participant fatigue, but
introduced a covert condition, in which words were imagined but not spoken, in the hope of having a reference version of the cortical signal free of articulatory artefacts.

The technical setup and apparatus were also changed as a result of our tests; the electronic microphone was replaced with an optical model, which generated much less noise in the scanner. We also fixed a standard for the positioning of subjects with respect to the stimulus projection screen, and behavioural scoring was done online, resulting in an improvement of standardisation of the procedure, and a substantial time saving in processing each participant respectively.

Learning from the signal recorded from pilot participants, we set a priority for the type of analysis we would focus on, and boundaries on the frequencies of filtration that excluded the majority of power generated by artefactual sources of noise. The construction of a pipeline for automatically processing trigger codes and importing raw data into Brainstorm, our analysis suite of choice, was another substantial saving of time. Over the course of the pilot investigation, we also changed the baseline period from one distant from stimulus (-1300ms to -800ms pre-stimulus) to be immediately before stimulus (-200ms to 0ms pre-stimulus), seeking to have the baseline as close to the data observed in the post-stimulus epoch as possible.

Confronting the complexity of our pilot data also refined our ideas about the choice of cortical anatomy that we would use for our analysis, moving us away from the recording of individual structural scans to using the COLIN 27 template for reasons of time and the simplification of our analysis. We also developed a data-driven pathway for the analysis of the recordings from the full experiment, allowing us to identify regions of significant difference between condition with no a priori regions of interest or other assumptions, while retaining strict statistical controls.

These refinements to our experiment lead us to the next section of this thesis, the execution of a full empirical investigation of the cortical response during picture naming.
4.10. References


5. Early semantic and late visual differences in the cortical response during picture naming.

5.1. Introduction:

While the production of spoken language has already been subjected to a good amount of neuroimaging analysis, the debate about whether the brain processes the functional components of speech serially or in parallel continues. Debate hinges on assigning particular functions to cortical areas, and observing the timing of their activity and interaction. Picture naming is often used as the experimental task in word production studies, as it thought to provide full engagement of most components used in normal speech (Glaser, 1992). The key components being: recognition of an external stimulus, access to semantic, lexical and phonological properties, followed by the conversion of these phonological codes into syllabic forms, articulatory motor programs and the ultimate articulation of a word (Dell, Nozari, & Oppenheim, 2014; Goldrick, 2007; Indefrey & Levelt, 2004).

For example, in an update to the WE AVER++ model of speech production (Levelt, Roelofs, & Meyer, 1999), Indefrey (2011) proposes a sequential relay of information between cortical areas each responsible for a discrete function in picture naming. Areas dealing with visual object recognition and “conceptual preparation” pass information to those which govern lemma selection, which in turn inform phonological code retrieval and so on through syllabification and articulatory preparation to motor output. The logic of this model states that a function, and thus the area processing that function, should not be critically involved until a ‘selection point’ at preceding stages of processing has been reached. The period of a function's critical involvement is associated with an increase in activity. Hickok presents an alternative in the hierarchical state dependent feedback [HSFC] model (Hickok, 2012). This more parallel explanation of speech production picks up after a lemma has been identified, and suggests simultaneous onset of activation in a parallel network of areas followed by sustained activity during articulation. Allocation of function to area in these models draws from studies that manipulate component functions of the rather than investigations of the global network, something seldom described in a single imaging study.

There are a few MEG papers that do provide descriptions of whole-brain cortical dynamics of word retrieval and articulation. The first was authored by Salmelin and colleagues (R. Salmelin, Hari, Lounasmaa, & Sams, 1994), who performed an exploratory study of picture naming (i.e. observing the global response and not performing a manipulation of stimulus properties). Their analysis used dipole fitting and a simplistic estimate of the cortical anatomy, which gave relatively poor information about anatomical structures involved in the response. Levelt, Praamstra, Meyer, Helenius, & Salmelin (1998) give another global description of the picture naming network, but attribution of function is applied a priori on the basis of when cortex is active, not by confirming their hypotheses with a contrast between conditions. Liljeström, Hultén, Parkkonen, & Salmelin (2009) give a full description of the picture naming network using a specific manipulation. In this study, the same protocol was performed in both fMRI and MEG, and the authors report different effect types (fMRI as well as MEG Minimum Norm Estimation [MNE]/Equivalent Current Dipoles [ECD]) with good spatial and temporal resolution in the whole brain. They also perform a semantic manipulation (actions vs. objects), enabling them to associate cortical areas with semantic function with good temporal resolution. Gehrig, Wibral, Arnold, & Kell (2012) perform an elegant analysis of oscillatory power in a whole-brain source level reporting...
of effects, reporting whole-brain connectivity data. Regrettably, they use 1000ms windows for their analysis, rendering it impossible to make inferences about the temporal dynamics of the involved areas during the analysis window. Lastly, (Klein et al., 2014), report a full time/frequency activation map of the cortical response to their masked phonological priming single word reading stimuli.

The studies reporting the whole-brain response describe an extensive bihemispheric pattern of activity in picture naming (and overt reading), showing a spread of activity from visual cortex forward through the dorsal and ventral streams (Creem & Proffitt, 2001; DeWitt & Rauschecker, 2013). Despite these repeated reports of bihemispheric activity, even some recent neurolinguistic proposals focus only on left hemisphere areas to describe the picture naming network (Indefrey, 2011), and effects in the right hemisphere are seldom discussed in these language studies. Equally, reports of the latency of activity in areas considered key to word production are varied: in IFG, Klein et al. (2014) find differences between phonologically primed and unprimed target utterances starting within 100ms of stimulus, while Salmelin et al. (1994), Sörös, Cornelissen, Laine, & Salmelin (2003), Vihla, Laine, & Salmelin, (2006), and Hultén, Vihla, Laine, & Salmelin (2009), report onsets of activity at 400ms, 200ms, 200ms and 500ms respectively for this area.

While they are a rich source of information, these studies do not provide appropriate data to attribute function to all of the structures seen to be active in the naming network. This is partly due to the scope and aim of the investigations reported, which are usually intended to study the dynamics of one function or another and not the whole network. Authors limit their investigations with a priori regions or time windows of interest (e.g. Clarke, Taylor, & Tyler, 2011; Gehrig et al., 2012; Levelt et al., 1998), are concerned by particular areas or functions, and exclude unrelated findings from their discussion. Even when data is collated from these different studies, they cannot unequivocally discriminate between the parallel and serial models (see Chapter 2, Munding, Dubarry, & Alario, 2015).

5.1.1. Aims:

There is diversity of timing reported for various component functions of speech, and extensive undiscussed active regions. Given this state of affairs, we wished to supplement the existing evidence with a data-driven, exploratory analysis of the network. Using manipulations of visual and semantic qualities of stimuli in a new picture naming protocol, we hope to recover information that will contribute to the serial/parallel debate.

We conducted a picture naming study employing a visuo-semantic manipulation of the stimuli. This allowed us to identify areas associated with these functions and investigate the dynamics of their response. We designed 3 sets of stimuli to differently require semantic engagement in the picture naming task.

1. Colour noise patches designed to have minimal semantic content, as discerning the colour relies only on sensory information.
2. A simple set of 8 black-and-white line drawings of vehicles, designed to have a small and predictable semantic 'target space'.
3. A complex set black-and-white line drawings of items from 4 semantic categories (vehicles, furniture, musical instruments and tools), to provide a larger and unpredictable semantic 'target space', thus requiring the most cognitive processing of our 3 stimulus sets.

We expected the coloured stimuli to be more stimulating to visual areas, and hypothesised that cortical activation would be more extensive in general, and more right-lateralized in the effects of the semantic manipulation than often reported. We further thought that LIFG and temporal pole would prove to be more strongly active in more semantically loaded conditions.

5.2. Methods:

We were guided by the advice laid out in the valuable 'MEG good practice' paper (Gross et al., 2012). The study received ethical approval from the Comité de Protection des Personnes Sud Méditerranée I (ID RCB : 2011-A00562-39).

5.2.1. Participants:

We recruited seventeen right handed native French speaking adults (mean age 26, SD 3.18 years, 13 females) with no reported speech or language deficits. All gave informed consent to their participation in the study, and were compensated EUR 10/hour. Participants all had normal vision or vision corrected to normal with contact lenses, wore no cosmetic products, metallic or magnetic jewellery or piercings, and had no surgical implants or metal inclusions in their clothing. All participants were subjected to a pre-scanning noise test in the scanner, and those whose noise levels were deemed abnormally high by the technical staff were not scanned. Two participants passed this screening process despite having dental implants.

5.2.2. Experimental Materials:

Stimuli were simple black line drawings of objects and randomly scrambled patches of colours. These were presented on a grey background, and participants were instructed to name them aloud quickly and accurately.

Different RGB colour values and candidate picture items were tested on 5 native French speakers (none of whom participated in the experiment) for name agreement before testing. As a first step, 8 colour names were established as being easily recognised and having 100% name agreement in our pool. Then, pictures of common (i.e. frequently encountered), inanimate, and easily identifiable objects were identified, their lexical frequency being paired with the previously established set of colour names, using occurrences per million in books, as listed in the LEXIQUE database (v3.55) (New, 2006).

