

Multi-modal imaging of brain networks subserving speech comprehension

A thesis submitted to the University
of Manchester for the degree of
Doctor of Philosophy in the Faculty of
Medical and Human Sciences

2013

Ajay D. Halai

School of Psychological Sciences

Table of Contents

LIST OF FIGURES	5
LIST OF TABLES	5
ABSTRACT	6
DECLARATION	7
COPYRIGHT STATEMENT	7
ACKNOWLEDGEMENTS	8
THE AUTHOR	9
CHAPTER 1	10
GENERAL INTRODUCTION	10
OVERVIEW	10
NEURAL BASIS OF SPEECH COMPREHENSION	10
1. <i>Spatial organisation of speech processing</i>	11
2. <i>Speech comprehension as part of the semantic network</i>	18
2.1. ATL	18
2.2. IFG	19
2.3. IPC	21
3. <i>Influence of neuroimaging methodology on speech models</i>	22
4. <i>Temporal evolution of speech processing</i>	23
SUMMARY	27
RESEARCH OBJECTIVES	29
CHAPTER 2	31
EXPLORING THE TEMPORAL AND SPATIAL CHARACTERISTICS OF SPEECH COMPREHENSION USING ELECTROPHYSIOLOGICAL EVIDENCE	31
ABSTRACT	32
1. INTRODUCTION	32
1.1. <i>Temporal dynamics of speech comprehension</i>	33
1.2. <i>Spatial organisation of speech comprehension</i>	34
1.3. <i>Outline and aim</i>	36
2. METHODS	37
2.1. <i>Subjects</i>	37
2.2. <i>Stimulus generation</i>	37
2.3. <i>Experimental procedure</i>	39
2.4. <i>Imaging acquisition and data analysis</i>	39
3. RESULTS	41
3.1. <i>Event-Related Potentials</i>	41
3.2. <i>Source analysis</i>	47
4. DISCUSSION	49
4.1. <i>The early component</i>	50
4.2. <i>The late component</i>	53
5. CONCLUSIONS	56
CHAPTER 3	58
A COMPARISON OF DUAL GRADIENT-ECHO AND SPIN-ECHO FMRI OF THE INFERIOR TEMPORAL LOBE	58
THIS CHAPTER IS CURRENTLY UNDER REVIEW IN HUMAN BRAIN MAPPING	59
ABSTRACT	59
1. INTRODUCTION	59
2. METHODS	63

2.1. <i>Subjects</i>	63
2.2. <i>Experimental paradigm</i>	63
2.3. <i>Image acquisition</i>	64
2.4. <i>fMRI data analyses</i>	65
Pre-processing.....	65
Statistical analysis.....	66
3. RESULTS.....	67
3.1. <i>Behavioural measures</i>	67
3.2. <i>Whole brain analysis</i>	68
3.3. <i>Region of Interest (ROI) analysis</i>	70
4. DISCUSSION.....	72
5. CONCLUSIONS.....	76
CHAPTER 4.....	77
DUAL-ECHO FMRI CAN DETECT ACTIVATIONS IN INFERIOR TEMPORAL LOBE DURING INTELLIGIBLE SPEECH COMPREHENSION.....	77
ABSTRACT.....	78
1. INTRODUCTION.....	78
2. METHODS.....	80
2.1. <i>Subjects</i>	80
2.2. <i>Experimental paradigm</i>	81
2.3. <i>Image acquisition</i>	81
2.4. <i>fMRI data analyses</i>	82
Pre-processing.....	82
Statistical analysis.....	83
3. RESULTS.....	83
3.1. <i>Whole brain analysis</i>	83
4. DISCUSSION.....	85
5. CONCLUSIONS.....	88
CHAPTER 5.....	89
COMBINING EEG-FMRI TO INVESTIGATE BRAIN NETWORKS INVOLVED IN SPOKEN WORD COMPREHENSION.....	89
ABSTRACT.....	90
1. INTRODUCTION.....	90
1.1. <i>Rationale for EEG-fMRI</i>	91
1.2. <i>MR related artefacts and correction procedures</i>	92
1.3. <i>Models of speech comprehension</i>	95
1.4. <i>Aims</i>	98
2. METHODS.....	99
2.1. <i>Subjects</i>	99
2.2. <i>Experimental paradigm</i>	99
2.3. <i>Image acquisition</i>	100
fMRI.....	100
EEG.....	101
2.4. <i>fMRI analysis</i>	102
Pre-processing.....	102
Statistical analysis.....	103
2.5. <i>EEG analysis</i>	103
fMRI artefact correction and pre-processing.....	103
Statistical analysis.....	106
Source analysis.....	106
Virtual electrode analysis.....	107
3. RESULTS.....	108
3.1. <i>fMRI analysis</i>	108

3.2. EEG analysis	110
4. DISCUSSION	112
4.1. Evaluation of fMRI analysis.....	112
4.2. Evaluation of virtual EEG electrode analysis	116
4.3. Evaluation of traditional EEG sensor level analysis	118
4.4. Implications for speech comprehension models	120
5. CONCLUSIONS	122
CHAPTER 6.....	124
GENERAL DISCUSSION	124
SUMMARY OF FINDINGS	124
IMPLICATIONS FOR CURRENT MODELS OF SPEECH COMPREHENSION	127
<i>Spatial organisation of speech processing.....</i>	<i>127</i>
<i>Speech comprehension as part of the semantic network.....</i>	<i>129</i>
<i>Influence of neuroimaging methodology on speech models</i>	<i>133</i>
<i>Temporal evolution of speech processing.....</i>	<i>135</i>
<i>Limitations.....</i>	<i>138</i>
<i>Future directions</i>	<i>139</i>
CONCLUSIONS	140
REFERENCES	143
APPENDICES.....	160
APPENDIX 1: WORD/PSEUDOWORD STIMULI	160
APPENDIX 2: SEMANTIC/LETTER CATEGORISATION STIMULI.....	161
APPENDIX 3: AUDITORY SENTENCE STIMULI.....	164

Word count: 56, 563

List of Figures

Figure 1: Comparing three spatial models of speech comprehension: a) Rauschecker and Scott (2009); b) Hickok and Poeppel (2007); and c) Price (2010).....	14
Figure 2: Comparing three temporal models of speech comprehension: left) Serial processing; middle) Parallel processing; and right) Two-pass parallel processing framework.....	26
Figure 3: The average topographical plots and event-related potentials (ERPs) during an early (a) and late (b) time window for real word conditions.	44
Figure 4: The average topographical plots and event-related potentials (ERPs) during an early (a) and late (b) time window for pseudoword conditions.	45
Figure 5: Event-related mean amplitudes averaged across 20 subjects for words and pseudoword conditions.	47
Figure 6: Beamformer results for early processing window (phonological difference) and late processing window (intelligibility difference) for real word conditions.....	48
Figure 7: Beamformer results for early processing window (phonological difference) and late processing window (intelligibility difference) for pseudoword conditions.....	49
Figure 9: Significant fMRI activations for semantic categorisation using various acquisition methods in the temporal lobe.....	69
Figure 10: Regions of interest analysis comparing group level z-scores between dual-echo and spin-echo.....	71
Figure 11: Significant fMRI activations for intelligible and unintelligible sentences: a) speech, b) intelligibility and c) phonology processing.....	85
Figure 13: EEG time series showing the effects of MR artefacts induced on EEG sensors and the effect of applying an artefact correction to produce relatively clean EEG signal.	105
Figure 14: Significant fMRI activations for intelligibility ($W > WRS$).	109
Figure 15: EEG virtual electrode results for intelligibility ($W > WRS$) depicted on MNI brain across 1 s epoch.	110
Figure 16: Topographical plot showing an intelligibility effect ($W > WRS$) in the theta frequency range between 0.45-0.65 s.	111
Figure 17: Source estimation of the theta effect during 0.45-0.65 s.....	112

List of Tables

Table 1. Psycholinguistic properties of the stimuli. Data obtained from CELEX, The English Lexicon Project and The Irvine Phonotactic Online Database.	39
Table 2. Showing the difference between fMRI methods activity within each region of interest (ROI) in terms of statistical distance (z-score).	72
Table 3: MNI peaks identified for speech, intelligibility and phonology contrast (sentences).....	84
Table 4: MNI peaks identified for the intelligibility contrast ($W > WRS$).	109

Abstract

Multi-modal imaging of brain networks subserving speech comprehension

Ajay D. Halai, The University of Manchester

For the degree of Doctor of Philosophy (PhD)

December 2013

Neurocognitive models of speech comprehension generally outline either the spatial or temporal organisation of speech processing and rarely consider combining the two to provide a more complete model. Simultaneous EEG-fMRI recordings have the potential to link these domains, due to the complementary high spatial (fMRI) and temporal (EEG) sensitivities. Although the neural basis of speech comprehension has been investigated intensively during the past few decades there are still some important outstanding questions. For instance, there is considerable evidence from neuropsychology and other convergent sources that the anterior temporal lobe (ATL) should play an important role in accessing meaning. However, fMRI studies do not usually highlight this area, possibly because magnetic susceptibility artefacts cause severe signal loss within the ventral ATL (vATL). In this thesis EEG and fMRI were used to refine the spatial and temporal components of neurocognitive models of speech comprehension, and to attempt to provide a combined spatial and temporal model.

Chapter 2 describes an EEG study that was conducted while participants listened to intelligible and unintelligible single words. A two-pass processing framework best explained the results, which showed comprehension to proceed in a somewhat hierarchical manner; however, top-down processes were involved during the early stages. These early processes were found to originate from the mid-superior temporal gyrus (STG) and inferior frontal gyrus (IFG), while the late processes were found within ATL and IFG regions.

Chapter 3 compared two novel fMRI methods known to overcome signal loss within vATL: dual-echo and spin-echo fMRI. The results showed dual-echo fMRI outperformed spin-echo fMRI in vATL regions, as well as extra temporal regions.

Chapter 4 harnessed the dual-echo method to investigate a speech comprehension task (sentences). Intelligibility related activation was found in bilateral STG, left vATL and left IFG. This is consistent with converging evidence implicating the vATL in semantic processing.

Chapter 5 describes how simultaneous EEG-fMRI was used to investigate word comprehension. The results showed activity in superior temporal sulcus (STS), vATL and IFG. The temporal profile showed that these nodes were most active around 400 ms (specifically the anterior STS and vATL), while the vATL was consistently active across the whole epoch.

Overall, these studies suggest that models of speech comprehension need to be updated to include the vATL region, as a way of accessing semantic meaning. Furthermore, the temporal evolution is best explained within a two-pass framework. The early top-down influence of vATL regions attempt to map speech-like sounds onto semantic representations. Successful mapping, and therefore comprehension, is achieved around 400 ms in the vATL and anterior STS.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see <http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf>), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see <http://www.manchester.ac.uk/library/aboutus/regulations>) and in The University’s policy on presentation of Theses.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Dr Stephen Welbourne for the continuous support of my PhD research, for his patience, motivation, enthusiasm, and immense knowledge to pursue this ambitious project. I could not have imagined having a better mentor for my PhD study.

I am also indebted to Dr Laura Parkes, who helped make the whole thing possible during countless meetings on technical aspects of the project. Thank you for the support and belief that we would be successful in achieving our goals during the many challenges we faced.

Thank you to my other supervisors and collaborators, Professor Geoffrey Parker, Professor Matthew Lambon-Ralph, and Dr Karl Embleton, who provided invaluable expertise, discussion and critical thought that helped me develop my ideas. I would also like to thank Dr Karen Mullinger (University of Nottingham), who provided expert technical advice on combining EEG-fMRI.

I would like to thank my fellow lab mates, aka Naruians, in the Neuroscience and Aphasia Research Unit (NARU) for stimulating discussions and welcomed distractions during my study. Although there are too many to name, I would like to specifically thank Dr Mark Drakesmith for his kindness and patience in guiding me through my early experiences with MATLAB, as with hindsight I know there is light at the end of the very dark tunnel. I am very grateful to my friends in other research units, Liam Mason, Jenna Roberts, Zoe-Lydia Tsivos and Keith Austin, who provided me with the perfect antidote to combat the stresses of the PhD.

I owe a huge debt to my oldest friends, Kanchan Varsani, Dipe Rajani, Kirti Halai, Bhavina Patel and Dhanshuk Vekaria, who have given me continual support during my academic pursuits. They have each been pillars of strength during the hard times, and provided encouragement when I need it the most - I love you all.

Last but not least, I want to thank my family, to whom this thesis is dedicated. They have provided me with unconditional love and support since I first left for undergraduate study. My parents have worked hard to afford me privileges they were never given, and for that I could never repay you. I love you, and I hope I have made you proud.

“It always seems impossible, until it is done” – Nelson Mandela

The Author

I graduated from University of York in 2007 (BSc in Psychology) having conducted an MEG project investigating how face recognition is affected by static and dynamic stimuli as my final year project. I later enrolled on a Cognitive Neuroscience MSc at the University of York, where I continued working on the neural basis of face recognition using fMRI. It was during this time I realised my passion for research and the cognitive neurosciences. I graduated from York in 2009 and registered as a full-time PhD student at the University of Manchester. I have very much enjoyed the opportunity to conduct novel multi-modal imaging research into the neural basis of speech comprehension, which has allowed me to be trained in the application of structural and functional MRI and EEG. I very much look forward to applying this training to new and challenging research endeavours.

Chapter 1

General Introduction

Overview

This thesis has been prepared in alternative format meaning that Chapters 2-5 are written in the style of journal articles. Each chapter is self-contained and includes a review of the literature relevant to the data presented. Chapter 2 contains an electroencephalography (EEG) study investigating the temporal and spatial profile of auditory word comprehension. Chapter 3 directly compares two functional magnetic resonance imaging (fMRI) methods known to overcome signal loss issues within magnetic susceptible regions. Chapter 4 makes use of the fMRI method of choice outlined in Chapter 3 to investigate intelligible and unintelligible sentence comprehension. Chapter 5 investigates the spatial and temporal organisation of auditory word comprehension, using a simultaneous EEG-fMRI approach. Finally, Chapter 6 provides a general discussion, which draws conclusions based on data from all the studies presented in the thesis and discusses the wider implications for neuroimaging research. The remainder of this first chapter attempts to achieve three purposes: 1) provide a broader theoretical context behind all the work in the thesis; 2) identify and set aims of the thesis; and 3) provide detail on how each Chapter attempts to meet those aims.

Neural basis of speech comprehension

Our understanding of the neurocognitive basis for language processing has progressed dramatically over the last decades. However, there is still much debate over the neural basis of speech comprehension. Speech comprehension can be defined as extracting and understanding meaning from incoming auditory signals. The task of breaking down these signals into subcomponents is very difficult and not yet fully understood. The motor theory of speech recognition suggests that we map incoming speech sounds onto the articulatory gestures required to produce them (e.g. reviewed in Galantucci, Fowler, & Turvey, 2006). The alternative acoustic processing theory suggests that speech sounds are determined purely on the basis of the spectrotemporal properties of the acoustic signal (e.g. Stevens, 1983). In either case the incoming sound must first be analysed at the sensory

level and then mapped onto stored meaningful representations; the overall meaning of the whole signal is extracted if the mapping has been successful. The act of understanding an utterance can be split into two broad categories; speech perception and speech comprehension. Speech perception is interested in how we extract sub-lexical features (i.e. phonemes and syllables) from sounds, which is yet to be resolved. In contrast, speech comprehension is the process of obtaining the meaning at a semantic level from the utterance. Two key goals in language neuroscience have been to identify regions of the brain that contribute to successful speech comprehension and to understand their temporal dynamics. Although these goals are clearly related they have been researched separately. For example, although prominent spatial models of speech comprehension suggest a general flow of information, they do not suggest which regions are activated at specific time points during the comprehension process. A similar but opposite argument can be put regarding temporal models of speech comprehension, where spatial information is generally lacking or at least vague. These issues are considered in detail in the following sections, which summarise four core aspects of speech comprehension research and highlight possible inconsistencies in our current understanding: 1) spatial organisation of speech processing; 2) speech comprehension as part of the semantic network; 3) influence of neuroimaging methodology on speech models; and 4) temporal evolution of speech processing.

1. Spatial organisation of speech processing

Before discussing the specific neural substrates of speech comprehension, it is important to understand how low-level features of the acoustic signal are represented in the brain. Single cell recording and lesion studies on non-human primates have outlined three main regions for sound processing within the temporal lobe. These are the core, belt and parabelt, which are arranged in a hierarchical manner starting from the core. Each area exhibits functional specificity, for example the core region shows tonotopic organisation to pure tones, whereas responses to band-pass noise bursts are seen laterally across the belt region (e.g. Rauschecker, Tian, & Hauser, 1995). This hierarchical organisation suggests discrete stages of processing for auditory perception. Invasive neuroimaging studies on human epileptic patients have found similar, frequency selective, tonotopic gradients within the transverse gyrus of Heschl

(HG) (Bitterman, Mukamel, Malach, Fried, & Nelken, 2008; Talavage, Ledden, Benson, Rosen, & Melcher, 2000), which is proposed to be the homologue of the primate core region. This region, also termed primary auditory cortex or Brodmann Area 41 (PAC, BA41) is located on the superior surface of the temporal lobe, bilaterally. The literature suggests that HG is not specific to speech processing and is activated by any type of sound (e.g. Mummery, Ashburner, Scott, & Wise, 1999). The area lateral to HG, which extends into superior temporal gyrus/sulcus (STG/STS) has been found to respond to the complex acoustic features found in phonemes (Binder et al., 2000) and spectral information in non-speech (Hall et al., 2002), suggesting that these regions neighbouring HG are not specialised for speech either. Furthermore, dorsolateral to HG sits the planum temporale (PT), which has been shown to respond to acoustically similar non-speech sounds, suggesting this region is also not speech specific (Wise et al., 2001; Zatorre, Bouffard, Ahad, & Belin, 2002). Taken together, the human literature is consistent with primate studies, whereby a hierarchical profile extends from core HG regions to the lateral STG/STS surfaces.

There is general agreement amongst models of speech comprehension that sounds and low-level acoustics are processed at HG and surrounding lateral STG/STS regions. The subsequent process in speech comprehension is to identify complex acoustic features that form phonemes (basic unit of speech). This process has been consistently localised to mid-STG/STS regions, lateral to HG (similar to parabelt cortex), which is primarily left lateralised. For example, an fMRI study showed selectivity in this site for consonant-vowel (CV) speech sounds as opposed to other natural sounds (Leaver & Rauschecker, 2010). Furthermore, meta-analyses have shown sub-lexical and complex acoustic processing related to phonemes to be localised around mid-STG (DeWitt & Rauschecker, 2012; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007b; Turkeltaub & Coslett, 2010). These studies suggest that the area lateral to HG possesses neurons optimized for the recognition of complex acoustic features related to phonemes. However, there is research opposing the left lateralised and antero-lateral locations that have been reported. For example, Hickok and Poeppel (2007) propose bilateral mid-posterior STS to be involved in phonological processing. Evidence for this comes from functional imaging studies that have reported bilateral STS activation during sub-lexical speech processing (Benson, Richardson, Whalen, & Lai, 2006; Jäncke, Wüstenberg, Scheich, & Heinze, 2002; Vouloumanos, Kiehl, Werker, & Liddle, 2001). Patient data have also shown that

damage to posterior STS results in word deafness as well as deficits in the perception of non-speech sounds (Pinard, Chertkow, Black, & Peretz, 2002). It is clear that although phonological processing occurs in the lateral vicinity of HG, the precise location and laterality is still a focus of debate.

It has been suggested that the auditory and visual processing pathways are organised in a similar way (Rauschecker & Tian, 2000): as two streams, a dorsal stream that computes location and a ventral stream that computes identification. For example, in the auditory domain, the ventral route is associated with specific vocalisations and auditory object identification in the anterior belt and ventro-lateral temporal areas, whereas the dorsal route is associated with sound localisation and motor mapping within posterior belt. Evidence for this comes from primates where species-specific communication sounds were found to activate neurons in the antero-lateral belt, regardless of spatial location. In contrast neurons in the dorsolateral belt were responsive to spatial location (Tian, Reser, Durham, Kustov, & Rauschecker, 2001). This suggests that the ventral route processes information about 'what' the auditory stimuli represents, whereas the dorsal route processes information about 'where' the sound is localised. Similar analogous distinctions have been proposed in human speech research. A ventral route has been suggested to project from HG to ventro-lateral temporal cortex mapping sensory stimulation to meaning ('what' route) (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), whereas a dorsal route projects to posterior temporal, parietal and frontal cortex and has been suggested to either represent spatial location or how auditory information mapped for articulation (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). The specific details of the dual route model are not central to this thesis (for further discussions see Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Saur et al., 2008; Saur et al., 2010), however it will focus on the ventral 'what' route, which maps sensation to meaning.

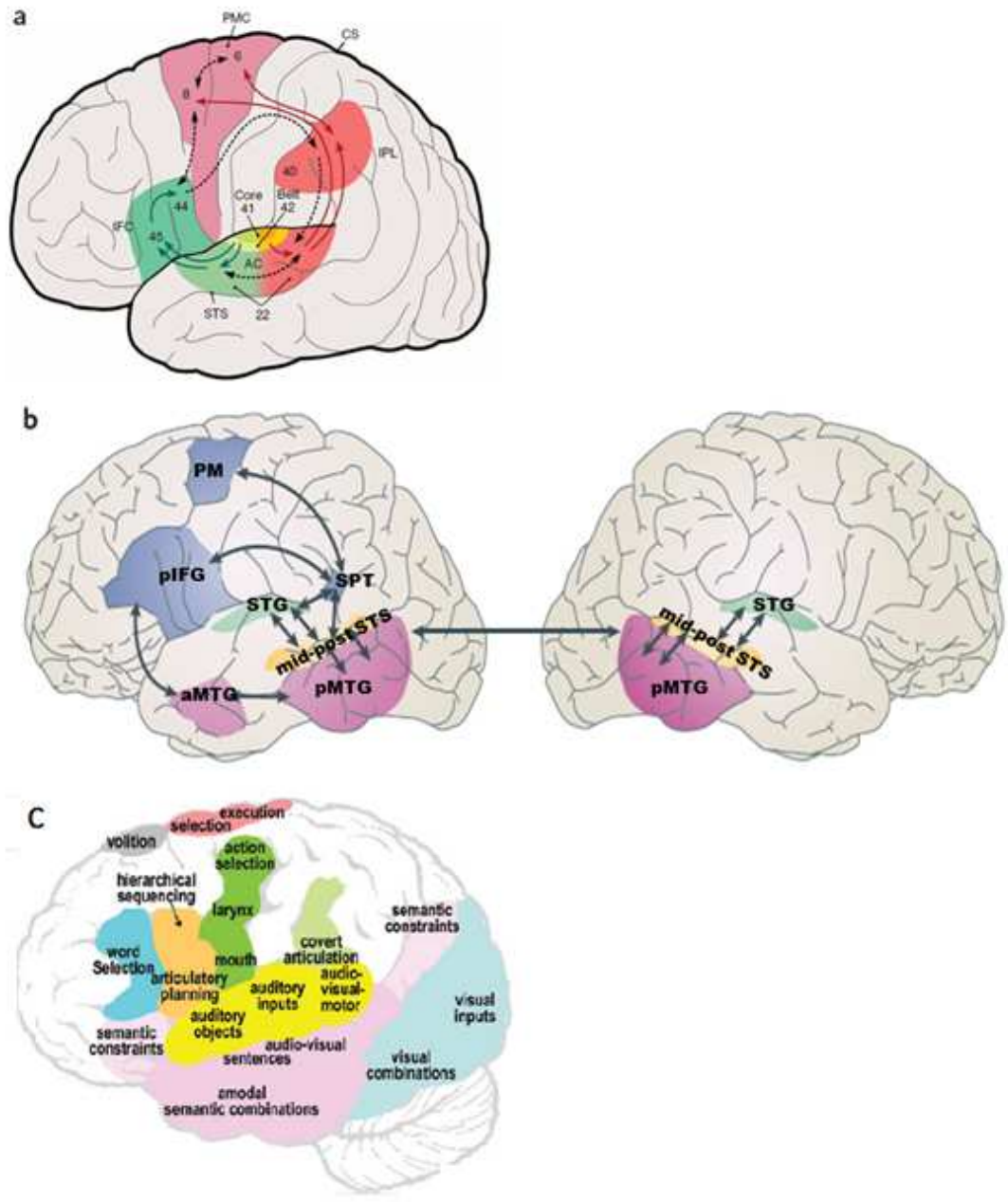


Figure 1: Comparing three spatial models of speech comprehension: a) Rauschecker and Scott (2009); b) Hickok and Poeppel (2007); and c) Price (2010). Footnote: The figure details are as follows: a) Rauschecker and Scott (2009), outline speech comprehension (green) and production (red) regions. b) Hickok and Poeppel (2007) outline spectrotemporal analysis (green, STG), phonological network (yellow, mid-post STS), lexical interface (pink, pMTG), combinatorial network (pink, aMTG), sensorimotor interface (blue, SPT), articulatory network (blue, pIFG). c) Price (2010) outlines a summary of functions attributed to language areas, in particular the STS/STG (yellow) is mainly involved in low-level processing (pre-lexical), whereas the anterior, middle, inferior temporal lobe (pink) is involved in word comprehension, in combination with the angular gyrus and pars orbitalis (light pink).

The debate over the next crucial stage of speech comprehension relates to ‘intelligibility’ or the access of meaning above and beyond complex acoustic

analysis. This issue forms a critical discussion within the thesis, and the remainder of this section will outline some of the prominent models of speech comprehension and highlight the inconsistencies between them. Although there are a number of theories for speech comprehension, this thesis focuses on the most relevant neurobiologically motivated models (Figure 1; Hickok & Poeppel, 2000, 2004, 2007; Price, 2010, 2012; Rauschecker & Scott, 2009). Other models include a more psycholinguistic element at the sentence and narrative level (Bornkessel-Schlesewsky & Schlewsky, 2013; Bornkessel & Schlewsky, 2006; Friederici, 2002; Friederici, 2011; Hagoort, 2009), which although pertinent to the general field, are not central to this thesis. The following paragraphs review the evidence for the organisation of speech in the temporal lobe, considering both posterior and superior temporal regions.

Historically, models of language have been predominately informed by lesion-symptom correlations derived from case and case series analysis of patients with aphasia. These showed the inferior frontal gyrus (IFG, discussed later) and posterior temporal cortex (PTC) to be important to language (Bates et al., 2003; Geschwind, 1970). Patients with PTC damage exhibit verbal comprehension deficits (Dronkers, Wilkins, Van Valin Jr, Redfern, & Jaeger, 2004; Hart & Gordon, 1990). However, lesion studies are subject to the criticism that brain damage does not respect cortical boundaries, and therefore regions such as posterior STS, middle STS, PT and inferior parietal cortex are all likely to be damaged (Bogen & Bogen, 1976; Wise et al., 2001). Therefore, it is difficult to draw specific conclusions from lesion studies alone. However, functional neuroimaging studies have also implicated the PTC in language comprehension (e.g. Binder, Desai, Graves, & Conant, 2009). Activation of this region is elicited by a range of comprehension tasks including semantic decision (Binder et al., 1997), semantic relatedness judgements (Bedny, McGill, & Thompson-Schill, 2008), semantic association judgements (Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996) and sentence comprehension (Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003). Hickok and Poeppel (2007) cite the posterior middle temporal gyrus/inferior temporal sulcus (pMTG/pITS) as the key locus for a sound-to-meaning interface that maps between phonological representations and distributed semantic representations (Martin & Chao, 2001). Although some researchers propose a bilateral temporal division for speech comprehension, there is evidence from multiple neuroimaging techniques to suggest a left lateralised, or

at least dominated, network (Binder et al., 2009; Price, 2010; Rauschecker & Scott, 2009; Rogers et al., 2004).

In contrast to the conclusion above, it has been proposed that the pMTG/pITS are not specific to comprehension. For example, Price (2010; 2012) suggests that the pMTG becomes involved as semantic content or task demands increase. Studies have found that although the pMTG region is activated during passive listening tasks (Awad, Warren, Scott, Turkheimer, & Wise, 2007; Jobard, Vigneau, Mazoyer, & Tzourio-Mazoyer, 2007), activation within this region increases with task demands (Jefferies & Lambon Ralph, 2006; Noppeney, Phillips, & Price, 2004; Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2011). For example, Whitney and colleagues (2011) had participants make semantic relatedness judgements involving strong and weak associations between cues and targets. Following repetitive TMS to pMTG, participants were disrupted on executively demanding semantic decisions but not on automatic semantic decisions (based on strong associations). Furthermore, pMTG is reported when integrating semantic information from auditory and visual inputs (Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Holle, Obleser, Rueschemeyer, & Gunter, 2010; Kircher, Sass, Sachs, & Krach, 2009a; Kircher et al., 2009b; Robins, Hunyadi, & Schultz, 2009). These studies suggest that pMTG is sensitive to task demands, which are related to semantic control or selection/retrieval processes, and not specifically for comprehension. In addition, some theories have suggested the PTC is primarily interested in pre-lexical semantic processes, for example retrieving event representations (Bedny, Caramazza, Grossman, Pascual-Leone, & Saxe, 2008; Tranel, Kemmerer, Adolphs, Damasio, & Damasio, 2003). For example, Bedny and colleagues conducted an fMRI experiment comparing semantic relatedness judgements on word pairs varying in the amount of visual-motion information. They found equally high activity within pMTG regions for action verbs (i.e. to run) and mental verbs (i.e. to think) and equally low activity to animal nouns (i.e. the cat) and inanimate nouns (i.e. the rock). The authors conclude that pMTG did not respond to visual-motion features of words but could reflect the retrieval of event concepts or grammatical information associated with verbs.

In contrast to the PTC, other models of speech comprehension suggest that the anterior superior temporal regions are crucial for accessing meaning (Price, 2010, 2012; Rauschecker & Scott, 2009). For example, Price (2010, 2012) suggests that the anterior temporal lobe (ATL); anterior superior temporal sulcus (aSTS), lateral anterior middle temporal gyrus (aMTG) and lateral anterior inferior

temporal gyrus (aITG) are involved in accessing increasingly specific semantic information, with activations for sentences and narratives spreading anteriorly into the temporal lobe. The model proposed by Rauschecker and Scott (2009) has similarities to Price (2010, 2012), whereby the former specifically propose the aSTS to be critically involved in intelligible speech comprehension (Scott, Blank, Rosen, & Wise, 2000). Evidence for aSTS can be seen in peak activation to word-forms in electrocorticography (ECoG) (Pei et al., 2011) and magnetoencephalography (MEG) (Marinkovic et al., 2003) studies. Furthermore, fMRI studies of acoustically complex stimuli comparing activity between word-form and pure tone stimuli report similar functional graded representation from the mid-STG to aSTS (Binder et al., 2000; Binder et al., 1997; DeWitt & Rauschecker, 2012). In addition, application of electrical cortical interference to anterior STG disrupts auditory comprehension, producing patient reports of speech as being like “a series of meaningless utterances” (Matsumoto et al., 2011). Finally, literature from the semantic dementia field provides convergent evidence for the role of the ATL regions in comprehension (discussed in detail below, Patterson, Nestor, & Rogers, 2007; Rogers et al., 2004).

There are still inconsistencies in regards to the specific function of the anterior STS/MTG, as studies have proposed their involvement in syntactical processing as well as semantics (Friederici, 2011; Hickok & Poeppel, 2007). For example, activity has been found in these regions for sentences compared to unstructured sentences or word lists, as well as for syntactically valid compared to syntactically invalid sentences (Friederici, Meyer, & von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006; Humphries, Love, Swinney, & Hickok, 2005; Vandenberghe, Nobre, & Price, 2002). This has led researchers to suggest that the anterior STS/MTG are not speech specific and instead support the construction of phrase structure (Friederici, 2011). These studies also suggest that these regions could be involved in combinatorial processes for syntax and semantics, a role which Hickok and Poeppel (2007) ascribe to the anterior MTG/ITS. This assertion would be incompatible with Rauschecker and Scott (2009) and Price (2010, 2012) who suggest a relatively specific role of anterior temporal regions for speech comprehension.

After discussing various studies that have attempted to identify a neural basis of speech comprehension, it is evident that inconsistencies remain. It can be determined that intelligibility is processed within the temporal lobe, following a lateral hierarchical projection of acoustic signal from HG. The extent to which

intelligibility is processed in the anterior, posterior or inferior temporal lobe is debated. Furthermore, it is not clear whether extra temporal regions play a central role in speech comprehension. For example, regions involved in semantic processing tasks that require access to meaning could be linked to successful speech comprehension. Therefore, one could ask, “How does speech comprehension fit into a general framework of semantic processing, and how does the semantic system contribute to speech comprehension?”

2. Speech comprehension as part of the semantic network

There are three main regions that have been implicated in semantic processing: 1) ATL, 2) IFG, and 3) inferior parietal cortex (IPC). This section will provide an overview of the functional role of each region, with attention to the role they could play in speech comprehension.

2.1. ATL

As mentioned above, there is growing evidence for the role of the ATL in comprehension and semantic cognition in general. The evidence comes in part from semantic dementia (SD) patients, where neuropsychological testing has shown an amodal impairment of semantic memory but well preserved perceptual and non-semantic cognitive abilities (Hodges, Patterson, Oxbury, & Funnell, 1992). They are known to be fluent and grammatical, but show comprehension and word-finding deficits (anomia, Hodges, Graham, & Patterson, 1995). Furthermore, extensive neuropsychological testing has revealed impaired semantic performance in a range of receptive and expressive tasks across all modalities, including spoken and written words, pictures, environmental sounds, smells, touch and taste (Bozeat et al., 2003; Bozeat, Ralph, Patterson, & Hodges, 2002; Lambon Ralph, Graham, Patterson, & Hodges, 1999; Luzzi et al., 2007; Piwnica-Worms, Omar, Hailstone, & Warren, 2010). Detailed brain imaging studies of SD patients have shown atrophy and hypometabolism of bilateral ATL regions (Galton et al., 2001; Nestor, Fryer, & Hodges, 2006), which has led to the conclusion that the ATLs are critical for amodal semantic representations (Lambon Ralph & Patterson, 2008; Patterson et al., 2007; Rogers et al., 2004). Although, the atrophy can extend into the boundaries of neighbouring cortical regions such as the ventral frontal cortex, it has been shown that the left ATL is the region most correlated with semantic impairments (Mummery et al., 2000). Compelling

convergent evidence comes from alternative neuroimaging methods which implicate the ATL in semantic processing, including positron emission tomography (PET) (Crinion et al., 2003; Rogers et al., 2006; Sharp, Scott, & Wise, 2004; Spitsyna, Warren, Scott, Turkheimer, & Wise, 2006; Vandenberghe et al., 1996), MEG (Marinkovic et al., 2003) and transcranial magnetic stimulation (TMS) on right and left ATL regions (Lambon Ralph, Pobric, & Jefferies, 2009; Pobric, Jefferies, & Lambon-Ralph, 2007, 2010).

Importantly, specific sub-divisions of the ATL affect semantic processing to different extents, with more recent evidence suggesting that the ventral ATL (vATL), which includes the anterior inferior temporal gyrus (ITG) and anterior temporal fusiform gyrus (TFG), contributes critically during amodal semantic access. For example, PET studies have identified the vATL during language and semantic processing tasks (Crinion et al., 2003; Devlin et al., 2000; Sharp et al., 2004; Spitsyna et al., 2006). This region has previously been termed the “basal temporal language area”, which was first identified by Luders and colleagues who demonstrated that direct electrical stimulation of the anterior TFG elicited selective language interference. This ranged from global expressive and receptive aphasia at high stimulus intensity to partial aphasia, with a predominantly receptive nature, at lower intensities (Luders et al., 1986; Luders et al., 1991). Recent novel fMRI studies have also implicated this vATL region during semantic processing (described in section 3) (Binney, Embleton, Jefferies, Parker, & Lambon-Ralph, 2010; Visser, Embleton, Jefferies, Parker, & Lambon-Ralph, 2010; Visser & Lambon-Ralph, 2011). Given this evidence for the role of vATL during semantic cognition, one might expect a larger presence of this region within models of speech comprehension than is currently afforded. More specifically, vATL might be expected to have the same role in successful speech comprehension as it does in other semantic tasks, namely providing meaning.

2.2. IFG

The role of the IFG in speech processing has historical roots from patient studies (Broca, 1861; Geschwind, 1970). Traditionally, damage to this region has been associated with production deficits, however recent neuroimaging, TMS and neuropsychological studies have indicated the IFG also plays a role in comprehension but perhaps with a function more akin to executive control processes (Binder et al., 2009; Devlin, Matthews, & Rushworth, 2003; Jefferies & Lambon Ralph, 2006; Noonan, Jefferies, Corbett, & Lambon Ralph, 2009;

Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). When patients with this lesion type complete tasks with semantic distractors, performance deteriorates as task difficulty increases (Corbett, Jefferies, Ehsan, & Lambon Ralph, 2009; Jefferies & Lambon Ralph, 2006; Noonan et al., 2009). This suggests that IFG contributes to selecting relevant aspects of meaning to a given verbal or nonverbal task (Corbett et al., 2009; Jefferies & Lambon Ralph, 2006; Noonan et al., 2009; Thompson-Schill et al., 1997; Wagner et al., 2001). This functional role of IFG is consistent with the proposal put forward by Price (2010, 2012), whereby it is suggested that IFG (*pars orbitalis/pars triangularis*) is involved in selecting and retrieving task related semantic attributes. In contrast, Rauschecker and Scott (2009) propose the IFG region to be involved in solving the 'invariance' problem (detecting a target against distortions in scale of time or frequency). It is likely that higher order systems are required to carry out such a task, and evidence has been seen from non-human primates where neurons were found within IFG that were coding complex species-specific communication sounds (Romanski & Averbeck, 2009; Romanski, Averbeck, & Diltz, 2005). Furthermore, an fMRI study in the human visual domain found that category training can lead to neurons in the prefrontal cortex responding selectively to the trained category (Jiang et al., 2007). Finally, Hickok and Poeppel (2007) propose that the IFG is not part of the comprehension network *per se*, but is involved in mapping sensory information onto articulatory motor representations. It should be noted that functional neuroimaging studies have also found the IFG to be involved in phonological processing (Fiez, 1997; Price, Moore, Humphreys, & Wise, 1997). Consequently, a line of research has started to investigate the division of labour within the specific parts of the IFG (Price, 2010, 2012) or the gradient of functional specificity (Friederici, 2011). For example, evidence from fMRI and TMS studies suggest that although parts of the IFG contribute to both semantic and phonological processes, they each do so to different extents (Badre & Wagner, 2007; Devlin et al., 2003; Dobbins & Wagner, 2005; Gough, Nobre, & Devlin, 2005; Poldrack et al., 1999); the anterior part (*pars orbitalis*) appears to be preferentially involved in semantic selection/retrieval, while posterior IFG (*pars opercularis*) is more involved in phonological processes. Mid-IFG on the other hand appears to play a role in more general post-retrieval/selection processes.

2.3. IPC

Finally the IPC, specifically supramarginal gyrus (SMG) and angular gyrus (AG), have recently been implicated in semantic cognition and comprehension studies (e.g. Binder et al., 2009). However, there is evidence to suggest that the IPC may process more domain general linguistic features; for example, working memory load (Buchsbaum & D'Esposito, 2008). Furthermore, Humphries and colleagues found that during auditory sentence comprehension, the AG showed late activation that began at the end of a sentence, but only when the constituent words could be combined to form a coherent meaning (Humphries, Binder, Medler, & Liebenthal, 2007). The IPC and IFG have been implicated in similar semantic/executive control functions (Corbett et al., 2009; Jefferies & Lambon Ralph, 2006; Noonan et al., 2009). Evidence in favour of this idea is illustrated by a study showing that difficulty in semantic selection led to increased IPC activity (Nagel, Schumacher, Goebel, & D'Esposito, 2008). Both Rauschecker and Scott (2009) and Price (2010, 2012) propose similar roles for the IPC, where the region is involved in integrating semantic features across online working memory stores in combination with prefrontal regions. Furthermore, this region is suggested to play a role in integrating motor-phonological information to predict the nature of semantic features (Jacquemot & Scott, 2006). The Hickok and Poeppel (2004, 2007) model does not ascribe a specific role to the IPC; however, they suggest that the Sylvian-parietal-temporal (Spt) area at the boundary between the temporal and parietal lobe is critical in mapping between auditory representations and motor representations of speech.