The final stimuli were drawn from those that had 100% name agreement within the test group. Images were black line drawings obtained either from copyright-free sources, or used under the provisions of fair use. Pictures were matched for size (to have a maximum height or width of 600 pixels), line thickness, and were adjusted to be of uniform contrast/colour intensity. Colour stimuli were
randomisations of the 8 vehicles in the SINGLE condition, 1 scramble for each picture coloured in each of the 8 colours, each scramble/colour combination presented twice. All stimuli were presented against a 50% grey background to reduce screen glare, and allow the projection of white in the COLOUR condition.

The final version of the experiment comprised the following materials:

- **TRAINING** - A set of 16 pictures, 8 animals, 2 examples of each animal, used to familiarise participants with the procedure and experimental task during setup and installation.
- **Colour naming [COLOUR]** – 8 colours, represented by random scrambles of coloured lines, these being the line drawings from the single condition (below) cut into 20 pixel squares and randomly rearranged.
- **Single semantic category [SINGLE]** – line drawings of vehicles. There were 8 items, each represented with a single repeated picture. Vehicle names were the same as those in the MULTI condition.
- **Multiple semantic categories [MULTI]** – line drawings of items from four semantic categories: vehicles, musical instruments, furniture and tools. There were eight different objects in each category, each represented by means of four different exemplars (i.e. 8 names and 32 different pictures per category and a total of 128 for the whole condition).

5.2.3. **Design:**

Participants were seated in the MEG scanner and familiarised with the procedure using a short training run of animal pictures. Each participant was then shown the COLOUR, SINGLE and MULTI conditions in three different blocks, whose order was counterbalanced across participants. Then, the MULTI condition was presented again as a covert naming condition. Lacking overt responses, this condition is not analysed further for this report. Each condition comprised 128 trials, the first 16 of which were to be considered as ‘warm-up’ trials. Trial order was pseudo-randomised using the Mix program (van Casteren & Davis, 2006), so that neither a response item nor items with identical onset phoneme were repeated in consecutive trials.

5.2.4. **Procedure:**

Participants were instructed to fixate a central cross presented for a jittered duration (range 1300-2700ms), before being replaced by the picture stimulus to be named presented for 2000ms, followed by another presentation of the fixation cross during a ‘recovery’ period for a jittered duration (jitter range 1300-2700 ms). Jitter durations were at steps of 200ms between 0 and 1400 ms (0, 200, 400, 600 etc.). Each jitter duration was presented 16 times, in a new random order for each condition and participant.

Participants were instructed to fixate on the central cross, and quickly and accurately name aloud the stimulus items (objects or colour patches) as soon as they appeared. In the additional covert condition, they were told to internally name items without speaking. Participants were instructed not to blink outside specified rest periods signified by the presence of the symbols o.o on screen for 3 seconds.
after every third trial. Blink periods were followed by a fixation cross during 1000ms before the next 3 experimental trials were presented. Participants were instructed to move as little as was consistent with natural speech. They were told that there would be four experimental conditions, one for each of the four different blocks.

5.2.5. Data Acquisition:

MEG data were recorded using a 4D Neuroimaging Magnes 3600 Whole Head 248 Channel scanner, the 248 magnetometers arranged in a helmet-shaped array. Data were sampled at 2034.15 Hz. Head shape and position coil location were recorded using a Polhemus Fastrak 3-D digitising stylus. Typically, more than 2000 points on the scalp surface were captured. Head position was measured at the beginning and end of each run. Real-time head tracking was not available.

Two EOG and two EMG electrode pairs were placed on participant’s faces. Vertical and horizontal EOG electrodes were placed on the left eye. EMG electrode pairs were placed across both sides of the orbicularis oris muscle of the top lip, and the depressor anguli oris muscle on the left side of the bottom lip/chin, recordings from these electrodes were also captured by the MEG system at 2034.15 Hz.

Participants were tested in a seated position, head positioned as close to the rear of the helmet as possible, and with the top of the head touching the helmet. Padding was placed between the rim of the helmet and participants' heads to improve comfort and minimise head movement. Distance between the participant’s eyes and the screen on which stimuli were displayed was 80.8cm (mean measurement, SD 2.1cm), stimuli thus subtending the same visual angle for each participant.

A trigger square invisible to the participant was projected onto a photodiode adjacent to the projector screen, to signal the presence of a stimulus picture on-screen, the signal from which was used to synchronise the MEG, EOG, EMG and sound recordings. Vocal responses were recorded using an MEG compatible optical microphone (a Sennheiser MO2000 optical microphone). Sound was amplified using a Roland Quad-Capture audio channel box, and recorded in stereo at 16kHz using Audacity 2.0.0 [http://www.audacity.sourceforge.net] one stereo channel recording the participant’s voice and the other the synchronisation signal generated by the photodiode.

5.2.6. Data pre-processing:

Continuous data were band-pass filtered between 0.5 and 35Hz using a fourth-order forward–backward Butterworth filter (AnyWave software, MEG laboratory, INS, http://meg.univ-amu.fr). All further processing and analysis was performed with the Brainstorm software suite (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011), which is documented and freely available for download online under the GNU general public license (http://neuroimage.usc.edu/brainstorm), or with specific Matlab code written for our project.
The eye-muscle EOG channels and a selection of frontal MEG channels at the helmet rim were used to manually detect and reject sections of the continuous data containing blinks, based on the amplitude of activity in the channel and the characteristic pattern of the artefact.

Behaviourally bad trials in which the participant gave an inappropriate response, did not respond or uttered any vocalisation other than a clear response were also rejected. The first 16 trials of each run were rejected as a 'warm-up' period. One participant was removed from the analysis at this stage due to the very small number of trials remaining after blink rejection, two due to very high levels of noise in a cluster of MEG sensors at the lateral-frontal helmet rim and a fourth due to a weak and noisy visual evoked potential. This left a total of 13 participants for full analysis. Finally, two MEG channels were removed entirely from the analysis, one being flat, and the other unacceptably noisy.

5.2.7. Source based analysis:

Continuous data were segmented into 1000ms epochs between -300ms and +700ms with respect to stimulus onset for each trial, and trials were averaged within conditions for each participant. Source level analysis was then performed on this within-condition averaged data for each participant.

Participant’s head shapes were co-registered to that in the MNI Colin27 template anatomy (Homes et al., 1998) in Brainstorm using the fiducial points, and the surfaces of the default anatomy were warped to the individual participant’s digitised scalp surface, ignoring the most aberrant 10% of digitised points.

These scalp surfaces were used to create an overlapping spheres model (Ermer, Mosher, Huang, & Leahy, 2000) generated in Brainstorm for each participant to project sources to the pial surface of the default anatomy using a mesh of 15002 vertices to represent the cortical surface.

Whitened Minimum Norm Estimation [wMNE] (Baillet, Mosher, & Leahy, 2001) was used to model unconstrained sources on the default anatomy for each participant. Noise covariance matrices were calculated for the wMNE using 10 seconds (approx. 20,000 samples) of artefact free data, captured in an eyes-open resting period immediately prior to each condition run.

For each participant and condition, the log of the norm of estimated magnetic field vector at each vertex (log(sqrt(x^2+y^2+z^2))) was calculated. These values were then used to compute Z-scores for the 700ms post stimulus period against the baseline period (-300ms to 0ms pre stimulus).

Z-scores for each vertex were averaged across participants, and a paired t-test was run concurrently on time (2034 samples) and space (15002 vertices) to find significant differences between conditions. The local false discovery rate [lFDR] correction was applied to account for the multiple comparisons problem (Efron, 2005). Clusters of > 6 vertices that passed an lFDR threshold of q = 0.2 for at least 20ms were taken to be statistically significant. In addition, we also report contrast values lying near these thresholds as “marginally significant”.
5.3. **Results:**

After behavioural and artefact rejection, 4188 trials remained across 3 conditions (see Table 1) and 13 participants, 804 trials having been rejected.

<table>
<thead>
<tr>
<th></th>
<th>COLOUR</th>
<th>SINGLE</th>
<th>MULTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good Trials (N)</td>
<td>1432</td>
<td>1361</td>
<td>1395</td>
</tr>
<tr>
<td>Mean RT (ms)</td>
<td>654</td>
<td>738</td>
<td>827</td>
</tr>
<tr>
<td>St. Dev of RT (ms)</td>
<td>165</td>
<td>205</td>
<td>243</td>
</tr>
</tbody>
</table>

Table 1. Trials remaining after behavioural and artefact rejection. All included participants had more than 80 good trials per condition remaining after rejection.

Despite a protocol that demanded speech, participants head movement (measured at the beginning and end of each run) were considered to be within acceptable bounds. The average movement over the course of a run was 0.36cm, with a S.D. of = 0.38cm.

5.3.1. **Descriptive statistics (vs. baseline changes)**

5.3.1.1. **Sensor level analysis:**

In all 3 conditions, stimulus locked activity at the sensor level begins with a peak in global field power around 100ms, concentrated over occipital areas, and a second clear peak close to 200ms subsisting mainly in occipital and occipito-temporal sensors. Thereafter there is less focused and more widespread activity across the sensor grid until a large, late deflection, principally present in frontal and fronto-temporal sensors (those around the frontal rim of the helmet). In the COLOUR condition the deflection is at ~450ms post stimulus (average voice onset in the colour condition is at 654ms), deflections occur in the SINGLE condition at ~550ms post-stimulus (average voice onset at 738ms), and the MULTI condition at ~600ms post-stimulus in the MULTI condition (average voice onset at 827ms).
Figure 1. A: Example evoked field at the sensor level for one representative participant in the 3 experimental conditions. Bracketed numbers below the condition name indicate the number of good trials used from a total of 128 per condition. Note the peak at ~100ms relating to the visual evoked potential and the large deflection in the COLOUR condition at ~600ms, possibly relating to artefacts of articulation. B: Source level statistics for group average data: Z scores vs -300 to 0ms pre-stimulus for log of norm of amplitudes in the 3 cardinal dimensions at each of 15002 vertices in the Colin27 template anatomy. Image thresholded at Z > 6.3 (approximately equivalent to p < 0.00001, uncorrected).

5.3.1.2. Source level analysis:

Overall, visual naming involved a widespread, bilateral network of cortical areas, with post-stimulus ERF activity (vs. the -300ms to 0ms) beginning in primary visual and occipital cortices before spreading forward to parietal and posterior temporal cortices by approximately 200ms, and reaching anterior temporal and frontal cortices by 300ms.

90-110ms window:

The data displayed in Figure 1 show that the first peak in sensor level power (around 100ms across participants) is associated with strong changes in source level activity, centred in visual cortex and visible on both the lateral and medial surfaces of the posterior occipital lobe, but also in extra-striate regions bilaterally.

190-210ms window:
The second peak in global field power (seen around 200ms across participants) shows a more disperse pattern of activity at the source level, the areas where change is present still being largely posterior, but with vs. baseline change spreading bilaterally forward into parietal and posterior temporal cortices, and even into pericentral areas. In the average time-window between 190ms and 210ms post-stimulus, there are notable vs. baseline differences in occipital cortices, presumably revealing a sustained visual response.

Posterior inferior temporal and posterior basal temporal areas (not shown in fig. 1) are active in all conditions, and there are isolated patches of change in the anterior temporal lobe bilaterally. Cortex in the left and right pre- and post-central gyri (potentially at the level of the lip, tongue and larynx areas) shows some patches of activity, predominantly in the right hemisphere.

290-310ms window:

Activity persists in inferior and basal temporal areas now spreading further forward from posterior to patches of the mid-region, but stronger and more widespread in bilateral temporal poles. TPJ is still active in all conditions, with activity stronger near the perisylvian region (Hickok’s area “Spt”). In the frontal lobe there is a large patch of vs. baseline activity in the frontal portion of insula in both L and R hemispheres, as well as inferior frontal gyrus and basal frontal areas.

390-410ms window:

The observed response is now dominated by basal frontal and basal temporal activity. Versus baseline changes in temporal cortices are bilateral and strongest at the temporal poles, but present in the mid temporal cortex for all conditions. There are clear patches of change in the left primary sensory and motor cortices around the level of orofacial, tongue and larynx areas, tailing away dorsally. Interestingly, there is far less vs. baseline change in right hemisphere sensorimotor areas. The LIFG, anterior insula and bilateral ventral BA11 (ventral frontal cortex) are now active in all conditions. The vs. baseline changes in activity observed in the left anterior supramarginal gyrus/parietal perisylvian region persist in this window.
5.3.2. **Significant differences between conditions**

The clusters passing the significance levels described in the methods (Efron’s suggested IFDR threshold of $q < 0.2$ for 20ms or more) are graphed in Figures 2 and 3, and discussed below.

**SINGLE-COLOUR contrast:**

Figure 2 shows the vertices that were significantly different between the SINGLE and COLOUR conditions. The first differences seen, M1 (right hemisphere) and M2 (left hemisphere), are spatially and temporally proximate, being mirrored across the inter-hemispheric cleft in medial frontal cortex (Brodmann area 8) and starting at 90 milliseconds after stimulus. The differences in M3 starting at 115ms, and M4 at 125ms present another “matched pair” of activations across the hemispheres in cingulate cortex.

![SINGLE vs. COLOR](image)

**Figure 2.** Significant differences between SINGLE (blue) and COLOUR (red) conditions. The areas of difference are labelled according to the cortical area in which they were seen, (O = Occipital, P = Parietal, T = Temporal, F = Frontal and M = Medial surfaces), followed first by a number indicating their rank in the order of onsets in that area, then by the post-stimulus latency of their onset in milliseconds. Z-scores of the log of the norm of each vertex were compared with t-tests and local false discovery rate corrected. Clusters of >6 vertices where $q < 0.2$ for 20ms or more are reported. X-axis = time Y = amplitude. Traces represent Root Mean Square average Z-score in clusters. Labels’ initial letter represents the region, the first number ranking onset order in that region, the second the ms latency of difference onset.
Occipital differences begin slightly later at 115 ms post-stimulus, with O1 approximately in the right collateral sulcus, between the lingual and fusiform gyri. Differences in temporal cortex for this contrast begin simultaneously with O1 at T1 in the right hemisphere (at 115ms post stimulus), bridging the inferior temporal sulcus between inferior temporal gyrus [ITG] and the fusiform gyrus. At area T2, cortex in the left anterior parahippocampal gyrus shows significantly greater activation for the SINGLE than COLOUR condition beginning at 402ms. This is followed by significant a difference in T3, in the left anterior superior temporal gyrus, beginning at 465ms post stimulus.

**SINGLE-MULTI contrast:**

The significant differences from the second contrast, SINGLE-MULTI, are displayed in Figure 4. Here the area label numbers again increase with the order of onset in their particular lobe, taking into account the area numbers from the previous contrast.

**Figure 3.** Significant differences between the SINGLE and MULTI conditions. Labelling and statistics as in Fig 2, taking into account the area numbers from that figure. The red trace represents the Z-scores in the SINGLE condition, the blue trace those in the MULTI condition, and the green line the time samples in which a significant difference is present.

The earliest difference observed in the SINGLE-MULTI contrast is at O2, in the left ventral occipital cortex in BA19, at the junction of posterior fusiform gyrus and cuneus, with an onset of 110ms post-stimulus. Second is T4 at 204ms post stimulus in the right medial face of the temporal pole, followed
by P1, an area in the right lateral fissure at the temporo-parietal junction, shows significant differences at 233ms post stimulus. A difference in T5, in the left posterior ITG, begins at 475ms post stimulus. In O3, right medial peri-calcarine cortex shows a between conditions difference where the MULTI condition is more active than the SINGLE starting at 527ms. The last observed difference in this contrast is at 580ms in F1, right anterior inferior frontal gyrus between IFG pars orbitalis (BA47) and anterior pars triangularis (BA45).

5.4. Discussion:

Our aim was to explore the extent and dynamics of the cortical network involved in visual naming. Using a picture naming protocol with semantic and visual manipulations in MEG, we succeeded in confirming a pattern of activity in a network of areas that has seldom been described in empirical studies. We also saw vs. baseline and between conditions differences in temporal, parietal and frontal cortex at latencies not predicted by existing serial models of speech production.

5.4.1. Dynamics of the picture naming network:

90-110ms post stimulus

In the first window, cortical activity vs. baseline in all conditions shows strong changes in activity centred in visual cortex bilaterally. There are also vs. baseline differences in the right ventromedial mid temporal lobe, responsible for complex visual analysis such as face, place, object and word recognition (Haxby et al., 2001; Scherf, Behrmann, Humphreys, & Luna, 2007; Vinckier, Dehaene, Jobert, Dubus, & Sigman, 2007). This pattern is similar to previous descriptions of the picture naming network. Less often reported are differences in the right insula. Similar activity was previously observed but not analysed in a MEG word/pseudoword reading task performed by Vartiainen, Liljeström, Koskinen, Renvall, & Salmelin, (2011). The insula has previously been implicated in motor control of speech and singing (Ackermann & Riecker, 2004; Dronkers, 1996). The insular/inferior frontal and temporoparietal areas are linked by the uncinate fasciculus, a possible pathway of communication between these two areas of activity. We also observe change vs. baseline in the limbic (cingulate) gyrus in medial frontal cortex, which Damasio (1998) associates with motor control, but which has also been associated with articulation (Robinson & Grossman, 1997).

Overall, the early state of the network reveals the engagement of visual areas (for stimulus identification) and motor areas (possibly generic, i.e. not modulated by stimulus) as well as areas involved in speech control.