This overview of the semantic network suggests that the vATL is implicated in semantic representation access, whereas IFG and IPC are implicated in semantic control, working memory or acoustic-motor integration processes. For speech comprehension to occur, it is proposed that the vATL would provide meaning and is therefore a critical component of speech comprehension. Furthermore, although the IFG and IPC would be useful in aiding speech comprehension during difficult conditions, they appear not to be critical to comprehension. The question remains – why is the vATL rarely reported in speech comprehension studies and almost always omitted in models of speech comprehension? To understand this, the following section provides a summary of the problems with neuroimaging methodologies that could account for the omission of vATL during speech comprehension.

3. Influence of neuroimaging methodology on speech models

Neuroimaging methodologies can bias our understanding of brain functions. As touched upon earlier, problems with conventional fMRI research can lead to the omission of magnetic susceptible regions, due to signal loss and distortions in the ATL regions. Although, PET imaging suffers from a number of drawbacks that limit sensitivity (limited number of scanning runs, behavioural designs and relatively low spatial resolution), it does provide whole brain signal coverage including the vATL. In contrast, fMRI is not subject to the same limitations, but provides a different problem whereby the sensitivity across the brain is not consistent. The fMRI signal near to air-filled cavities (e.g., sinuses) or air-bone interfaces is vulnerable to magnetic susceptibility artefacts, which occur in the form of signal loss and geometric distortions (Devlin et al., 2000). This is particularly problematic for imaging the orbitofrontal, temporopolar and vATL regions and is likely, at least partly, to account for the relative lack of functional neuroimaging studies that have detected ATL activation (Visser, Jefferies, & Lambon-Ralph, 2009). The effect of including fMRI studies in meta-analyses can bias our understandings of models of speech comprehension. For example, a recent meta-analysis suggested that core speech comprehension regions were confined to left dominated but bilateral STS regions (Adank, 2012a). Although specific design considerations (high or low baseline; sparse or continuous scanning; sentence or words) altered the core network to extend into anterior and posterior STS, frontal, parietal and motor regions, the vATL was not implicated. One critical factor was that the meta-analysis included 57 fMRI studies but only 3 PET studies (all of which found vATL activity). In addition to the signal loss in ATL regions using fMRI, Visser and colleagues found two other technical aspects found to affect the likelihood of detecting ATL activity; low-level baseline (i.e. rest) and small field-of-view (FOV) (Visser et al., 2009). Firstly, a low-level baseline can allow for subjects to daydream or produce inner speech during low-attention tasks, which are likely to activate semantic regions (Binder et al., 1999; Binder et al., 2011; Gusnard & Raichle, 2001; Stark & Squire, 2001). This is problematic in a subtraction analysis, where the regions related to semantic processing might be masked. Secondly, it is important that the FOV covers the anterior and inferior parts of the temporal lobe as they are often omitted using a small FOV. Visser and colleagues suggest studies with a FOV of 15 cm or less were not likely to find vATL activity.

It is important to overcome these methodological issues to be able to investigate the vATL reliably. Recently, a novel distortion corrected spin-echo fMRI method has been shown to alleviate the signal loss and distortion problems in the ATL and orbito-frontal regions (Embleton, Haroon, Morris, Lambon-Ralph, & Parker, 2010). Supporting evidence for the efficacy of the method has been seen in studies finding vATL activity during a variety of semantic tasks (Binney et al., 2010; Visser et al., 2010; Visser & Lambon-Ralph, 2011). The spin-echo technique is known to suffer from poor signal-to-noise ratio (SNR), where studies have found an approximate three-fold reduction in sensitivity compared to conventional gradient-echo (Bandettini, Wong, Jesmanowicz, Hinks, & Hyde, 1994; Jezzard, Matthews, & Smith, 2001; Norris, Zysset, Mildner, & Wiggins, 2002; Schmidt, Boesiger, & Ishai, 2005). An alternative approach is to use a multi-gradient echo technique (Poser & Norris, 2009; Poser, Versluis, Hoogduin, & Norris, 2006). This method collects an image volume relatively early after a radio frequency (RF) pulse, when the signal has not been lost, in combination with a conventional longer echo maximised for SNR across the rest of the brain. Recent fMRI studies have shown that multi-gradient EPI can be used to solve the problem of signal loss, whilst preserving the high-levels of SNR (Poser & Norris, 2009; Poser et al., 2006). The main advantage of this method would be the increase in SNR across the brain, which would allow for the study of neurocognitive functions using complicated designs and analysis methods.

4. Temporal evolution of speech processing

In order to understand brain functioning, it is important to understand both the spatial organisation (as described above) and the temporal dynamics of information processing. In order to understand the spatial organisation fMRI has been a critical tool, as it is known to provide a highly localised measure of brain activity (about 1-3 mm). However, the temporal resolution of fMRI is significantly longer than that needed to measure neural events, as each image is recorded every 1-3 s. Unlike EEG (or MEG), fMRI does not directly measure neural activity; instead, it relies on changes in blood oxygenation levels during stimulus presentations (Logothetis & Wandell, 2004). The amount of time it takes for local blood-oxygen levels to rise and peak is between 1-6 s (referred to as hemodynamic lag). Therefore, the temporal resolution of fMRI is not sufficient to detect rapid neural changes. M/EEG on the other hand provides a direct measure

of neural activity and can be sampled in the millisecond range (Coles & Rugg, 1995). The resulting data can be used to relate the physiological changes to properties of the stimulus (or task), in doing so increase our understanding of neuronal processing at a much finer temporal scale than fMRI methods allow.

EEG studies of speech processing have produced two competing frameworks to explain the neurophysiological data. One model suggests sequential processing of increasingly complex linguistic components (Friederici, 2002; Friederici, 2011). In this model the stages are: phoneme identification; word form identification (by combining phonemes); access to word meaning; reanalysis of semantic and syntactic structure. The model is supported by the fact that low-level sub-lexical effects emerge around 100-200 ms (Dehaene-Lambertz, 1997; Naatanen et al., 1997; Obleser, Lahiri, & Eulitz, 2003; Obleser, Scott, & Eulitz, 2006; Phillips et al., 2000). This suggests that the incoming auditory signal is initially parsed into language related complex acoustic features ready for subsequent combinations and processing. This is followed by modulation of a negative component around 400 ms (N400) – thought to reflect access to meaning. Furthermore, sentence level studies have shown, an early late anterior negativity (ELAN), which is thought to represent local phrase structure information (verb, noun or prepositional phrase; Friederici, 2002; Friederici, 2011). Finally, a positive component follows around 600 ms (P600) and is thought to reflect syntactic processes (Hagoort, Brown, & Groothusen, 1993).

As the N400 component is particularly relevant to the current thesis it will be discussed in more detail. This component has been widely investigated for words in context (sentence and discourse) (reviewed in Kutas & Federmeier, 2011) and less commonly for words in isolation. Studies that have reported similar N400 effects when using words in isolation include paradigms for naming (Stuss, Sarazin, Leech, & Picton, 1983), different types of lists, word – non-word pairs (Bentin, Kutas, & Hillyard, 1993), function-content words and high-low frequency words (Rugg, 1990; Smith & Halgren, 1987). Additionally, lexical priming tasks, where a target word was or was not related to a preceding word showed that related words reduced the N400 in comparison to unrelated words (reviewed in Kutas & Van Petten, 1988). This suggests that although context does seem to increase the amplitude of the N400, the underlying process is not tied to the context, and therefore is more likely to reflect an attempt to access the underlying semantic representations. This debate as to the precise circumstances that result in a N400 component has generated large amounts of research. For example, the

original publication used semantic violations to elicit the N400, which led to the misinterpretation that the N400 was a response in detecting linguistic anomalies or violations. It is clear however, that non-anomalous (albeit less expected) items also elicit moderate N400 responses (Kutas & Hillyard, 1980). The current understanding of the N400 is that all meaningful items elicit the N400 response, and have been found for all types of linguistic stimuli, including spoken, written and signed words, and word-like items such as pseudowords (Holcomb & Neville, 1990; Kutas, Neville, & Holcomb, 1987; Laszlo & Federmeier, 2008). The N400 response is not limited to linguistic stimuli as it has been observed in meaningful non-linguistic stimuli such as environmental sounds (i.e. animal sounds) (van Petten & Rheinfelder, 1995), line drawings and scenes (Ganis & Kutas, 2003), faces (Barrett & Rugg, 1989; Bobes, Valdessa, & Olivares, 1994), movies (Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008) and gestures (Kelly, Kravitz, & Hopkins, 2004). Although the N400 is often assumed to index some aspect of semantic integration, a number of different proposals have been put forward, including semantic memory retrieval and binding (Kutas & Federmeier, 2000), relating to orthographic/phonological analysis (Deacon, Dynowska, Ritter, & Grose-Fifer, 2004), semantic unification (Hagoort & van Berkum, 2007) and semantic inhibition (Debrulle, 2007).

In contrast to the serial processing framework some studies have shown semantic/syntactic affects modulating early components, around 100-200 ms (Pulvermüller, Assadollahi, & Elbert, 2001; Pulvermüller, Lutzenberger, & Birbaumer, 1995; Sereno, Cameron, & O'Donnell, 2003). This has led to the suggestion that sub-lexical, semantic and syntactical information are processed near-simultaneously in a parallel processing framework (Marslen-Wilson, 1987). This model can explain the early semantic/syntactic effects (around 100-200 ms) but the semantically modulated differences between late potentials become harder to explain, as do the lack of effects during late potentials. Both models can explain some findings but not all. Accordingly, Pulvermüller recently proposed a synthesis of the two models in which early first-pass parallel processes, are followed by a second-pass of parallel processes during which any inconsistencies during the first-pass are repaired (Pulvermüller, 2007).

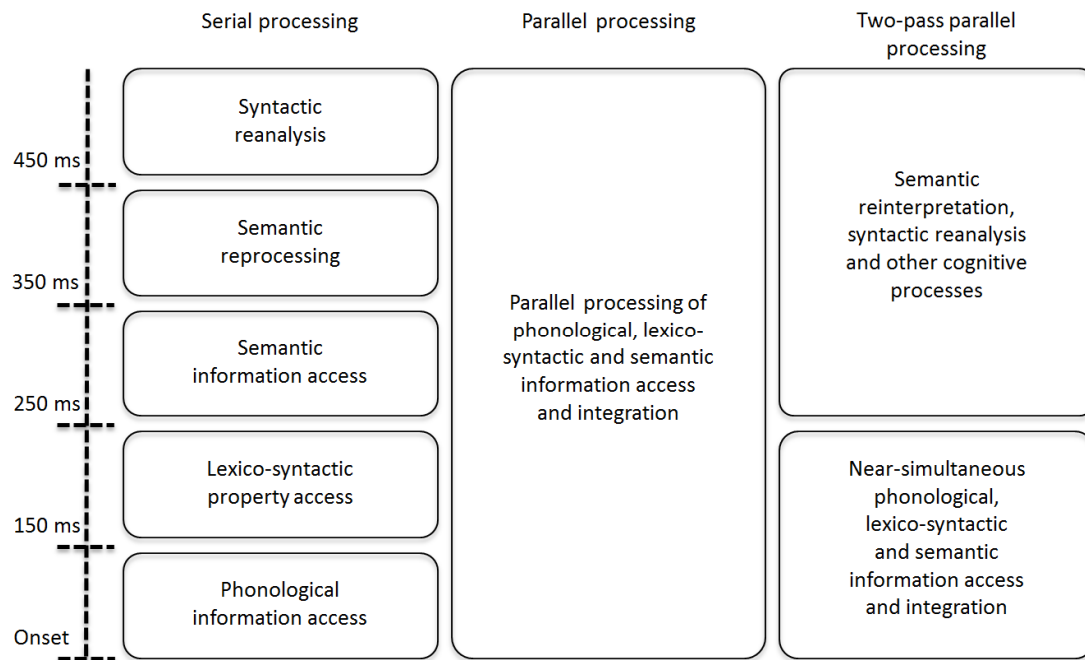


Figure 2: Comparing three temporal models of speech comprehension: left) Serial processing; middle) Parallel processing; and right) Two-pass parallel processing framework. Footnote: Estimation of linguistic processes after stimuli onset as predicted by three temporal speech processing models (adopted from Pulvermüller, 2007).

The overview of temporal processing models provides a useful insight into speech comprehension, however there is still a debate over the nature of the processing (Figure 2). There is evidence for a serial, bottom-up, framework that is consistent with the spatial organisation of the auditory ventral route stream. However, it has also been argued that comprehension processes occur in parallel or through a mixture of both serial and parallel processes. One consistent failing across these theories is that they fail to identify specific brain structures that could be involved in carrying out the specific processes. One of the main problems of recovering the source of neural activity from scalp M/EEG is that there are multiple solutions that can explain the observed data (Nunez & Srinivasan, 2006). One further problem with EEG specifically, is the issue of volume conduction where the electrical currents detected at the scalp have been deflected as the currents pass between different tissue densities (brain, skull and scalp). One alternative would be to use invasive cortical electrodes, which would provide extremely high spatial specificity, however the opportunities for such studies are rare. Therefore, a non-invasive alternative is to combine EEG-fMRI, which would provide both spatial and temporal information. Investigations into speech comprehension using this combination of neuroimaging methods have not yet been attempted. Although

recent models have attempted to draw conclusion about both spatial and temporal processing within language models (Bornkessel-Schlesewsky & Schlesewsky, 2013; Friederici, 2011), the inferences are largely made from separate uni-modal studies. The ability to use multi-model imaging methods to test speech comprehension tasks within the same subject should provide stronger statistical sensitivity than comparing across studies. Indeed, this platform would be suitable to test predictions of how speech comprehension is neurally organised in space and time.

Summary

The overview of the spatial organisation of speech comprehension leaves us with inconsistencies remaining. What we can determine is that intelligibility is processed within the temporal lobe, following a lateral hierarchical projection of acoustic signal from HG. While models of speech show strong agreement about the early stages of speech processing, there is no agreement concerning which parts of the temporal lobe (posterior or anterior superior) are responsible for accessing meaning of speech. Furthermore, it is not clear whether extra temporal regions play a central role in speech comprehension.

If we consider speech comprehension within a larger framework of semantic cognition, it becomes clear that the vATL is implicated in studies that look at access to meaning in other contexts (verbal and non-verbal). This suggestion has been shown to be consistent across a range of neuroimaging methodologies (neuropsychology, PET, MEG and TMS). However, this region is rarely implicated in models of intelligible speech processing, which is surprising as one might predict that access to representations of meaning would implicate the same areas regardless of how those stored representations are accessed. The overview of the semantic network suggested that IFG and IPC were not critical for comprehension, and are likely to play a more general role in linguistic processing, such as semantic control/selection, working memory load or acoustic-motor integration. These processes can be useful during difficult comprehension situations or can become active during explicit task requirements.

Careful analysis of studies that probe semantic activation reveal that there are imaging modality specific biases, which do not favour detecting activity within ATL regions. A large contributor for the lack of ATL evidence is the reliance on fMRI studies to inform speech comprehension models, which are affected by

magnetic susceptibility artefacts in the vATL regions. In contrast, research into semantic processing was largely reliant on neuropsychology studies, which has since been supported using other neuroimaging methods. The meaning of speech may be processed in the vATL, but previous fMRI studies have been blind to that region. This might also explain the lack of consensus on intelligibility processing in the speech models.

Finally, in order to understand speech processing fully it is important to use temporal as well as spatial information. One area of contention is whether speech processing is a bottom up or parallel process. Indeed, there is a suggestion that a hybrid theory of a two-pass parallel process could explain the available evidence. A non-invasive multi-modal method of combining EEG-fMRI, which can tie temporal and spatial processing, would be helpful in providing a more complete model of brain processing (e.g. Liebenthal et al., 2003; Menon, Ford, Lim, Glover, & Pfefferbaum, 1997).

Furthermore, the fundamental approach to cognitive neuroscience has subtly changed from structure-function relationships to dynamic brain networks and connectivity (Friston, 1994, 2003). This is often characterised in terms of either functional or effective connectivity. Definitions of these concepts tend to vary; however, here functional connectivity is defined as the quantification of interactions of multiple distinct brain regions that are engaged simultaneously in a given task. Effective connectivity is defined as the influence one neuronal group has over another, hence indicating a direction of processing influence. As fMRI alone cannot provide adequate insights into brain connectivity profiles during speech comprehension, using EEG in combination with fMRI would be critical in developing our understanding of the connectivity during comprehension.

A number of studies have combined fMRI and EEG using data recorded in separate sessions (e.g. Ball et al., 1999; Menon et al., 1997). An advantage of this approach is that the SNR of EEG data obtained outside the scanner is usually much better than that of data obtained inside the scanner. However, there may be significant subjective and experimental differences between two separate sessions. For example, there may be differences in subjects' levels of attention, motivation, and familiarity with the task. It has been shown that arousal levels are an important determinant of brain activation during cognitive tasks using simultaneous EEG-fMRI (Matsuda et al., 2002). In addition, we must consider that the two separate sessions do not provide the same environment, because during the fMRI session the subject is in a supine position and in a noisy environment,

whereas during the EEG session the subject generally sits in a comfortable chair in a quiet room. Simultaneous EEG-fMRI acquisition removes the methodological confounds arising from separate-session acquisition. In particular, for speech comprehension critical confounding variables such as practice and repetition effects, differences in task engagement, and differences in presentation orders would all be controlled.

Research objectives

Given this summary, it was clear there were a number of issues that the thesis should aim to address. The overall aim of this thesis was to use multi-modal imaging to refine our understanding of the spatial and temporal organisation of speech comprehension (Chapter 5).

However, before this could be achieved, it was critical to make sure that the methodology and task paradigm proposed was suitable for answering questions about speech comprehension within an EEG-fMRI setting. Therefore, Chapter 2 aimed to explore the temporal dynamics of intelligible speech comprehension, in particular with reference to how low-level complex acoustic features and intelligible speech processes modulated EEG components.

The second aim was motivated from the lack of evidence for vATL in speech comprehension models. Recent advances in fMRI have suggested two novel protocols that can overcome signal loss in the critical regions: distortion corrected spin-echo and dual gradient-echo. However, it was not known which of the two methods provide better detection capabilities within the vATL and the rest of the brain. In order to investigate this, Chapter 3 directly compared spin-echo and dual-echo fMRI during a semantic categorisation task known to activate the vATL.

Having established an fMRI method suitable for whole brain investigation, the next aim was to attempt to reconcile differences between speech comprehension and semantic comprehension tasks; in particular, to account for the omission of vATL within speech comprehension models. Chapter 4 aimed to identify intelligible speech regions in the brain using a previously established intelligible/unintelligible sentence comprehension paradigm.

Having established a suitable speech comprehension paradigm and fMRI protocol to use within an EEG-fMRI setting, we pursued the main aim of the thesis, which was to inform models of speech comprehension by outlining both spatial

and temporal dynamics (Chapter 5). This combined approach to speech comprehension is not currently implemented in the literature, and has the potential to inform and refine future iterations of models of speech comprehension in both spatial and temporal domains.

Chapter 2

Exploring the temporal and spatial characteristics of speech comprehension using electrophysiological evidence.

Ajay Halai¹, Laura M Parkes², and Stephen Welbourne¹

¹ Neuroscience and Aphasia Research Unit, School of Psychological Sciences,
University of Manchester,
Zochonis Building,
Brunswick Street.
Manchester, UK
M13 9PL

² Centre for Imaging Sciences, Institute of Population Health,
University of Manchester,
Stopford Building,
Oxford Road.
Manchester, UK
M13 9PL

Corresponding author:

Ajay Halai
3rd Floor Zochonis building,
Brunswick Street,
Manchester, UK
M13 9PL
Tel: +44 (161) 306 0451
Email: ajay.halai@postgrad.manchester.ac.uk

Keywords: Auditory; Speech; Comprehension; ERP; N400; Phonology; Intelligibility; Semantics

This chapter is currently under review in Brain and Language.

Abstract

Auditory speech comprehension extracts information from sounds and maps this onto stored semantic representations to obtain meaning. Three processing frameworks attempt to explain the temporal nature of word comprehension: serial, parallel and a mixture of both. Additionally, neuropsychological and neuroimaging studies have shown a hierarchical structural organisation of the temporal lobe for speech processing. Here, electroencephalography (EEG) is used to investigate how intelligible and unintelligible speech modulates event-related potentials (ERPs). Specifically, two ERPs were investigated in relation to low-level complex acoustic features and semantic processing. Larger amplitudes were found during an early peak (~150 ms) for conditions that contained complex acoustic features, which could plausibly be expected to generate a phonological code. Larger amplitudes were found during a late peak (~400 ms) for conditions with a valid phonological code that could plausibly be expected to map onto a corresponding semantic representation. The neural sources of these components were estimated, and the early component was localised to intermediate bilateral temporal regions, whereas the late component was localised to anterior temporal regions. In general, the results support a two-pass parallel processing framework. A general hierarchical framework is observed whereby phonological processing precedes comprehension; however, some top-down processing occurs during the early time window for distorted speech signals that contain features that are speech like. Furthermore, the localisation of these components provide converging evidence, as the early and late component show a hierarchical increase in complexity towards the antero-ventral temporal lobe, with executive function regions active during both early and late components.

1. Introduction

In recent years, there have been several attempts to outline brain networks for speech perception and production (e.g. Hickok & Poeppel, 2007; Price, 2010; Pulvermüller & Fadiga, 2010; Rauschecker & Scott, 2009). Alongside the spatial framework, a temporal framework has been proposed using EEG/MEG data, as

well as invasive single cell recording studies (e.g. Friederici, 2002; Marslen-Wilson, 1987; Pulvermüller, 2007; Rauschecker & Tian, 2000; Romanski et al., 1999). The present study aims to investigate the temporal evolution of single auditory word comprehension, whilst attempting to map how different brain regions become active as sounds become comprehensible. Brief overviews of the current temporal and spatial frameworks are discussed, followed by the specific aims and hypotheses.

1.1. Temporal dynamics of speech comprehension

Two broad frameworks have been proposed to explain neurophysiological psycholinguistic data. One model suggests sequential processing of linguistic information (Friederici, 2002). For example, the stages of comprehension can be split into the identification of phonemes, combining phonemes to produce word forms then mapping onto word meaning and finally re-passing and reanalysing the semantic and syntactic structure. The model is supported by simple pre-lexical effects around 100-200 ms, also known as P100/N200 (e.g. Naatanen et al., 1997). This is followed by modulation of a negative component around 400 ms (N400). This component has been widely investigated for words in context (sentence and discourse) (reviewed in Kutas & Federmeier, 2011), and less commonly for words in isolation (Bentin et al., 1993; Rugg, 1990; Smith & Halgren, 1987). Finally, a positive component follows around 600 ms (P600) which is thought to reflect syntactic processes (Hagoort et al., 1993). In contrast, studies have shown semantic/syntactic effects modulating early components, around 100-200 ms (Pulvermüller et al., 2001; Pulvermüller et al., 1995; Sereno et al., 2003). This 'higher' level information being processed very early is not consistent with a serial processing model. On the contrary, pre-lexical, semantic and syntactical information have been suggested to be processed near-simultaneously in a parallel processing framework (Marslen-Wilson, 1987). This model can explain the early semantic/syntactic effects (around 100-200 ms) but the semantic differences and the lack of phonological differences between late potentials become harder to explain. Furthermore, a parallel framework suggests that components reflect a mixture of processes and one individual process would be difficult to isolate. Both models can explain some findings but not all; therefore it may be helpful to view the models as complementary. Accordingly, a recent proposal suggested a combination of the models in which early first pass parallel processes, are followed

by a second pass of parallel processes during which any inconsistencies during the first pass are repaired (Pulvermüller, 2007). Figure 2 shows a graphical depiction of each temporal model of speech comprehension.

1.2. Spatial organisation of speech comprehension

The extensive data collected from patient studies on language impairment have been complemented with modern neuroimaging techniques. For example, the large number of published papers on language processing has allowed researchers to conduct meta-analyses (e.g. Price, 2010; Vigneau et al., 2011; Vigneau et al., 2006), develop models of speech comprehension and production (Hickok & Poeppel, 2007; Pulvermüller & Fadiga, 2010; Rauschecker & Scott, 2009) and extend to models of semantic memory/knowledge systems (e.g. Binder et al., 2009; Jefferies & Lambon Ralph, 2006; Pulvermüller et al., 2010; Visser et al., 2009).

Evidence from non-human primates supports a functionally distinct hierarchy for sound processing within the temporal lobe (Rauschecker et al., 1995). Studies on humans support similar sub-divisions, where an antero-lateral gradient is found for increasingly complex speech sounds (Binder, 2000; DeWitt & Rauschecker, 2012; Scott et al., 2000). In addition, the human speech system diverges into two distinct pathways, the ventral and dorsal routes. For the purpose of this report we focus on the ventral route, as this is predominately associated with meaning and identity (Hickok & Poeppel, 2007; Price, 2010; Rauschecker & Scott, 2009).

Hickok and Poeppel (2007), Rauschecker and Scott (2009) and Price (2010) propose models illustrating how sounds arrive at the cortex and the processing stages up to successful comprehension (Figure 1). The three theories are consistent on the origin of the incoming auditory signal, around Heschl's gyrus (HG), with information spreading via a ventral or dorsal route, depending on the task demands. However, the specific regions where sub-lexical and lexico-semantic representations are processed differ between these models (see Figure 1). For example, phonological features have been suggested to be processed within bilateral mid-posterior superior temporal gyrus (STG) (Hickok & Poeppel, 2007), and left mid-superior temporal sulcus (STS) antero-lateral to HG (Price, 2010; Rauschecker & Scott, 2009). Higher-level information such as lexical/semantic representations have been suggested to be processed within

bilateral but left dominated posterior middle temporal gyrus/inferior temporal sulcus (pMTG/pITS), while combinatorial processes for words, sentences and discourse occurs around anterior middle temporal gyrus/anterior inferior temporal sulcus (aMTG/aITS) (Hickok & Poeppel, 2007). In contrast, Rauschecker and Scott (2009) suggest that intelligible speech representations are confined to left anterior STS. Furthermore, Price (2010) suggests the anterior parts of the temporal lobe, including the superior, middle and inferior temporal gyri are involved during speech comprehension, with the addition of angular gyrus (AG) and inferior frontal gyrus (IFG) depending on task requirements.

The anterior temporal lobe (ATL) has been implicated in semantic processing based on neuropsychological research and recent neuroimaging data (Patterson et al., 2007). For example, an MEG study mapped the spread of activity for spoken and written words (Marinkovic et al., 2003). The activity for auditory and visually presented words extended down the ventral stream to the anterior temporal lobe and inferior prefrontal regions around 400 ms post stimulus (Marinkovic et al., 2003). Processing of written words was left lateralised; however, the spoken words did activate the bilateral ATL and IFG areas around 400 ms. Although the majority of neuroimaging and patient studies have shown left lateralised/dominated organisation for speech (Griffith, Rees, & Green, 1999), neuropsychological evidence from semantic dementia (SD) patients have pointed to bilateral ATL as 'semantic hubs' and crucial for comprehension (e.g. Jefferies & Lambon Ralph, 2006; Patterson et al., 2007). For example, SD patients show global amodal deficits, which do not improve even when patients are given a cue during a task, in contrast to transcortical sensory aphasia (TSA) patients who can improve after cueing. This implicates bilateral ATL regions as a store for semantic representations, which need to be accessed during successful comprehension.

In addition, the IFG has been implicated in speech comprehension. There is disagreement about the functional role of the IFG during speech comprehension. For example, Hickok and Poeppel (2007) suggest this region is primarily involved in articulatory processing (in combination with primary motor areas), whereas Rauschecker and Scott (2009) suggest there are high invariant properties available at this region to provide predictive constraints. Similarly, Price (2010) suggests the IFG imposes executive constraints on semantic selection to aid comprehension under difficult circumstances. Furthermore, the IFG has been suggested to be implicated in semantic control (e.g. Jefferies & Lambon Ralph, 2006; Whitney et al., 2011), which although is related to language processing in

general, it is not speech comprehension specific. One reason for a general role of the IFG might be the high-level of anatomical connectivity between this region and anterior STS, posterior STS and AG (e.g. Binney, Parker, & Lambon-Ralph, 2012; Parker et al., 2005; Saur et al., 2008).

1.3. Outline and aim

The aim of the current study was to measure the amplitude differences of event-related potentials (ERPs) generated by intelligible and unintelligible words, while also providing crude estimation of the spatial locations of any differences. The stimuli consisted of matched lexical and non-lexical words (pseudowords). Each of these conditions had acoustic transformations applied to them to manipulate phonological and semantic features, similar to the method of Scott and colleagues (Scott et al., 2000). The current study could not fully replicate the conditions (and therefore analyses) of Scott and colleagues due to limitations of applying noise vocoded transformation to single words. The transformation adds acoustic noise to speech, which results in the speech sounding like a harsh whisper. Sentences that have been noise vocoded are intelligible after short periods of training; however, single words do not become intelligible even after prolonged exposure. The type of analysis is affected by the exclusion of this condition. Ordinarily, a full factorial analysis (vocoded manipulation and rotated manipulation) could be performed along with a conjunction analysis for intelligibility (speech > rotated speech and noise vocoded speech > rotated noise vocoded speech). This conservative analysis provides a control against any interacting factors between the vocoded and rotated manipulations, isolating the intelligibility processes. However, we can still provide main effects of complex acoustic features and intelligible processes using single words. The complex acoustic features related to phonemes are preserved in spectrally rotated speech (i.e. intonation and pitch variation) and therefore can be isolated by observing differences between rotated speech and sounds where no phonological features are preserved. Semantic features are present in normal speech sounds and can be isolated by observing differences between intelligible and unintelligible speech. In addition, there are different levels of semantic processing that can be investigated, for example, differences between intelligible and unintelligible words are likely to be quite large as the latter does not contain any semantic information. However, one may compare words and pseudowords (lexicality differences),

where the latter does not contain specific semantic information. It could also be argued that word-like stimuli might engage the semantic system to a greater degree in a fruitless search for the correct lexical item. Finally, a more subtle difference might be expected for high and low imageability words, as these both contain semantic information, but the extent (or 'richness') of these representations are different.

The aim of the study was to explore how manipulating the phonological and semantic content of intelligible and unintelligible speech could modulate the early and late ERPs. The characteristics of the modulation by phonological and semantic features of the stimuli should allow us to adjudicate between serial and parallel models. We predicted that conditions with recognisable complex acoustic features would produce larger activity in the early time window compared to conditions without. Under the parallel processing view this early component should also be modulated by semantic content (indexed by lexicality or imageability). We also predicted that semantic content would modulate the late window (~400 ms). We expected this component to produce a larger negativity for conditions with semantic content than those without. In addition, we aimed to provide spatial estimation to describe approximately where these speech features are processed.

2. Methods

2.1. Subjects

Twenty-six healthy participants (7 male) took part in the study. All subjects were right handed, scoring at least 80 in the Edinburgh Inventory (Oldfield, 1971), native English speakers and had no known neurological or auditory problems. The research was approved by the University of Manchester Ethics Committee, UK.

2.2. Stimulus generation

There were two lexical conditions, words (W) and pseudowords (PW). These were presented in three auditory conditions: normal speech (W and PW), spectrally rotated (WRS and PWRS) and rotated noise vocoded (WRV and PWRV). Each lexical condition was presented in all three auditory modes, giving 6 conditions in total. In addition, the real words could be split into two groups based on high and low imageability (WH and WL), allowing for additional analysis of semantic effects. The spectrally rotated conditions were unintelligible but retained

variations in pitch and intonation (similar to phonological features) (Blessner, 1972). The rotated noise vocoded (6 bands) conditions were unintelligible and did not contain pitch variations (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). These manipulations allowed us to test for phonological and semantic processing (similar method to Scott et al., 2000). The psycholinguistic properties for the words used in this study were obtained from the CELEX word database, English Lexicon Project and The Irvine Phonotactic Online Dictionary (v2). In order to control the dataset as much as possible we imposed various constraints when selecting the word lists (see Table 1) - the lists were matched on word length ($t(39)= 0.338$, $p=0.700$), frequency ($t(39)= -0.947$, $p=0.349$), number of phonemes ($t(39)= -1.138$, $p=0.262$), phonological neighbourhood size (excluding and including homophones: ($t(39)= -0.262$, $p=0.795$), ($t(39)= -0.337$, $p=0.738$)), unstressed biphone probability average ($t(39)= 0.458$, $p=0.650$) and stressed biphone probability average ($t(39)= 1.747$, $p=0.088$). We obtained 80 words split equally between the two imageability conditions (40 items each, see Appendix 1). The mean imageability rating for the high imageability list was 554 ($SD=62.7$), and for low imageability list was 421 ($SD=59.4$). These mean ratings were significantly different ($t(39)= 10.789$, $p<0.001$). To create the pseudoword lists; we took the first phoneme from one of the real words and combined it with another word from the list to produce a phonologically legal pseudoword (80 items). This ensured the initial phoneme across groups was controlled. The word and pseudoword list were controlled on word length ($t(79)= -0.145$, $p=0.885$), phonological neighbourhood ($t(79)= -0.333$, $p=0.740$), unstressed biphone probability average ($t(79)= -0.118$, $p=0.907$), and stressed biphone probability average ($t(79)= 0.499$, $p=0.623$). An independent rater was given the pseudoword list to read aloud and confirm that the pseudowords did not sound like real words. Each word was recorded using Audacity software (<http://audacity.sourceforge.net>) at a sampling rate of 22050 Hz. The length of the words varied between 0.5-0.8 s, which was controlled between conditions. Each word was segmented using Praat editing software (<http://www.praat.org>), where the onset was determined by the spectrogram amplitude increase compared to baseline and the segment ended when the amplitude returned to baseline. Scripts were used in Praat to apply the rotated and rotated noise vocoded transformations to each word (W) and pseudoword (PW) to create the corresponding WRS, WRV, PWRS and PWRV conditions.

Stimuli	Freq	Image ability	Length	Phonemes	Phonological Neighbours	Un-stressed biphone probability average	Stressed biphone probability average
Word	166.88 (14.56)	487.83 (10.12)	4.10 (0.07)	3.14 (0.04)	30.55 (1.35)	2.36e-3 (2.41e-4)	1.28e-3 (1.24e-4)
Pseudo word			4.09 (0.06)	3.17 (0.04)	32.42 (1.83)	2.70e-3 (4.10e-4)	1.34e-3 (1.53e-4)
High Words	154.37 (17.15)	554.53 (9.91)*	4.13 (0.11)	3.10 (0.05)	30.25 (2.00)	2.45e-3 (3.10e-4)	1.48e-3 (2.20e-4)
Low Words	179.38 (23.60)	421.13 (9.40)*	4.08 (0.07)	3.18 (0.06)	30.85 (2.02)	2.27e-3 (3.80e-4)	1.07e-3 (1.11e-4)

Table 1. Psycholinguistic properties of the stimuli. Data obtained from CELEX, The English Lexicon Project and The Irvine Phonotactic Online Database.

*Footnote: There is no significant difference between any property, except between High and Low imageability words (marked *). Standard error shown in parentheses.*

2.3. Experimental procedure

Participants were asked to listen to audio clips and were told that they would later be asked questions about what they heard. E-Prime was used to present the stimuli. The experiment consisted of forty-eight 10 s blocks, where 10 items were presented in each block (rate of 1 s per item). The experiment lasted 16 minutes but was split into two 8-minute runs to allow the participant time to avoid tiredness. Each block consisted of items from one condition and the order of blocks was randomised for each participant. A 10 s rest gap was used between each experimental block. The experiment took place in a quiet, electronically shielded room that was designed for EEG acquisition. Participants were asked to close their eyes during the experiment to avoid visual contamination, and reserve any large movement during the silent gap or break.

2.4. Imaging acquisition and data analysis

Continuous EEG recording was obtained from 64 scalp electrodes using an ActiveTwo system, with 2 electrodes placed vertically above and below the right eye, and 2 electrodes placed horizontally at the corner of the right and left eye (BioSemi, Netherlands). The acquisition was digitised at 512 Hz in real time using Actiview software (BioSemi, Netherlands) and was written to file for offline analysis. The data was analysed using Brain Electrical Source Analysis 5.2 (BESA; Gräefeling, Germany). Trials were band-pass filtered (0.1-40 Hz) and re-

referenced using an average reference. Automatic artefact rejection was performed on the entire file using $\pm 100 \mu\text{V}$ detection threshold for both vertical and horizontal eye movements. Any outstanding artefacts (eye movements and other muscular activity) were removed manually, and any trials exceeding $100 \mu\text{V}$ were removed. Sensors were discarded if they displayed abnormally high amplitude fluctuations ($\pm 100 \mu\text{V}$) or high frequency changes across the epoch, which did not look consistent to neighbouring sensors. Epochs were defined with a 100 ms pre-stimulus baseline and 1000 ms post-stimulus duration. The remaining trials were exported into MATLAB (MathWorks, USA) and analysed using the Fieldtrip toolbox (<http://www.ru.nl/fcdonders/fieldtrip>) (Oostenveld, Fries, Maris, & Schoffelen, 2011). For each condition, two component peaks were identified and statistics were performed on subject averages. The early peak was identified as the maxima occurring within 50-200 ms post stimulus onset, with a 40 ms window taken around the peak. The late component was defined as the minima occurring during 300-600 ms and a 400 ms window was taken around the peak. As studies investigating language have noted laterality effects, we determined sensor clusters in the left and right hemispheres for each component. The ERP activity was averaged across the sensors within each cluster to increase SNR, and the components were identified for each cluster. The average amplitude of the peaks in each condition was included in statistical analysis. The data were analysed using separate $2 \times 2 \times 3$ repeated measures ANOVAs for the early and late components, where the independent variables were hemisphere (left and right), lexicality (word and pseudoword) and condition (normal, rotated and rotated noise vocoded speech). Two pre-planned contrasts were used to highlight complex acoustic processing related to phonemes and semantic processing. A contrast to identify phonological processing was determined by contrasting WRS with WRV (similarly for pseudowords PWRS against PWRV). A contrast to determine intelligible speech processing was determined by contrasting W with WRS (similarly for pseudowords PW against PWRS).

To determine source estimates, the linearly constrained minimum variance (LCMV) beamformer method (Van Veen, Van Drongelen, Yuchtman, & Suzuki, 1997) was applied to the time series data. This method produces a spatial distribution of the neuronal power. The pre-processed data from the ERP analysis was used to estimate the sources. Firstly, time windows of interest were extracted for the early and late components from the ERP data for each condition for each subject. A forward model was used to estimate the amplitude at the sensor level

for a given source distribution. The forward model was calculated using the leadfield at each location of a 3D grid with 1cm resolution (total of 670 points inside the brain). As the head distorts the EEG signal, it is necessary to use a head model that attempts to account for this distortion. In this case, a boundary element method (BEM) was used (Mosher, Leahy, & Lewis, 1999), which involved segmenting a T₁-weighted image (from MNI template) into white, grey and cerebral spinal fluid and using this information to obtain a brain boundary. The skin and skull boundary were determined by visually inspecting the T₁-weighted image and applying a threshold to separate them. The conductivity ratio for brain, skull and scalp was set at 1, 1/80 and 1 respectively. The LCMV algorithm calculates the power at each grid point during the time windows determined by the ERPs. The same pre-planned contrasts were used to isolate the phonological and semantic processing components as for the ERP analysis and the relative power change was calculated for each subject. The relative power change was calculated by first calculating the absolute change between conditions ($A - B$), and then expressing this as a ratio ($A - B / B$). This was then averaged across all subjects to obtain the overall relative power change for phonology and intelligibility. These power changes were then displayed on a brain template using an appropriate scale.