190-210ms post stimulus

Versus baseline activity in the second window spreads forward from the occipital cortex to inferior parietal and posterior superior temporal cortex bilaterally. This is associative cortex, specifically referred to in Hickok's HSFC model as responsible for the sensory/motor transformation step in syllable level representations (Hickok, 2012). The angular gyrus has also been found sensitive to semantic
manipulations in fMRI language tasks, active generally for performing actions (Price, 2012), is assigned as phonological retrieval and self-monitoring by Indefrey, (2011), and phoneme and word recognition by DeWitt & Rauschecker, (2013). In a review of the neuroimaging literature on the semantic system, Binder and colleagues state that “the AG likely plays a role in complex information integration and knowledge retrieval. Given its anatomical location adjoining visual, spatial, auditory, and somatosensory association areas, the AG may be “the single best candidate for a high-level, supramodal integration area in the human brain” (Binder, Desai, Graves, & Conant, 2009).

There is sparse activity in the anterior temporal lobe, an area associated with semantic knowledge/retrieval (Visser & Lambon Ralph, 2011). In a study of 64 aphasic patients (Schwartz et al., 2009) it was found that the left anterior temporal lobe plays a “specific and necessary role” for mapping concepts to words.

Overall activity in this second time-window shows activity spreading forward from perceptual areas to associative cortex and areas implicated in semantic and phonological processing.

290-310ms post stimulus:

The dissipation of versus baseline occipital activity in the 290-310ms post stimulus window for all conditions suggests that initial, intense visual processing is over, although some patches of difference remaining in caudal occipital cortex may reflect the maintenance of a visual representation communicating with other areas.

There is a vs. baseline increase in activity in ITG and basal temporal cortex, which (as previously mentioned) is high-level visual processing cortex often associated with reading and object recognition (Price & Devlin, 2003; Price, 2012). Increased activity in temporal poles during this period suggests more intensive semantic processing.

The increasing activity in the insula bilaterally can be explained in terms of motor control/preparation. This is more compatible with a parallel model of speech production, where multiple cortical areas contribute to response selection, information from all processes being integrated to arrive at a final decision. LIFG activity in this early window fits with the idea that it is involved in arbitrating response selection (Sahin, Pinker, Cash, Schomer, & Halgren, 2009; Thompson-Schill, Bedny, & Goldberg, 2005) than the late role in syllabification and articulation ascribed by serial models (Indefrey, 2011) (although this activity is still compatible with an early role in syllabification in a parallel system). The activity in basal-frontal cortex is also suggestive of semantic processing, but no significant between-conditions differences arise from our experimental manipulation in this area.

In summary, activity in this third time-window reveals a fading of low-level perceptual processing, the presence of high-level visual and semantic processing, as well as activity in response selection and motor control/articulation areas.

390-410ms post stimulus:

Widespread dorso-frontal and anterior temporal lobe/temporal pole activity is very pronounced in this later window, and a right lateral bias reassures us that this activity is not purely artefactual, as it should
not be present if this signal were principally due to muscle/movement noise from symmetrical articulators. Clarke et al. (2011) show similar medial anterior temporal activity in the window 200ms-258ms post stimulus and basal frontal activity at 170ms-258ms post stimulus, with a right lateral bias. We also observe activity in bilateral mid-temporal cortex and IFG, both in pars opercularis and triangularis. There is also a strong response in the left anterior supramarginal gyrus [aSMG] and parietal perisylvian regions, which have a role in associating different modalities of information in the speech network.

Overall activity in this fourth time-window reveals ongoing activity in semantic processing areas and basal temporal language processing regions, plus response selection and association cortex.

5.4.2. **Statistically significant between conditions differences:**

*Early visual differences:*

The differences observed in **O1** (115ms) in the COLOUR vs. SINGLE contrast and **O2** (110ms) in SINGLE vs. MULTI are during the visual evoked potential. These differences are easily ascribed to the different visual complexity of the stimuli across conditions. **O1** is explainable as a consequence of the visual differences in the COLOUR and SINGLE conditions. Activity is stronger in the for COLOUR stimuli, with two periods of significance the second around 240ms. This we attribute to the presence of colour information and an increased number of boundary edges in the COLOUR condition stimuli being more stimulating at a low level of visual processing. In **O2**, the limited number of pictures in the SINGLE condition may make it easy to recognise previously seen pictures, requiring less cortical activity.

**T1** (115ms) in SINGLE vs. COLOUR is contemporaneous with the aforementioned occipital-visual differences, suggesting a similar origin. Ventral temporal areas are important for higher level visual processing and object recognition (Binder et al., 2009; Binder & Desai, 2011; Dehaene, Cohen, Sigman, & Vinckier, 2005; Epstein & Kanwisher, 1998; Haxby et al., 2001; Price & Devlin, 2003; Vinckier et al., 2007).

*Early extra-visual differences:*

We found differences outside visual cortex far earlier than might be expected in a sequential model of processing. In the SINGLE vs. COLOUR contrast we found 2 paired patches of difference in medial frontal cortex.

The first pair of differences, **M1** and **M2** are close to the frontal-eye-field [FEF] in BA8, in the ventral part of the superior precentral sulcus, rostral to the hand motor area (Paus, 1996). These differences may thus be due to differences in how gaze is directed around the different stimulus types. More evidence to support this explanation is present in the average EOG traces, where subtracting the power in the SINGLE condition from that in COLOUR shows a difference in horizontal eye movements, beginning at 200ms. If the peak of the visual evoked potential occurs at 90ms post stimulus, the time to program and return a saccade (or micro-saccade) command might be in the same order of latency, meaning our FEF difference could match up with the observed muscle activation. An alternative explanation is provided by Binder and his colleagues, whose 2009 review of the semantic system
suggests a role in 'goal directed retrieval of semantic information' for the location of these differences (Binder et al., 2009). Lesions in this medial area produce difficulty in naming items when multiple possible responses are present, and Binder et al. (2009) link damage to the anterior dorsomedial prefrontal cortex (anterior to the supplementary motor area) to linguistic deficits affecting fluent semantic retrieval.

The second pair of medial differences, \textit{M3} (115ms) and \textit{M4} (125ms), are in the anterior cingulate, a region involved in motor planning, executive control, working memory tasks and error detection (Damasio, 1998). Damage in this region also impairs performance on verbal tasks (Kozlovskiy, Vartanov, Nikonova, Pyasik, & Velichkovsky, 2012). While evaluating these explanations for the observed medial differences, it is worth noting that for MEG, the signal to noise ratio is relatively poor at this depth (Attal & Schwartz, 2013; Goldenholz et al., 2009; Hillebrand & Barnes, 2002) and further assessment of the location of these effects is necessary before making strong claims about their functional role.

\textit{Fronto-temporal differences after 200ms}

Differences after the 200ms mark are predominantly in parietal and temporal cortex, in areas associated with complex visual processing, semantic knowledge, and motor control. \textit{T4} (207ms) in the \textit{SINGLE-MULTI} contrast shows greater activity for the \textit{SINGLE} than the \textit{MULTI} condition in the medial part of the temporal pole (BA38); an area associated with conceptual and fine grained semantic knowledge (Jefferies, 2013; Lambon Ralph, Pobric, & Jefferies, 2009; Visser, Jefferies, & Ralph, 2009). It is close to the anterior fusiform gyrus in the perirhinal cortex (BA36), where damage is also associated with deficits in semantic memory (Mion et al., 2010). This effect is in the opposite direction than might be expected for the semantic complexity of the respective conditions. It could be that the cortical response in the \textit{SINGLE} condition is more focal due to faster processing of previously seen stimuli, and more diffuse in the \textit{MULTI} condition, where recognition of new pictures may be more temporally variable. This would be manifest as a stronger signal in the \textit{SINGLE} condition.

\textit{P1} (233ms) shows greater source strength for the \textit{SINGLE} rather than the \textit{MULTI} condition in the right temporoparietal junction, an area associated with awareness of movement, self-monitoring and embodiment (Arzy, Thut, Mohr, Michel, & Blanke, 2006). The \textit{MULTI} condition featured stimulus pictures of tools and musical instruments, both highly manipulable types of item, while the \textit{SINGLE} condition featured only vehicles. This difference may thus correspond to greater embodiment of action during recognition of the more manipulable items, a difference of semantics and not self-monitoring.

There are 3 areas of difference in left temporal cortex between 400ms and 500ms post stimulus, in parts of the brain previously associated with complex visual processing, memory, semantic processing, and auditory speech processing. Given the late onset of these differences, their concurrence could be explained equally well by evolutions of sustained activity in a cascading system or by a parallel model in which all functions would be contemporaneously active.

\textit{T2} (402ms), greater for \textit{SINGLE} than \textit{COLOUR}, is in the anterior parahippocampal gyrus, which has been shown to have greater activity in fMRI for identifying of new visual scenes (landscapes, animals, traffic, buildings etc.) than for remembering previously seen items (Rombouts, Barkhof, Witter, Machielsen, & Scheltens, 2001). \textit{T5}(475ms), greater for \textit{SINGLE} than \textit{MULTI}, is in posterior inferior temporal gyrus,
reported as dealing with complex visual analysis such as: object recognition (Booth & Rolls, 1998); word and letter perception, (Vinckier et al., 2007) and listening and verb generation (Papathanassiou et al., 2000). In Price’s second review of the fMRI language literature, she mentions the inferior temporal and fusiform gyri as “the basal temporal language area” (Price, 2012), assigning these areas the role of processing words in speech perception, in reading, and in retrieving words from their semantics. A late onset for high level visual processing differences is incompatible with serial models, where visual processing has already taken place. Cascading and parallel models are more compatible with this finding, as differences may arise from unequal evolution of sustained visual responses in the different conditions, or communication between semantic and visual areas implementing a distributed representation.

Other than visual/semantic areas, T3 (465ms), where there is greater source strength for SINGLE than COLOUR, is part of STG (BA22) borders primary auditory cortex and is associated with speech processing (Buchsbaum, Hickok, & Humphries, 2001; Hickok, Houde, & Rong, 2011; Hickok & Poeppel, 2007; Indefrey, 2011), as well as the temporal pole (BA38). Different average reaction times between conditions mean both serial and parallel models predict differences, as the lemma (a prerequisite for accessing auditory and speech information) is selected later in semantically 'harder' conditions.

Occipital-frontal differences after 500ms:

According to sequential accounts of picture naming (Indefrey, 2011; Riitta Salmelin, 2007) the visual processing should be complete by this time. In a serial model, a difference in O3, where activity in the MULTI condition is greater than the SINGLE condition (527ms) right striate cortex could be related to a changing visual scene from post-stimulus micro-saccades, and a parallel model accounts for it with maintenance of a visual representation and/or ongoing communication between different functional areas. One possible candidate for communication is the right inferior frontal gyrus [RIFG], where differences begin at 580ms in F1, where activity in the MULTI condition is stronger than in the SINGLE condition. The RIFG is involved with inhibition of action and attentional control (Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003). In our case this may relate to the difference between perceptual and semantic discriminations, or perhaps to the participants restraining response in the MULTI condition.

It is conceivable that the speech specific functions so often observed and debated in the LIFG are not completely lateralised to that hemisphere. The right IFG has been shown to play a role in language recovery after stroke (Winhuisen et al., 2005), but if the difference is due to linguistic elements, it is puzzling that no difference should be detected in the left IFG, more usually associated with these language functions.

The pioneering MEG studies by Salmelin et al. (1994) and Levelt et al. (1998) provided the first insights into cortical activity during single word production. With respect to these papers, we provide a fuller description in both space and time, having more windows of analysis, with a tighter focus (20ms as compared to 200ms), and a projection of the sources onto a detailed template cortical surface giving a better idea of the sources' anatomical location. Liljeström et al. (2009) provide a very thorough investigation of the noun/verb production network with a picture naming task using ECDs, MNE and fMRI to describe the same network with excellent spatial and temporal reporting. The occipital to
temporofrontal nature of our cortical response is a general corroboration of their findings, to which we can add the modulation of the network by the different levels of visual and semantic engagement inherent in our experiment. We also show activation in cingulate cortex, and bilateral temporal poles, and activity in the right insula, left IFG, bilateral temporoparietal junction, and pericentral sensorimotor cortex earlier than reported by Liljeström et al. (2009), Klein et al. (2014) provide a rich description of the time/frequency response in delayed reading, showing a similar visual → motor / caudal → rostral progression of activity for this related task. Again, our results are complementary, as ERF and T/F data are largely different in the timings and spatial extents of reported activity (Laaksonen, Kujala, Hultén, Liljeström, & Salmelin, 2012), but both our study and Klein et al's (2014) report early (<100ms) extra-occipital activity in regions concerned with non-visual processing.

5.5. **Limitations:**

We made rigorous efforts to reject trials containing artefactual noise, but without online head tracking it is difficult to be certain that mouth and head movement during speech did not adversely affect source localisation. Using a bite bar to combat this source of uncertainty being impractical for reasons of participant comfort and the impairment of natural articulation. An alternative method for reducing head movement is presented in the work of Troebinger and colleagues, who recently describe a technique for 3D printing participant-specific inserts that fit between scalp and dewar helmet, holding the head firmly and comfortably in place and minimizing the scope for in-scanner movement while leaving the articulators unconstrained (Troebinger et al., 2014). As far as contamination by muscle movement is concerned, we have already attempted to address this with low pass filtration as suggested in the 'MEG good practice guide' (Gross et al., 2013), but a source reconstruction model that allows more freedom for the sources to be placed outside the brain in the scanner space might help identify the elements that have extra-cortical origins.

That different latencies of reaction time confusing the comparison between conditions is also important. Priming and repetition of stimuli tend to diminish the response in the brain (suppression). Low variance of RT's in the SINGLE and COLOUR conditions would tend to focus the energy of the response in a short duration. In the MULTI condition, more varied stimuli with greater semantic content should tend to increase the energy of the response, but the more varied RT's will spread the energy across a wider window of time. The expected consequences of these factors are not clear.

Some of the reported vs. baseline and between conditions differences are in deep brain areas. With relatively low signal-to-noise ratios in MEG, it would be useful to confirm the characteristics of these deeper sources in a protocol where one could capture more trials to combat this. However, our protocol required approximately 40 minutes of scanning, participant fatigue being an issue even on this timescale, and multiple-session recordings present additional technical challenges.

5.6. **Conclusions:**

Our study of picture naming in MEG replicates the generally reported pattern of vs. baseline activity for picture naming, beginning in occipital cortex and spreading forward to parietal, temporal and
inferior frontal areas. This widespread pattern has been described in a few other empirical papers and reviews, but seldom reported in full or in detail in a single stand-alone study. We show that right hemisphere areas are highly active during this complex language task, a phenomenon seldom discussed or analysed in the speech production literature.

We also observe early between conditions differences in bilateral BA8 (associated with goal directed semantic retrieval of words), and in anterior cingulate (associated with motor processing). We also observe differences in right anterior medial temporal cortex (207ms post stimulus) which processes semantic retrieval, and right temporo-parietal junction (233ms post stimulus), associated with self-monitoring and secondary motor function, which suggests parallel processing of articulatory and semantic task elements.

Late-onset between conditions differences in the right cuneus also suggest ongoing visual processing in our picture naming task after 500ms post-stimulus.

We further find early-onset (~100ms post-stimulus) between-conditions differences in novel frontal medial areas responding to visual and semantic manipulations. Between-conditions differences observed question the timing estimated for semantic and phonological processing suggested by current serial models of speech processing.
5.7. **References:**


5.8. Supplementary materials:

5.8.1. Sub-threshold discrepancies between conditions

We have placed restrictive limits on the size, duration and statistical tests for identifying significance in conditions differences. In addition to such statistically significant contrasts described above, we observed interesting sub-threshold discrepancies in activity between the experimental conditions. To meet the objective of fully exploring the observed network, we will describe and discuss the differences that were clearly apparent when the conditions were descriptively compared but that did not reach statistical significance.

90-110ms window:

In the COLOUR condition only, we observe small areas of change in the right temporoparietal junction in the very posterior part of the sylvian fissure. Also in the COLOUR condition, there are vs. baseline changes in the right insula and limbic (cingulate) gyrus, as well as the right ventromedial mid temporal lobe.

The activity in Right TPJ for the COLOUR condition alone may be attributable to signal spread from the large occipital sources, but may alternatively be a consequence of faster response identification and processing in the simpler condition provoking early activation in association or self-monitoring processes. The insula and cingulate gyrus are also active surprisingly early, the first dealing with motor control, the second with motor planning, so possibly to do with participants anticipating a motor response immediately after stimulus onset (perhaps before that response has been fully selected).

190-210ms window:

Bilateral inferior parietal lobules show stronger vs. baseline changes for MULTI than SINGLE, and stronger for SINGLE than COLOUR. In the MULTI condition alone, a patch of activity is visible in the left inferior frontal sulcus [LIFS] bordering and spreading slightly into LIFG pars triangularis.

Posterior inferior temporal and posterior basal temporal areas (not shown in fig. 1) are active in all conditions, and there are isolated patches of change in the anterior temporal lobe bilaterally. Cortex in the left and right pre- and post-central gyri (potentially at the level of the lip, tongue and larynx areas) shows some patches of activity, predominantly in the right hemisphere.

The gradient of strength for changes in bilateral inferior parietal lobules corresponds with the level of semantic engagement necessary to identify the names of each stimulus, this area being thought to manage semantic retrieval (Binder et al., 2009). An alternative explanation is that this area is integrating multimodal phonetic and articulatory information (Hickok & Poeppel, 2007), and that the greater variety of names in the MULTI condition requires more processing. This first explanation might corroborate our paradigm had we run a MULTI vs. COLOUR condition comparison, but no significant differences pass our thresholds in the SINGLE/MULTI contrast. Equally, Left inferior Frontal Sulcus [LIFS] activity in the MULTI condition may be a result of greater semantic engagement or greater phonological variety in the MULTI condition (semantics: Poldrack et al., 1999; Price, 2012; Roskies, Fiez,
Balota, Raichle, & Petersen, 2001, phonology: Burton, 2001; Roskies et al., 2001; Sahin et al., 2009)).
Even if this is early to expect semantic and/or phonological processing, there may be a block effect, where LIFS is more tonically active in while dealing with the more complex task, or more active in this early phase in anticipation of greater involvement in a later phase of the response.