3. Results

After inspection of the EEG data, 6 subjects were removed. The reasons for removing the participants were that two participants had large movements throughout the entire experiment; two participants had more than 6 sensors (>10%) removed; two participants' data was atypical of auditory EEG data and displayed very large high frequency changes throughout the session, which looked like electrical interference. The following results are from the remaining 20 subjects.

3.1. Event-Related Potentials

Following artefact rejection, the percentage of trials remaining per condition (from all subjects) was approximately 90% (WH, 92%; WL, 92%, WRS, 89%; WRV, 89%, PW, 92%, PWRS, 91% and PWRV, 90%). Sixteen subjects had no sensors removed from analysis. However, four subjects had some noisy sensors,

which were removed. This resulted in the following 4 sensors being removed from all subsequent analysis: O1, O2, Oz and Iz. The topographical plots of each component were visually inspected to determine clusters of sensors (see Figures 3 and 4). Sensors showing the largest amplitude components were included. During the early component, the left cluster consisted of F1, F3, FC1 and FC3 and the right cluster consisted of F2, F4, FC2 and FC4. During the late component, the left cluster consisted of C1, C3, F1, F3, FC1 and FC3 and the right cluster consisted of C2, C4, F2, F4, FC2 and FC4. The average peak amplitudes (and standard error) for each condition are shown in Figure 5. The averaged waveforms are shown in Figures 3 and 4 (for words and pseudowords respectively).

To examine the results of the early component, a 2 x 2 x 3 ANOVA was carried out with three within group factors; hemisphere (left and right), lexicality (word or pseudoword) and condition (normal, rotated and rotated noise vocoded speech). There was no main effect of hemisphere ($F(1,19) = 0.846, p=0.369$), however significant main effects of lexicality ($F(1,19) = 15.536, p<0.002$) and condition were observed ($F(2,38) = 40.268, p<0.001$). There were no interactions between any of the conditions (all p 's >0.371). This lexical main effect showed words to have higher amplitudes than pseudowords, with an average of 1.01 μV and 0.75 μV respectively. This result requires careful interpretation; however, as it appears to be driven by differences in the rotated and rotated noise vocoded conditions (Figure 5) and so cannot be interpreted as evidence for a classic lexicality effect (real words greater than pseudowords). Further investigation into this difference confirmed that there was no lexicality effect between the W and PW conditions ($t_{\text{word}}(39) = 1.138, p=0.262$), whereas both rotated and rotated noise vocoded conditions showed significant "lexicality" effects ($t_{\text{rotated}}(39) = 3.932, p<0.001$; $t_{\text{rotated-vocoded}}(39) = 2.814, p=0.008$). We also conducted a separate 3 x 2 ANOVA (conditions by hemisphere) on the word conditions only (more equivalent to the analysis in Scott et al., 2000). This revealed a main effect of condition ($F(2,38) = 6.082, p=0.005$), which was driven by the fact that the WRV condition produced lower activity than either the W or WRS conditions. There was no main effect of hemisphere ($F(1,19) = 1.361, p=0.258$) and no interaction ($F(2,38) = 0.354, p=0.704$).

A planned contrast on the main effect of condition in the full ANOVA showed greater amplitude for conditions with complex acoustic features compared to conditions without ($F(1,19) = 7.804, p<0.001$). However, a planned contrast for an intelligibility effect was not significant ($F(1,19) = 1.371, p=0.123$). In the word

only ANOVA the results were similar. The planned contrast of phonological features was significant ($F(1,19) = 8.952, p=0.007$), whereas it was not for intelligibility ($F(1,19)=2.917, p=0.104$). An additional analysis was carried between the WH and WL imageability words, but no significant effect of imageability was found ($t(39) = -0.992, p=0.362$), during the early window.

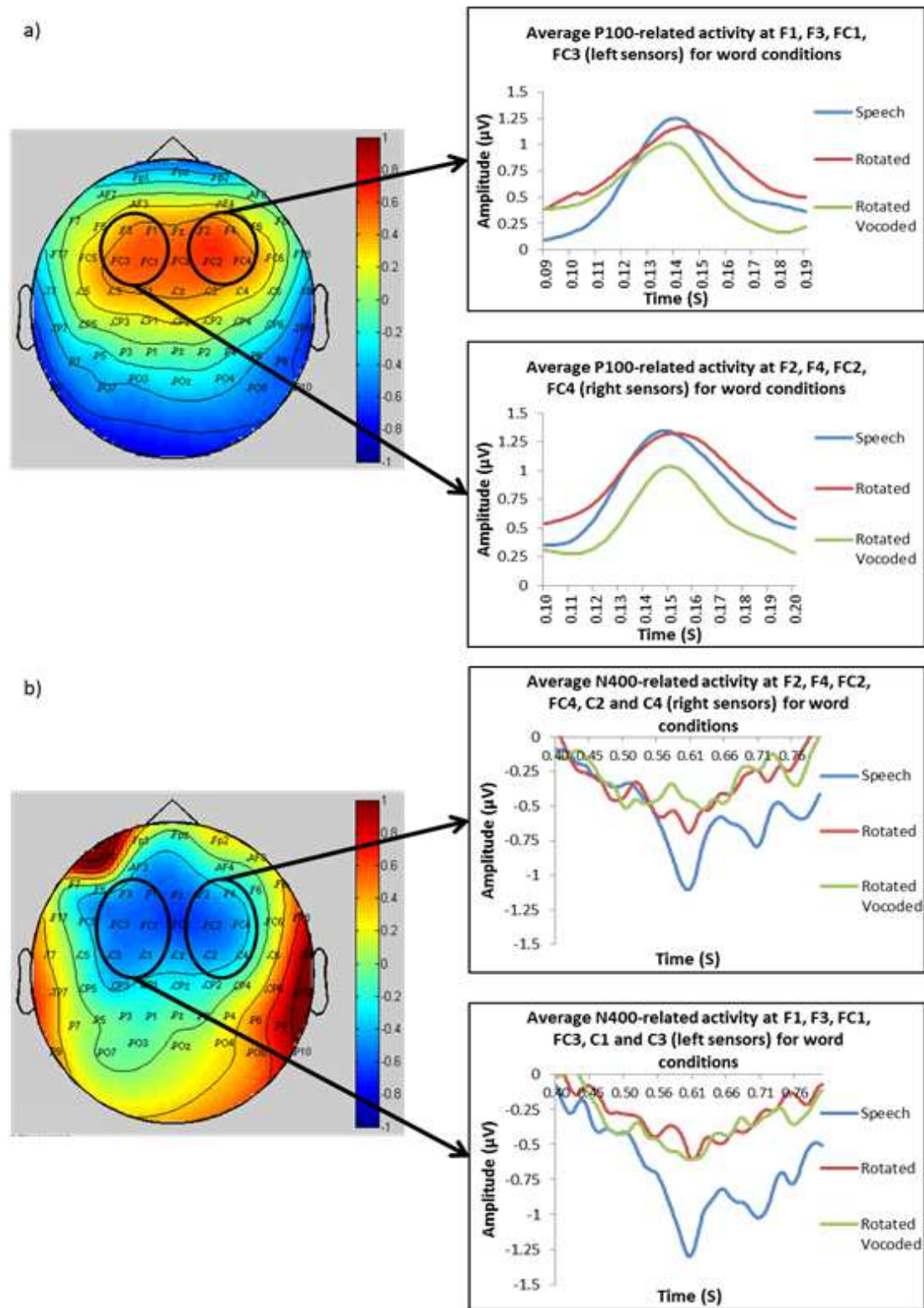


Figure 3: The average topographical plots and event-related potentials (ERPs) during an early (a) and late (b) time window for real word conditions. Footnote: The figures on the left outline the sensor clusters used for early and late time windows, with a scale of $\pm 1 \mu\text{V}$. The graphs on the right show the mean event-related potentials (ERPs) over all subjects within the cluster of sensors.

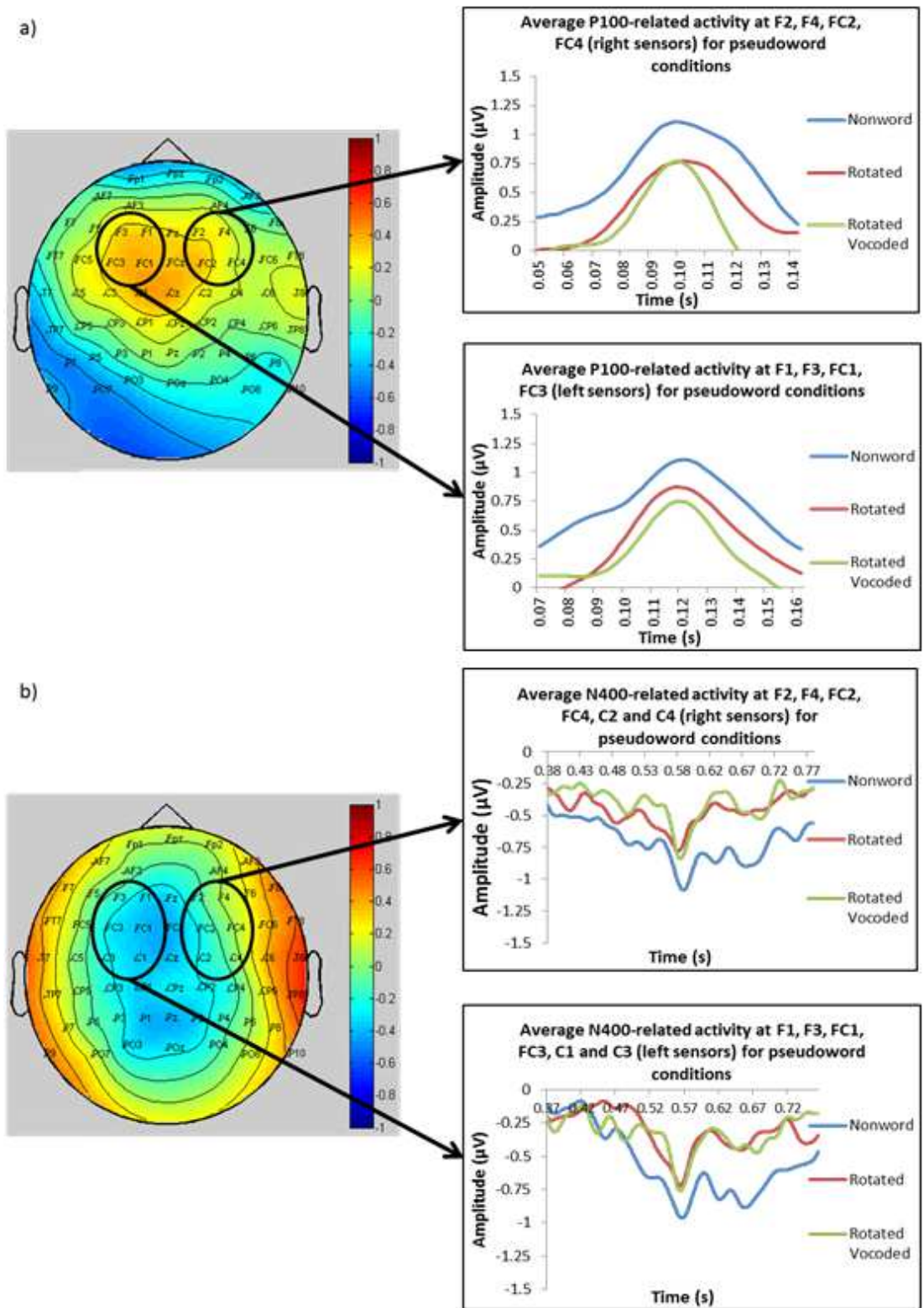


Figure 4: The average topographical plots and event-related potentials (ERPs) during an early (a) and late (b) time window for pseudoword conditions. *Footnote: The figures on the left outline the sensor clusters used for early and late time windows, with a scale of $\pm 1 \mu\text{V}$. The graphs on the right show the mean event-related potentials (ERPs) over all subjects within the cluster of sensors.*

To examine the results of the late component, a 2 x 2 x 3 ANOVA was carried out with three within group factors; hemisphere (left and right), lexicality (word or pseudoword) and condition (normal, rotated and rotated noise vocoded speech). There was no main effect of hemisphere ($F(1,19) = 0.226, p=0.640$) or lexicality ($F(1,19) = 0.154, p=0.650$), however a main effect of condition ($F(2,18) = 5.794, p=0.006$) emerged. There was a marginally significant interaction between hemisphere and condition ($F(2,18) = 2.653, p=0.083$) with intelligible stimuli producing larger negative responses in the left hemisphere, but no other significant interactions (all $p>0.786$). The planned contrast for intelligibility showed greater negative amplitude for intelligible conditions than unintelligible conditions (intelligibility effect; $F(1,19) = 10.626, p=0.004$). Furthermore, the planned contrast for phonological features was not significant ($F(1,19) = 1.806, p=0.195$). The hemisphere by condition interaction was further explored using the intelligibility contrast. This interaction revealed a significant intelligibility effect in the left hemisphere ($t(39) = -4.032, p<0.001$); however, the effect in the right hemisphere was not significant ($t(39) = -1.590, p=0.120$). There is a significant difference between these two effects ($t(39) = -2.067, p=0.045$), indicating a left lateralisation during the late component. An additional analysis was carried out just on the word condition, which could be split into high and low imageability groups, but there was no significant effect of imageability ($t(39) = -0.838, p=0.407$). An ANOVA analysis of just the normal word conditions showed a similar outcome: no main effect of hemisphere ($F(1,19) = 0.262, p=0.615$), main effect of condition ($F(2,38) = 5.144, p=0.011$) and no interaction ($F(2,38) = 0.809, p=0.453$). A planned contrast for intelligibility showed greater negative amplitude for intelligible over unintelligible conditions ($F(1,19) = 8.046, p=0.011$). Furthermore, the planned contrast for phonological features was not significant ($F(1,19) = 0.375, p=0.548$).

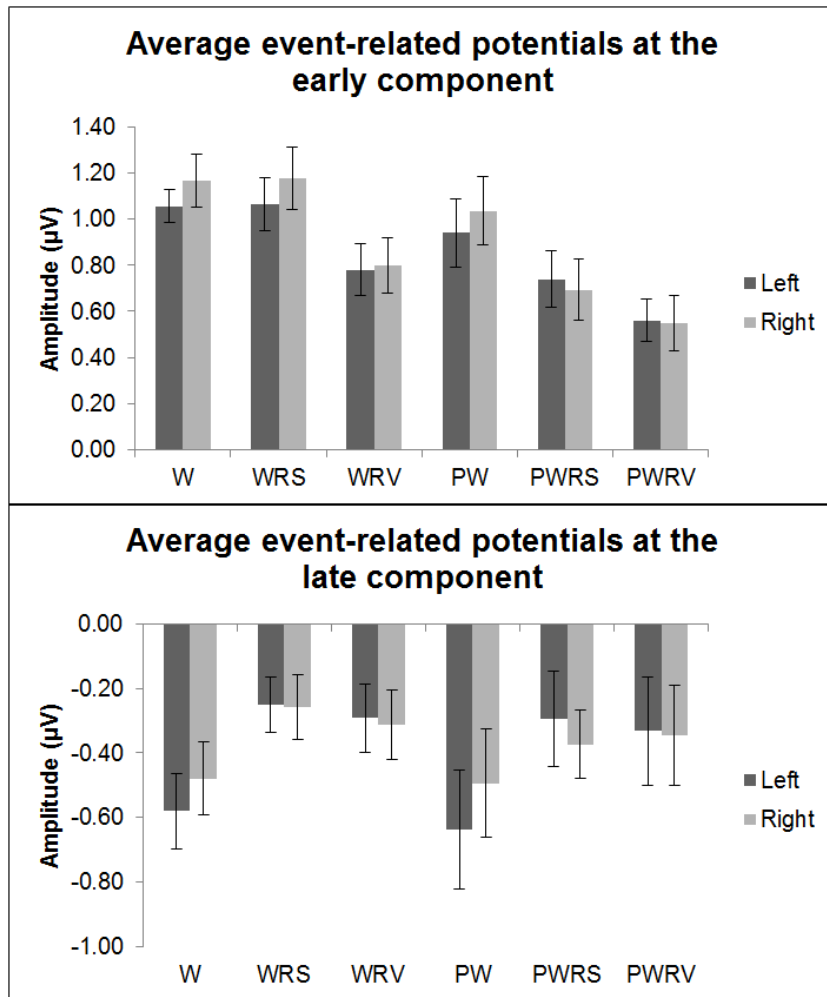


Figure 5: Event-related mean amplitudes averaged across 20 subjects for words and pseudoword conditions. Footnote: Event-related potentials were measured in each hemisphere for the early and late time window. The x-axis refers to the conditions: words (W), rotated speech (WRS), rotated noise vocoded speech (WRV), pseudoword speech (PW), pseudoword rotated speech (PWRS) and pseudoword rotated noise vocoded speech (PWRV).

3.2. Source analysis

The power was calculated for all 670-grid points for each condition, as described in the Methods section (2.4). Figure 6 shows the relative power changes for the phonology and intelligibility contrasts for real words, during early (Figure 6a) and late (Figure 6b) time windows. During the early window, there is increased power for phonological features in bilateral intermediate temporal regions, including mid-superior and inferior regions, although critically not extending down to the ATL. There is also increased power in posterior temporal regions bordering the parietal and occipital lobes, with the addition of bilateral activity in the frontal lobes. During the late window, there is increased power for intelligibility in bilateral ATL regions. Importantly, in the left temporal lobe, it spans across superior, middle and inferior

regions, whereas in the right temporal lobe it is mainly confined to superior regions. Although there is some evidence for increased power in intermediate and posterior temporal regions, it is largely diffuse. There is activity within the frontal lobe during the late window, although it is mainly within anterior and inferior regions. Figure 7 shows the relative power changes for pseudowords during the early (Figure 7a) and late (Figure 7b) window. The relative change during the early window is less evident compared to the real words, but there are hints of increased power for phonological features in the right STS and IFG regions. During late window, there is increased power for PW within left ATL regions and a very small portion of the right temporal lobe. The temporal lobe activity is similar to the activity seen for intelligibility in real words. Furthermore, power changes in the IFG for PW are not as widespread as was found in the intelligible real word contrast.

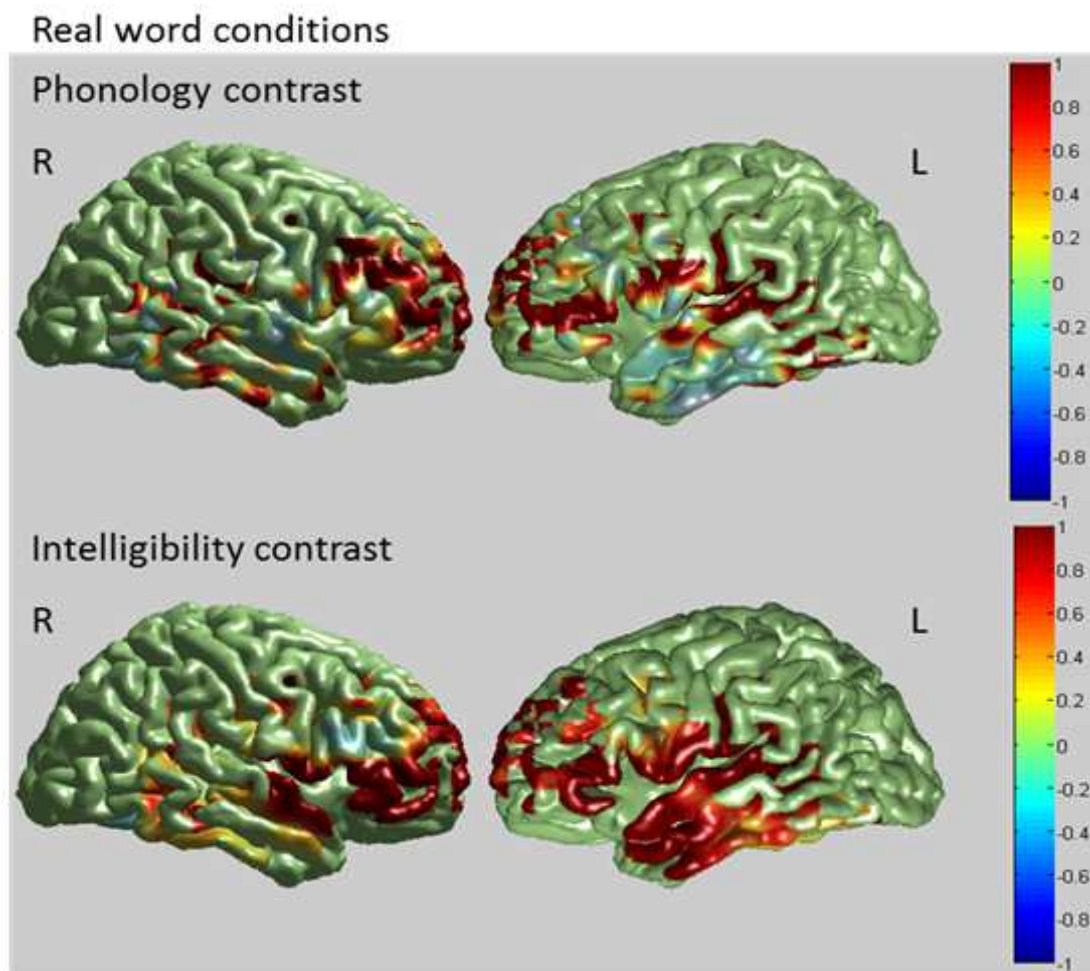


Figure 6: Beamformer results for early processing window (phonological difference) and late processing window (intelligibility difference) for real word conditions. *Footnote: The scale is ratio change (A-B/B) at ± 1 .*

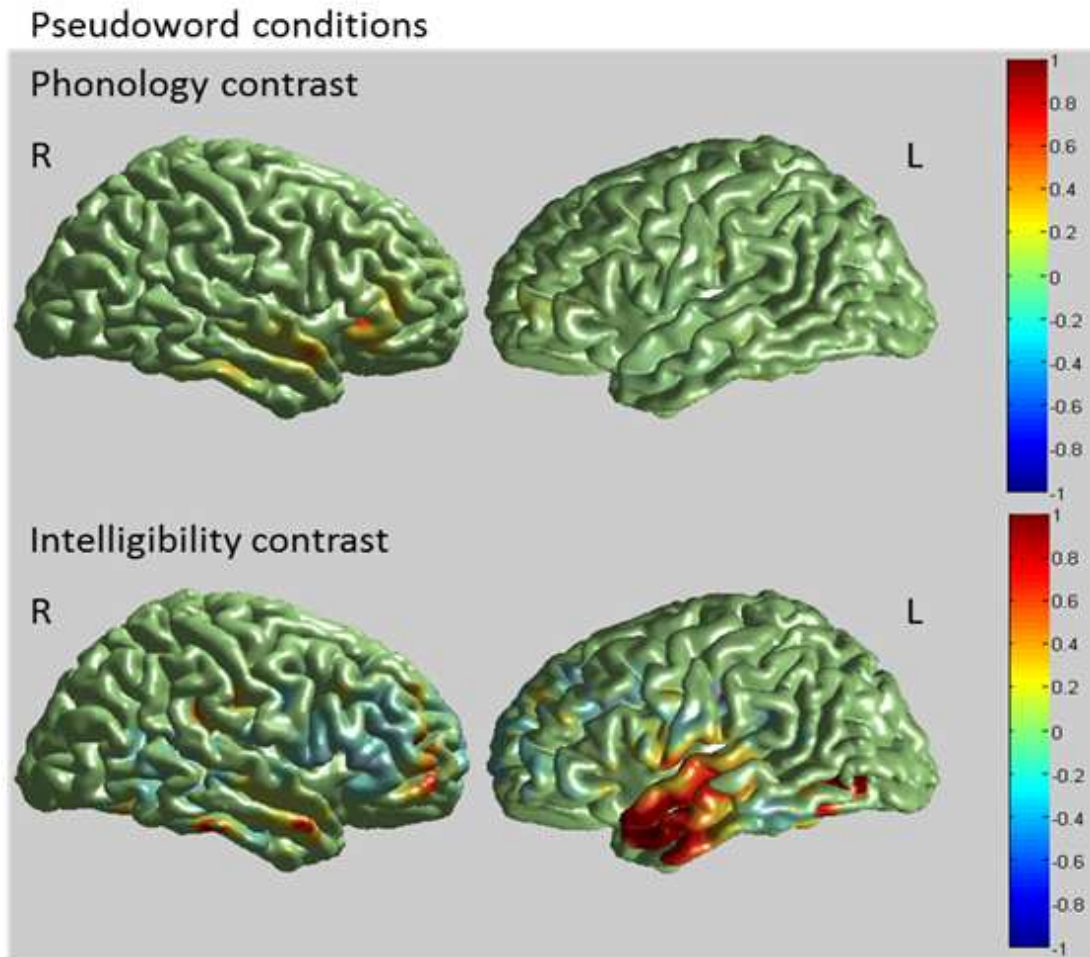


Figure 7: Beamformer results for early processing window (phonological difference) and late processing window (intelligibility difference) for pseudoword conditions. *Footnote: The scale is ratio change (A-B/B) at ±1.*

4. Discussion

The aim of this study was to measure the amplitude differences of ERPs generated by intelligible and unintelligible words, while also providing estimates of the spatial locations of these differences. Specifically, investigating how early and late ERPs were modulated by manipulating phonological and semantic content in speech could help determine if the processing occurs in serial, parallel or via a mixture of both processes. In addition, localising the early and late components could help distinguish between current models of speech comprehension. It is noted however, that the spatial resolution obtained using EEG source reconstruction is generally poor (across cm not mm). The fundamental cause for this is that electrical signals are deflected as they cross different tissue densities (volume conduction). Therefore, it is important to use caution when interpreting

source results from EEG. However, one can still use the results as approximate estimations, which is the approach we take in the current study. Participants listened to auditory clips of words and transformed words whilst electrophysiological data was recorded. The results suggested a complex answer for the temporal evolution of speech comprehension. For normal word processing the evidence points to a relatively simple serial mechanism with complex acoustic features related to phonemes being processed in the early window and semantic processing occurring during the late window. However, when the stimuli are distorted there is evidence for some additional top-down processing with distorted stimuli that have some similarity to known words generating a higher signal than distorted pseudoword stimuli. Therefore, taken as a whole, the results support a two-pass parallel processing framework. During the first time window, the incoming signal is assessed for general speech-like qualities based on existing stored information (top-down constraints), but in general the lower-level features (i.e. complex acoustic variations) are dominant. The prefrontal cortex is heavily associated with executive control or selection processes (i.e. Price, 2012), and so the co-activation of this region during the early component suggests a similar influence is applied to distorted stimuli. During the second window, detailed information from the signal is mapped onto stored semantic representations (possibly related to a N400-type component). The early component was localised around the mid-STS, HG and frontal regions. In contrast, the late component was localised, anteriorly to the early component, in bilateral ATL regions, with some additional frontal activity. The results for each component will be discussed in more detail below.

4.1. The early component

The ERP findings are largely consistent with a hierarchical framework for single word comprehension with processing in this window focused on identifying the complex acoustic features related to phonemes within the stimuli. However, given the clear lexical effect within the distorted stimuli, it is likely that some higher-level information is processed within this window. This cannot be based on semantic/lexical information as the effect is present in the rotated and rotated noise vocoded stimuli, which are unintelligible. Therefore the best explanation is that there is an additional top-down phonological/lexical process that enhances

processing for those distorted stimuli that appear to bear some resemblance to known phonological forms.

The serial processing framework suggests that low-level complex acoustic features related to phonemes are processed before higher-level (semantic and/or syntactic) features and our results are largely consistent with that. The conditions that contained complex acoustic features (including variations in pitch and intonation) all produced a larger early peak compared to conditions without. This suggests that phonemes, or specifically phoneme like pitch variation is processed relatively early after stimulus onset. It is well established that processing for complex acoustic features related to phonemes is likely to occur around 100-150 ms after stimulus onset (i.e. Naatanen et al., 1997) so this result is not surprising. However, the serial processing framework cannot account for the early lexical effect observed in the rotated and rotated noise vocoded conditions. Therefore, the two pass parallel framework best explains the pattern of results. This framework can also account for studies that have shown high-level processes being modulated during the early component around 100-200 ms when comparing words and pseudowords (Rugg, 1983), word types such as function, content, visual, and multimodal nouns (Pulvermüller et al., 2001; Pulvermüller et al., 1995; Sereno et al., 2003) and syntactic violations (Friederici & Kotz, 2003; Neville, Nicol, Barss, Forster, & Garrett, 1991). Our results did not find differences between word and pseudowords, intelligible and unintelligible words, or high and low imageability words during early window. It should be noted however, that this absence of effect does not mean that subtle top-down effects are not present. It could be argued that the current study may not have been sufficiently sensitive. However, for distorted stimuli, there was a 'lexical' effect with the distorted words generating a stronger response than distorted pseudowords. This suggests that at the early level of analysis these distorted items appear in some way familiar and that this phonologically based familiarity produces enhanced processing in an attempt to resolve these stimuli to a known form. Overall these results suggest that a relatively quick early pass is conducted on the incoming signal, where predominately lower-level complex acoustic features are processed with some top-down influences from existing abstract phonological representations.

Further investigation revealed that the early component attributable to the word list peaked slightly later than the pseudowords (approximately 140 ms vs. 110 ms respectively). This was not expected, as the stimuli length was obtained by determining when the spectrogram energy deviated from baseline. It is possible

that this latency difference could reflect differences in the articulation of the pseudowords compared to words. When the latency to the first pitch variation was compared, it emerged that pseudowords had significantly earlier onsets than real words. Therefore, it is possible the early component is sensitive to pitch variation as well as spectrogram energy.

There has been general criticism towards studies that fail to find lexical or semantic effects during early ERP (or related magnetic) component (reviewed in Pulvermüller, 2007). It is suggested that differences during early components are small in comparison to the late components, and therefore there is a need to keep stimulus variance low by controlling as many psycholinguistic variables as possible. For example, Penolazzi and colleagues did find that word frequency and close probability (the likelihood of a word ending a sentence) interact with word length for early potentials (100-200 ms) but not for late potentials (Penolazzi, Hauk, & Pulvermüller, 2007). In order to keep stimulus variance as low as possible in the current study, the stimuli were controlled on psycholinguistic variables such as the word length, frequency, phonological neighbourhood size (including homophones), unstressed/stressed biphone probability average and phoneme onset. However, in spite of this, intelligible semantic information did not modulate the early component significantly. However, we cannot rule out the possibility that increasing the statistical sensitivity of the current study, for example by increasing the number of subjects or number of items, might reveal a small early semantic effect.

The localisation of the early component provides support for a two-pass parallel framework, with a hierarchical structure within the temporal lobe. The temporal lobe is thought to be organised in a functionally hierarchical manner based on human neuroimaging (Binder, 2000; DeWitt & Rauschecker, 2012) and non-human primate studies (Rauschecker & Tian, 2000; Tian et al., 2001). In addition, a recent meta-analysis reviewing the anatomy of language showed pre-lexical processing (acoustic analysis of frequency spectrum and integration of features over time) to be largely confined to the intermediate part of the temporal lobe around HG (Vigneau et al., 2006). The current results provides support to Price (2010), who suggests that early processing is confined around HG extending down the STS anteriorly and laterally, with the addition of IFG resources. Price suggests that top-down processes can influence the location of pre-lexical processing along the STS. For example the posterior STS can be activated during a speech production task, whereas the ventral STS can be activated during a

speech comprehension task (Price, 2010). The activity for phonological features provides partial evidence for Rauschecker and Scott (2009) as although activity is seen within the STS, the current data does suggest a widespread activation to be involved. Furthermore, the Hickok and Poeppel (2007) model which suggests that pre-lexical phonological features are processed in bilateral mid-posterior STS is also not entirely consistent with the current data, as the majority of the activity is found around the PAC in the superior temporal sulcus.

It has been shown that the frontal system becomes more active during explicit language tasks, which is not generally seen during passive tasks (Mummery et al., 1999; Poeppel, 1996). However, for the current study, although a passive design was used there was still indication of widespread activity within bilateral frontal regions. There is evidence for the IFG being involved during executive control and selection processes (Badre & D'Esposito, 2009; Corbett et al., 2009; Jefferies & Lambon Ralph, 2006; Noonan, Jefferies, Visser, & Lambon Ralph, 2013; Thompson-Schill et al., 1997). These processes could be applied as top-down effects, in which distorted speech-like signals produce control and selection constraints in attempt to aid comprehension. For example, it has been shown that when incoming stimuli are distorted there is a greater reliance on regions involved in executive processes (IFG and AG) to extract meaningful information (i.e. Obleser, Wise, Alex Dresner, & Scott, 2007a). Such a mechanism would fit with a parallel or two pass parallel framework. However, an alternative possible explanation for the inferior frontal activity can be provided by a meta-analysis, which found that frontal-motor regions were involved in overt or covert articulation of phoneme, syllables, letters, pseudowords or word repetition (reviewed in Vigneau et al., 2006). During the passive task, although they were instructed not to, participants may have covertly articulated some items, resulting in frontal-motor region involvement.

4.2. The late component

There has been considerable effort invested in researching the N400 component of language processing. This has generally focused around words in context within sentence or discourse (Kutas & Federmeier, 2011). Some studies have found similar N400 effects when using words in isolation for naming tasks (Stuss et al., 1983), or in different types of lists, word – non-word pairs (Bentin et al., 1993), function-content words and high-low frequency words (Rugg, 1990;

Smith & Halgren, 1987). Additionally, lexical priming tasks, where a target word was or was not related to a preceding word showed that related words reduced the N400 in comparison to unrelated words (reviewed in Kutas & Van Petten, 1988). This suggests that although context does seem to increase the amplitude of the N400, the underlying process is not tied to the context, and therefore is more likely to reflect an attempt to access the underlying semantic representations.

As well as considering data from previous studies on spoken word recognition, it is also instructive to consider data from reading studies, which might be expected to activate similar networks. One startling aspect of these studies is that the N400 response is not limited to words, but is also generated in response to plausible sounding pseudowords and is modulated by orthographic neighbourhood (Laszlo & Federmeier, 2011). Thus it seems that, for reading at least, the N400 amplitude is not a direct metric of semantic processing, but rather a marker of attempted semantic access (regardless of whether this attempt is ultimately successful in generating a representation). Perhaps a better way of think about the N400 in the context of speech or reading would be to regard it as an inverse measure of how easy it is to reject the possibility that an item may have a semantic representation.

Based on neuroimaging and neurophysiological studies, we hypothesised those auditory items that had plausible phonological codes (undistorted words and pseudowords) would generate the strongest N400 responses and that these would localise to anterior temporal regions. The N400 is thought to be modulated by amodal semantic content, which converges with evidence from semantic dementia patients where the ATLS are thought to house amodal semantic representations (e.g. Patterson et al., 2007). During the late window, an effect of intelligibility was found with larger negativity for intelligible words and pseudowords compared to acoustically matched unintelligible words. Although this contrast suggests that the component is related to a language processing system with semantic properties, the paradigm was not specific enough to distinguish between whether this component is specific for semantic memory access (Kutas & Federmeier, 2000), semantic integration (Hagoort, 2009), neural binding processes (Federmeier & Laszlo, 2009) or inhibiting language-related neighbours that are not related to the task (Barber & Kutas, 2007). Similar to data from reading studies, the undistorted pseudowords modulated the N400 component and produced activity within the anterior temporal lobe, which are thought to process semantic representations. It is possible that these stimuli initiate a fruitless lexical search, which requires

semantic memory access. There is also the possibility that the brain is attempting to integrate the pre-lexical features of the pseudowords to a semantic representation, where a meta-analysis has shown that pseudowords increase the demand on pre-lexical features that are being integrated (Davis & Gaskell, 2009). These results suggest that the N400 is generated at the moment where stimulus processing has progressed to a point where there is an attempt to try and map the phonological code to a semantic representation. It appears to be this attempted semantic access that generates the signal. This is supported by the fact that conditions with pitch variation only and noise bursts do not produce a N400 component (Figure 5), but phonologically intact words and pseudo words both generate a robust N400 response. This result is difficult to reconcile with a fully parallel processing framework, as it would predict phonological and semantic differences during both early and late processing windows.

Although the majority of the studies on N400 use a mismatch, oddball or word pairing paradigm (e.g. Bentin et al., 1993; Rugg, 1990; Smith & Halgren, 1987), the current work does show that the component can be elicited by using random single intelligible/unintelligible words presented in a list. Overall these data appear to support a two pass parallel processing framework, where after initial phonological processing has been achieved during the early window with the help of top-down attention or selection constraints from prefrontal regions, the late window attempts to automatically access a corresponding semantic representation.

The spatial localisation of the N400 activity is found anterior to the activity estimated for phonological feature processing (Figures 6 and 7), which is consistent with an antero-lateral gradient shift for increased speech sound complexity and hierarchical organisation of the temporal lobe (Binder, 2000; Scott et al., 2000). An interaction between hemisphere and condition revealed that an intelligibility effect was present in the left hemisphere sensors, while this effect in the right hemisphere was not significant. It has been argued in the past that intelligibility is processed bilaterally, with a bias towards the left hemisphere, which is consistent with the current results (e.g. Griffith et al., 1999). For example, Lambon-Ralph and colleagues suggest that although both hemispheres play a part in comprehension, there is stronger verbal to semantic coupling in the left hemisphere, based on neuropsychological and computational modelling data (Lambon-Ralph, McClelland, Patterson, Galton, & Hodges, 2001). This is further supported by MEG data showing bilateral anterior temporal activity for spoken

words during the early stages, but after 300 ms it is biased to the left hemisphere (Marinkovic et al., 2003). The localisation of the late component in this dataset provides supporting evidence for these studies, which indicate the anterior temporal pole to be involved in semantic memory retrieval (Jefferies & Lambon Ralph, 2006; Patterson et al., 2007; Pobric et al., 2010). Similarly, the anatomical model of language suggested by Price (2010) does explain the current set of results for the word and pseudoword N400 activity. For example, the model suggests that words activate ventral, anterior and posterior borders of the perisylvian regions. This distributed activation is not entirely consistent with Rauschecker and Scott (2009), as they propose that speech is confined to the anterior STS. Additionally, Hickok and Poeppel (2007) suggest bilateral posterior MTG for phonological processing, which is not supported. They also propose that the aMTG is involved in combinatorial processes at the sentence and discourse level, but the current data provides evidence against such a role as the anterior temporal pole can be activated at the single word level.

Finally, the IFG was active during the N400 even though the current design used a passive listening task. As stated previously, studies have suggested that explicit tasks increase the activations of IFG whilst when conducting passive tasks the activity generally does not appear (Mummary et al., 1999; Poeppel, 1996). However, based on the current results it is suggested that IFG activity plays an active role in language comprehension, where it is likely this region exerts top-down attention or selection influences during adverse conditions for successful comprehension (Price, 2010). Therefore, for single auditory word comprehension, the current data suggests that the anterior/inferior temporal regions and IFG regions are related to speech comprehension, supporting the anatomical model proposed by Price (2010).

5. Conclusions

This study investigated the temporal and spatial organisation of speech comprehension. The temporal data were best explained by a two pass parallel processing framework (Pulvermüller, 2007), where the early component is generated by attempts to identify a phonological code corresponding to the stimulus with the aid of stored representations, whereas the N400 component was generated by attempts to use the phonological code to access a semantic representation. A hierarchy was observed for the spatial organisation of word

comprehension. During the early component, activity was found in bilateral intermediate temporal, whereas during the late component this activity moved towards the anterior and inferior temporal cortex. Both processing windows observed frontal involvement, which could indicate top-down attention or selection influences to aid distorted signals during each processing stage. The regions estimated to process pre-lexical and semantic features are supported by evidence of anatomical and functional models of language comprehension and semantic dementia (Jefferies & Lambon Ralph, 2006; Price, 2010); where the anterior portion of the temporal lobe is mainly involved, but there are contributions along the STS and from IFG regions.

Chapter 3

A comparison of dual gradient-echo and spin-echo fMRI of the inferior temporal lobe.

Ajay Halai¹, Stephen Welbourne¹, Karl Embleton² and Laura M Parkes³

¹ Neuroscience and Aphasia Research Unit, School of Psychological Sciences,
University of Manchester,
Zochonis Building,
Manchester, UK
M13 9PL

² Cognition and Cognitive Neuroscience Group, School of Psychological Sciences,
University of Manchester,
Zochonis Building,
Manchester, UK
M13 9PL

³ Centre for Imaging Sciences, Institute of Population Health,
University of Manchester,
Stopford Building,
Manchester, UK
M13 9PL

Corresponding author:

Ajay Halai

3rd Floor, Zochonis building,

Brunswick Street,

Manchester, UK

M13 9PL

Tel: +44 (161) 306 0451

Email: ajay.halai@postgrad.manchester.ac.uk

Keywords: dual echo; spin echo; EPI; magnetic susceptibility; inferior temporal lobe

This chapter is currently under review in Human Brain Mapping.

Abstract

Magnetic susceptibility differences at tissue interfaces lead to signal loss in conventional gradient-echo EPI. This poses a problem for fMRI in language and memory paradigms, which activate the most affected regions. Two methods proposed to overcome this are spin-echo EPI and dual gradient-echo EPI, where two EPI read-outs are serially collected at a short and longer echo time. The spin-echo method applies a refocusing pulse to recover dephased MR signal due to static field inhomogeneities, but is known to have a relatively low blood oxygenation level dependant (BOLD) sensitivity. In comparison, gradient-echo has superior BOLD sensitivity, and by employing an additional shorter echo, in a dual gradient-echo sequence, it can reduce signal loss due to spin dephasing. We directly compared dual gradient-echo and spin-echo fMRI during a semantic categorisation task, which has been shown to activate the ventral temporal region and is a region known to be affected by magnetic susceptibility artefacts. A whole brain analysis showed that the dual gradient-echo resulted in significantly higher activation within the left ventral anterior temporal lobe (vATL), compared to spin-echo. The inferior frontal gyrus (IFG) was activated for dual gradient-echo, but not spin-echo. A region of interest analysis was carried out on the vATL, IFG and part of the cerebellum. Dual gradient-echo outperformed spin-echo in the anterior and posterior portions of the vATL and bilateral IFG regions, whilst being equal in the cerebellum. Hence, dual gradient-echo should be the method of choice for fMRI studies of ventral temporal regions.

1. Introduction

Single shot gradient-echo (GE) echo planar imaging (EPI) is the method of choice for the majority of fMRI research. Although this method offers high contrast-to-noise ratio (CNR), it is affected by magnetic field inhomogeneity artefacts, which include signal loss, image blurring and geometric distortion. Signal loss is caused by phase dispersion within the voxel and is predominately due to the through-plane (z-direction) local susceptibility-induced field gradients due to the interface of tissues with different magnetic susceptibilities such as brain and air. This is a particular problem in the ventral anterior temporal lobes (vATL), which include the inferior temporal gyrus (ITG) and temporal fusiform gyrus (TFG), as the specific

anatomy at the base of the brain results in larger field gradients in the z-direction. Furthermore, signal loss is exacerbated at higher field strengths (Deichmann, Josephs, Hutton, Corfield, & Turner, 2002; Jezzard et al., 2001; Schmithorst, Dardzinski, & Holland, 2001; Weiskopf, Hutton, Josephs, & Deichmann, 2006). The frequency shift of the MR signal by in-plane local field gradients, on the other hand, results in pixel shift and hence distortion of the images. These distortion artefacts are more prominent in the phase-encoding direction, because the data sampling rate is much lower than in the frequency-encoding direction, and small frequency shifts will lead to larger shifts in distance in the eventual image (Deichmann et al., 2002). Parallel imaging techniques such as SENSE (Pruessmann, Weiger, Scheidegger, & Boesiger, 1999) can reduce, but not solve, distortion and signal loss by reducing the time needed for data sampling.

Signal loss in ventral regions presents a particular problem for fMRI as these regions become inaccessible for statistical analysis and may be interpreted as false negative activation. There have been many attempts to overcome this problem for fMRI, using methods such as tailored radiofrequency (RF) pulses (Cho & Ro, 1992; Stenger, Boada, & Noll, 2000), z-shimming (Constable & Spencer, 1999; Cordes, Turski, & Sorenson, 2000; Deichmann et al., 2002; Frahm, Merboldt, & Hänicke, 1988; Ordidge, Gorell, Deniau, Knight, & Helpert, 1994), optimisation of the slice orientation (De Panfilis & Schwarzbauer, 2005; Deichmann, Gottfried, Hutton, & Turner, 2003), reduced slice thickness (Bellgowan, Bandettini, van Gelderen, Martin, & Bodurka, 2006), passive shimming using mouth inserts (Cusack et al., 2005), alternative k-space sampling strategies (Glover & Law, 2001), dual gradient and spin-echo acquisition (Schwarzbauer, Mildner, Heinke, Brett, & Deichmann, 2010; Schwarzbauer & Porter, 2010) and multi-echo acquisitions (Poser & Norris, 2007; Poser et al., 2006; Posse et al., 1999; Volz, Hattingen, Preibisch, Gasser, & Deichmann, 2009).

In contrast to a gradient-echo approach, a spin-echo EPI is robust against signal loss caused by magnetic field inhomogeneities due to the 180° refocusing pulse. Image distortion problems remain but spin-echo is also amenable to distortion correction of the recovered signal (Embleton et al., 2010). A study using spin-echo fMRI showed frontopolar and ventral frontomedial activity during a Stroop colour-word matching task, which could not be detected using gradient-echo (Norris et al., 2002). Furthermore, a comparison of fMRI activation measured with gradient-echo and spin-echo during visual perception of faces outlined similar regions related to face perception; however only the spin-echo method detected

activity within the orbital-frontal cortex, which is known to be affected by signal dropout (Schmidt et al., 2005). The authors noted that within the regions that were detected by both methods, the size of the clusters and strength of activity was larger for the gradient-echo method. Indeed, many studies have identified an approximate three-fold blood oxygenation level dependant (BOLD) sensitivity reduction for spin-echo in comparison to gradient-echo at 3 T (Bandettini et al., 1994; Jezzard et al., 2001; Jones, Schirmer, Lipinski, Elbel, & Auer, 1998; Lowe, Lurito, Mathews, Phillips, & Hutchins, 2000; Norris et al., 2002; Parkes et al., 2005; Schmidt et al., 2005; Stroman, Krause, Frankenstein, Malisza, & Tomanek, 2001). For spin-echo, the trade-off is to be able to detect signal in regions that are normally inaccessible by gradient-echo but at a lower detection sensitivity over the whole brain. There is also evidence that spin-echo can provide better spatial localisation than gradient-echo using 3 T fMRI (Norris et al., 2002; Parkes et al., 2005; Thulborn, Chang, Shen, & Voyvodic, 1997), due to reduced sensitivity to large vessels (Boxerman, Hamberg, Rosen, & Weisskoff, 1995; Kennan, Zhong, & Gore, 1994).

An alternative approach to reduce the signal dropout is to use a multi gradient-echo technique (Poser & Norris, 2007; Poser & Norris, 2009), where multiple gradient echoes are collected following each RF pulse. Maximum BOLD contrast is achieved with a long echo time (long_TE) that is equal to the T_{2^*} of a region (Bandettini et al., 1994). However, a long_TE results in more time for spin dephasing and hence increased signal loss in regions of variable magnetic susceptibility. Therefore, one can combine this with a shorter TE (short_TE) to reduce signal dephasing, but at the cost of reduced contrast sensitivity. In addition, T_{2^*} varies across brain regions and between subjects, therefore sampling at only one echo time may not be optimal across the brain. Multi-echo techniques can improve the fMRI data by reducing signal dropout, whilst improving or at least maintaining BOLD contrast sensitivity at 3 T (Poser et al., 2006). Furthermore, a substantial improvement was shown when comparing multi-echo and standard gradient-echo methods at 7 T, during a Stroop task (Poser & Norris, 2009). The signal at different echo times can be combined, with the optimum strategy found to be weighted CNR compared to a linear summation (Poser et al., 2006). These studies suggest that the multi-echo method provides a viable alternative to the spin-echo method; however, these approaches have not been directly compared, which is the principle aim of this work.

There is growing evidence that the anterior and inferior temporal regions, which are affected by signal dropout, are critical for semantic cognition; with evidence based on TMS (Pobric et al., 2007, 2010), MEG (Marinkovic et al., 2003), PET (Mummery et al., 2000; Tranel, Grabowski, Lyon, & Damasio, 2005; Vandenberghe et al., 1996) and neuropsychology (Hodges et al., 1992; Patterson et al., 2007). However, most fMRI research fails to identify these regions, and focuses mainly on posterior superior temporal sulcus (pSTS), posterior medial temporal gyrus (pMTG) and inferior frontal gyrus (IFG) regions (Devlin et al., 2000; Thompson-Schill et al., 1997). A recent meta-analysis has suggested that this inconsistency between fMRI and other neuroimaging methods may lie in technical deficiencies of the standard fMRI acquisition protocol (Visser et al., 2009). For example, they found three main factors that influence anterior temporal lobe (ATL) detection: 1) the modality, PET is more likely to find ATL activity than fMRI (no signal loss in anterior/inferior regions); 2) the field of view (studies with field of view <15 cm in the head-foot axis may miss the edges of the brain space including the ATL); and 3) the control task used (low-level control tasks or rest can allow subjects to generate inner speech during this period). The impact of signal dropout on the ability to detect ATL activations was demonstrated by a study that compared PET and fMRI (Devlin et al., 2000). They used a semantic categorisation task, and directly compared 3 T gradient-echo fMRI with H₂O¹⁵ PET. Both techniques found activity within the prefrontal regions for the semantic task over a control letter task, however, ATL activations (especially in the left vATL) were only found using H₂O¹⁵ PET. Therefore, PET offers one possible solution for investigating semantic cognition, but it does have a number of drawbacks in comparison to fMRI. Further support comes from work using spin-echo fMRI on similar language tasks. Visser and colleagues (2010) replicated Devlin and colleagues (2000) paradigm and showed significant activity within vATL. Whereas, Binney and colleagues used a synonym judgement task contrasted against a number judgement task to show activity within the vATL (Binney et al., 2010). The conclusions from these studies point to this region being involved in semantic knowledge, and theories investigating semantic cognition would need to use an fMRI acquisition that allows for suitable detection of this critical region.

Our principle aim was to compare dual gradient-echo and spin-echo fMRI for investigating the vATL. Specifically, the dual gradient-echo consisted of a short_TE (12 ms) to minimise signal loss and a long_TE (35 ms) to maximise

BOLD sensitivity for the whole brain. The results of these individual echoes are presented for purposes of replication of the effects of field inhomogeneities. The echoes were combined using two methods; a linear average (dual_sum) and CNR-weighted average (dual_weighted). A spin-echo acquisition was used based on parameters used in previous fMRI studies (Binney et al., 2010; Visser et al., 2010). We used a semantic categorisation task (Devlin et al., 2000; Visser et al., 2010), which allowed us to focus on particular regions of interest (ROI); the ventral temporal, frontal and cerebellum regions. The questions that we considered were: 1) which acquisition protocol is the method of choice for detecting activity within vATL and/or extra temporal regions? and 2) which method of combining the two gradient echoes is superior? The semantic contrast (words over letters) was compared using statistical parametric mapping on a voxel-wise basis across the brain and also within regions of interest (vATL, IFG and cerebellum).

2. Methods

2.1. Subjects

Seventeen participants (mean age = 24 years, SD = 3.4 years, 6 males) took part in the study. All subjects were right handed, scoring at least 85 in the Edinburgh Inventory (Oldfield, 1971), native English speakers and had normal or corrected-to-normal vision. The research was approved by a local National Health Service (NHS) ethics committee.

2.2. Experimental paradigm

E-PRIME software was used for the presentation of the stimuli. The participants completed semantic and letter categorisation tasks as described by Devlin and colleagues (Devlin et al., 2000). The task was an ABCX judgement task, where participants were presented with three words sequentially from a single semantic category and then had to decide if a fourth word was related to that category (e.g., taxi, boat, bicycle followed by COACH vs. saucer, mug, bowl followed by TROUSERS; see Appendix 2 for full stimuli list). All concepts and categories were drawn from the manmade domain and controlled for psycholinguistic properties (such as word frequency, familiarity and letter length). The semantic categorisation task was contrasted with a control letter categorisation task, in which subjects viewed a series of lower-case, letter strings

matched to the length of the words. Participants saw three strings of a specific letter sequentially and then had to decide if a fourth letter string (in capitals) was the same letter (e.g., rrrr, rrr, rrrrr followed by RRR or DDDD). The study was presented in a random block design, and was made up of experimental blocks and rest blocks each lasting 15 s. There were 15 semantic judgement blocks and 15 letter judgement blocks, and a rest block in between each task block, where subjects were presented with a fixation cross. Within each semantic and letter block there were 4 epochs. Each epoch lasted 3750 ms, where each item was present for 200 ms with an inter stimulus interval of 400 ms. The fourth item was presented in uppercase for 1750 ms and was underlined to indicate that a decision was required. A fixation was present for 200 ms after the target item, before the next epoch started. The participants were asked to respond by pressing one of two buttons held in the right hand.

2.3. Image acquisition

All imaging was performed on a 3T Philips Achieva scanner using an eight-element SENSE head coil with a SENSE factor of 2.5. The functional protocols were set to be replicable on basic clinical scanners, with emphasis on the FOV to cover the whole brain (sacrificing some resolution). The dual-echo EPI acquisition used a short_TE and long_TE of 12 and 35 ms, respectively and a TR of 3680 ms. The short_TE (12 ms) serves as a compromise between reduced signal loss and sufficient contrast sensitivity. Poser and colleagues (2006) used 11 ms for the shortest echo in a functional study, which produced sensible results. A long_TE (35 ms) was used as it has been shown to be optimum for whole brain sensitivity (Bandettini et al., 1994). The functional parameters were: 42 slices, 96 × 96 acquisition matrix, 240 × 240 × 126 mm FOV, in-plane resolution 2.5 mm × 2.5 mm, and slice thickness 3 mm (no gap) and an anterior-posterior (A-P) phase encoding direction. 249 volumes were collected giving an acquisition time of 15 minutes. The dual gradient-echo was not specifically optimised to reduce signal loss in terms of slice orientation or slice thickness (e.g. thinner slices; Robinson, Windischberger, Rauscher, & Moser, 2004), but instead kept similar to the spin-echo for comparison. In fact, use of thinner slices and a thicker slice gap could have been counter-productive as this would have reduced the SNR and may not have achieved any greater reduction in signal loss than the use of the short echo gradient-echo.

The spin-echo sequence included 42 slices covering the whole brain with echo time (TE) = 75 ms (used previously by Visser et al., 2010), repetition time (TR) = 4150 ms, 96 × 96 acquisition matrix, 240 × 240 × 126 mm FOV, in-plane resolution 2.5 mm × 2.5 mm, and slice thickness 3 mm (no gap). 218 volumes were collected, giving a total duration of 15 minutes. The spin-echo was collected in the left-right (L-R) phase encoding direction, as pilot data showed severe ghosting artefacts on the ventral temporal lobes when using an A-P encoding. Previous studies have also used this encoding direction and demonstrated robust activation in the ventral temporal regions (Binney et al., 2010; Visser et al., 2010). Prior to the spin-echo functional run, two identical spin-echo images were acquired except that the phase encoding directions were reversed, in order to achieve sets of images matching the functional time-series but with opposing direction distortions (left-right and right-left) for image distortion correction (Binney et al., 2010).

A high resolution T₁-weighted structural image was acquired using a 3D MP-RAGE pulse sequence, with in-plane resolution of 0.94 mm and slice thickness of 0.9 mm, TR = 8.4 ms, TE = 3.9 ms and 200 slices. This was collected as a reference to provide a qualitative indication of distortion correction accuracy and for co-registration purposes. The order of the spin-echo and dual-echo acquisition was randomised across participants.

2.4. fMRI data analyses

Pre-processing

First the spin-echo images were corrected for distortion using the two pre-scans as has been described and used elsewhere (Binney et al., 2010; Visser et al., 2010). In brief, the pre-scans were used to compute a spatial remapping matrix that is then applied to the functional time-series following registration. The dual-echo data was not corrected for distortion as the reference scan for the spin-echo was acquired in L-R phase encoding direction, and no B₀ map was collected prior to the gradient-echo scan. Pre-processing steps were applied to the dual gradient echo images to extract and combine the short_TE and long_TE images for each TR; using in-house MATLAB code (available upon request). We had two methods of combination; dual_sum and dual_weighted. The dual_sum volume was created by simply adding together the images at each echo-time. The dual_weighted volume was created using the method of Poser and colleagues (2006), where the

CNR at each voxel was determined as a product of the temporal SNR and TE. A weight was determined for the short_TE and long_TE at each voxel, and was used to combine the corresponding image volumes.

Subsequent processing and statistical analysis was carried out using statistical parametric mapping (SPM8) software (Wellcome Trust Centre for Neuroimaging). The same procedures were applied to short_TE, long_TE, dual_sum, dual_weighted, and the distortion-corrected spin echo data. Functional image volumes were corrected for motion artefacts by registering to the first image volume using a rigid body spatial transformation (for the combined echo data, realignment was done after combining the short and long echo images). Functional image volumes were then transformed into standard stereotaxic space, according to the Montreal Neurological Institute (MNI) protocol. However, before this, we considered the masking threshold that is imposed by default in SPM, to remove non-brain and low signal voxels. The default mask is calculated per individual and only includes voxels whose mean signal is at least 80% of the global signal across all raw image volumes. For the long_TE, this results in voxels within the vATL and orbito-frontal region being removed. Hence, we analysed the long_TE data in two ways, one without application of a mask, and one using this default masking approach. The mean image for each technique was coregistered to the subjects' T₁-weighted image. The T₁-weighted image was then segmented (grey and white matter) and normalised to MNI space using a 12-parameter affine transformation and a nonlinear discrete cosine transformation. The normalised grey matter image is converted into a binary mask and used as the statistical mask for analysis. The same MNI transformations were then used to normalise each functional volume to MNI space. The EPI volumes were re-sampled to 3 x 3 x 3 mm voxel size using tri-linear interpolation. Images were then smoothed with an 8 mm full-width half maximum (FWHM) Gaussian filter. Images were temporally filtered using a high-pass filter with a cut-off of 90 s (0.01 Hz).

Statistical analysis

Data were analysed using the general linear model approach. At the individual subject level, each task was modelled as a boxcar function (resting blocks were modelled implicitly), and subsequently convolved with the canonical hemodynamic response function. Contrasts were calculated to assess differences in activations between the word and letter task. The whole brain multi-subject analysis was carried out using a random effects model with a one-sample t-test.

The group t-map was assessed for cluster-wise significance using a voxel height threshold of $p=0.005$, and a FWE cluster threshold of $p=0.05$ with a cluster size of 150 voxels. An a priori ROI analysis was performed using the MarsBar toolbox (<http://marsbar.sourceforge.net/>). Anatomical region of interest masks for the left and right IFG and left anterior and posterior portions of the vATL (based only on the TFG region as activity was not seen in ITG regions) were obtained from the Oxford centre for Functional MRI of the Brain (FMRIB; <http://www.fmrib.ox.ac.uk/fsl/>) based on the Harvard-Oxford atlas. The anatomical mask for the cerebellum, however, was much larger than the activated region, so we instead used a sphere (10 mm radius) centred on the peak voxel from Visser and colleagues (2010) study. MarsBar was used to extract mean z-scores within the a priori regions for the words > letter contrast and each acquisition method. Methods were considered to be significantly different if their z-scores differed by 2.58 or more, corresponding to $p=0.01$.

3. Results

3.1. Behavioural measures

The behavioural data based on the semantic categorisation task is summarised in Figure 8. A composite score (reaction time/accuracy) was used to determine if participants differed in behaviour during dual-echo and spin-echo acquisition. A 2 (dual-echo and spin-echo) x 2 (word and letter) ANOVA showed no main effect of method ($F(1,27) = 1.417, p=0.244$), but a main effect of condition was shown ($F(1,27) = 48.006, p<0.001$). Post-hoc analysis showed that the word condition had a higher composite score (i.e. worse performance) in comparison to the control letter task ($p<0.001$). No interaction effects were observed ($F(1,27) = 0.346, p=0.591$). This shows that although the letter task was easier than the word task, participants were performing them equally during dual-echo and spin-echo, removing any potential performance bias between methods.

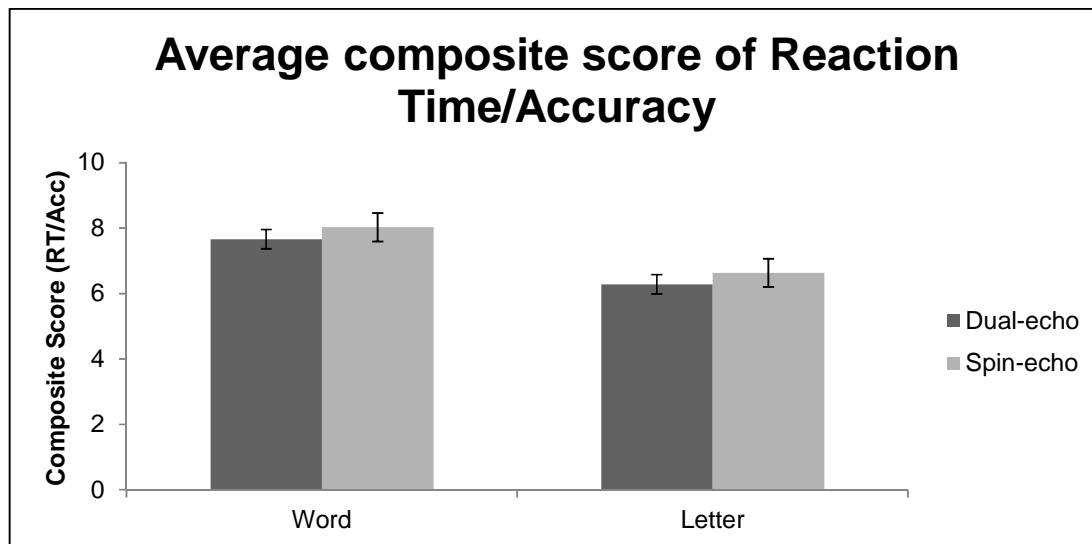


Figure 8: Behavioural results during categorisation task for each acquisition method, dual-echo and spin-echo. Footnote: Amplitude shows composite score (Reaction Time/Accuracy) and error bars represent standard error.

3.2. Whole brain analysis

Before group level analysis was carried out, three subjects were removed for excessive movement during the scan (>5 mm in any direction). The subsequent analysis was done on 14 subjects. Figure 9 shows the contrast ‘word > letter’, thresholded at a voxel height of $p=0.005$ (FWE cluster corrected at $p=0.05$). Figure 9a shows a replication of previous data (Visser et al., 2010), where the spin-echo activations (red) are compared with conventional long_TE (green). However, Figure 9b, shows the same comparison of spin-echo (red) with long_TE without a masking threshold (green), which exposes the limitation of the default masking procedure in parts of the ventral temporal lobe. Figure 9c, shows the spin-echo activation (red) compared to the short_TE (green), where large ventral temporal activity can be seen for the short_TE. Lastly, Figures 9d and 9e show the spin-echo activations (red) compared to dual_sum and dual_weighted, respectively, where large activations can be seen in the ventral temporal lobe for the dual gradient-echo. Surprisingly, the spin-echo activation does not extend anteriorly as far as the long_TE activation when the mask was removed. However, when using a masking procedure that removes the low-signal voxels, the spin-echo is comparatively better within this region. Interestingly, the spin-echo shows activity within the cerebellum, with no gradient-echo method showing activity here surviving correction at the current threshold. However, cerebellum activity for dual

gradient-echo methods is seen at less stringent threshold levels (voxel height threshold $p=0.005$, FWE cluster corrected $p=0.08$; figure not shown).

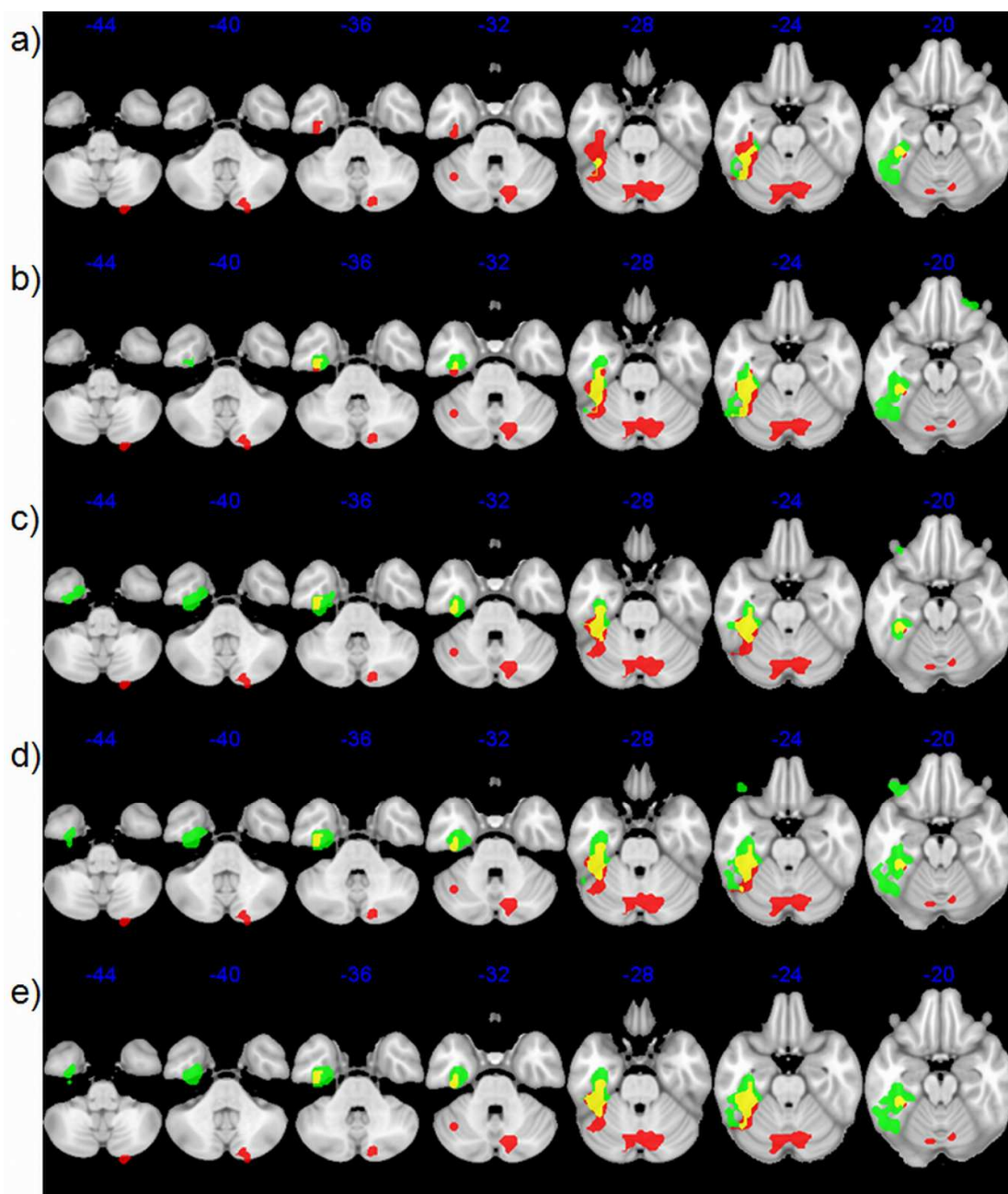


Figure 9: Significant fMRI activations for semantic categorisation using various acquisition methods in the temporal lobe. Footnote: Activity threshold using cluster corrected FWE $p=0.05$, at a voxel height threshold $p=0.005$. Each row shows the spin-echo (red) clusters with different acquisition methods (green) and overlapping regions (yellow). 9a) default-mask long gradient-echo, 9b) modified-mask long gradient-echo, 9c) short gradient-echo, 9d) dual gradient-echo SUM and 9e) dual gradient-echo wCNR.

3.3. Region of Interest (ROI) analysis

Binary anatomical masks were used based on the Harvard-Oxford atlas. Figure 10, shows where the vATL, IFG and cerebellum masks are located for the subsequent analysis. Although the main comparison is between the spin-echo and dual gradient-echo, the results from the individual echoes from the dual gradient-echo are shown for illustration of how each echo contributes to the combination. Figure 10 also shows the averaged z-scores within each ROI across methods (bar colours represent z-scores within the ROIs). Table 2 shows the difference in z-scores between the methods, where differences that are significant at $p=0.01$ are marked in red (z-score difference > 2.58), and $p=0.05$ are marked in orange (z-score difference > 1.96). For Figure 10 and Table 2, the long_TE refers to the analysis without a mask, as it produced better activations within the ventral temporal lobe than the long_TE with a default-mask.

In the anterior vATL, the long_TE was no better than spin-echo ($\Delta z=1.06$). The short_TE was superior to the long_TE ($\Delta z=3.76$) and spin-echo ($\Delta z=2.70$). Both dual-echo methods (sum and wCNR) performed better than the long_TE ($\Delta z=3.82$ and 3.67) and spin-echo ($\Delta z=2.76$ and 2.61). In summary, the dual-echoes or the short_TE are superior for imaging this region.

For the posterior vATL portion, the results are similar to the anterior portion except that the dual-echoes and the short_TE no longer have an advantage over the long_TE, making any of these methods suitable for imaging in this region.

The results in the IFG showed that spin-echo produced little activation within these regions. All gradient-echo methods produced significantly higher values than spin-echo in LIFG, but there were no differences between gradient-echo methods (see Table 2). The same pattern emerges from the RIFG, although it is noted that the overall values are lower than LIFG (see Figure 10). Interestingly, at the whole brain analysis the short_TE did not show activation surviving cluster correction in the RIFG except at the ROI level. There were no differences between short_TE and long_TE, but there was a weak trend for short_TE to be worse than long_TE ($\Delta z=1.97$). After further analysis, when testing whether the z-score in the RIFG were significantly above zero, the short_TE was not ($\Delta z=0.80$), but the other gradient-echo methods were greater than zero (long_TE $\Delta z=2.59$, dual_sum $\Delta z=1.92$, dual_weighted $\Delta z=2.06$). On the contrary the spin-echo shows a significant reduction in the RIFG ($\Delta z=2.73$), when testing against no overall change.

In the cerebellum, the comparison of z-scores showed the spin-echo signal trending to be higher than short_TE ($\Delta z=2.38$). Furthermore, the long_TE showed a similar trend ($\Delta z=2.16$). Therefore, there appears to be some detriment of using short_TE alone for activity within the cerebellum. However, there were no differences between spin-echo and dual_sum ($\Delta z=1.43$) or dual_weighted ($\Delta z=1.26$).

In summary, taking all regions into account, the dual-echo approach is superior. The choice of combination method is not important, as there was no significant difference between the two combined methods in any region.

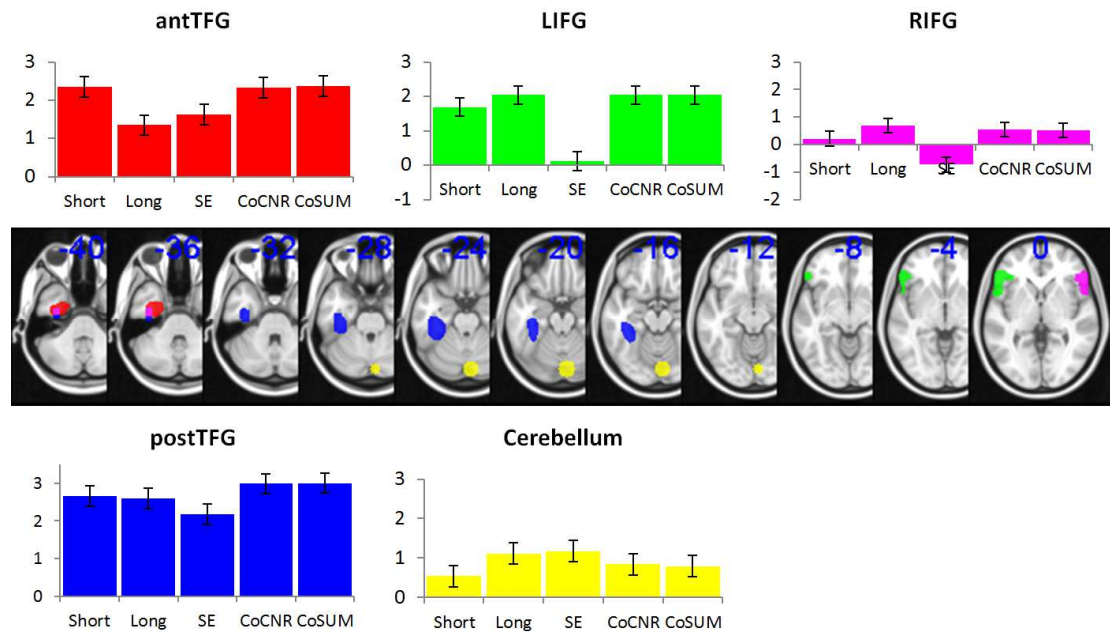


Figure 10: Regions of interest analysis comparing group level z-scores between dual-echo and spin-echo. Footnote: The ROI masks are shown in standard MNI space; left posterior inferior temporal fusiform gyrus (postTFG, blue), left anterior inferior temporal fusiform gyrus (antTFG, red), left inferior frontal gyrus (LIFG, green), right inferior frontal gyrus (RIFG, purple) and cerebellum (yellow). Error bars represent standard error over subjects.

Region of interest		Long	SE	CoCNR	CoSUM
ITF anterior	Short	-3.76	-2.70	-0.09	0.06
	Long		1.06	3.67	3.82
	SE			2.61	2.76
	CoCNR				0.15
ITF posterior	Short	-0.22	-1.79	1.23	1.30
	Long		-1.58	1.45	1.51
	SE			3.03	3.09
	CoCNR				0.06
LIFG	Short	1.33	-5.87	1.33	1.32
	Long		-7.20	-0.01	-0.01
	SE			7.20	7.20
	CoCNR				-0.01
RIFG	Short	1.78	-3.53	1.26	1.11
	Long		-5.31	-0.53	-0.67
	SE			4.79	4.64
	CoCNR				-0.15
Cerebellum	Short	2.16	2.38	1.12	0.95
	Long		0.22	-1.04	-1.21
	SE			-1.26	-1.43
	CoCNR				-0.17

Table 2. Showing the difference between fMRI methods activity within each region of interest (ROI) in terms of statistical distance (z-score).

*Footnote: Z-differences greater than ± 2.58 ($p=0.01$) are with **, whilst values greater than ± 1.96 ($p=0.05$) are marked with *. A positive value indicates that the method in the column is superior to the method in the row.*

4. Discussion

The aim of this study was to compare dual gradient-echo and spin-echo fMRI for investigating parts of the vATL that suffer signal loss. A semantic categorisation paradigm was used that is known to activate the anterior/inferior temporal lobe when using PET (Devlin et al., 2000) or distortion corrected spin-echo fMRI (Visser et al., 2010). In the first instance, results from the long_TE acquisition replicated those from Devlin and colleagues (2000), whereby activation in the anterior vATL was not found due to signal loss. It is somewhat misleading to

compare the activity from dual-echo and spin-echo methods to the long_{TE}, as it was not optimised to reduce signal loss in ventral regions. However, this comparison does provide an illustration of how the two novel methods perform in comparison to a gradient-echo acquisition that is optimised for whole brain analysis. For example, the spin-echo acquisition performed significantly better than the long_{TE} acquisition for detecting activity within anterior vATL regions. As reviewed earlier, the spin-echo method has been used successfully to investigate these problematic regions (e.g. Binney et al., 2010; Norris et al., 2002). It should be noted however, when removing the default mask for the long_{TE} data analysis; there was no difference between this and the spin-echo in the anterior vATL. Therefore, we suggest that removing or modifying the default mask in SPM for fMRI analysis could be useful when investigating this region. In order to reduce the number of voxels considered for statistical analysis, each subjects' T₁-weighted image can be used to inform which brain space voxels should be included (for example restricting to grey matter only), as we have done here. Figure 9a and 9b show activation for the long_{TE} with and without the mask (green), where activity is shown in the vATL that would have otherwise been missed. Although there is still a possibility of false positives due to the data in this region having low signal, we point out that there were no such false positives within other poor signal regions like the orbitofrontal and ventro-frontal region. This suggests that when using a long_{TE} method, it may be beneficial to include low signal voxels in a manner shown in this study, or at least to reduce the default-masking threshold in SPM.

The dual gradient-echo outperformed the long_{TE} in the same region. This is due to the effect of the individual short_{TE} of the dual gradient-echo protocol, where a comparison of this echo with spin-echo and long_{TE} shows superior detection of activity within the anterior vATL. It should be noted that the poor contrast sensitivity of the short_{TE} resulted in poor detection of activity within the cerebellum and right IFG, whereas the superior contrast sensitivity for the long_{TE} resulted in large activations in bilateral IFG and no difference in the cerebellum with spin-echo. Therefore, the comparison of the individual echoes from the dual gradient-echo provides us with an illustration of how they both contribute to the dual method. Although the combination of these individual gradient-echoes (dual_{sum} and dual_{weighted}) did not differ, they performed better (or equally) than the spin-echo method at the whole brain and ROI level of analysis. The spin-echo detected little frontal activity, whereas the dual gradient-echo method had

high bilateral activity. It should be noted that other studies using spin-echo and similar paradigms have found IFG activity (Binney et al., 2010; Visser et al., 2010).

There is evidence that spin-echo can provide better spatial localisation (Norris et al., 2002; Parkes et al., 2005; Thulborn et al., 1997). Our data does show a smaller cluster within the anterior vATL for the spin-echo in comparison to the gradient-echo methods, which could be the result of the spatial sensitivity towards smaller vessels. However, even when smaller ROIs were used, based on peak activations from Visser and colleagues (2010) and Binney and colleagues (2010), the results did not change.

The results obtained were compared within the same participants in the same session. The behavioural scores for the semantic categorisation task were not significantly different for the dual gradient-echo and spin-echo methods, suggesting any differences in neural activity between methods are unlikely to be related to task performance. We note that the control letter task was easier than the word task (reflected in quicker accuracy for letters). However, this did not have a detrimental effect on the overall outcome of detecting activity within the ventral temporal lobe as also found in Devlin and colleagues (2000) study. It is possible that a control task that is more closely matched to the difficulty of the word category condition would exaggerate the effect seen in the vATL. For example, Binney and colleagues (2010) used a number judgement task that was matched behaviourally to a synonym judgement task and produced activity with large statistical significance.