290-310ms window:

There are patches of extra-striate and occipito-temporal cortex where activity is sustained more in the COLOUR than SINGLE and MULTI conditions particularly in left hemisphere. There is also vs. baseline change above the arbitrary threshold in vBA6 and S1/M1, but the active areas are not consistent between conditions.

The temporal poles are associated with semantic processing, which may explain their stronger activity in the SINGLE and MULTI conditions relative to the COLOUR condition at this time. Additionally, the ongoing activity in TPJ is weaker in the COLOUR condition than the others, which fits better with the theory that it serves an associative role in processing words (Hickok & Poeppel, 2007; Price, 2012) than a purely phonological one (Indefrey, 2011), as the phonological properties of the COLOUR and SINGLE conditions are balanced while semantic properties differ. Greater sustained left extra-striate/occipito-temporal activity in the COLOUR condition is most simply explained as an effect of the visual noise in the colour stimuli causing an ongoing search in pattern recognition areas. The inconsistent activity in vBA6 and motor areas may be a consequence of either different phonological qualities between the conditions, more complex conditions requiring more communication between areas in a parallel network (Hickok, 2012), or the staggered response onsets expressed in cortical processes associated with articulation. Since it is stronger in this window for the phonologically simpler conditions with the earlier response onsets, a motor rather than phonological function seems more plausible.

390-410ms window:

Occipital activity has dropped back below vs. baseline threshold apart from in the COLOUR condition, where activity persists in a posterior ventral patch of extrastriate cortex. In the temporal poles intensity is highest for the SINGLE condition, slightly weaker in the MULTI condition, and dramatically weaker in the COLOUR condition, with a right hemisphere bias in all conditions. Mid temporal changes are weak in COLOUR, where there is almost no activity in the left hemisphere, while there are large, strong patches of activity in the mid temporal regions in superior, middle and inferior gyri in the SINGLE and MULTI conditions. LIFG pars triangularis and orbitalis are clearly active in SINGLE and MULTI, but changes are patchier in the COLOUR condition, In the COLOUR condition, the right inferior and middle frontal gyri are active in the mid and frontal regions, while in the SINGLE condition changes are more focused in the inferior frontal gyrus, but are spread across pars orbitalis, triangularis and opercularis with even consistency.

That activity in temporal poles is strongest in the SINGLE condition could be due to a combination of opposing factors: the reaction times in the MULTI condition have the highest variance, indicating that component processes may have more variable onsets and thus less temporally focused cortical activity, while the semantic content of the COLOUR and SINGLE conditions is less, but the reaction times, and thus cortical responses, are less temporally variable, leading to greater focus of energy in the averaged responses.
The strength of mid temporal changes in activity follows the gradient of semantic complexity in our conditions, but in cortex associated with phonological processing, we would expect similar responses for the SINGLE and COLOUR conditions, where reaction times are more similar and the number of responses is matched. Possibly this gradient results from the integration of semantic and phonological information.

The stronger activity in LIFG (associated with syntactic and semantic processing (Baumgaertner, Buccino, Lange, McNamara, & Binkofski, 2007; Cai, Kochiyama, Osaka, & Wu, 2007; Friederici, Rüschemeyer, Hahne, & Fiebach, 2003; Sahin et al., 2009), and RIFG (associated with attentional control (Aron et al., 2003) for the SINGLE and MULTI conditions supports the successful implementation of our experimental manipulation, but the contrasts we make between adjacent levels do not reveal a significant between conditions difference.

The relatively weak dorso-frontal response in the COLOUR condition is a counter to the idea that it is of artefactual origin, as COLOUR has the earliest responses, and the activity would therefore be strongest in this period if it were artefactual. Relatively weak activity in left mid-temporal cortex for the COLOUR condition against SINGLE/MULTI lends weight to the idea that there is a semantic component to the observed response. COLOUR stimuli have elicited a stronger response in other visual areas, and the phonological elements of the response are similar between conditions. This explanation fits with the attribution of function provided by Indefrey's (2011) review/model update, but not with the timing he suggests (lemma selection and retrieval between 200 and 275ms). A similar explanation can be given for the robust activity in LIFG pars opercularis and triangularis for the SINGLE and MULTI conditions compared to the patchy difference against baseline in COLOUR.

The response in the left aSMG and perisylvian cortex is stronger for more semantically demanding conditions. This trend suggests a multifaceted role for aSMG, as a purely phonological/articulatory role would elicit a similar response in this area for the COLOUR and SINGLE conditions and stronger activity in the MULTI condition where a greater variety of words have to be processed. A purely semantic role would show a gradient of activity from weakest in COLOUR to strongest in MULTI, which is not apparent.
6. Summary and General Discussion

This thesis attempts to contribute to the magnetoencephalography [MEG] language production literature in two ways. (i) An appreciation and critique of the general method of approach used for researching the language network using MEG, with a particular focus on the distinction between serial and parallel models of word production, and (ii) the contribution of empirical evidence concerning the dynamics of the language network during picture naming, from an MEG investigation of picture naming.

6.1. Introduction and background:

Firstly, we discuss the evolution of psycholinguistic models from the first ‘boxological’ or ‘verbal’ models of the language network to modern computationally implemented models, we explore in more depth the hierarchical state feedback control [HSFC] model (Hickok, Houde, & Rong, 2011; Hickok & Poeppel, 2007; Hickok, 2012) and Indefrey and Levelt’s [I+L] model (Indefrey & Levelt, 2004; Indefrey, 2011) as examples of the parallel and serial type respectively. The equal validity of proposed serial and parallel computational models of language motivated a review of the MEG language production literature. MEG being of particular interest as its whole head coverage, good spatial and excellent temporal resolution make it a candidate to provide useful evidence in this debate.

6.2. Literature review:

This review presents an analysis of data collected from 17 empirical MEG studies of word or speech production. Wishing to assess the evidence as it relates to the parallel and serial schools of thought, we collated the statistically significant effects from all of these papers and classified them according to their cognitive function and anatomical location. We used the authors own attribution of function and location as well as the functional/anatomical associations described in Price’s (Price, 2012) and Indefrey’s (Indefrey, 2011) recent reviews of the language production literature to frame the timing of the found effects. These functional/anatomical frames show that a large proportion of effects reported go undiscussed by authors, that the I+L frame casts a general trend towards seriality in data, but this is not apparent in interpretations of other data. We also show evidence supportive of a parallel model present in the data, such as early effects in conceptually late functional areas, such as the inferior frontal gyrus [IFG] and motor areas, and late effects in visual areas. We also find that not only are event related potential [ERP] and time/frequency [T/F] analysis complementary sources of information (as also shown by Laaksonen, Kujala, Hultén, Liljeström, & Salmelin, 2012), but they may have generally different temporal characteristics, with T/F effects having earlier onsets than ERP effects.

This work led us to several general recommendations about how MEG studies might be conducted in future so as to make comparing and collating evidence between studies easier.
and more fruitful. We suggest that results for the whole brain across the response period be reported by all studies, so that readers and researchers can validate and compare the presence of activity in all areas to other similar reports. Studies focussing on \textit{a priori} defined regions of interest (ROI's) should provide the standard co-ordinates of such points where possible, and the method by which these were defined if not. Large variation in pre-defined time windows of interest is common practice, but has the effect of blurring temporal data, as the edges on an activity’s onset may be displaced in time. Given that the timing of functional onsets is crucial for discriminating between the two types of model we discuss, this is a key point. Functional and anatomical lines are also blurred by the wide range of tasks that are used by the different studies. While this may be unavoidable given the different objectives of research groups, some form of standardised testing would allow the verification of certain effects across multiple populations and validate analysis procedures should the expected response patterns be observed.

Using different versions of tasks also blurs the issue of attributing function to area, as neighbouring functions can invest neighbouring areas (Price, 2010, 2012), and MEG does not usually possess the spatial resolution to separate these. Specialised procedures (e.g. a bespoke 3D printed fitted insert for individual participants to prevent movement) and refined techniques (e.g. distinguishing between sources in different cortical laminae, or setting a number of trials based on the structure one wishes to investigate) can markedly improve the spatial resolution of MEG (Gross et al., 2013; Troebinger, López, Lutti, Bestmann, & Barnes, 2014; Troebinger, López, Lutti, Bradbury, et al., 2014), and these should be employed where viable. Different protocols and analysis techniques, while equally valid, make comparing the similar data between studies treacherous. An ‘industry standard’ for such procedures would also be beneficial for comparative purposes. In all cases, correction for multiple comparisons and individual structural MRI scans should ideally be used.

Of currently existing neuroimaging techniques, MEG is ideally posed to provide evidence discriminating between the serial and parallel models of language production. It provides whole-brain coverage, which is sadly under-reported, excellent temporal resolution, which is compromised by the use of \textit{a priori} time windows of interest, and good spatial resolution, which is susceptible to improvement by using advanced practical and analytical techniques. Were all these elements to be properly employed in a protocol where functional localisation was performed in individual participants before performing an analysis of their response to the task proper, one could hope to produce strong evidence for the timing of functional onset, and thus discern between competing models.