The current dual gradient-echo method was implemented as modification of studies that have proposed a multi-echo procedure to overcome signal loss in susceptible regions and obtain image volumes across a wider range of T_2^* -weightings (Poser et al., 2006; Speck & Hennig, 1998; Yang, Dardzinski, Li, Eslinger, & Smith, 1997). We made use of a dual gradient-echo to ensure we had whole brain coverage and the shortest TR possible for the scanner used. However, a multi-echo protocol can be employed with the trade off on increasing TR time or reduce overall coverage (slices and volume dimensions). We propose, therefore, that the dual gradient-echo protocol can be easily implemented on basic clinical scanners. An alternative approach to the dual gradient-echo method has been explored by collecting a spin-echo and gradient-echo in a single shot acquisition (Schwarzbauer et al., 2010) as well as a quicker variant; single shot partial dual echo (SPADE; Schwarzbauer & Porter, 2010). This method aims to accomplish the same task of reducing magnetic susceptibility artefacts and retain

a high CNR across the whole brain. Although, the current study cannot directly compare a dual gradient-echo with the dual gradient/spin-echo, the comparison of the short_TE with spin-echo in the magnetic susceptible region (anterior vATL) showed a benefit for the short_TE. This could suggest that a dual gradient-echo may outperform a dual gradient/spin-echo method, although formal testing would be needed.

It is important to dissect the dual-echo data further to determine how each echo performed alone and how they contributed after combining them together. As we know, the long_TE suffers from signal loss within vATL and orbitofrontal regions but has good CNR in general. Whereas the short_TE detects signal within these regions before it has had time to dephase, but does so at a relatively poor CNR compared to the long_TE. Each echo shows a distinct pattern of activity, which is capitalised on when used in combination. For example, in the anterior vATL, the short_TE produced higher z-scores than the long_TE. In contrast, the long_TE showed a trend to higher z-scores in the frontal regions and cerebellum than the short_TE. Interestingly, the whole brain analysis did not produce a cluster in the RIFG for the short_TE. The ROI analysis only pointed to a possible trend in favour of the long_TE and further analysis showed that activity in the RIFG using short_TE showed no change (i.e. not significantly different to zero), whereas the long_TE did. This demonstrates the necessity of combining the gradient-echo data to produce superior detection sensitivity within vATL (short_TE) and frontal regions (long_TE). The dual_sum and dual_weighted methods of combining the gradient-echoes were compared to determine which method was superior. There was no significant difference between these combinations, in contrast to the findings by Poser and Norris (2006), where they found larger average gains for dual_weighted than dual_sum. The echo-time dependence of BOLD contrast at 3 T is higher than at 7 T, therefore dual_weighted averaging is less important at 7 T (Poser & Norris, 2009).

Multiple echoes can result in extending the TR, thus reducing the temporal resolution, which could be problematic for some applications such as functional connectivity. The spin-echo method is not optimum for event-related fMRI designs, as it has low CNR and requires relatively long TR, hence one would have to scan for a considerably longer time to achieve the same statistical power in comparison to using a dual-echo method.

5. Conclusions

Most fMRI studies have failed to report vATL activity in semantic memory tasks, even though there is converging evidence from neuropsychology, rTMS, PET and MEG studies. However, multi-echo EPI can be utilised as a powerful tool for investigating brain regions that are normally affected by magnetic susceptibility artefacts. In comparison to spin-echo, dual gradient-echo EPI provides superior sensitivity within regions that are affected by magnetic susceptibility artefacts (vATL), as well as greater sensitivity within unaffected areas (i.e. IFG). The results shown here suggest that combining gradient-echo images collected at both a short and long echo time in a dual echo sequence is the method of choice for investigating activation over the whole brain, with specific attention to ventral temporal regions.

Chapter 4

Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension.

Ajay Halai¹, Laura M Parkes², and Stephen Welbourne¹

¹ Neuroscience and Aphasia Research Unit, School of Psychological Sciences,
University of Manchester,
Zochonis Building,
Brunswick Street.
Manchester, UK
M13 9PL

² Centre for Imaging Sciences, Institute of Population Health,
University of Manchester,
Stopford Building,
Oxford Road.
Manchester, UK
M13 9PL

Corresponding author:

Ajay Halai

3rd Floor, Zochonis building,

Brunswick Street,

Manchester, UK

M13 9PL

Tel: +44 (161) 306 0451

Email: ajay.halai@postgrad.manchester.ac.uk

Keywords: dual gradient-echo; intelligible speech; inferior temporal lobe; magnetic susceptibility

This chapter is currently under review in Neuroimage.

Abstract

The neural basis of speech comprehension has been investigated intensively during the past few decades. Incoming auditory signals are analysed for speech-like patterns and meaningful information can be extracted by mapping these sounds onto stored semantic representations. Investigation into the neural basis of speech comprehension has largely focused on the temporal lobe, in particular the superior and posterior regions. The ventral anterior temporal lobe (vATL), which includes the inferior temporal gyrus (ITG) and temporal fusiform gyrus (TFG) are consistently omitted in fMRI studies. In contrast, PET studies have shown the involvement of these ventral temporal regions. One crucial factor is the signal loss experienced using conventional echo planar imaging (EPI) for fMRI, at tissue interfaces such as the vATL. One method to overcome this signal loss is to employ a dual-echo EPI technique. The aim of this study was to determine if the vATL could be detected during a passive speech comprehension task using a dual-echo acquisition. A whole brain analysis for an intelligibility contrast showed bilateral superior temporal lobe activations and a cluster of activation within the left vATL. Converging evidence implicates the same ventral temporal regions during semantic processing tasks, which include language processing. The specific role of the ventral temporal region during intelligible speech processing cannot be determined from this data alone, but the converging evidence from PET, MEG, TMS and neuropsychology strongly suggest that it contains the stored semantic representations that are activated by the speech decoding process.

1. Introduction

Speech comprehension can be defined as extracting and understanding meaning from incoming auditory signals. This can be differentiated from speech perception, which can be defined as processing of sub-lexical features (i.e. phonemes and syllables). A recent meta-analysis of sub-lexical speech processing has located processing of sub-lexical components in bilateral superior temporal gyrus (STG) and left posterior middle frontal gyrus (MFG), with complex acoustic features related to phoneme processing being located in left middle superior temporal sulcus (STS) (Turkeltaub & Coslett, 2010). In contrast there is dispute

over the neural basis of speech comprehension, although it is mainly focused on the temporal lobe. In a seminal study, Scott and colleagues (2000) used PET to localise intelligible speech to the left anterior STS, when adequately controlling for speech-like variation. However, other researchers have proposed the left middle temporal gyrus (MTG) (Davis & Johnsrude, 2003), bilateral STS (Oleser & Kotz, 2010), and posterior temporal gyrus (Hickok & Poeppel, 2007; Okada et al., 2010).

A recent meta-analysis attempts to reconcile the variation seen across speech comprehension studies (Adank, 2012a). The author identifies 58 experiments (of which 3 were PET) that investigated speech comprehension (words, sentences, narratives). The results showed, independent of the design choices, bilateral superior temporal lobe regions with dominance on the left hemisphere as the core network. It was noted that closely matched control items and a sparse scanning procedure were more likely to activate anterior STS regions, whereas sentence or narrative paradigms were more likely to activate left IFG regions. Despite the effect of these design choices, the meta-analysis failed to identify activity in the ventral anterior temporal lobe (vATL), which includes the inferior temporal gyrus (ITG) and temporal fusiform gyrus (TFG). However, six studies did find activity in left posterior portion of the TFG (Bozic, Tyler, Ives, Randall, & Marslen-Wilson, 2010; Davis, Ford, Kherif, & Johnsrude, 2011; LoCasto, Krebs-Noble, Gullapalli, & Burton, 2004; Okada et al., 2010; Orfanidou, Marslen-Wilson, & Davis, 2006; Rodd, Davis, & Johnsrude, 2005). It is suggested that the lack of activity in the anterior portions of the vATL is likely to be due to the dominance of fMRI studies (54 vs. 3 PET), where the reliability of detecting blood oxygenation level dependent (BOLD) changes in anterior vATL regions is reduced (Devlin et al., 2000).

Indeed, there is evidence implicating the ventral temporal lobe in speech comprehension. The theoretical premise comes from models of semantic cognition, where the anterior temporal lobe (ATL) (in particular the vATL region) serves to integrate sensory information onto amodal semantic representations (Lambon-Ralph, Sage, Jones, & Mayberry, 2010; Patterson et al., 2007; Rogers et al., 2004). This model is based on converging evidence for the role of ATL in semantic processing arising from studies using TMS (Pobric et al., 2007, 2010), MEG (Marinkovic et al., 2003), PET (Mummery et al., 2000; Tranel et al., 2005; Vandenberghe et al., 1996) and neuropsychology (Hodges et al., 1992; Patterson et al., 2007). In particular, PET imaging studies have shown activation of the vATL for speech comprehension in neurologically intact volunteers (Crinion et al., 2003;

Scott, Rosen, Lang, & Wise, 2006; Sharp et al., 2004; Spitsyna et al., 2006). Although, these studies show the importance of this region in speech comprehension, functional models can be biased by the fMRI literature, resulting in omission of the vATL.

The fundamental disadvantage of conventional fMRI acquisition protocols is the signal loss experienced at the interface of tissue and fluid/air due to local differences in magnetic susceptibility. The regions most affected by this are the vATL and orbito-frontal gyrus. Devlin and colleagues (2000) demonstrated the problems of using fMRI to study semantic tasks, by showing that data collected using PET was able to detect regions of activation in vATL regions, whereas fMRI could not. One possible alternative to the conventional echo planar imaging (EPI) acquisition used for fMRI is to use a multi-echo technique to overcome magnetic susceptibility problems (Poser & Norris, 2009; Poser et al., 2006). The strategy is to use more than one EPI readout, one with a short echo time that reduces signal loss in regions with local differences in magnetic susceptibility but at a cost of BOLD sensitivity, and one with a longer echo time to retain superior BOLD sensitivity. The combination of these images at short and long echo times provides superior detection power in semantic processing regions (e.g. vATL regions), in comparison to the short and long echo images alone (Halai et al., chapter 3). Furthermore, the dual echo method outperformed a spin-echo protocol, which has been recently used to detect activity within the vATL during semantic tasks (Binney et al., 2010; Embleton et al., 2010; Visser et al., 2010; Visser & Lambon-Ralph, 2011).

The aim of the current study was to attempt to detect ventral temporal lobe activity using dual echo fMRI during a passive speech comprehension task. This would reconcile the discrepancies between the existing fMRI and PET literatures, and form a consistent picture with converging evidence from neuropsychology.

2. Methods

2.1. Subjects

Twenty participants (mean age = 23 years, SD = 5.4 years, 10 males) took part in the study. All subjects were right handed, scoring at least 85 in the Edinburgh Inventory (Oldfield, 1971), native English speakers and had no known

hearing problems. The research was approved by a local National Health Service (NHS) ethics committee.

2.2. Experimental paradigm

E-PRIME software was used for the presentation of the stimuli. The participants were asked to passively listen to audio clips and were told that they would be asked questions about what they heard. To avoid confounding visual contamination, we asked participants to close their eyes during the experiment. We used three audio conditions; normal intelligible speech (SP), rotated speech (RS) and rotated noise vocoded speech (RV) kindly provided to us by Scott and colleagues (2000) (See Appendix 3 for full list of sentences). Due to constraints on time and resources, we excluded the noise vocoded speech condition, originally used by Scott and colleagues (Scott et al., 2000), which reduced the scanning time from approximately 30 minutes to 22.5 minutes. The sentences have a simple 'subject-verb-object' structure, and were matched for complexity. The experiment consisted of 90 blocks lasting 15 s each, with 15 blocks per condition separated by a 15 s rest block (only scanner noise). Each experimental block contained 5 items, where each item lasted 2000 ms, and was prefixed with a random ISI from a pool of 800, 900, 1000, 1100, 1200 ms (maintaining the overall length of the block at 15 s). Each block consisted of items from one condition and the order of blocks was pseudo-randomised using Optseq (<http://surfer.nmr.mgh.harvard.edu/optseq/>). The experiment lasted 22.5 minutes split into five 4.5 minute runs to help the participant avoid tiredness. The auditory stimuli were delivered using a MR compatible noise-cancelling headphone unit developed by MR Confon (<http://www.mr-confon.de/>).

2.3. Image acquisition

All imaging was performed on a 3 T Philips Achieva scanner using an eight-element head coil with a SENSE factor of 2.5. The dual-echo sequence included 42 slices covering the whole brain with a short and long TE of 12 ms and 35 ms, TR = 3.7 ms, 96 × 96 acquisition matrix, FOV = 240 × 126 × 240 mm, in-plane resolution 2.5 mm × 2.5 mm, and slice thickness 3 mm (no gap). The experiment consisted of 5 runs each with 78 volumes (390 volumes in total). A high-resolution T₁-weighted structural scan with in-plane resolution of 0.94 mm and slice thickness

of 0.9 mm was also obtained for co-registration purposes. The T_1 -weighted scan was acquired using a 3D MP-RAGE sequence with the following parameters: TR = 8.4 ms, TE = 3.9 ms, FOV = 240 x 180 x 191 mm and 256 x 256 acquisition matrix.

2.4. fMRI data analyses

Pre-processing

The original images from the scanner were initially analysed using in-house MATLAB code (available upon request). This extracts an image volume for each echo and subsequently combines the short and long echo for each TR using simple linear summation. Subsequent processing and statistical analysis was carried out using statistical parametric mapping (SPM8) software (Wellcome Trust Centre for Neuroimaging). Single subject EPI volumes were corrected for motion artefacts by registering them to the first image volume using a rigid body spatial transformation. The mean functional volume was co-registered to the subjects' T_1 -weighted image. DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra; Ashburner, 2007) was used to improve inter-subject registration and precision in anatomical localisation (Klein et al., 2009). The T_1 -weighted image of each subject was partitioned into grey matter, white matter and CSF tissue classes using SPM8's 'new segment' toolbox. Following this, we used the DARTEL toolbox to create a group template derived from all of the subjects in our dataset. The grey matter component of this group template was then registered to the SPM8 grey matter probability map (MNI standard stereotactic space), by estimating a 12-parameter affine transform. In the process of creating the template brain, DARTEL also outputs 'flow fields' for each subject, which contain the transform from the original T_1 -weighted image space to that of the group template. Each subject's flow field was combined with the group template-to-MNI transform using SPM8 deformation utility. This transformation from subject native space to MNI space was then applied to all functional image volumes, with a smoothing kernel of 8 mm full-width half maximum (FWHM) Gaussian filter and re-sampled to 3 x 3 x 3 mm voxel size using tri-linear interpolation. Functional images were also temporally filtered using a high-pass filter with a cut-off of 128 s (~0.0078 Hz). The normalised segmented grey matter image was converted into a binary mask and used as the statistical mask for analysis.

Statistical analysis

Due to the omission of the noise vocoded speech condition, we could not replicate the analysis method of Scott and colleagues (2000), which used a full factorial analysis. This would have provided a conservative analysis controlling for interaction effects between the vocoded and rotated manipulations. Furthermore, conjunction analysis would have allowed for a more controlled way to isolate the intelligibility process. In spite of these limitations, the data were analysed using the general linear model approach investigating simple main effects. At the individual subject level, each task was modelled as a boxcar function (resting blocks were modelled implicitly), and subsequently convolved with the canonical hemodynamic response function. Contrasts were calculated to assess activations for speech (SP > RV), intelligibility (SP > RS) and phonological feature processing (RS > RV). The whole brain multi-subject analysis was carried out using a random effects model with a one-sample t-test. The group t-map was assessed for cluster-wise significance using a FDR cluster threshold of $p=0.05$ based on a voxel height threshold of $p=0.001$.

3. Results

3.1. Whole brain analysis

Before group level analysis was carried out, two subjects were removed for excessive movement during the scan (>5 mm in any direction). The subsequent analysis was done on 18 subjects and the figures shown are FDR cluster corrected at $p = 0.05$ based on a voxel height threshold of $p=0.001$. Figure 11 shows the results for speech, intelligibility and phonology contrasts. As predicted, for speech and intelligibility, there was significant activation in the vATL, including the ITG and TFG, as well as the expected activation along the STS. However, the contrast for phonological features did not show significant activity within the vATL, but did show much smaller bilateral STS activity. The peak activations and cluster sizes are shown in Table 3.

Contrast	Cluster size	Anatomical Region	t-value	z-value	MNI co-ordinate			
					x	y	z	
Speech	837	pSTG	10.84	5.86	63	-21	0	
		TP	9.19	5.44	56	7	-12	
		aSTG	9.32	5.48	63	-6	-3	
		pMTG	7.66	4.97	51	-30	-3	
	1335	aSTG	10.61	5.81	-54	-6	-6	
		PT	8.86	5.35	-57	-18	0	
		pSTG	8.40	5.21	-57	-27	3	
		TP	8.11	5.12	-50	11	-14	
		pSTG	7.81	5.02	-63	-36	6	
		IFG pOp	6.20	4.42	-45	18	18	
		IFG pTri	5.08	3.91	-48	27	0	
		pITG	4.35	3.52	-45	-36	-12	
		77	aTFG	7.09	4.77	-33	-12	-39
			aITG	6.40	4.51	-42	-12	-42
	pTFG		5.56	4.14	-36	-18	-33	
	55	Amygdala	6.50	4.55	30	0	-18	
		Hippocampus	4.26	3.46	32	-11	-29	
	49	Cerebellum	5.05	3.90	21	-78	-45	
	40	Amygdala	4.72	3.72	-24	0	-12	
Putamen		4.38	3.53	-24	3	-3		
Intelligibility	642	pMTG	10.82	5.85	51	-30	-6	
		aSTG	8.30	5.18	63	0	-9	
		TP	7.26	4.83	54	12	-18	
		TP	5.40	4.07	48	21	-30	
	1280	pSTG	8.65	5.29	-57	-18	-3	
		aSTG	8.12	5.12	-60	-3	-9	
		TP	6.71	4.63	-50	14	-17	
		IFG/pOrb	6.28	4.46	-45	27	-6	
		pMTG	4.41	3.55	-42	-30	-6	
		IFG pTri	4.23	3.45	-54	24	6	
		aITG/aTFG	5.20	3.97	-42	-12	-45	
	32	aSTG	5.54	4.13	60	-3	-3	
		pSTG	5.02	3.88	72	-12	3	
pSTG		4.97	3.85	66	-18	3		
pSTG		4.63	3.67	72	-24	9		
58		PT	4.78	3.76	-60	-12	3	
		PT/aSTG	4.76	3.75	-57	-9	0	
Phonology		137	aSTG	5.54	4.13	60	-3	-3
	pSTG		5.02	3.88	72	-12	3	
	pSTG		4.97	3.85	66	-18	3	
	pSTG		4.63	3.67	72	-24	9	
	58	PT	4.78	3.76	-60	-12	3	
		PT/aSTG	4.76	3.75	-57	-9	0	

Table 3: MNI peaks identified for speech, intelligibility and phonology contrast (sentences).

Footnote: The Harvard-Oxford atlas was used to identify the corresponding anatomical region for each peak. Abbreviations: superior temporal gyrus (STG); middle temporal gyrus (MTG); inferior temporal gyrus (ITG); temporal fusiform gyrus (TFG); inferior frontal gyrus (IFG); pars triangularis (pTri); pars orbitalis (pOrb); temporal pole (TP); planum temporale (PT).

a) SP > RV
(Speech)

b) SP > RS
(Intelligibility)

c) RS > RV
(Phonology)

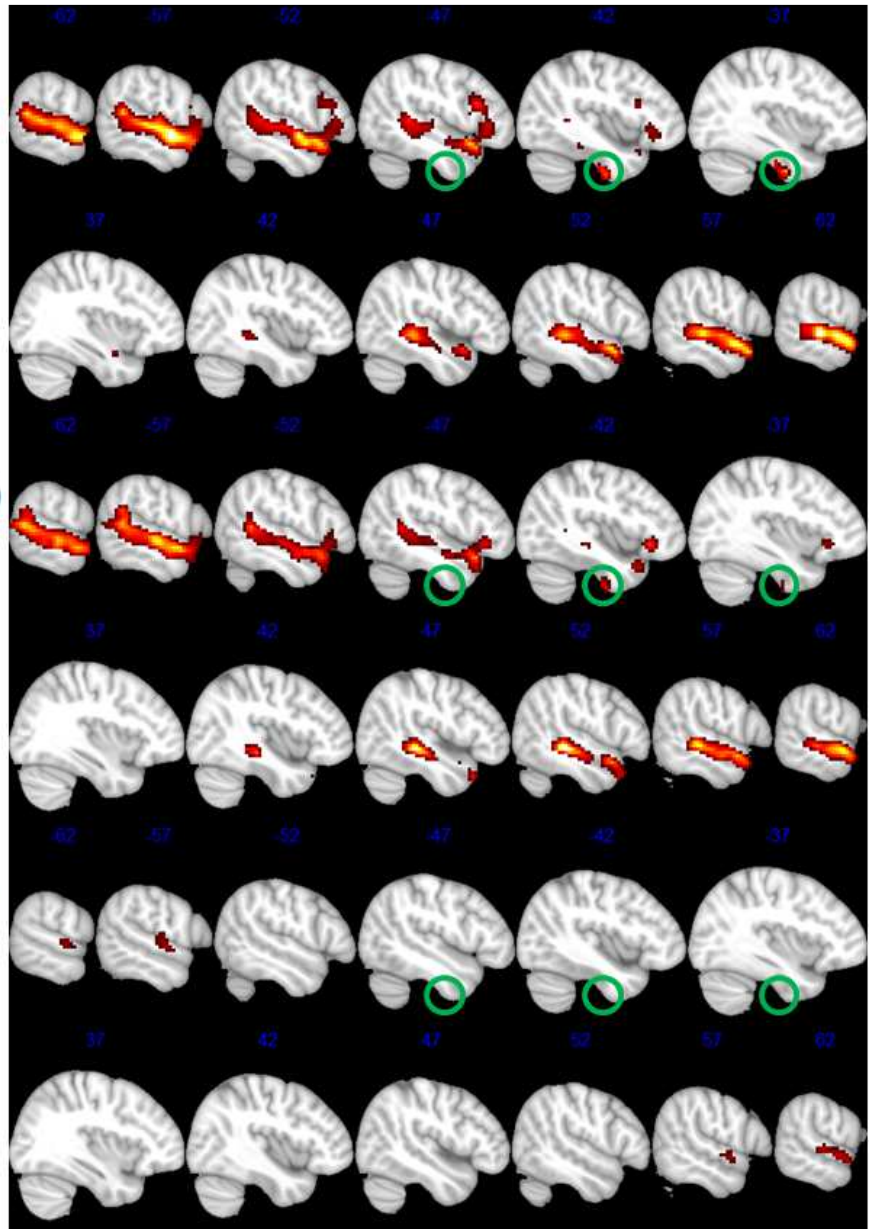


Figure 11: Significant fMRI activations for intelligible and unintelligible sentences: a) speech, b) intelligibility and c) phonology processing.

Footnote: Activations are FDR cluster corrected at $p=0.05$ based on an uncorrected voxel-threshold of $p=0.001$. Sagittal slices are shown centred around the left (negative slices) and right (positive slices) temporal lobes. SP (normal speech); RS (rotated speech); RV (rotated noise vocoded speech).

4. Discussion

The aim of the study was to determine if dual-echo fMRI could detect activity in ventral temporal regions during speech comprehension. Particularly in relation to the conflicting evidence from fMRI studies in comparison with PET (Crinion et al., 2003; Scott et al., 2006; Sharp et al., 2004; Spitsyna et al., 2006), MEG (Marinkovic et al., 2003) and neuropsychology (Hodges et al., 1992;

Patterson et al., 2007). The current dual-echo fMRI result provides evidence that the vATL region is involved during semantic processing as suggested by converging PET (Mummary et al., 2000; Tranel et al., 2005; Vandenberghe et al., 1996), TMS (Pobric et al., 2007, 2010), MEG (Marinkovic et al., 2003) and neuropsychological studies (Hodges et al., 1992; Patterson et al., 2007). Furthermore, it provides supporting evidence that the protocol is sensitive enough to detect activity in the vATL during a passive speech comprehension task. Therefore, current models of speech comprehension which omit this region need to be updated (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009).

We believe this study is the first to show activation in the vATL during passive speech comprehension using an fMRI protocol. However, there have been recent studies using a spin-echo fMRI method that have shown the amodal sensitivity of the ventral anterior temporal region (Binney et al., 2010; Embleton et al., 2010; Visser et al., 2010). Further investigations into the specific role of sub-regions of the vATL have begun to emerge. For example, it has been shown that the left and right vATL respond significantly to pictures, auditory words and environmental sounds, where auditory words produce a higher response in the left than right vATL (Visser & Lambon-Ralph, 2011). Furthermore, in the same study, the aSTS only responded to both auditory stimuli but not to pictures. This suggests that the aSTS processes aspects of high-order auditory information, which works with the vATL to achieve intelligibility (Sharp et al., 2004; Visser & Lambon-Ralph, 2011). The vATL might be expected to have the same role in speech comprehension as it does in other semantic tasks, namely providing meaning. Therefore, one would expect activation of vATL during successful comprehension, be it at the word, sentence or discourse level. However, the current models of speech processing do not implicate the vATL as a component and instead focus on aSTS (Rauschecker & Scott, 2009) or pMTG/ITS (Hickok & Poeppel, 2007). This could be largely due to the types of studies used to develop the models, where the fMRI literature has heavily dominated. In contrast, Spitsyna and colleagues (2006) used PET to outline how the comprehension network can incorporate the vATL from sensory auditory and visual inputs.

The intelligibility results showed activation along the STS, extending down to the temporal pole. This provides supporting evidence for a number of neuroimaging studies (McGettigan et al., 2012; Narain et al., 2003; Scott et al., 2000; Scott et al., 2006) and speech comprehension models (Price, 2010; Rauschecker & Scott, 2009). This activity along the STS contradicts the model

proposed by Hickok and Poeppel (2007), which suggests that intelligibility is processed at pMTG/pITS and combinatorial semantics occur at aMTG/aITS. The STS results are seen bilaterally, and therefore does not fit entirely into a left lateralised framework (Rauschecker & Scott, 2009). One possible confounding factor could be related to the use of continuous fMRI scanning as opposed to sparse sampling, which results in high-levels of competing background noise (although this will have been reduced by the use of noise cancelling headphones). Peelle and colleagues (2010a; 2010b) suggest that background noise introduces additional task components such as increased listening effort. They found that increasing the listening effort resulted in increased BOLD response in left supramarginal gyrus and posterior STS, which is not part of the core comprehension network (Adank, 2012a). Finally, a meta-analysis showed that sparse fMRI scanning promoted activity within bilateral aSTG, left pSTG and left angular gyrus (AG), while continuous fMRI scanning promoted activity within supplementary motor area (SMA) (Adank, 2012a). As the meta-review of sparse fMRI scanning results show bilateral superior temporal activation, it is likely that speech is supported bilaterally, with a left dominance. It is noted that a recent study showed increased functional integration across distributed regions, which included AG and prefrontal cortex, during adverse listening conditions (Obleser et al., 2007a). This study suggested that during difficult speech comprehension a distributed network of regions become involved to aid comprehension in which semantic context becomes important. The AG has been implicated in verbal working memory (Scott, Leff, & Wise, 2003; Sharp et al., 2004) and auditory-motor transformation processes (Hickok & Poeppel, 2007; Jacquemot & Scott, 2006). While the fronto-parietal network has been implicated in monitoring and selection processes related to attention (Lebedev, Messinger, Kralik, & Wise, 2004).

The activity in IFG seen in the current data supports the finding that studies using sentence/narratives are more likely to identify left pars orbitalis and pars triangularis. It can be argued that these regions are involved in processing longer segments of data, and this could involve syntax or semantic integration (e.g. Newman, Just, Keller, Roth, & Carpenter, 2003; Thompson-Schill et al., 1997). Furthermore, it has been shown that the pars orbitalis makes strongest contribution to semantic tasks, whereas posterior regions (pars opercularis) play a greater role in phonological tasks (Noonan et al., 2013). It could be argued that semantic control processes were required to resolve speech comprehension within the noisy fMRI environment, which involves the pars orbitalis (Badre & D'Esposito,

2009). The speech comprehension model proposed by Hickok and Poeppel (2007) does not consider the IFG as part of the speech comprehension network (instead as part of the production network), and therefore our results are inconsistent with that suggestion. However, Rauschecker and Scott (2009) propose the IFG to be actively involved during speech comprehension by mapping speech sounds onto abstract forms in the mental lexicon. One possible line of supporting evidence for this idea comes from a recent tractography study that revealed anatomical connections between the anterior temporal pole and the pars orbitalis via the uncinate fasciculus (Binney et al., 2012).

Finally, it should be noted that an fMRI method that is sensitive to BOLD changes in the ventral temporal region could help better inform brain connectivity analysis. For example, Leff and colleagues (2008) performed a dynamic causal modelling (DCM) analysis that investigated how three regions involved in intelligible speech comprehension were functionally connected (aSTS, pSTS and IFG). This network of regions does not include the vATL, which is suggested to be critically involved during speech comprehension. In contrast, dual-echo fMRI would increase the likelihood of detecting BOLD changes in ventral temporal regions, which then can be included in analyses, such as DCM to provide a better understanding of the connectivity of the speech comprehension network. Although, there are technical difficulties in investigating temporal/effective connectivity using DCM of fMRI data (i.e. outcome biased by a priori nodes selected, slow temporal change of fMRI signals), it can be a useful exploratory tool.

5. Conclusions

The current study solves a longstanding contradiction between fMRI-derived models of speech comprehension, which typically do not include ventral temporal regions and convergent evidence from other modalities, which suggest that the vATL is the repository of semantic representations that should be central to comprehension. Furthermore, the study has highlighted the necessity of using a dual-echo fMRI method to investigate the functional role of regions with a high vulnerability to signal loss due to changes in magnetic susceptibility at tissue boundaries.

Chapter 5

Combining EEG-fMRI to investigate brain networks involved in spoken word comprehension.

Ajay Halai¹, Laura M Parkes², and Stephen Welbourne¹

¹ Neuroscience and Aphasia Research Unit, School of Psychological Sciences,
University of Manchester,
Zochonis Building,
Brunswick Street,
Manchester, UK
M13 9PL

² Centre for Imaging Sciences, Institute of Population Health,
University of Manchester,
Stopford Building,
Oxford Road,
Manchester, UK
M13 9PL

Corresponding author:

Ajay Halai
3rd Floor, Zochonis building,
Brunswick Street,
Manchester, UK
M13 9PL

Tel: +44 (161) 306 0451

Email: ajay.halai@postgrad.manchester.ac.uk

Keywords: simultaneous EEG-fMRI; speech comprehension; superior temporal sulcus; inferior temporal gyrus; temporal; spatial

Abstract

Simultaneous EEG-fMRI recordings have the potential to probe whole brain networks, due to complementary spatial (fMRI) and temporal (EEG) sensitivities. MR related gradient and cardiac artefacts affect the EEG signal; however recent advances in hardware and software can filter out these artefacts to produce relatively clean EEG signals. Neurobiological models of speech comprehension generally identify either the spatial or temporal profile and tend not to provide both pieces of information. The current study recorded simultaneous EEG-fMRI whilst participants were exposed to intelligible or unintelligible words to identify a spatial and temporal profile for spoken word comprehension. The fMRI results showed activity within bilateral superior temporal sulcus (STS) and left inferior frontal gyrus (IFG); furthermore, ROI analysis showed significant involvement of the left ventral anterior temporal lobe (vATL). The temporal profile for these nodes showed most activity between 300-400 ms, whilst interestingly the vATL was active throughout the epoch. Functional connectivity analysis showed that the middle STS was connected to posterior, anterior STS and IFG, while the anterior STS was connected to IFG and vATL regions. These results illustrate a novel approach to understanding brain networks of speech comprehension. The localisation of intelligible speech within anterior STS is consistent with current models, whereas the activity within the vATL is consistent with models based on neuropsychology and PET data. The temporal profile revealed maximal network activation around 300-400 ms which fits with theories of a semantically related N400 component, whereas vATL involvement at early time periods could reflect top-down processes during comprehension.

1. Introduction

Understanding the neural basis of brain functioning requires knowledge about the spatial and temporal aspects of information processing. Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) are two non-invasive techniques widely used to investigate human brain function. Neither of these alone, however, can provide the information necessary to understand the spatio-temporal aspects of information processing in the human brain. fMRI does provide a highly localised measure of brain activation (about 2-3 mm), however; it has a temporal resolution significantly longer than the time needed for most perceptual and cognitive processes. EEG has the necessary temporal resolution

to study the dynamics of brain function, but its poor spatial resolution due to volume conduction precludes accurate identification of underlying neural sources. Speech comprehension has been the focus of neuroimaging research for the past few decades, with many spatial models of speech comprehension being derived predominately from fMRI studies (Halai et al., Chapter 4; Hickok & Poeppel, 2007; Price, 2010; Rauschecker & Scott, 2009). EEG studies have also been successful in developing temporal models of speech (e.g. Friederici, 2002; Marslen-Wilson, 1987; Pulvermüller, 2007), but the spatial and temporal models have developed in isolation. Therefore, EEG and fMRI represent complementary imaging techniques, and combining information from them should be a particularly useful way to examine the spatial and temporal dynamics of brain processes (e.g. Liebenthal et al., 2003; Menon et al., 1997). A detailed overview of the current state of knowledge about the organisation of speech comprehension is given in Section 1.3. Before that, it is important to explore the rationale for simultaneous EEG-fMRI data collection (1.1), as well as the methodological issues related collecting high quality EEG-fMRI data (1.2).

1.1. Rationale for EEG-fMRI

The integration of EEG and T_{2^*} -weighted fMRI signals is complex with a number of substantial challenges to overcome. Our understanding of the relationship between T_{2^*} -weighted signal and underlying neuronal activity is limited (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001), despite a growing literature investigating the correlation between T_{2^*} -weighted signals and EEG oscillatory activity in specific frequency bands (Mukamel et al., 2005). Despite this, a number of studies have combined EEG-fMRI using data recorded in separate sessions (e.g. Ball et al., 1999; Menon et al., 1997). An advantage of this approach is that the SNR of the EEG data obtained outside the MRI scanner is usually much better than that of data obtained inside the scanner. However, there are three main problems associated with collecting the EEG and fMRI data separately. Firstly, there are potential confounds related to the tasks that include practice, repetition and order effects. For the current study, this is particularly problematic for distorted stimuli, as repeated exposure could allow subjects to learn to understand them. Secondly, the experimental context is not the same for both sets of data. For example, there may be differences in subjects' level of attention, motivation and familiarity with the task, which could affect the neural

response to the same stimuli. Furthermore, during the fMRI session the subject is in a supine position within a noisy environment, whereas during the EEG session the subject generally sits comfortably in a quiet room. Any differences in acoustic environment are important in the current study, as this could lead to compensatory mechanisms of speech comprehension to become involved (i.e. Adank, 2012b; Peelle et al., 2010a). Finally, and most importantly, the combination of separate EEG and fMRI datasets do not accurately reflect the temporal and spatial response of any given stimuli. As there are many confounding factors between sessions, it is difficult to conclude that a separate measure of neural activity for any given stimuli reflects the same neural process. In contrast, during simultaneous EEG-fMRI, effects of practice, repetition and presentation order do not apply between imaging modality as the data is collected during a single exposure. Furthermore, the experimental context effects (i.e. attention, motivation and environment) are kept the same across imaging modality. As these confounding factors are constant for any given item, the conceptual combinations of the EEG-fMRI data reflect the same neural event. This allows the temporal information from EEG to be paired with the spatial information from fMRI to produce a combined analysis that is not confounded by the effects described above. It is noted, however, that the trade-off for collecting simultaneous EEG-fMRI datasets is a reduced SNR of the EEG data (due to MR-related artefacts). There are a number of techniques to minimise and correct for these artefacts, which are discussed in the following section.

1.2. MR related artefacts and correction procedures

Despite the obvious advantages of simultaneous EEG-fMRI acquisition, there are a number of methodological hurdles that need to be surmounted to ensure high quality data collection as the combination of EEG with fMRI generates some EEG data artefacts, which would not be present in EEG alone. The largest artefacts are produced as the magnetic field gradients shift during fMRI acquisition. This generates a voltage in the conducting material (subject head and EEG wires) by means of electromagnetic induction (Allen, Josephs, & Turner, 2000; Allen, Polizzi, Krakow, Fish, & Lemieux, 1998; Yan, Mullinger, Brookes, & Bowtell, 2009). The resulting gradient artefact (GA) is recorded in the magnitude of mV, in comparison with typical EEG amplitude of $<100 \mu\text{V}$ (Figure 13a). In addition to the large amplitude, there are two characteristics of this artefact; the topography

and latency. The topography of the artefact depends on the position of the head and EEG cap with respect to the magnetic field gradients, in which case any head movements can alter the artefact profile. However, because the MR gradient waveforms are highly reproducible across repeated executions of the fMRI sequence the latency of the artefact is highly predictable. Therefore, although the GA manifests differently at different electrodes, the profile at each electrode stays constant if the entire apparatus (electrodes, connecting leads, amplifiers, etc.) remain immobile, and this inherent predictability can be exploited when applying correction algorithms. The most common approach is to calculate a mean artefact template for each EEG channel and subtract this at each GA occurrence (average artifact subtraction, AAS; Allen et al., 2000). In order for this to be effective, the EEG amplifiers must have high dynamic range to prevent saturation by large artefact voltages, thus allowing an undistorted artefact template to be obtained. Furthermore, the artefact must be accurately and reproducibly sampled to provide an accurate template for each subtraction. Traditional EEG data is acquired at around 500Hz. However, the induced voltage from MR-related artefacts typically change by 1000 μ V over 1 ms. Sampling data at 500 Hz would lead to a poor estimation of the artefact at each occurrence (every 2 ms). To avoid this problem simultaneous EEG-fMRI is typically collected at 5000 Hz or higher (<0.2 ms). To improve the removal of the GA artefact further it has been shown that significant improvements are possible when the EEG sampling is exactly synchronised with the MR clock, compared to when it is not synchronised (Mandelkow, Halder, Boesiger, & Brandeis, 2006; Mullinger, Morgan, & Bowtell, 2008c). Finally, there is also evidence to suggest the preparation method can affect the amplitude of the GA artefacts. For example, the GA amplitude can be reduced by up to 30% by positioning subjects further out of the scanner (a shift of 3 cm in the axial plane was found to be optimal; Mullinger, Yan, & Bowtell, 2011). This effect is achieved by positioning the electrodes in the iso-centre of the scanner bore where field homogeneity is highest. It has also been shown that EEG artefacts can be reduced by placing sandbags on the cables and by mounting the EEG amplifier away from the scanner bore on a temporary plinth, which reduces currents generated by micro vibrations (Mullinger, Brookes, Stevenson, Morgan, & Bowtell, 2008a). Although these methods provide considerable improvement of the EEG signal quality, there are still some slow changes in the artefact form (for example, due for example to subject movement) that mean residual GA artefacts remain.

The second artefact is caused by movements of the head and electrodes, within the magnetic field, that are linked to the subject's cardiac cycle (Yan, Mullinger, Geirsdottir, & Bowtell, 2010). The exact origins of this ballistocardiogram artefact (BCG) is not well understood, but potential causes include: a) a voltage generated by movement associated with pulse driven expansion of the scalp (Debener, Mullinger, Niazy, & Bowtell, 2008); b) a voltage induced by rotation of the head in the MR field (rotation driven by momentum changes in blood as it is pumped into the head) (Bonmassar et al., 2002; Huang-Hellinger et al., 1995); or c) the Hall voltage generated by blood flow in arteries in the brain and scalp. The unpredictability of this artefact arises from variations in the cardiac cycle and current limitations in the precise identification of the source of the artefact (Debener et al., 2008; Yan et al., 2010). The standard method to correct this artefact requires recording of the cardiac cycle (usually with an ECG electrode), which is used to identify the QRS complex and mark the R peaks (largest and sharpest positive peak, Figure 13e). An algorithm is used to detect the manifestations on the EEG channels and calculate a time lag between the two events (i.e. time for the blood to leave the heart and enter head). An average template is built, similar to GA, however due to the variability across samples a sliding window template is formed. Typically 21 samples are used, which means the template adjusts to changes in the artefact profile approximately after 20 s. The quality of the ECG trace is degraded at high fields (Debener et al., 2008) due to larger MR induced potentials, which can result in poor estimation of R peaks based on recordings from ECG electrodes. An alternative method is based on the use of vector-cardiogram (VCG) recordings for accurately identifying the timing of the R peak (Chia, Fischer, Wickline, & Lorenz, 2000; Mullinger et al., 2008c). The four electrode VCG is not as affected by gradient artefacts as ECG, which allows for more accurate R peak identification (Figure 13e). Following the accurate identification of R peaks, the AAS method can be used to correct for BCG, however as with the GA correction, residual artefacts remain due to subject movement and abrupt cardiac changes. A number of artefact correction methods have been developed in order to clean the residual artefacts in the EEG data, many of which draw on blind source separation techniques such as ICA and optimal basis sets (OBS; e.g. Debener et al., 2007; Freyer et al., 2009; Niazy, Beckmann, Iannetti, Brady, & Smith, 2005). In most cases, the standard GA and BCG corrections are used first followed by ICA or OBS to remove any residual artefacts.

1.3. Models of speech comprehension

As stated earlier, understanding brain processes requires knowledge about the spatial and temporal aspects of information processing. There is considerable potential in using simultaneous EEG-fMRI studies to investigate brain networks and this combined information is lacking in the current models of speech comprehension. For example, three of the most influential anatomical models of speech comprehension rarely mention the temporal dynamics between regions of cortex (Hickok & Poeppel, 2007; Price, 2010; Rauschecker & Scott, 2009). They do, however, suggest a direction of information flow. For example, it is generally agreed that auditory processing is organized hierarchically, with responses to sounds of increasing complexity and intelligibility progressing from the superior temporal plane into anterior, posterior and inferior temporal gyri and sulci (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Binder et al., 2000; Davis & Johnsrude, 2003; Scott et al., 2000; Spitsyna et al., 2006). The models of speech comprehension do not agree on the specific location where intelligibility processing occurs. Three separate regions have been implicated for intelligible speech processing: 1) bilateral posterior middle temporal gyrus/inferior temporal gyrus (pMTG/pITG) for lexical access (Hickok & Poeppel, 2007); 2) left lateralised anterior superior temporal sulcus (aSTS) (Rauschecker & Scott, 2009); and 3) anterior temporal lobe (ATL) including aSTS, lateral anterior MTG and ITG (Price, 2010, 2012). This highlights the inconsistencies in our understanding of the locus of speech comprehension.

During speech comprehension, the outputs of auditory perceptual processing are mapped to stores of semantic knowledge where the specific semantic representations allow meaning to be accessed. Semantic processing has been known to involve the ventral anterior temporal lobe (vATL), the temporal poles (TP), the left angular gyrus (AG), and regions within the left inferior frontal gyrus (IFG) (Martin & Chao, 2001; McClelland & Rogers, 2003; Sharp et al., 2004; Spitsyna et al., 2006; Vandenberghe et al., 1996). Furthermore, a recent meta-analysis on speech comprehension has shown that bilateral (but left dominated) STS regions are common to all studies (Adank, 2012a). However, design choices (matched vs. non-matched baseline; sparse vs. continuous scanning; and sentence vs. words) can influence the localisation of activity to an extended network of regions including the posterior, anterior STS, left IFG and motor

regions. Interestingly, the meta-analysis included mainly fMRI studies (57), with only a few PET studies (3). It is known that fMRI studies are heavily affected by signal loss in vATL and orbito-frontal regions that can lead their omission. In contrast, PET imaging does not suffer from signal loss, and studies have found vATL regions to be involved during speech and semantic processing in general (Crinion et al., 2003; Sharp et al., 2004; Spitsyna et al., 2006). One model which attempts to integrate the vATL within a speech comprehension network is proposed by Spitsyna and colleagues (2006). They outline a functional model of speech processing that shows the spread of information from the primary auditory cortex (PAC) to the anterior and posterior STS, including the vATL, specifically the anterior temporal fusiform gyrus (TFG) (Figure 12). This follows a hierarchical structure; with low-level auditory information being processed at PAC, while increasingly complex forms are processed anteriorly down the STS. It is suggested that the anterior regions engage in mutual information processing to achieve comprehension. This highlights the importance of using convergent evidence when building models of brain processing.

The importance of signal loss within vATL regions experienced by conventional fMRI studies should not be overlooked. This methodological issue is likely to be an important factor in explaining the discrepancy between speech comprehension and semantic memory models, and the role ascribed to the vATL. Recent fMRI advances have shown that studies using distortion corrected spin-echo can identify fMRI signal changes within vATL regions (Binney et al., 2010; Embleton et al., 2010; Visser et al., 2010; Visser & Lambon-Ralph, 2011). However, it has been demonstrated that a dual-echo protocol is superior to spin-echo within vATL and IFG regions (Halai et al., Chapter 3; Chapter 4). Therefore, it is suggested that studies investigating language processes that involve the vATL should use a method known to adequately overcome signal loss and distortion due to magnetic susceptibility artefacts.

In addition, the IFG has been implicated in speech comprehension; however the specific role of this region is disputed. Hickok and Poeppel (2007) suggest this region is primarily involved in articulatory processing (in combination with primary motor areas), whereas Rauschecker and Scott (2009) suggest high-level invariant properties available at this region provide constraints to comprehension and Price (2010) suggests the IFG is involved in selective/executive control processes (Jefferies & Lambon Ralph, 2006; Noonan et al., 2009; Thompson-Schill et al., 1997). Primary evidence for this comes from

patients with IFG lesions who perform poorly when task demands or task difficulty increases (Corbett et al., 2009; Jefferies & Lambon Ralph, 2006; Noonan et al., 2009). The IFG is routinely sub-divided into three broad regions, pars orbitalis, pars triangularis and pars opercularis. A recent meta-analysis suggested that the anterior IFG (pars orbitalis) contributes to semantic control processes, while the posterior IFG (pars opercularis) contributes to tasks involving phonological control and the portion in between (pars triangularis) is involved with both processes (Noonan et al., 2013). These studies suggest that the IFG contributes to selecting relevant aspects of meaning, rather than being necessary for retrieving intelligible speech representations.

Two opposing frameworks have been proposed to explain the temporal evolution of language comprehension. One model suggests sequential processing of linguistic information (Friederici, 2002). For example, the stages can be split into the identification of phonemes, combining phonemes to produce word forms then mapping these onto word meaning and finally re-passing and reanalysing the semantic and syntactic structure. The model is supported by evidence showing that increasingly complex features become apparent at later intervals of time after stimuli onset; for example pre-lexical effects around 100-200 ms (e.g. Dehaene-Lambertz, 1997), are preceded by semantic effects around 400 ms (N400 reviewed in Kutas & Federmeier, 2011), and a positive component suggested to be related to syntactic processes around 600 ms (P600, Hagoort et al., 1993). However, studies have shown semantic/syntactic effects modulating early components, around 100-200 ms (Pulvermüller et al., 2001; Pulvermüller et al., 1995; Sereno et al., 2003), and this 'higher' level information being processed very early is not consistent with a serial processing model. On the contrary, pre-lexical, semantic and syntactical information have been suggested to be processed near-simultaneously in a parallel processing framework (Marslen-Wilson, 1987). This model can explain the early semantic/syntactic effects (around 100-200 ms) but the large semantic differences and the lack of sub-lexical differences during late components become harder to explain. Furthermore, a parallel framework suggests that components reflect a mixture of processes and one individual process would be difficult to isolate. Both models can explain some findings but not all; therefore it may be helpful to view the models as complementary. Accordingly, a recent proposal suggested a combination of the models in which early first pass parallel processes, are followed by a second-pass of parallel processes during which any inconsistencies during the first pass are repaired

(Pulvermüller, 2007). There have been efforts to localise temporal components of speech processing, however, due to the relatively poor spatial resolution of M/EEG in comparison with fMRI, conclusions drawn from such studies should be treated cautiously. One example used MEG to show the temporal flow of comprehension during auditory and visual stimulation (Marinkovic et al., 2003). For auditory processing, they found that information generally flowed in a hierarchical fashion, originating from the PAC and spreading mainly towards the ATL. Large regions of ATL were active including superior and inferior portions, as well as extra temporal regions such as IFG. The integration of this temporal information with accurate spatial localisation is necessary to provide a better understanding of the speech comprehension network.

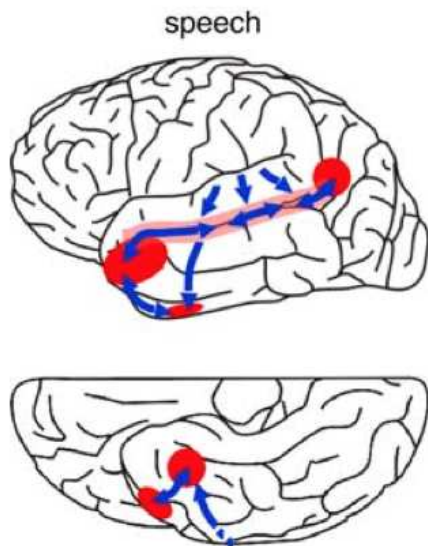


Figure 12: Functional model of speech processing proposed by Spitsyna and colleagues (2006). Footnote: Auditory information arrives at primary auditory regions and spreads to posterior and anterior superior temporal regions, with particular note given to the inclusion of the anterior temporal fusiform gyrus (TFG). Figure from Spitsyna et al., (2006).

1.4. Aims

The current study aims to explore the spatial and temporal profiles of intelligible word comprehension in a multi-modal setting. The study focuses on intelligible speech comprehension, and not sub-lexical processing as it is expected that intelligibility will activate a wider semantic network including vATL for semantic access, which would not be a part of the sub-lexical network (Spitsyna et al., 2006). Firstly, dual-echo fMRI data will be used identify the spatial profile of

intelligible speech comprehension. This method will allow for signal coverage over the vATL (as well as orbito-frontal) regions, which are generally omitted from speech comprehension models. The aim is to replicate and identify specific regions related to intelligibility within the temporal lobe that include the vATL, which will help adjudicate between current spatial models of speech processing. Secondly, temporal sequencing information can determine when specific regions of this model become active during intelligible speech comprehension. Therefore, the EEG data will be used to determine when intelligible speech regions become active across the epoch. This can provide information about the relative contribution of each region across time. It has been proposed that semantic processing occurs around 400 ms post stimulus onset (N400, Kutas & Federmeier, 2011), therefore it is expected that the majority of the regions involved in intelligible speech should be active around this time, in particular regions known to process semantic information (i.e. vATL).

As the current study uses a passive comprehension paradigm, we would expect no specific contribution of articulatory processes and therefore no motor or IFG activity. However, as the study involves single word comprehension within a relatively noisy environment, this could engage the IFG as a compensatory mechanism to aid comprehension through semantic control processes. Therefore, the extent to when this region is involved across time can help determine if the passive speech comprehension engages attentional or selection resources at relatively early (top-down) or late (bottom-up) periods.

2. Methods

2.1. Subjects

Twenty participants (mean age = 23 years, SD = 5.4 years, 10 males) took part in the study. All subjects were right handed, scoring at least 85 in the Edinburgh Inventory (Oldfield, 1971), native English speakers and had no known hearing problems. The research was approved by a local National Health Service (NHS) ethics committee.

2.2. Experimental paradigm

E-PRIME software was used for the presentation of the stimuli. This was a passive task where the participants listened to audio clips and were told that they

would subsequently be asked questions about what they heard. To avoid visual contamination, we asked participants to close their eyes during the experiment. Single words were presented in three audio conditions; normal word (W), rotated speech (WRS) and rotated noise vocoded speech (WRV). The spectrally rotated condition was unintelligible but retained pitch variations (similar to phonological features) (Blessner, 1972), whereas the rotated noise vocoded (6 bands) condition was unintelligible and did not contain pitch variations (Shannon et al., 1995). These manipulations were used to test for intelligibility processing by contrasting the intelligible condition with the unintelligible rotated speech condition. The single words used in this study were obtained from the CELEX word database (80 words; see Appendix 1), which were high frequency (average 117) and one syllable long. Each word was recorded using Audacity software (<http://audacity.sourceforge.net>) at a sampling rate of 22050 Hz. The length of the words varied between 0.5-0.8 s. Each word was segmented using Praat editing software (<http://www.praat.org>), where the onset was determined by the spectrogram amplitude increase compared to baseline and the segment ended when the amplitude returned to baseline. Scripts were used in Praat to apply the rotated and rotated noise vocoded transformations to each word to create the corresponding WRS and WRV conditions. The experiment consisted of 60 blocks of 16 s each, with 10 blocks per condition separated by a 16 s rest block (only scanner noise). Each block contained 8 items, where each item lasted 1000 ms, and was prefixed with a random ISI from a pool of 650, 750, 850, 950, 1050, 1150, 1250 and 1350 ms (maintaining the overall length of the block at 16 s). Each block consisted of items from one condition and the order of blocks was pseudo-randomised using Optseq (<http://surfer.nmr.mgh.harvard.edu/optseq/>). The experiment lasted 16 minutes split into three 5 minute and 20 second runs to help the participant avoid tiredness. The auditory stimuli were delivered using a MR compatible noise-cancelling headphone unit developed by MR Confon (<http://www.mr-confon.de/>).

2.3. Image acquisition

fMRI

All imaging was performed on a 3T Philips Achieva scanner using an eight-element head coil with a SENSE factor of 2.5. The dual-echo sequence included 42 slices covering the whole brain with a short and long TE of 12 ms and 35 ms, TR = 3.696 ms, 96 × 96 acquisition matrix, FOV = 240 × 126 × 240 mm, in-plane

resolution 2.5 mm x 2.5 mm, and slice thickness 3 mm (no gap). The experiment consisted of 3 runs each with 90 volumes (270 volumes in total). A high-resolution T₁-weighted structural scan with in-plane resolution of 0.94 mm and slice thickness of 0.9 mm was also obtained for co-registration purposes. The T₁-weighted scan was acquired using a 3D MP-RAGE sequence with the following parameters: TR = 8.4 ms, TE = 3.9 ms, FOV = 240 x 180 x 191 mm and 256 x 256 acquisition matrix. Participants were unconventionally positioned in the MRI scanner, approximately 3 cm towards the foot direction. The iso-centre was set at the position of the frontal EEG electrodes (FP1 and FP2).

EEG

EEG data were collected using an MR-compatible cap equipped with 63 Ag/AgCl electrodes positioned on the scalp according to the extended international 10-20 system (Braincap MR, Brain Products, Germany), with an additional ECG electrode. Impedances at each electrode were below 10 kOhm (scalp) and 30 kOhm (ECG) at the time of entering the scanner. A high sampling rate of 5 kHz was used, with FCz as the reference and Iz as the ground electrodes. The EEG sampling and fMRI gradient waveforms were synchronised by driving the EEG amplifier's clock cycle using a 5 kHz signal derived from the 10 MHz reference signal from the MR scanner (Mandelkow et al., 2006; Mullinger et al., 2008c). The times taken to collect an fMRI volume (TR) and slice number were selected to aid gradient artefact removal, such that each slice was collected at an integer number to avoid errors in the timing detection. For example, although the fMRI image acquisition was synchronised with the EEG acquisition, it does not necessarily mean the slice onsets are synchronised. In order to make sure the slice onsets are synchronised, one must make sure the acquisition of each slice is a multiple of the sampling rate. This was achieved by making sure the slice onset was an integer ($3.696/42 = 0.088$ ms/slice), which is a multiple of the sampling rate (0.0002 ms). The electrodes were routed to amplifiers by means of long ribbon cables to allow the amplifiers to be set away from the scanner bore. A plinth was set up at the back of the scanner for the amplifiers to rest on; this limits micro vibrations and reduces the amount of interference. Furthermore, the cables were placed between two sandbags at the back of the scanner to dampen vibrations on the cables. Two methods were used to collect cardiac cycles: an ECG electrode (connected to the EEG cap) and a four electrode VCG sampled at 500 Hz (Chia et al., 2000). The R peak markers derived from the VCG trace were aligned with the EEG data in

MATLAB. The VCG is based on recording signals from four electrodes, three placed along the mid-line of the chest and the fourth to the left of the heart forming a “T”-shaped arrangement on the surface of the chest. The dipole vector representing the magnetic-field related flow artefact recorded at the chest surface has been shown to be approximately perpendicular to that reflecting the R peak (Fischer, Wickline, & Lorenz, 1999), so that the trace aligned with the R peak vector is not strongly confounded by flow effects and therefore allows optimal detection of the time of onset of the R peak (Chia et al., 2000; Fischer et al., 1999).

2.4. fMRI analysis

Pre-processing

The original images from the scanner were pre-processed using in-house MATLAB code (available upon request). This extracted an image volume for each echo and then combined the two echoes using simple linear summation. Subsequent processing and statistical analysis was carried out using statistical parametric mapping (SPM8) software (Wellcome Trust Centre for Neuroimaging). Single subject EPI volumes were corrected for motion artefacts by registering them to the first image volume using a rigid body spatial transformation. The mean functional volume was co-registered to the subjects' T₁-weighted image. DARTEL (Diffeomorphic Anatomical Registration using Exponentiated Lie algebra; Ashburner, 2007) was used to improve inter-subject registration and precision in anatomical localisation (Klein et al., 2009). The T₁-weighted image of each subject was partitioned into grey matter, white matter and CSF tissue classes using SPM8's 'new segment' toolbox. Following this, we used the DARTEL toolbox to create a group template derived from all of the subjects in our dataset. The grey matter component of this group template was then registered to the SPM8 grey matter probability map (MNI standard stereotactic space), by estimating a 12-parameter affine transform. In the process of creating the template brain, DARTEL also outputs 'flow fields' for each subject, which contain the transform from the original T₁-weighted image space to that of the group template. Each subject's flow field was combined with the group template-to-MNI transform using SPM8 deformation utility. This transformation from subject native space to MNI space was then applied to all functional image volumes, with a smoothing kernel of 8 mm full-width half maximum (FWHM) Gaussian filter and re-sampled to 3 x 3 x 3 mm

voxel size using tri-linear interpolation. Functional images were also temporally filtered using a high-pass filter with a cut-off of 128 s (~ 0.0078 Hz). The MNI normalised grey matter image for each subject was converted into a binary mask and used as the statistical mask for analysis.

Statistical analysis

Data were analysed using the general linear model approach. At the individual subject level, each task was modelled as a boxcar function (resting blocks were modelled implicitly), and subsequently convolved with the canonical hemodynamic response function. Contrasts were calculated to assess activations for intelligibility ($W - WRS$). The whole brain multi-subject analysis was carried out using a random effects model with a one-sample t-test. The group t-map was assessed for cluster-wise significance using a FDR threshold of $p=0.05$ based on a voxel height threshold of $p=0.001$.

2.5. EEG analysis

fMRI artefact correction and pre-processing

BrainVision Analyzer 2.0 (Braincap MR, Brain Products, Germany) was used for pre-processing, including cleaning of the gradient (GA) and ballistocardiogram (BCG) artefacts. A high-pass filter (0.1 Hz) was applied to the raw EEG trace to reduce the effect of drifts on the subsequent correction steps. The time taken for the fMRI volume to be collected was an integer divisible by $2 \mu\text{s}$ ($\text{TR}/\text{slice number} = \text{integer divisible by sampling rate}$; $3.696/42 = 0.088$ s), therefore we implemented an automatic GA correction method. This places an onset trigger (voltage acceleration $> 600 \mu\text{V}/\text{ms}$) on the raw EEG trace, which yielded the correct number of volume acquisitions (plus dummy volumes). A visual inspection of the onset trigger was made to make sure the triggers were sensible. A local average artefact subtraction (AAS; Allen et al., 2000) was performed with a sliding window of 21 samples to reduce the effect of temporal discontinuities. The data was low-pass filtered at 40 Hz and down sampled to 500 Hz in order to reduce the data size (using a cardinal spline interpolation). The subsequent BCG correction can be achieved by using this lower sampling rate, as the voltage acceleration during each artefact is not as rapid as the GA.

A two-stage procedure was required to correct the BCG artefact. The four electrode VCG collected cardiac data directly from the participants' chest. This yielded a VCG trace that was less affected by the magnetic field artefact, which

produced a clean and accurate observation of the heart response (see Figure 13e). The VCG and EEG recordings were aligned in time based on the triggers generated by the scanner at the start of each volume acquisition, so that the R peak times identified from the VCG could be directly applied to correct the EEG data. The next stage required the accurate labelling of the QRS complex on the VCG vector. To do this, some in-house MATLAB code was used to identify the R peak across the trace. This code searched the VCG vector for maximum amplitudes in sliding time windows (for example, 500-1200 ms), and marked each peak location. The VCG did contain spikes and random fluctuations; therefore a second manual pass was carried out to make sure the labelling was as accurate as possible. Having obtained a list for the time of each R peak, we imported pseudo-triggers into the original EEG file to represent each R peak. These peaks were subsequently used for correction using the AAS method with 21 samples per local average. Unfortunately, due to technical difficulties, not all subjects had a usable VCG trace (4/20). In order to protect against any possible failures, we attached an ECG channel to the subject that connected directly into the EEG bundle. For these subjects an automatic detection of the QRS cardiac cycle was performed on the ECG channel, implemented in BrainVision Analyzer. The placement of the R peak markers were checked and amended by visual inspection to ensure appropriate R peak identification. However, the R peak was not as easy to identify in these cases due to MR contamination (see Figure 13e, for a comparison between VCG and ECG trace).

The GA and BCG corrected data was subsequently marked for bad segments. These included large muscle artefacts or anomalies. Finally, ICA (informax extended) was used to detect and remove residual artefacts. As cleaning the data using ICA can be subjective, we limited the amount of components to be removed to the top three artefact components in terms of energy and kurtosis. The clean data was then re-referenced to an average reference and then exported (continuous trace) into MATLAB for further trial rejection and analysis (ERP, frequency and source) using the FIELDTRIP toolbox (Oostenveld et al., 2011). All participants had at least 50 trials per condition in the final averages.

The FIELDTRIP toolbox was used to extract epochs from the continuous trace, at the onset of each condition (-100-1000 ms). The resulting trials were baseline corrected and band-pass filtered at 1-40 Hz. All trials were processed using the artefact rejection tool to remove noisy trials/channels. This tool plots

descriptive information for each trial and channel (i.e. max, min, variance, z-score, kurtosis etc.). The artefact-rejected data was subsequently used for ERP averaging and statistical comparisons. The same trials were used for conducting frequency analysis, however a larger epoch segment was taken to avoid edge effects at lower frequencies (-750-1250 ms).

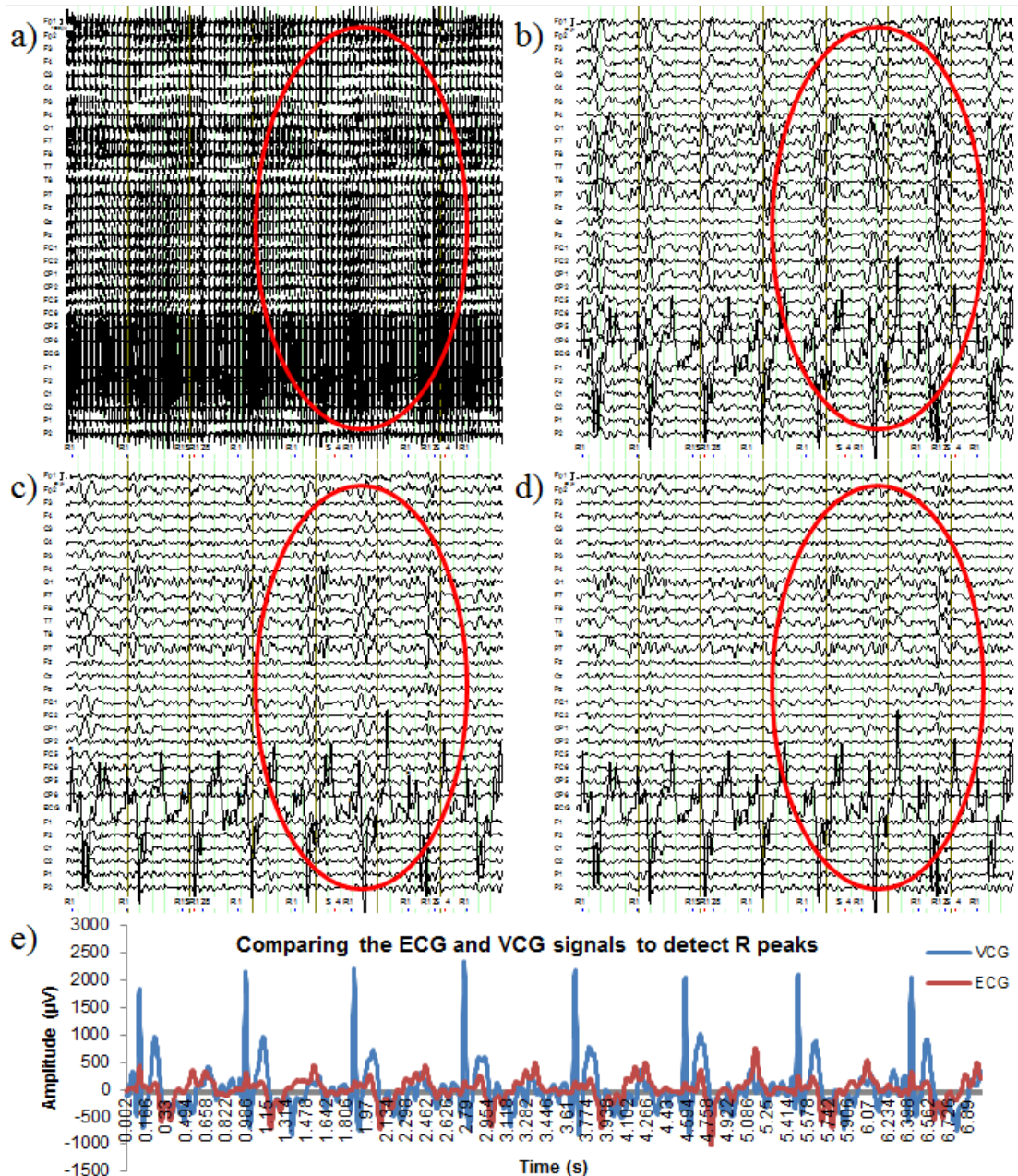


Figure 13: EEG time series showing the effects of MR artefacts induced on EEG sensors and the effect of applying an artefact correction to produce relatively clean EEG signal. Footnote: a) raw EEG signal with gradient artefact (GA) and ballistocardiogram artefact (BCG), characterised by large amplitudes; b) GA correction removes the large amplitudes however the BCG still remains; c) BCG correction using ECG electrode to identify R peaks, although residual artefacts remain; d) BCG correction using VCG to identify R peaks, removes more pulse related artefact with smaller residuals; e) comparison of the quality ECG and VCG trace across 7 s, the peaks are clear on the VCG trace and much harder to identify on ECG.

Statistical analysis

The statistical difference between two conditions was evaluated by a cluster-based random permutation test (Maris & Oostenveld, 2007), which was implemented in the Fieldtrip toolbox. This approach controls the Type-1 error rate, which involves multiple comparisons (one comparison for each electrode). Here is a brief description of the general procedure. First, for every data sample (electrode or electrode * time point) a simple dependent-samples t-test is performed. All adjacent data samples (spatial or temporal) exceeding a preset significance level (5%) are grouped into clusters. For each cluster the sum of the t-statistics is used in the cluster-level test statistic. Then a null distribution that assumes no difference between conditions is created. This distribution is obtained by randomly assigning the conditions in subjects 1000 times, and calculating the largest cluster-level statistic for each randomization. Finally, the observed cluster-level test statistics are compared against the null distribution, and clusters falling in the highest or lowest 2.5th percentile are considered significant. For each subject, condition specific averages were computed, which were in turn used for group-level statistics using the permutation method described above. Furthermore, the frequency data was analysed in the same way but used an average across the frequency band of interest (i.e. 4-8 Hz theta, 8-13 Hz alpha, 13-30 Hz beta and 30-40 Hz gamma). The same contrast for intelligibility was constructed as the fMRI analysis, whereby we compared $W > WRS$.

Source analysis

For the estimation of the neuronal sources, we used a beamforming method: linear constrained minimum variance (LCMV) (Van Veen et al., 1997) and placed virtual electrodes to extract EEG time series. Each subjects' brain volume was divided into a 3-D grid. A canonical MNI 0.75 cm grid was used as a template to obtain a grid for each subject. We applied an inverse non-linear transformation to the grid positions of the canonical grid to obtain subject-specific dipole grids. This procedure facilitates group analysis, because no spatial interpolation of the reconstructed activity is required.

For each grid position, we constructed spatial filters. These filters have the property that they optimally pass activity from the location of interest, while other activity, which is present in the data, is suppressed. To compute the leadfield matrices, we used a 3-layer BEM model, which to some degree assumes realistic

information about the interface between skin, skull and brain surfaces (Oostendorp & van Oosterom, 1989). The layers were derived from each individual subject's structural T₁-weighted image, which was spatially aligned to the EEG sensors. The spatial filters were obtained using a common filter strategy, where all conditions are appended to calculate a common spatial filter. The advantages of this method are that more data is used to construct the spatial filters, which lead to more reliable estimates of the filters. Furthermore, as the same filters are used to estimate the source in all conditions, one can ascribe differences in source power to conditions, rather than differences between the filters.

To analyse the location of the sources accounting for the significant sensor level effect, we band-pass filtered the pre-processed trial data at the significant frequency band. A covariance matrix between all EEG sensor pairs was calculated between the significant time windows. For each subject, the intelligibility contrast was calculated at the source level using common spatial filters. The statistical difference between conditions was evaluated by using the cluster-based random permutation test described above (Maris & Oostenveld, 2007).

Virtual electrode analysis

Virtual electrodes were placed in the EEG source space based on the location of the peak activity using the fMRI intelligibility contrast. The group level MNI peaks were converted into subjects' native space, in which the source grid was created, by applying the inverse transformation described above to create the 3-D grid positions (MNI to native space). The grid position that was closest to the peak location was determined by searching for the grid point with the smallest Euclidean distance. Subsequently, trial-by-trial data was projected through the spatial filters at each peak position for each subject, which resulted in a virtual electrode time series measured in three directions (x, y and z). In order to reduce the three dimensions to a singular representation, we obtained the regional power by calculating the root mean square ($RMS = \sqrt{(x^2 + y^2 + z^2)}$). Furthermore, due to the poor spatial resolution of EEG and the fact that spatial filters that are close to each are highly correlated, we clustered the virtual electrodes based on the Euclidean distance between them. This resulted in averaging virtual electrodes to form regional electrodes of posterior, mid and anterior STS (pSTS, mSTS and aSTS), IFG and vATL. The time series were normalised and baseline corrected. An intelligibility contrast was obtained ($W > WRS$) based on 100 ms time bin partitions. Statistical analyses were carried out to determine how intelligibility

differences changed across time. Finally, in order to investigate the functional correlations between ROIs, the post stimulus epoch (0-1000 ms) was used to calculate a Pearson's correlation between all ROIs combinations (5 x 5) for each subject. The correlation values were converted into a Fisher Z-score. The group Z-score was tested against a $p=0.005$ significance level to determine if two regions were functionally correlated.

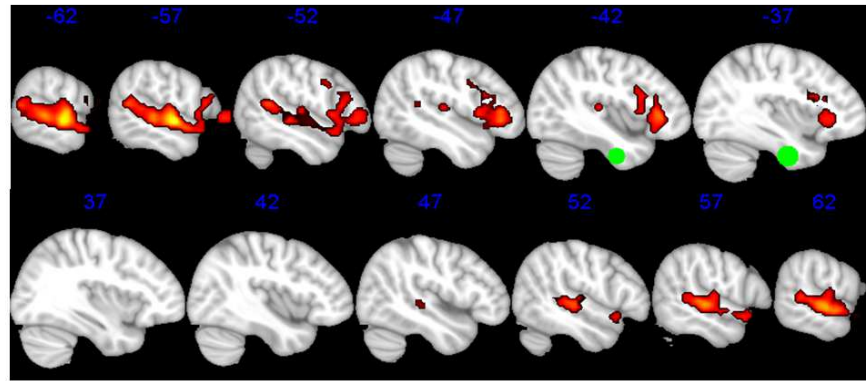
3. Results

3.1. fMRI analysis

Before group level analysis was carried out, two subjects were removed for excessive movement during the scan (>5 mm in any direction). The subsequent analysis was done on 18 subjects and the figure shown is FDR cluster corrected at $p=0.05$ based on a voxel height threshold of $p=0.001$. Figure 14 shows the results for the intelligibility contrast. As predicted, there was significant activation along bilateral STS. Furthermore, the left hemisphere cluster encompassed peaks within the IFG. The MNI peak activations and cluster sizes are shown in Table 4. Activation in the vATL was not seen at the current whole brain threshold. However, as vATL activity has been shown in previous studies (e.g. Halai et al., Chapter 3 and 4; Visser & Lambon-Ralph, 2011), an independent ROI was used to determine if any differences can be seen in the current data. This analysis showed that the fMRI signal for intelligibility ($M=0.329$, $SE= 0.135$) was significantly greater than zero ($t(17)=2.434$, $p=0.026$, Figure 14b), indicating an effect was present but not strong enough to survive the current whole brain threshold.

W > WRS
(Intelligibility)

● = apriori region
of interest



fMRI signal from inferior
temporal ROI [-36 -9 -36]

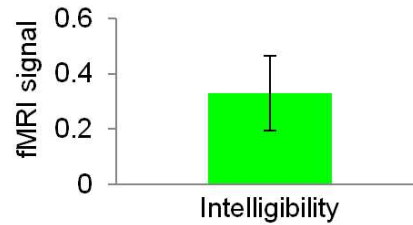


Figure 14: Significant fMRI activations for intelligibility (W > WRS). Footnote: Activations are FDR cluster corrected at $p=0.05$ based on an uncorrected voxel-threshold of $p=0.001$ (hot scale, 1-10 t-value). The figure is centred on the left and right temporal lobes. The ventral temporal ROI (green) is a 12 mm sphere at an apriori peak (MNI: -36 -9 -36) from Visser and Lambon-Ralph (2011). The bar graph shows the averaged fMRI signal across the group for intelligibility (including standard error), which is significantly active.

Contrast	Cluster size	Anatomical Region	t-value	z-value	Peak MNI co-ordinate				
					x	y	z		
Intelligibility	1256	aSTG	8.76	5.32	-57	-12	-3		
		pSTG	6.57	4.58	-60	-33	3		
		IFG pTri	6.27	4.45	-45	30	3		
		toMTG	5.50	4.11	-68	-54	9		
		IFG pOrb	5.20	3.97	-33	24	-3		
		HG	5.03	3.89	-45	-24	9		
		IFG pTri	4.90	3.81	-54	18	12		
		IFG pTri	4.87	3.80	-57	36	0		
		TP	4.84	3.79	-54	9	-12		
		IFG pOp	4.43	3.56	-41	14	20		
		Precentral gyrus	4.14	3.39	-51	3	36		
		423	423	pSTG	6.95	4.72	62	-18	-2
				pMTG	4.97	3.85	63	-39	6
				TP	4.71	3.72	57	12	-12
Postcentral gyrus	4.28			3.48	69	-9	15		

Table 4: MNI peaks identified for the intelligibility contrast (W > WRS).

Footnote: The Harvard-Oxford atlas was used to identify the corresponding anatomical region for each peak. Abbreviations: superior temporal gyrus (STG); middle temporal gyrus (MTG); inferior frontal gyrus (IFG); pars orbitalis (pOrb); pars triangularis (pTri); pars opercularis (pOp); temporal pole (TP); Heschl's gyrus (HG).

3.2. EEG analysis

The same subjects as above were used to conduct the EEG analysis. The peak fMRI co-ordinates were used to place a virtual dipole to obtain an EEG signal from the closest corresponding grid point in EEG source space. Due to the poor spatial resolution of EEG and the number of close peaks in the fMRI results, the peaks were clustered into 5 larger regions; pSTS, mSTS, aSTS, IFG and vATL. The same intelligibility contrast was applied to the EEG signal, which was analysed in 100 ms time bins. Figure 15 depicts the location of the 5 ROIs across the whole epoch (0-1000 ms). Each ROI is colour coded to highlight the statistical significance level for intelligibility (red: $p < 0.01$, yellow: $p < 0.05$, green: not significant). The functional connectivity between ROIs was calculated across the whole epoch, and significant connections are shown (red line: $p < 0.005$).

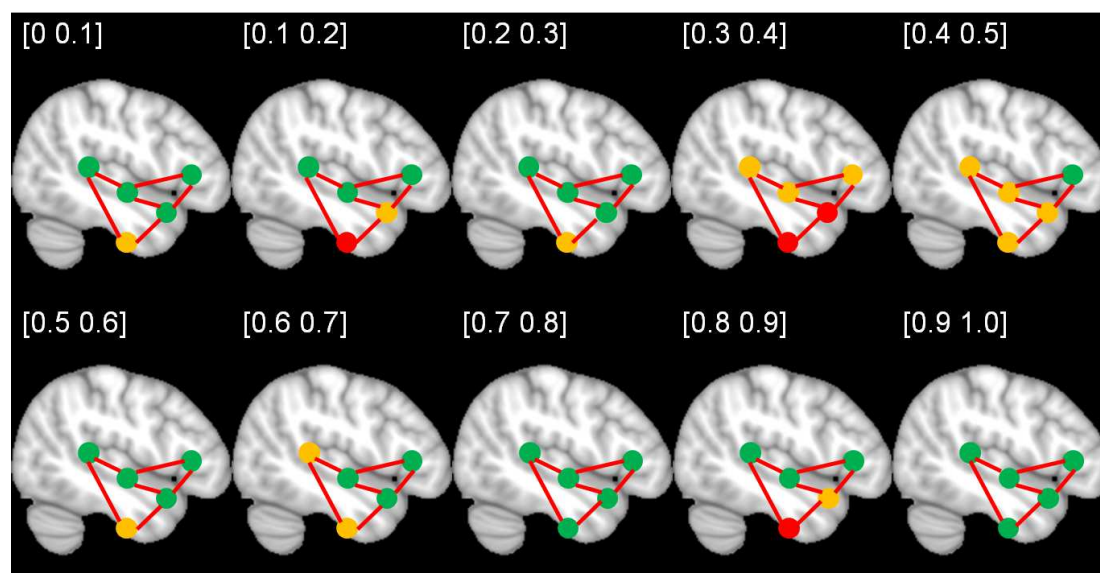


Figure 15: EEG virtual electrode results for intelligibility (W > WRS) depicted on MNI brain across 1 s epoch. Footnote: The five ROIs (pSTS, mSTS, aSTS, IFG and vATL) are colour coded to show significant activations; red: $p < 0.01$, yellow: $p < 0.05$, green: not significant. Functional connectivity was calculated across the entire epoch between all ROIs for each subject. Significantly correlated ROIs are shown; red line: $p < 0.005$.

Event-related potential and frequency analyses were carried out on the original sensor data, using the same intelligibility contrast. The ERP analysis did not show any significant differences at group level after multiple comparison correction. However, the frequency analysis showed a significant effect of theta between 0.45-0.65 s, after multiple comparison correction (Figure 16). There were no differences seen in the alpha, beta or gamma range. The topographical distribution of the theta difference indicated a bilateral source over the

anterior/middle part of the brain; therefore, in order to get a better understanding of the neural generators, a source analysis of the theta effect was carried out. The group source analysis is cluster corrected for multiple comparisons and showed that the activity is likely to originate from the left vATL and IFG (Figure 17). The Figure shows the raw t-scores across the top-row, revealing a large area of activation along the left fusiform cortex, extending back towards the posterior temporal-occipital junction. However, the activity observed in the posterior temporal-occipital region does not survive correction, whereas the anterior fusiform activity is robust. Activity in the right hemisphere was diffuse at the uncorrected threshold, and no regions in this hemisphere survived correction. The Figure only shows positive values (effect of intelligibility), as no activity in the negative range (effect of rotated speech) survived multiple comparison correction.