6.3. \textbf{Experimental design and technical development:}

Having assessed the contributions necessary to further the debate and possible to provide using MEG, we worked to adapt a picture naming protocol to have a manipulation that would highlight different elements of the task network in MEG. Experimental aims were identified,
and stimuli were developed to fit the technical constraints of the scanner and to allow both stimulus and response locked analysis.

The protocol was piloted on a small population of participants in the MEG scanner, and both it and the technical elements were refined to maximise usable data. This involved (among other time-consuming iterative modifications) changing the task to avoid having a predictable onset phoneme, the instructions regarding blinking, reducing by half the number of trials in each condition, and replacing the original condenser microphone with an optical one to eliminate electromagnetic noise.

6.4. Empirical study:

After a long process of design and refinement, our investigation into the cortical response to naming pictures succeeded in employing most of the ideas proposed after our review of the MEG literature. We used 3 levels of semantic and visual manipulation in a picture naming paradigm, a semantically poor colour naming condition, a middle level of semantic processing in a picture naming condition where 8 items from a single semantic category were used, and a semantically rich condition, where 4 different instances of 32 items from 4 different semantic categories were named. The cortical response in the whole brain was reported in 20ms windows at 100ms steps post-stimulus, and those statistically significant (local false discovery rate correction q < 0.2) between-conditions differences from a whole brain comparison that passed arbitrary thresholds for size (>6 adjacent vertices in a 15,002 vertex template anatomy) and duration (20ms) were reported.

The cortical response to picture naming showed the characteristic strong occipital activation, followed by a forward-spreading tide of activation, with a stronger response in visual areas for the colour condition, and in semantic processing areas such as the anterior temporal lobe and LIFG for the more semantically rich conditions, followed by strong frontal and motor-area activity. These observations mirror those commonly seen in other picture naming and speech production experiments, and validate our protocol.

Among other statistically significant differences (see Chapter 5 for details), the between conditions comparisons revealed the following key findings:

- Colour vs. single category picture naming:
  - A strong and sustained difference in medial superior frontal gyrus and anterior cingulate cortex [ACC] of both hemispheres beginning at 90ms and 115ms respectively, before the first significant difference in visual areas argues against serial involvement of cognitive functions. ACC has been associated with conflict monitoring and continuous speech monitoring. These differences suggest that colours might require more top-down control over perceptual information to identify than imagined, and that control processes are active very early after stimulus presentation.
Markedly stronger activity for colour over picture naming in posterior collateral sulcus at 115ms bears witness to early visual differences between the colour and single category picture naming conditions, confirming expectations. This onset time is shared with a difference in the posterior inferior temporal sulcus, a site previously associated with higher order visual processing. This latter suggests that noise may be stimulating to pattern recognition areas.

Greater response strength for pictures than colours in the left anterior temporal lobe, beginning in entorhinal cortex at 402ms and superior temporal sulcus at 465ms speaks to late semantic processing, a standard prediction of the neuroimaging literature that also validates our methods.

**Single vs. multiple semantic category picture naming:**

A stronger response in occipital cortex for multiple over single condition at 110ms confirms the expected difference in visual processing.

Differences at 207ms in the right anterior temporal lobe confirm rapid transfer of information to anterior areas, and distributed semantic processing.

Greater response strength for the single condition in the right parieto-temporal junction at 233ms confirms bi-hemispheric involvement in visuo-semantic processing, outside of occipital cortex.

A stronger response at 475ms in inferior temporal cortex suggests either differences in maintaining representations of repetitive or changing pictures, or a difference in communication between the conceptually ‘early’ areas and the rest of the network due to semantic or articulatory factors. Neither option is supportive of serial models.

Significantly larger response in the right cuneus at 527ms for the multi condition confirms the idea that visual cortex is continuing to work either on maintaining a more complex representation in the more varied condition, or communicating more with other network areas.

The multi condition shows a stronger response in right inferior frontal gyrus, implying its implication in either semantic or articulatory processing, as with LIFG.

### 6.5. **General interpretation:**

While the early differences in anterior areas and the late differences in perceptual area may appear to be more in line with parallel models like the HSFC, they are also explicable in the more interactive serial-cascading models, where information is passed to later areas before a ‘decision point’ is reached at earlier levels, and feedback is possible from later to earlier functions. This calls into question the validity of strongly serial accounts. Multiple areas seem to be concurrently active to manage the articulatory element of the response (even before considering self-monitoring and feedback processes), this could be the case for conceptual element too. Our data show multiple sites concurrently active after 100ms, and both the primarily visual and primarily semantic manipulations have early effects in supposedly
semantic areas and late effects in perceptual areas. It is conceivable then, that the perceptual and conceptual representations are managed by parallel areas, being maintained over the course of preparation for, and perhaps during, articulation. This certainly makes sense from the perspective of fluent speech, where a set of concepts must be maintained concurrently with articulation. Once we know we want to say "The cat sat on the mat", the conceptual mat must wait patiently in memory until the end of the sentence to be sat upon by the cat.

6.6. **Limitations and potential developments:**

6.6.1. Protocol:

Participants were screened for handedness on a self-assessment basis; in future studies, this would be augmented with the use of a more detailed questionnaire to assess the degree of hand-dominance. This would be of particular interest if analysing the results on an individual basis.

Our visual stimuli were only balanced for complexity on a subjective measure; balancing could be done on a more formal basis, with more similar stimuli used for the colour and picture naming conditions, while maintaining the lack of coherent form in the colour condition.

Mixing the manipulation of factors between the conditions may also have confused the interpretability of our data. An elegant alternative to our three-level approach was performed in the MEG speech domain by Miozzo, Pulvermüller, & Hauk, (2014) who use visual complexity (the number of lines and their intersections in a drawing), the number of specific semantic features, the number of action features (a measure of how people interact with objects, (see Magnié, Besson, Poncet, & Dolisi, (2003); and McRae, Cree, Seidenberg, & McNorgan, (2005)) and word form complexity as continuous predictors of cortical activity in a multiple regression study. Their data also show strong anterior responses within the first 150ms post stimulus at the peaks of activity associated with their 4 factors.

Given that the ultimate aim of our experimental task is to allow us to gather data discriminating between serial/parallel models, it would be useful to localise the component functions of picture naming before the task proper. In a more elaborate protocol, separate tasks to isolate functional areas in individuals responsible for the factors to be later tested could be employed. These would then allow any peaks of power or between conditions differences in the response to be explicitly associated with those functions. In the visual domain this might employ increasingly complex pictures with the same name, the semantic, the difference between domain and basic level naming (Clarke, Taylor, & Tyler, 2011), and the phonological/articulatory domain a manipulation of word length (always being controlled for frequency).

6.6.2. Technical elements:
We were forced to remove one participant due to excessive blinking, and rejection rates could have been improved in other participants by pre-testing outside the scanner for an optimal blink strategy, maximising the number of trials within the recommended 15 minutes for a scanning session at the same time as participant comfort by managing rest period duration and frequency, and the instructions given to participants.

We also experienced difficulties with high levels of head movement within scanning sessions. Troebinger, López, Lutti, Bradbury, et al., (2014) provide a novel technical solution to this that does not require installing head tracking, where 3D-printed inserts are made to fill the space between an individual’s head and the Dewar helmet, effectively immobilising the participant without affecting their comfort. This is obviously more cost-effective for habitual or repeat participants!

6.6.3. Data processing:

While we low-pass filtered our data for ERP analysis, a T/F analysis of the data would ideally include the gamma band, susceptible to pollution by artefactual noise from articulators. By identifying the power band where most of the energy is contained using EMG on the major articulators of participants, one could isolate the worst offending frequencies and filter them out specifically for each participant at the time that noise was produced.

6.6.4. Analyses:

Much still remains to be done with the data recovered in this study; thus far, we have only performed a stimulus locked analysis of ERP's, while the time/frequency and response locked analyses go unexplored. These would be of particular interest given the complementary nature of ERP and T/F data, and our critical interest in functional dynamics and the order of functional onsets.

It is also regrettable that individual anatomical scans and parcellations were not available, as these would not only have improved the accuracy of the source model, but also allowed investigation of the response at the individual level, perhaps allowing better definition of functional areas and the onset of differences in those areas, whose latencies could have been meaning after identification at the individual level.

Lastly, an analysis of the data from the task engaged period of the scanning session could be compared against a baseline in non-task-engaged period. In the pre-stimulus period of the experimental recording session, participants are aware of the task they are engaged in, and expecting the next stimulus to appear on screen shortly. This is a very different state to that where the participant is at rest before the task begins, and such a comparison may give insight into the areas of cortex that are active for this task, particularly in conjunction with an analysis of connectivity.

6.7. Summary and future directions/perspectives:
This thesis has been concerned with evaluating the available evidence in the literature, and in particular the MEG evidence, on speech production. It assesses the capability of MEG to further explore this domain, and to provide evidence that could help researchers discriminate between serial and parallel models of speech production. Additionally, it suggests ways in which MEG could be fruitfully employed in investigating speech production, and describes the implementation of some of these suggestions in an empirical study.