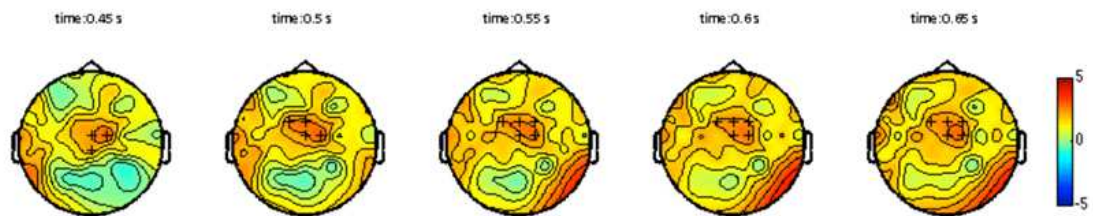


Figure 16: Topographical plot showing an intelligibility effect ($W > WRS$) in the theta frequency range between 0.45-0.65 s. Footnote: Sensors that are significant (corrected for multiple comparisons) are marked (+).

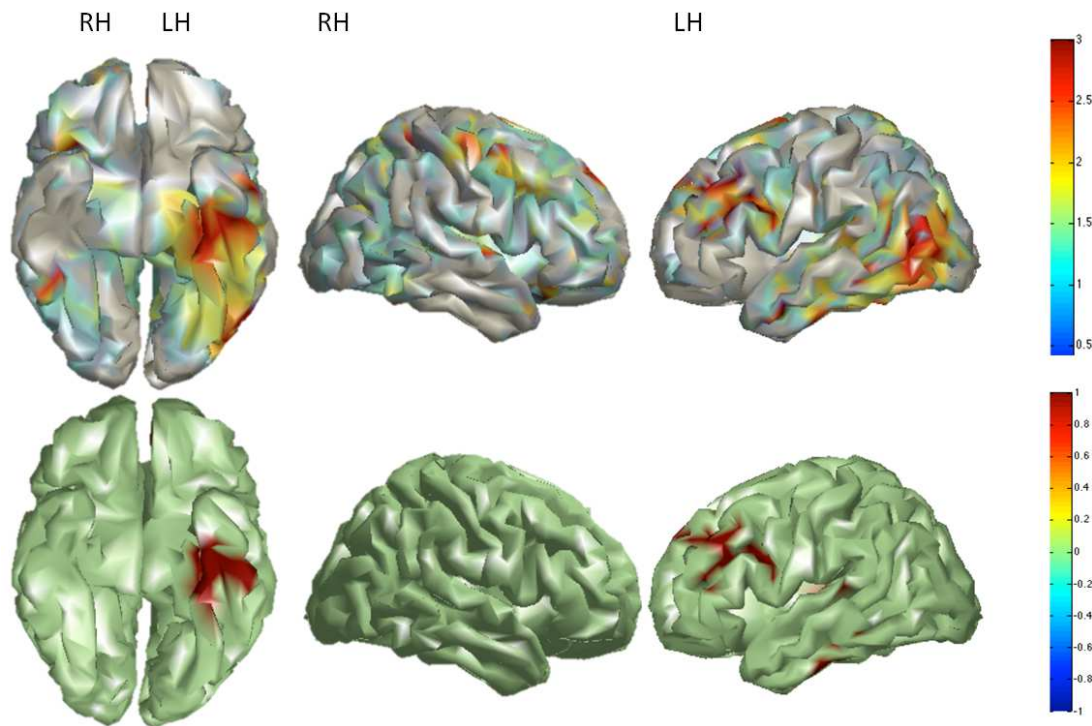


Figure 17: Source estimation of the theta effect during 0.45-0.65 s. Footnote: The top row shows an overlay of uncorrected *t*-values, whereas the bottom row shows the clusters that survive a multiple comparison correction. Two clusters are found; left temporal fusiform gyrus (TFG) and inferior frontal gyrus (IFG).

4. Discussion

The aim of this study was to explore the spatial and temporal profile of passive intelligible word comprehension using simultaneous EEG-fMRI. Research into speech comprehension has largely been investigated using neuroimaging methods in isolation (e.g. fMRI, M/EEG, PET, TMS), but has not focused on combining spatial and temporal information to provide a comprehensive model. The results of the spatial and temporal analyses from the current study will be discussed, ending with an overview of how they both fit into current spatial and temporal models of speech comprehension.

4.1. Evaluation of fMRI analysis

The fMRI results showed activation related to intelligibility along the entire length of the STS bilaterally, as well as in the left IFG and left vATL (the vATL activation was weaker than the other activations and could only be detected using an ROI analysis). The bilateral activation supports models proposing that speech comprehension is bilaterally organised (Hickok & Poeppel, 2007; Okada et al.,

2010), although it should be noted that the network was more extensive in the left hemisphere.

It is important to note that the role of vATL has been non-existent in many models of speech processing. Neuropsychological literature has provided strong evidence for the involvement of the ATL for semantic cognition. In addition, recent neuropsychological (Patterson et al., 2007), fMRI (Visser et al., 2010), TMS (Pobric et al., 2010) and MEG (Marinkovic et al., 2003) studies have all provided converging evidence for the role of ATL, and in particular vATL (including ITG and TFG), during semantic memory tasks. The act of comprehending speech requires access to semantic representations, and therefore, it is likely that the vATL is involved during speech comprehension. PET studies have implicated the vATL during narrative comprehension (Sharp et al., 2004; Spitsyna et al., 2006), and novel fMRI methods are producing similar results for sentence and word comprehension paradigms (Halai et al., Chapter 3 and 4; Visser et al., 2010; Visser & Lambon-Ralph, 2011). The current findings provide supporting evidence for the involvement of vATL during passive word comprehension. However, it is noted that this activation was relatively weak, as it was observed using ROI analysis and did not survive whole brain correction. It has been suggested that passive tasks might mask semantic processes as the idle moments during the experiment can allow subjects to generate internal speech or task-unrelated thoughts, thereby activating brain regions related to semantic processing (Binder et al., 1999; Gusnard & Raichle, 2001; Stark & Squire, 2001). Within a subtraction analysis, the semantic processing regions would not appear due to this masking effect. It is possible, in the context of the current study that during the unintelligible speech conditions; subjects might generate stimuli independent thoughts, which would result in masking activation of semantic processing when performing contrast analysis. Therefore, leading to the activity within the vATL to be diluted and not survive whole brain correction. This suggests that a passive speech comprehension paradigm can be used to detect changes in the semantic processing system, although the true effect might be greater when controlling the baseline with an explicit unrelated task (e.g. a number task; Binder et al., 2011; Binney et al., 2010). It should be noted that previously a passive sentence comprehension paradigm was sufficient to produce robust vATL activity for intelligibility (Halai et al., Chapter 4). A potential explanation could be that subjects pay more attention to sentences as a meaningful message is being conveyed, whereas words in isolation are relatively arbitrary. In addition, there could be a

methodological cause to the relatively low activity in the vATL. The presence of EEG electrodes within the MRI can cause increased load on the RF coils used for signal reception, which reduces the SNR of the recorded data. For example, Scarff and colleagues found that increasing the number of sensors on the EEG cap within a 3 T MRI decreased the SNR (Scarff et al., 2004). However, recent evidence has suggested that 64 sensors at 3 T do not significantly reduce temporal SNR, which is particularly relevant for fMRI studies (Mullinger, Debener, Coxon, & Bowtell, 2008b).

A large region around the prefrontal cortex was activated during intelligible word comprehension. Specifically, the pars orbitalis and pars triangularis were activated with a small portion of activity extending posteriorly towards pars opercularis. This activity largely supports studies that suggest that the anterior portions of the IFG (pars orbitalis and pars triangularis) contribute to tasks requiring semantic control (Noonan et al., 2013). In contrast, posterior IFG (pars opercularis) is suggested to contribute to tasks involving phonological control processes (Noonan et al., 2013), which is not supported by data as semantic processing is found to encroach this region. However, an alternative hypothesis could be that this posterior region is involved in selection between alternatives and inhibition of irrelevant processes, which are processes that semantic and phonological domains are likely to share (Badre & D'Esposito, 2009). The acoustic noise from the fMRI could make the comprehension of the word more challenging, which could explain the activity observed in pars orbitalis and pars triangularis for semantic control (Jefferies & Lambon Ralph, 2006; Price, 2010; Rauschecker & Scott, 2009; Thompson-Schill et al., 1997). Furthermore, Hickok and Poeppel (2007) suggest this region is primarily involved in articulatory processing (in combination with primary motor areas), which is not supported due to the lack of activity seen in other speech motor regions. It could be argued that listening to random single words in a noisy environment is actually harder than listening to sentences because the sentence has context that aids the comprehension process. Therefore, the activation of widespread IFG might be explained by an increased demand on the semantic system to obtain a correct representation having been given a noisy speech signal (e.g. Jefferies & Lambon Ralph, 2006; Thompson-Schill et al., 1997).

The bilateral nature of the STS, in particular the posterior STS, activity could be due to methodological factors concerning continuous fMRI scanning, which add scanner noise during the stimulus presentations. For example, it is

argued that difficult speech comprehension recruits extended components to the core and necessary speech-related network. Increases in listening effort have been associated with increases in posterior STS and supramarginal gyrus (Pelle et al., 2010a), which suggest that the posterior STS regions detected in the current study are possibly related to increased listening effort (Adank, 2012b; Pelle et al., 2010a). A recent meta-analysis, did however, find that studies using sparse sampling (quiet environment) were more likely to find bilateral anterior STS/MTG activity compared to continuous scanning, whereas continuous scanning promoted activity within pre-SMA and SMA (Adank, 2012a). Adank suggested that the presence of scanner noise causes difficulty in resolving speech, which leads to speech motor areas becoming involved to aid in comprehension. Furthermore, it could be possible that the contralateral hemisphere becomes engaged as a compensatory mechanism (Just & Carpenter, 1992; Prat & Just, 2008; Prat & Just, 2011). For example, a right hemisphere 'spill-over' hypothesis suggests that the right hemisphere serves as a resource for language processing, with similar but coarser and less efficient capabilities than the left hemisphere. The right hemisphere becomes more active when processing demands of a task outstrip the resources in the left hemisphere, therefore processing spills-over into right hemisphere (Prat & Just, 2011). Finally, increased functional integration between distributed regions has been observed during degraded speech (Obleser et al., 2007a). This study showed that when the speech signal is compromised or degraded, the functional integration between AG and prefrontal regions increases if predictability of the speech is high. In other words, high-order semantic information can be used to aid degraded speech comprehension. This could be possible through working memory (Scott et al., 2003; Sharp et al., 2004) or auditory-motor transformation processes (Hickok & Poeppel, 2007; Jacquemot & Scott, 2006) found within AG, or increases in attention processes related to monitoring or selection found within the fronto-parietal network (Lebedev et al., 2004). The anterior location of STS activity for intelligibility supports models that implicate activity along the anterior STS (although not as anterior as for sentences) (Davis & Johnsruide, 2003; Narain et al., 2003; Price, 2010; Rauschecker & Scott, 2009; Scott et al., 2000; Scott et al., 2006). Furthermore, the model proposed by Hickok and Poeppel (2007) predicts that a lexical interface is located at the posterior MTG/ITG; however, this is not supported by the current results, where activity is found along the STS and vATL.

4.2. Evaluation of virtual EEG electrode analysis

The outcome from the virtual EEG electrode analysis was quite complex, however there are some interesting features that have emerged. It is clear from the evolution across the epoch that the majority of the nodes become significantly active between 300-400 ms (5/5 nodes) and 400-500 ms (4/5 nodes), in comparison to other time windows which generally show zero, one or two active nodes. The latency of this effect suggests that it could be related to the N400 effect seen in previous EEG studies. Considerable research has focused on the N400 effect, which usually focuses around words in context within sentences or discourse (Kutas & Federmeier, 2011). However, some studies have found similar N400 effects when using words in isolation for naming tasks (Stuss et al., 1983), or in different types of lists, word – non-word pairs (Bentin et al., 1993), function-content words and high-low frequency words (Rugg, 1990; Smith & Halgren, 1987). Additionally, lexical priming tasks, where a target word was or was not related to a preceding word showed that related words reduced the N400 in comparison to unrelated words (reviewed in Kutas & Van Petten, 1988). This suggests that although unrelated context does seem to increase the amplitude of the N400, the underlying process is not tied to the context, and therefore is more likely to reflect an attempt to access the underlying semantic representations. An intelligibility effect is seen strongly within vATL and aSTS, whilst weaker effects are seen in pSTS, mSTS and IFG, during 300-400 ms. In combination with knowledge of the functional role of these regions, it is suggested that the aSTS and vATL are involved in the comprehension of speech during this time window. The weaker effects from pSTS, mSTS and IFG could be down to the fact they are not necessary for comprehension and are likely to be by-products of predictive constraints imposed by the semantic system that is aiding comprehension. Evidence for this comes from the involvement of pSTS and mSTS during 400-500 ms, after the aSTS and vATL have been active, with the addition of the functional connection between pSTS and vATL, which could be applying some semantic control to aid semantic access within vATL.

Interestingly, the vATL node was found to be active consistently across the epoch, unlike other nodes. There is intracranial EEG evidence to suggest that during visual stimulation, the posterior portion of the vATL becomes active at very early time periods (e.g. less than 200ms; Liu, Agam, Madsen, & Kreiman, 2009). However, to our knowledge there are no intracranial studies presenting auditory

stimuli that address the question of how early the differential parts of the vATL are active. However, Marinkovic and colleagues used MEG to show activity for auditory words within the ATL, including the vATL at around 400 ms, which is relatively late compared to the early involvement of vATL in the current study. We speculate that the anterior portion of the vATL region could be active very early, in similar fashion to the posterior portion of the vATL during visual stimulation due to the role the anterior vATL plays in auditory speech processing. The early activation can be explained by two possible frameworks; a parallel processing (Marslen-Wilson, 1987) and two-pass framework (Pulvermüller, 2007). The former suggests that incoming signals are processed along multiple dimensions, including low-level acoustic features and high-level semantic information, whereas the latter suggests that an initial crude first pass of parallel processing occurs, followed by a secondary pass, which refines and corrects the initial predictions. The two-pass framework may fit the appearance of the vATL better as the activation is sustained, which suggests that the region is constantly accessing semantic information into late time windows. In contrast, the parallel approach does not account for the late effects of semantic processing; due to the fact that these are likely to be resolved earlier.

Functional connectivity analysis was conducted to determine which of the ROIs were correlated with each other during intelligible speech comprehension. Firstly, it is noted that the most connected nodes were mSTS and aSTS. This could provide an insight into their role within the speech comprehension system as central nodes. The mSTS node is the closest to the PAC, which is known to receive incoming auditory signals, which then spread towards pSTS and aSTS regions as complexity increases (Rauschecker & Tian, 2000; Spitsyna et al., 2006). The mSTS connections with pSTS and IFG could be related to predictive constraints being applied in order to successfully and reliably process the incoming stimulus, and divert attentional resources in the correct manner. The aSTS node is connected with mSTS, IFG and vATL, which could be further evidence for a critical role of aSTS in speech comprehension. The aSTS is well known to be involved in intelligible speech comprehension, with the addition of the vATLs role in semantic knowledge; therefore, it is likely that these two regions are both key to achieving comprehension. The IFG connection with aSTS is also likely to be due to the role of IFG in semantic control and predictive constraints, which can aid successful comprehension. In addition, the pSTS is correlated with vATL, which could suggest a more direct connection of the auditory system with the

semantic system. In order to understand these functional connections, one could start to think about which white matter tracts these processes could communicate via. Diffusion tensor imaging provides a non-invasive method to obtain anatomical pathway information, which could add a further dimension to the model of speech comprehension (e.g. see Saur et al., 2010).

4.3. Evaluation of traditional EEG sensor level analysis

Traditional EEG analysis focuses on determining differences between conditions at the sensor level, in the time or frequency domain, which are designed to investigate either evoked or induced activity. Evoked activity is directly driven by stimuli onset, and is time and phase locked to an event. Therefore evoked activity can be extracted from the on-going EEG by straightforward averaging of the event-related EEG epochs, as long as the epochs are free from artefacts and latency shifts. Due to the nature of simultaneous EEG-fMRI, the EEG data is affected by MR related artefacts (although we have taken all possible measures to reduce these). Any residual noise in the EEG data affects ERP analysis at the group level, by reducing the SNR. This could explain the outcome of the ERP analysis in the current dataset, where no differences were seen for intelligibility across sensor and time. This is in contrast to previous work using the same paradigm and EEG outside the scanner where a significant difference was seen during early and late time windows, relating to low-level and semantic level processing respectively (Halai et al., Chapter 2). In order to boost the SNR of the EEG data, we could consider increasing the number of trials within the EEG-fMRI setting or increasing the number of participants. However, there is a trade-off with overall fMRI scanning time and resources. It could be argued that in order to obtain clean EEG data, future studies should collect EEG and fMRI data separately (Ball et al., 1999; Menon et al., 1997). However, as we have argued in the introduction this is not desirable as it removes the possibility of some types of joint analysis and introduces other potentially confounding factors such as repetition/order effects.

In contrast to the ERP analysis, frequency analysis is used to investigate induced activity, which is largely rhythmic (oscillatory) in nature. Induced activity is best considered as rhythmic background activity that is modulated, rather than evoked, by external (i.e. some sort of stimulation) or internal (e.g., motor preparation, cognitive processing) events. The amplitude (or power) of oscillatory

activity is indicative of (de)synchronization, which can be detected at each sensor. This can occur in different frequency bands and evidence for the effect of oscillations during language processing have been found in theta (4-7 Hz; based on Bastiaansen, van Berkum, & Hagoort, 2002; Hagoort, Hald, Bastiaansen, & Petersson, 2004), alpha (8-12 Hz; based on the studies reviewed in Klimesch, 1999), low-beta (13-18 Hz based on the studies reviewed in Weiss & Mueller, 2003) and gamma bands (above 30 Hz; based on Hagoort et al., 2004; Pulvermüller, Lutzenberger, & Preissl, 1999). The results from the current study showed a significant effect of the theta for intelligibility between 0.45-0.65 s. There were no differences in any other frequency band, although an absence of effects due to poor SNR cannot be ruled out. There is evidence to suggest that theta is involved in lexical retrieval (Bastiaansen & Hagoort, 2006; Bastiaansen, Van der Linden, ter Keurs, Dijkstra, & Hagoort, 2005), where it was found that the scalp topography of theta power for words with auditory or visual properties were distinct. The word with auditory properties showed differences in electrodes above temporal cortices, whereas words with visual properties showed differences in electrodes above occipital cortices. This indicate that theta power increases are related to lexico-semantic retrieval (Bastiaansen et al., 2005). It is suggested that hippocampal neural activity reflects oscillations within the theta frequency range (4-8 Hz; reviewed in Klimesch, 1999), which is known to be related to memory processing. Studies have shown theta increases during encoding and retrieval of episodic information from long-term memory (e.g. Burgess & Ali, 2002; Fell et al., 2001; Klimesch, 1999). Furthermore, it has been shown that theta is more strongly involved in retrieval than during encoding, which indicates that theta activity is not involved in memory processes in general, but for retrieving or accessing stored information (Klimesch et al., 2001). It must be noted, however, that synchronisation in theta oscillations are not suggested to be the sole mechanism by which lexical retrieval is achieved, as alpha, beta and gamma have also been implicated in language processes (Hagoort et al., 2004; Klimesch, 1999; Pulvermüller et al., 1999; Weiss & Mueller, 2003).

The source estimates of the theta difference during intelligibility were found within the left vATL and IFG. This is an important finding as the localisation of the theta power change is in-line with the fMRI results. The source localisation of the theta changes was spatially unconstrained, and therefore there would no bias in the source estimation analysis towards these regions. This provides strong converging evidence for the role of the vATL and IFG for speech comprehension.

As discussed above, the vATL has been suggested to be a core component of semantic knowledge (e.g. Patterson et al., 2007; Rogers et al., 2004; Visser & Lambon-Ralph, 2011), therefore it is likely that this region is involved in the core aspect of speech comprehension. The source estimation of theta power changes also implicated the IFG, which is known to be involved in general language processing related to semantic control systems (Jefferies & Lambon Ralph, 2006; Noonan et al., 2009; Thompson-Schill et al., 1997). The subjects' were exposed to auditory stimuli within a relatively noisy fMRI environment. Furthermore the words were random and presented in isolation, which is likely to be difficult to comprehend compared to sentences that have neighbouring words and context to aid comprehension. Therefore, the localisation of the IFG might not be entirely surprising, as the experimental conditions might require additional executive control to achieve comprehension. It would be interesting to see how the role of theta changes during difficult comprehension periods, and whether the theta power increased at vATL, IFG or both regions or indeed if other regions of the brain are recruited (right hemisphere homologues or angular gyrus). Further research is needed to better understand the role of theta for speech comprehension and semantic knowledge in general.

Overall, although careful consideration should be given to the study design to minimise artefacts related to simultaneous EEG-fMRI data collection, it is proposed that future studies would benefit from collecting simultaneous EEG-fMRI by being able to directly compare the results from both modalities. However, the need to obtain simultaneous EEG-fMRI data will ultimately depend on the questions being asked. For example, is there a requirement to directly link fMRI and EEG data in the same area and time course? If the experiment does require such integration, the researchers need to be aware that ERP analysis (evoked response) is particularly vulnerable to noise induced by MR-gradients, whilst frequency analysis (induced response) seems to be relatively robust to the noise.

4.4. Implications for speech comprehension models

The current study used simultaneous EEG-fMRI to investigate how regions related to intelligible word comprehension change over time. The fMRI evidence indicated bilateral (but left dominated) anterior STS activity with the addition of left IFG and vATL activity. In general, this provides supporting evidence for models of speech comprehension proposing anterior STS regions to be involved during

speech comprehension (Price, 2010, 2012; Rauschecker & Scott, 2009). However, the vATL activity is almost entirely omitted from speech comprehension models, despite convergent evidence for its role in semantic processing tasks for comprehension (Hodges et al., 1992; Marinkovic et al., 2003; Pobric et al., 2010; Spitsyna et al., 2006; Visser et al., 2010). The current results suggest an update of the models is needed to include the vATL region during intelligible speech comprehension. In addition, it is crucial to understand the mechanism by which semantic representations are accessed. One possibility could be that low frequency theta oscillations, which are related to memory retrieval processes, and involved during intelligible speech comprehension to access/retrieve specific semantic representations; although further research is needed to understand the mechanisms.

The temporal profile of intelligibility processing provides supporting evidence for semantic processes occurring around 400 ms, which is consistent with a large body of research into the N400 (Friederici, 2002; Kutas & Federmeier, 2011). In particular, this effect around 400 ms is strongly observed in aSTS and vATL regions, which further support the functional role of these regions during speech comprehension. The temporal profile also provides some indication that the pSTS, mSTS and IFG are not critical to comprehension, and has been suggested to relate to some aspect of semantic control (Jefferies & Lambon Ralph, 2006; Price, 2010, 2012). For example, both pSTS and IFG are weakly activated around 400ms, whereas the pSTS activity persists into 500 ms. In addition, the pSTS was functionally connected with vATL, which could indicate some communication between the nodes and semantic control being applied to identify the correct semantic representations. It should be noted, however, that the vATL shows intelligibility related activity from around 100/200 ms, which is too early to be explained within a hierarchical framework. The early activation can be explained by two possible frameworks; a parallel processing (Marslen-Wilson, 1987) and two-pass framework (Pulvermüller, 2007). However, the late effects seen around 400 ms are difficult to explain with a parallel processing framework, as this model suggests that semantic processing starts from the onset of the presentation and would be resolved by this time. On the other hand, the two-pass framework can account for both late and early effects. It proposes that the early effects are reflective of a relatively crude pass of the speech information, which includes semantic features. This is evident by seeing vATL activity from around 100/200 ms. The late effects are suggested to be a more detailed reanalysis of the

crudely resolved speech information, which refines and corrects initial predictions. This reanalysis could be occurring around 400 ms, as during this time a number of nodes become active for intelligible speech comprehension, including parts of the speech network that are related to executive semantic control functions. As the vATL activity was sustained, it is suggested that the region is involved during the whole process of comprehension. Therefore, it is suggested that semantic representations provide top-down influences on incoming sensory signals that possess speech-like information, which could potentially be mapped onto semantic representations.

5. Conclusions

The current study provides preliminary evidence that simultaneous EEG-fMRI can be useful in probing brain networks of speech comprehension. The spatial profile of intelligible speech comprehension showed activity within bilateral STS, left IFG and left vATL using dual-echo fMRI. The spatial estimation of theta power increases during speech comprehension, also implicated the vATL and IFG regions. This spatially unconstrained analysis provides important convergent evidence for these regions during speech comprehension. The temporal ordering of these regions suggests that the majority of the activation occurred around 300-400 ms (possibly related to an N400 effect). Furthermore, an early sustained response was seen at the vATL, which could be related to top-down predictive and corrective processing of low-level auditory information using stored semantic representations. Current models of speech comprehension do not implicate the vATL; given the results from the current study (and Chapter 4) it is important to update these models to include the vATL. It is noted that the aSTS region is also critically involved during comprehension, which has been highlighted in previous models (Price, 2010, 2012; Rauschecker & Scott, 2009), however speech comprehension should include regions known to be critical for semantic representations (Patterson et al., 2007). In addition, the temporal evolution of auditory word comprehension is suggested to occur in a two pass framework (Pulvermüller, 2007), with early contributions from vATL providing top-down semantic information to predict low-level features and a wider contributions from all network nodes around 400 ms to resolve/repair any inconsistencies. It is proposed that using simultaneous EEG-fMRI could be a useful tool to help

integrate spatial and temporal information, which will be important in obtaining a more detailed understanding of speech comprehension.

Chapter 6

General Discussion

Advances in hardware and software have allowed two key techniques to be combined; EEG and fMRI to provide high temporal and spatial resolution, respectively. The aim of this research was to investigate the spatial and temporal characteristics of speech comprehension, using this novel multi-modal neuroimaging approach (Chapter 5). In order to achieve this goal, there were other considerations that needed to be taken into account. Firstly it was critical to make sure that the methodology and task paradigm was suitable for answering questions about speech comprehension within an EEG-fMRI setting. Therefore, Chapter 2 aimed to explore the temporal (and spatial) dynamics of intelligible speech comprehension. Secondly, the lack of evidence for vATL in speech comprehension models led us to question the fMRI methodology. Therefore, Chapter 3 aimed to select an fMRI method that is capable of imaging the vATL. The method of choice would help reconcile differences between speech comprehension and semantic comprehension tasks; in particular, accounting for the omission of the vATL (Chapter 4). This chapter starts by summarising the key results from each study, followed by a discussion of the implications of the findings for models of speech comprehension. The discussion is organised using the core research themes outlined in the introductory chapter: spatial organisation of speech processing; speech comprehension as part of the semantic network; influence of neuroimaging methodology on speech models; and temporal evolution of speech processing. Finally, there will be comments on methodological limitations and outstanding questions for future research.

Summary of findings

In Chapter 2, electrophysiological responses were measured using EEG while participants listened to intelligible and unintelligible words. The main aim of this study was to determine the suitability of the speech comprehension paradigm for research within an EEG-fMRI setting. Therefore, we wanted to determine how phonological and intelligible manipulations modulated ERP components. Specifically, the study investigated how the ERPs were modulated by manipulating

phonological and semantic content in speech, which could help determine if word comprehension was processed in serial, parallel or by using a mixture of both processes. The results showed a quite complex answer to this question. For normal word processing the evidence pointed to a relatively simple serial mechanism with phonological feature processing occurring early and semantic processing occurring later. However, when the stimuli were distorted there was evidence for some additional top-down processing. Distorted stimuli that were word like generated a higher signal than distorted pseudoword stimuli during the early window. Therefore, during the early time window (~150 ms), the incoming signal appeared to be assessed for general speech-like qualities based on existing stored information using executive control or selection processes (top-down constraints), but in general the lower-level feature processing (i.e. complex acoustic variations) was dominant. During the second window (~400 ms), detailed information from the signal was mapped onto stored semantic representations. The early component was localised around HG and frontal regions. In contrast, the late component was localised in ATL regions, with some additional frontal activity. In sum, the results suggest a two-pass processing framework for auditory word comprehension, where the processing complexity increases down the temporal lobe and intelligibility processing occurs in the anterior regions, which unfortunately, are highly susceptible to signal loss in fMRI paradigms.

Chapter 3 described an fMRI study, which was aimed at comparing two methods (distortion corrected spin-echo and dual gradient-echo) designed to overcome signal loss within vATL and orbito-frontal regions. The study used a semantic categorisation paradigm, which is known to activate the vATL using PET (Devlin et al., 2000) and distortion corrected spin-echo fMRI (Visser et al., 2010). The results replicated findings from conventional gradient-echo fMRI protocols that showed severe signal loss in vATL regions (Devlin et al., 2000). Although, spin-echo performed better than conventional gradient-echo, the dual-echo outperformed the spin-echo within the anterior and posterior portion of the vATL and bilateral IFG regions, whilst being equal in the cerebellum.

In Chapter 4, the dual-echo fMRI protocol was used to identify regions related to speech comprehension. The aim was to replicate activity found in speech-specific regions based on previous studies using the same stimuli (Narain et al., 2003; Scott et al., 2000 etc), while observing if activity in the vATL would be detected using the novel fMRI protocol. This is important as the lack of fMRI evidence for anterior temporal activity during comprehension is in contrast with the

results from Chapter 2 as well as converging evidence from PET (Crinion et al., 2003; Scott et al., 2006; Sharp et al., 2004; Spitsyna et al., 2006), MEG (Marinkovic et al., 2003) and neuropsychology (Hodges et al., 1992; Patterson et al., 2007). The dual-echo result showed bilateral STS activation for intelligibility, with the crucial addition of a cluster within the vATL. The vATL was not identified using a contrast for phonological feature processing. This provided evidence for the role of the vATL during speech comprehension, which is likely to engage access of stored semantic representations. It is noted, however, that the STS activity (primarily anterior portions) was much greater than the vATL activity. The dual-echo reduces the signal loss within the vATL region; however it does not eliminate it entirely, which could explain the lower activity within the susceptible region. This study provided further evidence for the sensitivity of the dual-echo method in detecting reliable fMRI signal changes in the vATL. The results from Chapter 3 and 4 suggest that the vATL is involved in amodal semantic processing, as the former used visual word presentations while the latter used auditory sentence stimuli. The results from Chapter 4 suggest that current models of speech comprehension that omit this vATL region need to be updated.

In general, speech comprehension studies have been investigated with neuroimaging techniques in isolation (e.g. fMRI, M/EEG, PET, TMS), and have not focused on combining these techniques to provide a comprehensive model of speech. Chapter 5 reports a study using simultaneous EEG-fMRI recordings to explore the spatial and temporal characteristics of passive auditory word comprehension. The fMRI data highlighted intelligibility related activity within bilateral STS and left IFG. An additional ROI analysis also showed significant involvement of the vATL. The localisation of intelligible speech along the STS is consistent with existing models of speech comprehension, whilst the involvement of the vATL is consistent with models of semantic comprehension based on neuropsychology and PET data. The spatial profile of intelligible speech processing was used to extract EEG time-series in 5 broad ROIs: pSTS, mSTS, aSTS, IFG and vATL. The EEG-derived temporal profile of this data showed a relatively complex picture. The most active time window for intelligibility was between 300-400 ms when 5 out of the 5 nodes were active. This supports the idea that semantic processing has a distinctive electrophysiological signature that occurs at around 400 ms. The anterior STS and vATL node were both strongly activated for intelligibility during this time window, which is consistent with their roles for speech comprehension. Interestingly, the vATL was active throughout the

majority of the trial including the very early time periods. This suggests that an early pass of semantic processing occurs within these regions, which could reflect a crude first pass analysis of the incoming stimuli (Pulvermüller, 2007) . Functional connectivity analysis across the epoch suggested that two of the nodes act like hubs with the wider network. This pattern of connectivity suggests that the mSTS and aSTS are central nodes during speech comprehension. The connectivity profile of each node could provide an insight into the types of processing that occurs in smaller parts of the network. For example, the mSTS is connected to pSTS, aSTS and IFG, which could reflect these regions engaging each other to refine low-level speech information arriving at HG. In contrast, the aSTS is connected with mSTS, IFG and vATL, which could be related to engaging processes specifically related to accessing or selecting semantic content to achieve comprehension. These results provide a novel approach to understanding brain networks of speech comprehension, where the results showed how the spatial network becomes active across time indicating that a two-pass processing framework could explain speech comprehension.

Implications for current models of speech comprehension

Spatial organisation of speech processing

There is a large body of evidence to suggest auditory signals arrive at HG and propagate antero-laterally, with increasing complexity (Binder, 2000; DeWitt & Rauschecker, 2012). The primary auditory regions are not speech specific, as they process all sound types. However, complex acoustic features related to phoneme processing have been found to activate left lateralised mid-STG/STS, lateral to HG (i.e. Leaver & Rauschecker, 2010; Obleser et al., 2007b; Turkeltaub & Coslett, 2010; Vigneau et al., 2006). EEG results from Chapter 2 showed phonological features were localised around left mid-STS and bilateral IFG regions, whereas fMRI results, from Chapter 4, showed bilateral STS activity (with IFG activity apparent at slightly lower thresholds). In general the evidence from the thesis supports the notion that complex acoustic features related to phonological processing are processed laterally to HG, in mid-STG/STS regions. This can be seen to be extending anteriorly, and possibly bilaterally when the listening environment has some added noise.

In addition to activity in mid STG/STS there was also evidence of activity related to low-level complex acoustic features in the IFG. Some studies have also reported IFG activity during phonological processing (Fiez, 1997; Price et al., 1997), and it has been suggested that IFG or primary motor regions could be activated due to articulatory re-coding that helps to disambiguate speech sounds. This seems plausible for the stimuli used in Chapter 2 and 4 as, although the rotated speech does not contain real phonemes, it does contain pitch and intonation variations that are similar.

Overall, these results support the speech models proposed by Price (2010, 2012) and Rauschecker and Scott (2009), as both suggest a role of lateral STG to be involved in speech, specifically for phonological feature processing. In contrast, Hickok & Poeppel (2007), suggest a posterior STS involvement for phonological processing, which is not supported.

One of the central issues in this thesis was to identify the regions related to intelligible speech comprehension. Existing models all agree that intelligibility is processed within the temporal lobe, following projections of the acoustic signal from HG. However, there is poor consensus regarding which specific parts of the temporal lobe are involved. A number of key regions have been put forward: bilateral pMTG/pITS (Hickok & Poeppel, 2007); left aSTS (Rauschecker & Scott, 2009) or left ATL (Price, 2010, 2012). The results from three experiments (Chapter 2, 4, and 5) on speech comprehension implicate bilateral aSTS, left IFG and left vATL regions. This provides supporting evidence to a number of neuroimaging studies that have suggested an antero-lateral gradient of increasing acoustic complexity (Binder et al., 2000; DeWitt & Rauschecker, 2012), from low-level acoustic processing (HG) and phonological processing (lateral mid-STG/STS) to intelligible speech processing (aSTS). Secondly, the aSTS location for intelligible speech processing provides supporting evidence for two models of speech comprehension (Price, 2010, 2012; Rauschecker & Scott, 2009). For example, Price (2010, 2012), suggested the ATL is involved in accessing semantics to understand the meaning of a word or sentence. Furthermore, the temporal poles are involved in accessing increasingly specific semantic associations; words that have similar meanings or when concepts occur together to form discourse level processing. Although, the current work did not directly test the hypothesis that sentences are located anterior to words, in general the peak aSTS activity for sentences (Chapter 4) was found to be anterior to the peak aSTS activity for words (Chapter 5), providing some informal support to such a proposal. The

results provide partial evidence for Rauschecker and Scott (2009), where aSTS activity was predominately seen, however, there was widespread spread bilateral STS and left vATL activity, which is not explicitly suggested in this model. The results obtained in this thesis do not provide evidence for the involvement of pMTG/pITS during lexical retrieval/access (Hickok & Poeppel, 2007). It has recently been shown that activity within the pMTG increases with task demand increases (Jefferies & Lambon Ralph, 2006; Noppeney et al., 2004; Whitney et al., 2011) and when combining semantic information from auditory and visual inputs (Dick et al., 2009; Holle et al., 2010; Kircher et al., 2009a; Kircher et al., 2009b; Robins et al., 2009). This suggests that pMTG could be more sensitive to task demands that are related to semantic control (selection/retrieval) processes and not specifically for access/retrieval of meaning. Finally, the aSTS has been implicated in processing syntax as well as semantic information, which has led researchers to suggest this region (ATL/aMTG) is involved in semantic and syntax combinatorial processes (Friederici, 2011; Hickok & Poeppel, 2007). Although the experiments within this thesis did not directly investigate syntax, activity was found for sentences (Chapter 4) and words (Chapter 2 and Chapter 5) during speech comprehension in anterior temporal regions. Although, this might suggest that the ATL is involved in semantic and syntax processing, the activity for sentences could be a by-product of multiple activations of each word within the sentence. It has been suggested that syntax or combinatorial processes do not occur in anterior temporal regions, but instead occur primarily within inferior frontal gyrus and temporoparietal cortex (Ben-Shachar, Palti, & Grodzinsky, 2004; Caplan, 2001; Friederici & Kotz, 2003). The current thesis provides evidence for the involvement of the aSTS and inferior temporal lobe for intelligible speech (word and sentence) comprehension; however the specific role of the aMTG remains to be determined.

Speech comprehension as part of the semantic network

A recent model of semantic cognition has implicated the ATL for processing amodal semantic representations (Patterson et al., 2007; Rogers et al., 2004). This was originally based on the types of impairments observed in SD patients (multi-modal semantic deficits) after bilateral ATL atrophy (Hodges et al., 1992). Subsequently, converging evidence from neuroimaging methods have found similar contributions of ATL during semantic processing tasks across multiple sensory modalities; TMS (Pobric et al., 2007, 2010), MEG (Marinkovic et al., 2003)

and PET (Crinion et al., 2003; Mummery et al., 2000; Sharp et al., 2004; Spitsyna et al., 2006; Tranel et al., 2005; Vandenberghe et al., 1996). However, in general these methods provide relatively poor spatial resolution so cannot determine precisely which parts of the ATL contribute to semantic processing. There is evidence to suggest that the vATL, specifically, plays a key role in semantic processing. For example, recent fMRI studies have implicated the vATL during amodal semantic processing using a distortion corrected spin-echo protocol to overcome signal loss and image distortion problems prevalent in conventional fMRI studies (Binney et al., 2010; Visser et al., 2010; Visser & Lambon-Ralph, 2011). It has been suggested, therefore, that the vATL regions plays a critical role in semantic cognition, and acts as a computational hub, which maps transformations between sensory modality specific information (e.g. visual or verbal) stored within a distributed network of regions (Patterson et al., 2007). This idea has come to be known as the hub-and-spoke model (Lambon-Ralph et al., 2010).

Overall, the results from the thesis provide supporting evidence for the role of vATL as an amodal semantic hub; activation was found here during auditory speech comprehension (Chapter 2, 4 and 5) and visual semantic categorisation (Chapter 3) tasks. It has been known that semantic categorisation activates the vATL (i.e. Devlin et al., 2000), however, results from the thesis revealed that passive sentence (Chapter 4) and word (Chapter 5) comprehension paradigms also activate the left vATL. It is noted, however, that activations for sentences were generally higher than words. ATL activity for words, in Chapter 5, was not apparent at the level of whole brain analysis, but was confirmed by ROI-level analysis. This could reflect two possibilities: firstly, the computational effort required to integrate words into a coherent sentences is larger than accessing semantic information for a single word. Secondly, in order to comprehend a sentence, the semantic representation of each word need to be identified and this could produce larger vATL activity. The finding that the vATL is active during passive speech comprehension tasks is indicative of its involvement during naturalistic and relatively automatic speech comprehension processes in comparison to explicit comprehension judgement tasks that engage the an extended semantic system (i.e. attention/selection processes), which are likely to include extra regions that can aid but are not necessary for comprehension (i.e. IFG and IPC regions).