With regard to the literature in general, it seems that relatively few studies evaluate right hemisphere activity or the contribution non-cortical brain structures. Likewise, is it often neglected that since participants are constantly task engaged, the whole task network may be online in some way throughout the analysis window. A lack of vs. rest baseline may this be depriving us of interesting information about changes in brain activity between these two states.

The fast and parallel nature of information distribution in the speech production network (Clarke et al., 2011; Klein et al., 2014; Miozzo et al., 2014), and the late differences demonstrated in perceptual areas by our own and other studies do not mesh easily with strictly feedforward, modular cognitive models (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), instead being a more comfortable fit for interactive or connectionist models like the PDP or triangle family of models (Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989) or the HSFC model more fully explored in this thesis (Hickok, 2012). Such models are a good match for a neurologically plausible system in which communication between anatomically distant nodes of the cortical network via oscillatory communication settle into an "attractor state" (Rueckl, 2002) that codes for a word, and a gestalt representation of all its qualities, semantic, phonological, articulatory etc. This system evolves over time, and learns with repeated exposure to language, the strengths of connections being adjusted depending on what components of the system are co-active, in a way that owes much to Hebb’s, "fire together, wire together" principle (Hebb, 1949). The system can thus be adapted in the short (as for priming and habituation) and the long (as for the integration of new words into the lexicon (Davis & Gaskell, 2009; Davis, Maria, Betta, Macdonald, & Gaskell, 2009)). In this way the system adapts itself to respond faster and more accurately to previously encountered (and especially, frequently encountered) words. MEG, and in particular connectivity analysis in the time/frequency domain, is ideally placed to provide data on such a system, reliant as it is on oscillatory communication and the maintenance of quasi-stable, repeated and identifiable network states.

6.8. Conclusion:

Theories of speech production have traditionally been built on psycholinguistic models, and adapted to suit behavioural evidence. Diligent scientific and technical development has now brought us to the point where informed by decades of functional PET, EEG, intracranial recordings, and MRI, MEG can provide a global perspective at an appropriate temporal
resolution to outline the word production system from the ground up, on a data-driven basis, using a primary measure of neural function.

The review and empirical results presented in this thesis show a generally serial trend in the activity of areas associated with the component functions of speech production, but also show strong reasons to doubt that the system is entirely serial. In particular early differences in anterior areas dealing with conceptually late functions such as syllabification and articulation, and late differences in early perceptual and semantic areas seem to point at a parallel or highly interactive organisation of the network. Ultimately, the evidence available at this moment is insufficient to conclusively dismiss either the serial or parallel school of thought, and it seems likely that elements of both are present in the brain.

6.7b. Résumé et perspectives:

Cette thèse explore la capacité de la MEG à discriminer entre les modèles sériel et parallèle de la production de la parole. Elle propose aussi des moyens pour standardiser et améliorer l’implémentation des études en MEG, dans le but d’effectuer de meilleures comparaisons entre des différentes études.

Au regard de la littérature générale, peu d’études évaluent la contribution de l’hémisphère droit ou celle des structures non-corticales. Le contraste direct entre l’activité corticale pendant la tâche et celle durant une vraie période de repos est aussi peu fréquemment reporté. Cela pourrait nous priver d’informations intéressantes sur les différences entre les deux états; la période pré-stimulus quand les participants sont engagés dans la tâche et une période de repos hors tâche.

La distribution d’informations dans le réseau cortical de la production de la parole est rapide et semble être en parallèle plutôt que sérielle (Clarke, Taylor, & Tyler, 2011; Klein et al., 2014; Miozzo, Pulvermüller, & Hauk, 2014). Les effets tardifs dans les aires perceptuelles (Munding et al., 2015; Salmelin, Schnitzler, Schmitz, & Freund, 2000; Sörös, Cornelissen, Laine, & Salmelin, 2003), ainsi que les différences précoce dans les aires antérieures ne sont pas compatibles avec les modèles strictement sériels (e.g. Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Au contraire, ces données soutiennent plutôt des modèles interactifs ou ‘connexionnistes’, par exemple les modèles PDP (Seidenberg & McClelland, 1989; Zevin & Seidenberg, 2006). De tels modèles proposent le concept d’un réseau cortical dans lequel les régions cérébrales se coordonnent dans différents « états d’attraction » quasi-stables (Rueckl, 2002). Ces états pourraient représenter l’ensemble des propriétés d’un mot : sémantique, phonologique, articulatoire/motrice, etc., distribuée dans le réseau. Ce système serait capable d’évoluer grâce à l’expérience avec le langage, le poids des connexions étant modulé selon la coactivité des différent(e)s régions/neurones. Un tel système peut donc être modifié à court-terme (amorçage et habitation) ainsi qu’à long terme (comme pour l’apprentissage et l’intégration de nouveaux mots dans le lexique mental (Davis & Gaskell, 2009; Davis, Maria,
Betta, Macdonald, & Gaskell, 2009)). De cette manière le système est capable de s’adapter et de s’améliorer pour mieux reconnaître les stimuli familiers, et en particulier les mots et formulations fréquemment rencontrés.

Avec sa capacité à observer les dynamiques de la fonction neuronale dans les différentes bandes de fréquences, la MEG est un bon outil de report d’un système fait d’états d’activités oscillatoires quasi-stables dans différentes régions cérébrales.

6.8b. Conclusion:

Jusqu’à l’avènement de la neuroimagerie, les modèles de la production de la parole ont été construits sur la base de données comportementales et de théories psycholinguistiques. Après quelques dizaines d’années de travaux soigneux, des études en PET, IRM, EEG, les enregistrements intra-crâniens et ceux réalisés avec la MEG fournissent une perspective plus ancrée sur les bases biologiques de la production de la parole. Alors, que les méthodes électrophysiologiques ont une résolution temporelle appropriée pour nous informer sur les changements rapides dans le système de production des mots, la MEG nous permet d’observer la réponse de tout le cortex.

La revue de littérature et les résultats empiriques présentés dans cette thèse montrent une séquence d’activités séries dans les régions fonctionnelles qui soutiennent la production de la parole. Néanmoins, ils montrent aussi de bonnes raisons de douter que le système est entièrement sériel. En particulier, nos études montrent des différences précoces dans les aires frontales, et tardives dans les aires perceptuelles et sémantiques. Ces données pourraient suggérer une organisation du système fortement interactive. En résumé, les informations disponibles à l’heure actuelle ne suffisent pas pour rejeter définitivement ni une explication sérielle, ni parallèle. Il semblerait en fait que des éléments des deux types soient présents dans le système linguistique.
6.9. **References:**


7. Abstract:

This thesis concerns the use of magnetoencephalography [MEG] as a tool for investigating the dynamics of the cortical response during word production. The evidence gathered is considered in the context of existing psycho- and neuro-linguistic models of word and speech production. An exploration of the evolution of psycholinguistic models is performed, motivating a review of the MEG literature. The strengths and limitations of the technique and existing evidence are considered, and used to guide the design of a picture naming protocol compatible with MEG. An empirical MEG study is then developed and implemented using a visuo-semantic manipulation to explore the dynamics of the cortical response. This study demonstrates a broad, bi-hemispheric response with early (~100ms) between-conditions differences in bilateral BA8 and anterior cingulate cortex, in right anterior medial temporal cortex at 207ms, and a difference in right temporo-parietal junction at 233ms post stimulus. Late between conditions differences in the right cuneus also suggest ongoing visual processing. Our findings question the timing estimated for semantic and phonological processing suggested by current serial models of speech processing. In the light of the review and empirical study, a contextual evaluation of existing models is performed and potential future avenues of investigation are discussed.

7b. Résumé:

Cette thèse porte sur l'utilisation de la magnétoencéphalographie (MEG) comme outil d'étude de la dynamique des réponses corticales durant la production de mots. Les données empiriques accumulées dans la littérature sont évaluées au regard des modèles psycho- et neuro-linguistiques de la production des mots et du langage. Nous réalisons une exploration de l'évolution des modèles psycholinguistiques et effectuons en ce sens une revue de la littérature MEG. Les forces et limites de la technique et des données empiriques existantes sont considérées et utilisées pour établir un protocole de dénomination d'images qui soit compatible avec la MEG. Nous développons ensuite une étude empirique réalisée en MEG, en utilisant une manipulation visuo-sémantique pour explorer la dynamique des réponses corticales. Cette étude démontre une large réponse bi-hémisphérique avec des différences inter-conditions précoces (~100ms) dans la BA8 et des différences dans le gyrus cingulaire antérieur, le cortex médial temporal antérieur droit à 207ms et dans la jonction temporo-pariétale à 233ms après apparition du stimulus. Des différences entre les conditions apparaissent tardivement dans le cuneus droit et suggèrent également un traitement visuel en cours. Nos résultats questionnent le timing estimé pour les traitements phonologiques et sémantiques suggérés par les modèles sériels actuels de production du langage. A la lumière de la revue de la littérature et de l'étude empirique conduite, nous évaluons les modèles existants et discutons des directions potentielles pour les recherches futures.