There is a debate within the literature about how semantic cognition is organised in the brain, particular in terms of laterality. For example, patients who have undergone unilateral ATL resection for the treatment of temporal lobe epilepsy (TLE) do not show severe impairment on semantic tasks (Simmons & Martin, 2009). In contrast, there is strong evidence from SD patients, who show severe semantic deficits after bilateral atrophy of the ATL. Furthermore, TMS experiments have shown that healthy participants show reduced efficiency during verbal and non-verbal semantic tasks, after being stimulated on the left or right ATL (Lambon Ralph et al., 2009; Pobric et al., 2010). Therefore, it is suggested that bilateral ATL regions contribute to semantic memory. It is argued that this difference of unilateral or bilateral damage to the ATL is a crucial factor in understanding the nature in which semantic representations are stored. It is possible that semantic representations are stored, at least in part, redundantly across the two hemispheres (Marshall, 1984), which could allow the intact ATL to functionally compensate the loss of the damaged/removed ATL. If, however, there is bilateral ATL damage (as in SD), the resulting semantic impairment is severe. In a recent fMRI study, the bilateral vATL regions were found to respond to visual and verbal tasks (Visser & Lambon-Ralph, 2011), however left hemisphere dominance was seen. It has been proposed that this dominance of language on the left temporal lobe could reflect hemispheric asymmetry in the connection between semantic and phonological systems (Lambon-Ralph et al., 2001). According to this model, the phonological system is represented primarily in the language-dominant (usually left) hemisphere and receives more input from the dominant-hemisphere semantic system than from the non-dominant hemisphere. Furthermore, this model suggests that semantic representations are distributed in a semi-redundant manner across two layers corresponding to the left and right temporal lobes, which renders the semantic representations robust to damage. In order to elicit severe semantic impairment, the model requires bilateral damage. This is in-line with recent detailed imaging studies of SD patients that find a left bias of atrophy in ATL regions (Nestor et al., 2006). Due to the passive nature of the auditory comprehension tasks, it is not surprising left lateralised activity was observed, as this is suggested to be the core region for semantic representations. The right ATL might possibly become involved during more computationally demanding tasks. This was not confirmed in Chapter 3, which used an explicit semantic categorisation task, however other studies that have used similar explicit

tasks have reported left dominated bilateral ATL activity (Binney et al., 2010; Visser et al., 2010; Visser & Lambon-Ralph, 2011).

Studies in the thesis do provide evidence for the ventral IFG to be involved in speech comprehension. In particular, Chapter 4 and 5 indicate that the pars orbitalis and pars triangularis are involved during speech comprehension. The overall literature suggests that these regions are not specific for speech comprehension (i.e. accessing meaning); instead they play a role in semantic control or selection during time in which the task of comprehension may be compromised. For example, Obleser and colleagues provide evidence to suggest that there is an increase in functional integration between IFG and AG during degraded speech comprehension, which provides a mechanism of top-down control to aid comprehension (Obleser et al., 2007a). Recent neuroimaging studies have indicated that the IFG plays a role in comprehension, but with an executive control function rather than accessing meaning (Binder et al., 2009; Devlin et al., 2003; Jefferies & Lambon Ralph, 2006; Noonan et al., 2009; Thompson-Schill et al., 1997; Wagner et al., 2001). The three speech comprehension models discussed provide alternative suggestions for the role of the IFG during comprehension. Price suggests that the ventral inferior frontal areas (pars orbitalis and pars triangularis) are involved in selecting and retrieving task related semantic attributes, whilst the dorsal superior frontal gyrus is also involved, albeit less consistently, in constraining semantic processing (Price, 2010, 2012). The need for selection/retrieval processes may be increased during the experiments in this thesis as environmental noise induced by the fMRI could increase the difficulty to comprehend the auditory stimuli. Rauschecker and Scott (2009) propose the IFG to be involved in solving the 'invariance' problem (the ability to detect a target against distortions in scale of time or frequency). This mechanism could be used to aid speech comprehension when stimuli are distorted. Finally, Hickok and Poeppel (2007) suggest a different role of the IFG in speech processing, namely that it is part of a speech production network. The current results provide evidence against a view of the IFG being involved only in the articulatory network. Although one could argue that a motor theory of speech perception would allow the dorsal network to become active during speech comprehension (Galantucci et al., 2006); if this were the case a larger involvement of the motor and supplementary motor area (SMA) would be expected. In contrast to the fMRI results from Chapter 5, the EEG results suggest a weak role for IFG during comprehension (between 300-400ms), with functional connections from the aSTS. This suggests that the IFG is

likely to be involved in speech comprehension, but as a functional aid rather than being necessary for comprehension. Further work would be needed to isolate the specific processes of IFG in relation to speech comprehension.

Finally, although IPC has been implicated in semantic cognition and comprehension studies (Binder et al., 2009), it is not regarded as central to intelligible speech comprehension in many models (Jefferies & Lambon Ralph, 2006; Noonan et al., 2009; Price, 2010; Rauschecker & Scott, 2009). The results from this thesis do not provide evidence for a direct role of IPC during comprehension. Further evidence comes from neuroimaging studies that suggest this region is modulated by working memory demands (Buchsbaum & D'Esposito, 2008; Rauschecker & Scott, 2009) or during the combination of concepts to form a coherent whole (e.g. Price, 2010, 2012). During sentence comprehension, one must combine multiple concepts to form a coherent understanding of the message; therefore, one might expect IPC activation. However, activity within this region was not observed in experiments for sentence comprehension (Chapter 4). The stimuli were simple 'subject-verb-object' sentences, and therefore it is likely that this imposed very small demands in both working memory and in combining the constituent parts to understand the whole. Furthermore, the experiments were conducted using a passive comprehension approach, which again are not likely to require explicit online retention or combination of parts.

Influence of neuroimaging methodology on speech models

The impact of neuroimaging methodology used to investigate speech comprehension and semantic cognition cannot be overlooked. The omission of the vATL in speech comprehension models could largely be down to the dominance of inadequate fMRI methods that have signal loss in this critical region (Visser et al., 2009). Recent advancements have shown distortion corrected spin-echo fMRI to be useful in overcoming this problem and detecting activity within vATL during semantic tasks (Binney et al., 2010; Embleton et al., 2010; Visser et al., 2010; Visser & Lambon-Ralph, 2011). Results from Chapter 3 suggest that although spin-echo is better than conventional fMRI protocols, dual gradient-echo is better at detecting BOLD changes within vATL and extra temporal regions. This provided supporting evidence for studies that have identified a superior SNR for gradient-echo protocols compared to spin-echo (Bandettini et al., 1994; Jezzard et al., 2001; Norris et al., 2002; Schmidt et al., 2005). Furthermore, the dual-echo

technique provides a fairly simple solution to the signal loss problem, which has affected previous fMRI studies (Poser & Norris, 2009; Poser et al., 2006). Evidence from Chapter 4 and 5 support the idea that the vATL is involved in speech comprehension for both sentences and words. These experiments highlight the importance of using an fMRI method capable of detecting signal changes within these critical regions, otherwise they will inevitably be omitted in models of speech comprehension (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009).

In addition to a sensitive fMRI protocol, two factors can influence the likelihood of detecting ATL activity (Visser et al., 2009): 1) baseline contrast; 2) field of view (FOV). It has been suggested that a low-level baseline (i.e. rest) can allow for subjects to daydream during conditions in which no comprehension is required. This can lead to activity related to comprehension being subtracted at the contrast level (Binder et al., 1999; Binder et al., 2011). Therefore, it is generally recommended to use a high-level non-semantic baseline that attentionally engages the subject (i.e. mental arithmetic) (Binder et al., 2011). The FOV should be kept large enough to provide full coverage of all anterior and ventral temporal regions during imaging experiments investigating vATL. It was found that studies using a FOV less than 15 cm were less likely to detect ATL activity (Visser et al., 2009), therefore experiments in the current thesis used a large FOV of 24 cm centred along the AC-PC line, to be certain of capturing the whole brain. Future studies investigating the vATL should be aware of these technical limitations of brain imaging studies, and the solutions available to provide a platform to overcome them.

The result from Chapters 3, 4 and 5, provides evidence for the reliability of dual-echo fMRI to detect activity within the vATL. The tasks employed in the experiments were visual and verbal, which suggest that this region represents amodal semantic information (Lambon-Ralph et al., 2010; Patterson et al., 2007; Rogers et al., 2004). It is proposed that future studies interested in probing the specific role of vATL (and orbito-frontal) regions should employ a dual-echo fMRI protocol to enhance the chances of detecting BOLD changes, which can help refine our understanding of the language network and speech comprehension (i.e. Hickok & Poeppel, 2007; Price, 2010, 2012; Rauschecker & Scott, 2009).

Temporal evolution of speech processing

It is well known that fMRI provides an indirect measure of neural activity via changes in the levels of (de)oxygenated blood. This hemodynamic response is known to lag (order of seconds) behind the electrical neural activity, which leads to the poor temporal resolution of fMRI. However, M/EEG provides a direct measure of magnetic/electric neural properties from which neuronal firing occur, which can be sampled in the millisecond range. This high temporal resolution has been used to relate the physiological changes to properties of the stimulus (or task), thus providing a means to understand how brain processes unfold over time. There are two broad frameworks developed to explain electro/magneto neurophysiological data on language studies. They fall within a hierarchical (Friederici, 2002) or parallel (Marslen-Wilson, 1987) framework, however recently a combination of the two has been suggested with two-pass parallel processing stages (Pulvermüller, 2007). The results from the EEG data within this thesis are quite complex. In Chapter 2, for normal word processing the evidence points to a relatively simple serial mechanism with low-level complex acoustic feature processing occurring early and semantic processing occurring later. However, when the stimuli are distorted there is evidence for top-down processing which provides executive control or selection aid to the distorted stimuli that have some similarity to known words than distorted pseudoword stimuli. There were no differences between low-level phonological features in the late window, which a parallel processing framework would predict; therefore a two-pass framework best suited the pattern of results. During the early time window (~150 ms), the incoming signal is assessed for general speech-like qualities, such as phonological features, however if the signal is difficult to process additional top-down executive control or selection constraints may be imposed by prefrontal regions (Thompson-Schill et al., 1997). This early component was localised around the primary auditory cortex (PAC) and frontal regions, which is in general agreement with models proposing phonological processing in lateral mid-STG/STS regions (Price, 2010, 2012; Rauschecker & Scott, 2009). Furthermore, a two-pass framework can account for studies that have shown semantic and syntactic effects during early time windows (~100-200 ms) (Friederici & Kotz, 2003; Neville et al., 1991; Pulvermüller et al., 1995; Rugg, 1983). It could be argued that the current study did not have enough statistical power to observe similar semantic differences (using a direct comparison of words with pseudowords), although there were no statistical trends

from 20 subjects. During the second later window (~400 ms), it is suggested that detailed information from the signal is mapped onto stored semantic representations as this showed differences between intelligible and unintelligible only (Chapter 2 and 5). There has been a considerable and sustained effort focused on researching the N400. It has been suggested that the N400 is related to an attempt to access semantic representations (Kutas & Federmeier, 2011). It is likely that the N400 is not a direct measure of semantic processing, as it has been shown to be modulated by other factors (i.e. orthographic neighbourhood, Laszlo & Federmeier, 2011). Instead, it might be a marker of attempted semantic access (regardless of whether this attempt is successful). Based on these psycholinguistic properties, it is proposed that N400 should in part be localised within regions interested in processing semantic information, such as ATL. The results from Chapter 2 did show the N400 component to be localised, anterior to the early component, around ATL regions, with some additional frontal activity. This provides converging evidence for the role of the ATL for semantic access during speech comprehension. Furthermore, the results provide evidence for a somewhat hierarchical spatial projection within the temporal lobe (Binder et al., 2000; DeWitt & Rauschecker, 2012).

The EEG-fMRI study has the potential to provide a novel insight into the spatial and temporal processing of spoken word comprehension (Chapter 5). The intelligibility contrast was the main focus of Chapter 5, due to the expected outcome that the fMRI contrast for intelligibility showed vATL, in addition to STS, activity, whereas the contrast for phonological features did not. The primary aim was to determine when these intelligibility-sensitive speech regions (mSTS, pSTS, aSTS, IFG and vATL) become involved during speech comprehension. The majority of the activity occurred between 300-400 and 400-500 ms. These time windows are consistent with suggestions of a semantically related N400 component. The intelligibility effect was located mainly in the vATL and aSTS regions during this time window, whilst weaker effects were also seen in pSTS, mSTS and IFG. These results provide a consistent picture of semantic processing both in time and space, where the semantic effect is seen around 400ms (Kutas & Federmeier, 2011), localised mainly within aSTS (Price, 2010, 2012; Rauschecker & Scott, 2009) and vATL regions (Lambon-Ralph et al., 2010; Patterson et al., 2007). The weaker effects from the other speech intelligible nodes (pSTS, mSTS and IFG) could reflect processing that is not necessary for comprehension per se but is related to semantic control or predictive constraints imposed by a system

that is aiding comprehension (Jefferies & Lambon Ralph, 2006; Price, 2010, 2012; Rauschecker & Scott, 2009). Speculative evidence for this is provided by the involvement of pSTS, mSTS and IFG during the 300-400ms window, with a sustained effect of pSTS and mSTS during the 400-500ms window, which could be imposing semantic selection biases on the semantic representation region (vATL). In addition, the functional connectivity profiles of the nodes provided interesting and complementary evidence. For example, the pSTS is functionally correlated with vATL, which could suggest information being shared between these two regions. As the pSTS has been implicated in semantic control processing, and the vATL during semantic knowledge, these two regions could be engaging each other to achieve comprehension by applying executive control to identify specific semantic representations. Interestingly, there was a connectivity profile involving the anterior fronto-temporal nodes (aSTS, vATL, IFG and mSTS). The aSTS has long been implicated in intelligible speech comprehension, while vATL is related to semantic processing; it is likely that these two regions are engaged in mutual information processing to achieve comprehension. The IFG connection with aSTS may reflect the engagement of semantic control processes, which can aid successful comprehension. Finally, the connection between mSTS and aSTS could be linked to the antero-lateral projection of primary auditory processing regions to higher speech specific regions. Taken together these results suggest the anterior fronto-temporal network is important for successful comprehension.

The vATL was shown to be active consistently across the epoch, including during the early time windows (<200ms). This early involvement of the vATL suggests some influence of higher-level semantic information, in support of a parallel or two pass frameworks (Marslen-Wilson, 1987; Pulvermüller, 2007). The two-pass framework may fit the appearance of the vATL better as the activation is sustained, which suggests that the region is reinterpreting the accessed semantic information and repairing errors into late time windows. In contrast, the parallel approach would expect phonological related processing to occur through to late time windows as all features are continually processed. Although, Chapter 5 did not specifically investigate phonological feature processing, Chapter 2 did not find any evidence for an effect of phonological features in the late window, which does not support a parallel framework. Evidence for an early semantic effect within temporal fusiform cortex has been seen in an intracranial EEG study, where the posterior portion of the vATL was active during visual stimulation (e.g. less than

200ms, Liu et al., 2009). This area is also highlighted within models of higher visual processing (Deco & Rolls, 2004; Felleman & Van Essen, 1991; Rolls, 1991), which describe a hierarchy of processing stages that proceed anteriorly along the ventral temporal lobe, with representations becoming increasingly complex and invariant around the point of the ventral ATL (Halgren et al., 2006). However, to the knowledge of the author, there are no such invasive studies presenting auditory speech stimuli to determine how early parts of the FG are active during comprehension.

Finally, converging evidence for the role of vATL in semantic access was obtained using frequency analysis of the EEG sensor data (Chapter 5). The results showed a significant effect in the theta band for intelligibility occurring between 0.45 – 0.65 s. This was suggested to be related to a N400-type component of semantic processing (Kutas & Federmeier, 2011). It is noted that this time window occurs later than the observed effect in the temporal domain (~300-400ms), which could partially be explained by the fact that theta oscillations would not be adequately captured within a 100ms time bin (theta oscillations need approximately ~140-250ms window size). One may expect any effect in theta to span across a larger time window. There were no differences in any other frequency band, although an absence of effects due to poor SNR (i.e. residual MR related artefacts) cannot be ruled out. Previous studies have implicated the theta band with lexical retrieval (Bastiaansen & Hagoort, 2006; Bastiaansen et al., 2005). It must be noted, however, that synchronisation in theta oscillations are not necessarily the sole mechanism by which lexical retrieval is achieved. Importantly, the source estimation of this theta effect was found within left vATL and IFG. This converges with existing evidence of vATL involvement in semantic processing (Patterson et al., 2007; Rogers et al., 2004; Visser & Lambon-Ralph, 2011) and it reinforces the finding that this region is key for speech comprehension (Chapter 2, 4 and 5).

Limitations

A key goal for speech comprehension research, and indeed investigations into all brain processes, is to identify both spatial and temporal processing principles. Recent efforts have been made within a speech comprehension domain (Bornkessel-Schlesewsky & Schlewsky, 2013; Friederici, 2011). However, they draw largely on spatial and temporal data that is collected in isolation. A combined

EEG-fMRI approach is likely to play a critical part in achieving the goal of providing a more complete model, as these techniques provide complementary high spatial (fMRI) and temporal (EEG) resolution, as well as eliminating subject and session effects. This method is not without its limitations, as the EEG signal suffers from MR related artefacts, which lead to poor SNR sensitivity on the EEG data. Careful preparation and design implementation can help achieve good EEG signal quality (i.e. Mullinger et al., 2008c; Mullinger et al., 2011), as well as time spent applying post-hoc correction algorithms to clean the EEG signal (Allen et al., 2000; Mullinger et al., 2008c; Niazy et al., 2005). It has been suggested to collect EEG and fMRI data in separate sessions to avoid contaminated EEG data. However this design comes with other problems such as poorly controlled session effects (attention, motivation and environment) as well as practice and repetition effects of the task. Despite the relatively poor SNR for the EEG data, future EEG-fMRI studies should be used with larger sample sizes to further develop our understanding of brain processes as part of a larger brain network. Functional connectivity analyses are key in pursuing such a goal, where both functional and effective connectivity measures could provide detailed insights into speech that are currently unavailable. EEG is suited to study brain dynamics at the neural level, while the poor resolution of fMRI data is not ideal. A recent study comparing connectivity measures using simulated fMRI data, suggests that in theory the connectivity measures can perform well, but when these measures are applied to real fMRI data there are many problems and performance of the measures are generally poor (Smith et al., 2011). This is not to say fMRI temporal dynamics should not be studied, as there have been attempts to understand the dynamics of speech using fMRI (e.g. Leff et al., 2008); however, some amount of caution should be used when interpreting the conclusions. Future studies into the connectivity profiles of speech comprehension should focus on using EEG-fMRI to provide detailed spatial and temporal dynamic information.

Future directions

The current thesis provides evidence to refine the current models of speech comprehension to include spatial and temporal characteristics; however there are some issues that would need to be addressed in future iterations of speech models. Having identified the vATL region as critical to speech comprehension, it is important for future fMRI studies of semantic processing to use an fMRI protocol

suitable to detect signal changes within vATL regions. It is important to understand how the surrounding ATL regions are involved during semantic and speech processing. For example, early neuropsychology studies of SD patients showed that atrophy in large parts of the ATL produces semantic impairments, whereas recent fMRI studies have suggested that the vATL is the critical locus of semantic processing. If the aSTS and vATL are collaborating to achieve comprehension, how is the intermediate aMTG region involved? Furthermore, the information processing from HG to antero-lateral temporal region appears to follow a graded specialisation, therefore what functions do the MTG and aMTG carry out if they are part of this gradient?

There have been attempts to distinguish how different parts of the IFG (pars opercularis, pars triangular and pars orbitalis) are involved during speech comprehension (Badre & D'Esposito, 2009; Hickok & Poeppel, 2007; Noonan et al., 2013; Price, 2010, 2012; Rauschecker & Scott, 2009). However, future work would need to improve the spatial accuracy to specific sub-regions of the IFG, rather than providing a general anatomical label of IFG.

The brain network of speech comprehension was explored within the high temporal resolution of EEG. Further work on brain networks would provide a novel addition to the current understanding of speech comprehension. Some attempts were made in the current work, whereby temporal sequencing and functional connectivity analyses were carried out. An important addition would be to provide effective connectivity results between each node in the network, to provide a unique understanding of information flow. Such analyses using fMRI data would not capture the high temporal dynamics of brain processes, and so a combined EEG-fMRI approach would be important to investigate brain networks. Finally, the current thesis found theta to be involved during intelligible speech comprehension; however the exact role of theta is not clear. For example, does difficulty affect the role of theta during comprehension, if so it is a local change (increase in power) or global change (recruit larger brain networks)? Further research is needed to better understand the role of theta for speech comprehension and semantic knowledge in general.

Conclusions

The research in this thesis has provided an insight into the spatial and temporal organisation of speech comprehension, by making use of EEG and fMRI

methods. The conclusions reached are critically dependant on two methodological considerations; 1) using an fMRI method capable of detecting BOLD changes within inferior temporal regions and 2) combining simultaneous EEG recordings with fMRI. The first issue has a direct effect on the current models of speech comprehension, whereby the inferior temporal regions are regularly omitted (Hickok & Poeppel, 2007; Price, 2010, 2012; Rauschecker & Scott, 2009). Therefore, in an attempt to redress the imbalance of evidence within the fMRI literature, the thesis provides evidence to propose that a dual-echo fMRI should be used for future studies investigating semantic processing and language comprehension. This method can detect BOLD changes within inferior temporal regions and preserves a high SNR across the whole brain.

The findings from this research suggest that speech comprehension is related to a network of brain regions, primarily located in the temporal and frontal lobes. Comprehension is largely supported by aSTS regions, but also requires vATL to access semantic representations, while pSTS and IFG regions provide executive semantic control processes. Current models of speech comprehension omit the vATL (Hickok & Poeppel, 2007; Price, 2010, 2012; Rauschecker & Scott, 2009), and it is proposed that that the vATL is important part of the speech comprehension network which needs to be included in future iterations. Spoken word comprehension proceeds in a somewhat hierarchical manner; however higher-order processes such as executive control or selection are involved to aid distorted signals during successful comprehension, therefore a two-pass parallel process is likely to best explain the results. We propose a model of speech comprehension in which phonological processing is observed within mid-STG regions antero-lateral to HG (Price, 2010, 2012; Rauschecker & Scott, 2009). Critically, however, we propose that intelligible speech comprehension is located in aSTS and vATL regions. Furthermore, we propose that the temporal evolution for speech comprehension occurs within a two pass processing framework. In general, low-level phonological features are processed within early time components; however top-down constraints from high-level semantic or executive control processes are imposed to identify stimuli that resemble speech to aid comprehension. The late time window (~400 ms) is likely to reflect processing of semantic representations in a similar way to the N400 effect. It might be wrong to think of the N400 as a direct measure of semantic processing. However, it could be regarded as in inverse measure of how easy it is to reject the possibility that an item may have a semantic representation. The N400-type response was observed

in regions known to process intelligible speech (aSTS and IFG), which provides converging evidence from EEG and fMRI that this particular time and space are critical during comprehension.

Interest in time frequency analysis has gathered momentum in the past decade or so and novel proposals about how oscillatory activity could potentially provide insights into mechanisms of how information is communicated across the brain has reinvigorated interest in analysis of the frequency domain (Varela, Lachaux, Rodriguez, & Martinerie, 2001). A recent review of how frequency bands are related to different language processes has implicated theta, alpha, beta and gamma bands (reviewed in Bastiaansen & Hagoort, 2006). The current thesis found theta to be involved during intelligible speech comprehension (Chapter 5) between 450-650 ms, within the left vATL. This provides converging evidence for the importance of vATL during speech and semantic processing, as well as its role during the critical window around 400 ms, which is implicated in semantic access/retrieval. It is proposed that the speech comprehension network may rely on low frequency theta oscillations during the act of comprehension, which centres primarily on vATL regions.

References

- Adank, P. (2012a). Design choices in imaging speech comprehension: An Activation Likelihood Estimation (ALE) meta-analysis. *NeuroImage*, 63(3), 1601-1613. doi: <http://dx.doi.org/10.1016/j.neuroimage.2012.07.027>
- Adank, P. (2012b). The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. *Brain and Language*, 122(1), 42-54. doi: <http://dx.doi.org/10.1016/j.bandl.2012.04.014>
- Allen, P. J., Josephs, O., & Turner, R. (2000). A Method for Removing Imaging Artifact from Continuous EEG Recorded during Functional MRI. *NeuroImage*, 12(2), 230-239. doi: <http://dx.doi.org/10.1006/nimg.2000.0599>
- Allen, P. J., Polizzi, G., Krakow, K., Fish, D. R., & Lemieux, L. (1998). Identification of EEG Events in the MR Scanner: The Problem of Pulse Artifact and a Method for Its Subtraction. *NeuroImage*, 8(3), 229-239. doi: <http://dx.doi.org/10.1006/nimg.1998.0361>
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1), 95-113. doi: <http://dx.doi.org/10.1016/j.neuroimage.2007.07.007>
- Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2007). A Common System for the Comprehension and Production of Narrative Speech. *The Journal of Neuroscience*, 27(43), 11455-11464. doi: <http://dx.doi.org/10.1523/jneurosci.5257-06.2007>
- Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci*, 10(9), 659-669. doi: http://www.nature.com/nrn/journal/v10/n9/supinfo/nrn2667_S1.html
- Badre, D., & Wagner, A. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, 45(13), 2883-2901. doi: <http://dx.doi.org/10.1016/j.neuropsychologia.2007.06.015>
- Ball, T., Schreiber, A., Feige, B., Wagner, M., Lücking, C. H., & Kristeva-Feige, R. (1999). The Role of Higher-Order Motor Areas in Voluntary Movement as Revealed by High-Resolution EEG and fMRI. *NeuroImage*, 10(6), 682-694. doi: <http://dx.doi.org/10.1006/nimg.1999.0507>
- Bandettini, P. A., Wong, E. C., Jesmanowicz, A., Hinks, R. S., & Hyde, J. S. (1994). Spin-echo and gradient-echo epi of human brain activation using bold contrast: A comparative study at 1.5 T. *NMR in Biomedicine*, 7(1-2), 12-20. doi: <http://dx.doi.org/10.1002/nbm.1940070104>
- Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, 53(1), 98-123. doi: <http://dx.doi.org/10.1016/j.brainresrev.2006.07.002>
- Barrett, S. E., & Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. *Neuropsychologia*, 27(7), 913-922. doi: [http://dx.doi.org/10.1016/0028-3932\(89\)90067-5](http://dx.doi.org/10.1016/0028-3932(89)90067-5)
- Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, 159, 179-196.
- Bastiaansen, M., van Berkum, J. J. A., & Hagoort, P. (2002). Event-related theta power increases in the human EEG during online sentence processing. *Neuroscience Letters*, 323(1), 13-16. doi: [http://dx.doi.org/10.1016/S0304-3940\(01\)02535-6](http://dx.doi.org/10.1016/S0304-3940(01)02535-6)
- Bastiaansen, M., Van der Linden, M., ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta Responses Are Involved in Lexical—Semantic Retrieval during Language Processing. *Journal of Cognitive Neuroscience*, 17(3), 530-541. doi: <http://dx.doi.org/10.1162/0898929053279469>
- Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience*, 6(5), 448-450. doi: <http://dx.doi.org/10.1038/nn1050>
- Bedny, M., Caramazza, A., Grossman, E., Pascual-Leone, A., & Saxe, R. (2008). Concepts Are More than Percepts: The Case of Action Verbs. *The Journal of Neuroscience*, 28(44), 11347-11353. doi: <http://dx.doi.org/10.1523/jneurosci.3039-08.2008>

- Bedny, M., McGill, M., & Thompson-Schill, S. L. (2008). Semantic Adaptation and Competition during Word Comprehension. *Cerebral Cortex*, *18*(11), 2574-2585. doi: <http://dx.doi.org/10.1093/cercor/bhn018>
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-312. doi: <http://dx.doi.org/10.1038/35002078>
- Bellgowan, P. S. F., Bandettini, P. A., van Gelderen, P., Martin, A., & Bodurka, J. (2006). Improved BOLD detection in the medial temporal region using parallel imaging and voxel volume reduction. *NeuroImage*, *29*(4), 1244-1251. doi: <http://dx.doi.org/10.1016/j.neuroimage.2005.08.042>
- Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *NeuroImage*, *21*(4), 1320-1336. doi: <http://dx.doi.org/10.1016/j.neuroimage.2003.11.027>
- Benson, R. R., Richardson, M., Whalen, D. H., & Lai, S. (2006). Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *NeuroImage*, *31*(1), 342-353. doi: <http://dx.doi.org/10.1016/j.neuroimage.2005.11.029>
- Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, *30*(2), 161-169. doi: <http://dx.doi.org/10.1111/j.1469-8986.1993.tb01729.x>
- Binder, J. R. (2000). The new neuroanatomy of speech perception. *Brain*, *123*(12), 2371-2372. doi: <http://dx.doi.org/10.1093/brain/123.12.2371>
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex*, *19*(12), 2767-2796. doi: <http://dx.doi.org/10.1093/cercor/bhp055>
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., & Cox, R. W. (1999). Conceptual Processing during the Conscious Resting State: A Functional MRI Study. *Journal of Cognitive Neuroscience*, *11*(1), 80-93. doi: <http://dx.doi.org/10.1162/089892999563265>
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, *10*(5), 512-528. doi: <http://dx.doi.org/10.1093/cercor/10.5.512>
- Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, *17*(1), 353-362.
- Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., . . . Weaver, K. E. (2011). Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study. *NeuroImage*, *54*(2), 1465-1475. doi: <http://dx.doi.org/10.1016/j.neuroimage.2010.09.048>
- Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J. M., & Lambon-Ralph, M. A. (2010). The Ventral and Inferolateral Aspects of the Anterior Temporal Lobe Are Crucial in Semantic Memory: Evidence from a Novel Direct Comparison of Distortion-Corrected fMRI, rTMS, and Semantic Dementia. *Cerebral Cortex*, *20*(11), 2728-2738. doi: <http://dx.doi.org/10.1093/cercor/bhq019>
- Binney, R. J., Parker, G. J. M., & Lambon-Ralph, M. A. (2012). Convergent Connectivity and Graded Specialization in the Rostral Human Temporal Lobe as Revealed by Diffusion-Weighted Imaging Probabilistic Tractography. *Journal of Cognitive Neuroscience*, *24*(10), 1998-2014. doi: http://dx.doi.org/10.1162/jocn_a_00263
- Bitterman, Y., Mukamel, R., Malach, R., Fried, I., & Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature*, *451*(7175), 197-201. doi: <http://dx.doi.org/10.1038/nature06476>
- Blessner, B. (1972). Speech Perception Under Conditions of Spectral Transformation: I. Phonetic Characteristics. *Journal of Speech and Hearing Research*, *15*(1), 5-41.
- Bobes, M. A., Valdessa, M., & Olivares, E. (1994). An ERP Study of Expectancy Violation in Face Perception. *Brain and Cognition*, *26*(1), 1-22. doi: <http://dx.doi.org/10.1006/brcg.1994.1039>

- Bogen, J. E., & Bogen, G. M. (1976). Wernicke's region – where is it? *ANNALS OF THE NEW YORK ACADEMY OF SCIENCES*, 280, 834-843.
- Bonmassar, G., Purdon, P. L., Jääskeläinen, I. P., Chiappa, K., Solo, V., Brown, E. N., & Belliveau, J. W. (2002). Motion and Ballistocardiogram Artifact Removal for Interleaved Recording of EEG and EPs during MRI. *NeuroImage*, 16(4), 1127-1141. doi: <http://dx.doi.org/10.1006/nimg.2002.1125>
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal–ventral stream model of sentence comprehension. *Brain and Language*, 125(1), 60-76. doi: <http://dx.doi.org/10.1016/j.bandl.2013.01.010>
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787-821.
- Boxerman, J. L., Hamberg, L. M., Rosen, B. R., & Weisskoff, R. M. (1995). MR contrast due to intravascular magnetic susceptibility perturbations. *Magnetic Resonance in Medicine*, 34(4), 555-566.
- Bozeat, S., Lambon Ralph, M. A., Graham, K. S., Patterson, K., Wilkin, H., Rowland, J., . . . Hodges, J. R. (2003). A duck with four legs: Investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cognitive Neuropsychology*, 20(1), 27-47. doi: <http://dx.doi.org/10.1080/02643290244000176>
- Bozeat, S., Ralph, M. L., Patterson, K., & Hodges, J. (2002). When objects lose their meaning: What happens to their use? *Cognitive, Affective, & Behavioral Neuroscience*, 2(3), 236-251. doi: <http://dx.doi.org/10.3758/CABN.2.3.236>
- Bozic, M., Tyler, L. K., Ives, D. T., Randall, B., & Marslen-Wilson, W. D. (2010). Bihemispheric foundations for human speech comprehension. *Proceedings of the National Academy of Sciences*. doi: <http://dx.doi.org/10.1073/pnas.1000531107>
- Broca, P. (1866). Remarques sur le siege de la faculte du langage articule suivies d'une observation d'aphemie. *Bulletins Société anatomique de Paris*, 6, 330-357.
- Buchsbaum, B. R., & D'Esposito, M. (2008). The Search for the Phonological Store: From Loop to Convolution. *Journal of Cognitive Neuroscience*, 20(5), 762-778. doi: <http://dx.doi.org/10.1162/jocn.2008.20501>
- Burgess, A. P., & Ali, L. (2002). Functional connectivity of gamma EEG activity is modulated at low frequency during conscious recollection. *International Journal of Psychophysiology*, 46(2), 91-100. doi: [http://dx.doi.org/10.1016/S0167-8760\(02\)00108-3](http://dx.doi.org/10.1016/S0167-8760(02)00108-3)
- Caplan, D. (2001). Functional Neuroimaging Studies of Syntactic Processing. *Journal of Psycholinguistic Research*, 30(3), 297-320. doi: <http://dx.doi.org/10.1023/A:1010495018484>
- Chia, J. M., Fischer, S. E., Wickline, S. A., & Lorenz, C. H. (2000). Performance of QRS detection for cardiac magnetic resonance imaging with a novel vectorcardiographic triggering method. *Journal of Magnetic Resonance Imaging*, 12(5), 678-688. doi: [http://dx.doi.org/10.1002/1522-2586\(200011\)12:5<678::aid-jmri4>3.0.co;2-5](http://dx.doi.org/10.1002/1522-2586(200011)12:5<678::aid-jmri4>3.0.co;2-5)
- Cho, Z. H., & Ro, Y. M. (1992). Reduction of susceptibility artifact in gradient-echo imaging. *Magnetic Resonance in Medicine*, 23(1), 193-200. doi: <http://dx.doi.org/10.1002/mrm.1910230120>
- Coles, M. G. H., & Rugg, M. D. (1995). *Electrophysiology of mind: event-related brain potentials and cognition*. Oxford: Oxford University Press.
- Constable, R. T., & Spencer, D. D. (1999). Composite image formation in z-shimmed functional MR imaging. *Magnetic Resonance in Medicine*, 42(1), 110-117. doi: [http://dx.doi.org/10.1002/\(sici\)1522-2594\(199907\)42:1<110::aid-mrm15>3.0.co;2-3](http://dx.doi.org/10.1002/(sici)1522-2594(199907)42:1<110::aid-mrm15>3.0.co;2-3)
- Corbett, F., Jefferies, E., Ehsan, S., & Lambon Ralph, M. A. (2009). Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. *Brain*, 132(9), 2593-2608. doi: <http://dx.doi.org/10.1093/brain/awp146>
- Cordes, D., Turski, P. A., & Sorenson, J. A. (2000). Compensation of susceptibility-induced signal loss in echo-planar imaging for functional applications. *Magnetic Resonance Imaging*, 18(9), 1055-1068.

- Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. S. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, 126(5), 1193-1201. doi: <http://dx.doi.org/10.1093/brain/awq104>
- Cusack, R., Russell, B., Cox, S. M. L., De Panfilis, C., Schwarzbauer, C., & Ansorge, R. (2005). An evaluation of the use of passive shimming to improve frontal sensitivity in fMRI. *NeuroImage*, 24(1), 82-91.
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does Semantic Context Benefit Speech Understanding through "Top-Down" Processes? Evidence from Time-resolved Sparse fMRI. *Journal of Cognitive Neuroscience*, 23(12), 3914-3932. doi: http://dx.doi.org/10.1162/jocn_a.00084
- Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3773-3800. doi: <http://dx.doi.org/10.1098/rstb.2009.0111>
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *The Journal of Neuroscience*, 23(8), 3423-3431.
- De Panfilis, C., & Schwarzbauer, C. (2005). Positive or negative blips? The effect of phase encoding scheme on susceptibility-induced signal losses in EPI. *NeuroImage*, 25(1), 112-121.
- Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, 41(1), 60-74. doi: <http://dx.doi.org/10.1111/1469-8986.00120>
- Debener, S., Mullinger, K. J., Niazy, R. K., & Bowtell, R. W. (2008). Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 T static magnetic field strength. *International Journal of Psychophysiology*, 67(3), 189-199. doi: <http://dx.doi.org/10.1016/j.ijpsycho.2007.05.015>
- Debener, S., Strobel, A., Sorger, B., Peters, J., Kranczioch, C., Engel, A. K., & Goebel, R. (2007). Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: Removal of the ballistocardiogram artefact. *NeuroImage*, 34(2), 587-597. doi: <http://dx.doi.org/10.1016/j.neuroimage.2006.09.031>
- Debruille, J. B. (2007). The N400 potential could index a semantic inhibition. *Brain Research Reviews*, 56(2), 472-477. doi: <http://dx.doi.org/10.1016/j.brainresrev.2007.10.001>
- Deco, G., & Rolls, E. T. (2004). A Neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, 44(6), 621-642. doi: <http://dx.doi.org/10.1016/j.visres.2003.09.037>
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8(4), 919-924.
- Deichmann, R., Gottfried, J. A., Hutton, C., & Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. *NeuroImage*, 19(2), 430-441.
- Deichmann, R., Josephs, O., Hutton, C., Corfield, D. R., & Turner, R. (2002). Compensation of Susceptibility-Induced BOLD Sensitivity Losses in Echo-Planar fMRI Imaging. *NeuroImage*, 15(1), 120-135.
- Devlin, J. T., Matthews, P. M., & Rushworth, M. F. S. (2003). Semantic Processing in the Left Inferior Prefrontal Cortex: A Combined Functional Magnetic Resonance Imaging and Transcranial Magnetic Stimulation Study. *Journal of Cognitive Neuroscience*, 15(1), 71-84. doi: <http://dx.doi.org/10.1162/089892903321107837>
- Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., . . . Tyler, L. K. (2000). Susceptibility-Induced Loss of Signal: Comparing PET and fMRI on a Semantic Task. *NeuroImage*, 11, 589-600.
- DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*. doi: <http://dx.doi.org/10.1073/pnas.1113427109>
- Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., & Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*, 30(11), 3509-3526. doi: <http://dx.doi.org/10.1002/hbm.20774>

- Dobbins, I. G., & Wagner, A. D. (2005). Domain-general and Domain-sensitive Prefrontal Mechanisms for Recollecting Events and Detecting Novelty. *Cerebral Cortex*, 15(11), 1768-1778. doi: <http://dx.doi.org/10.1093/cercor/bhi054>
- Dronkers, N. F., Wilkins, D. P., Van Valin Jr, R. D., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1-2), 145-177. doi: <http://dx.doi.org/10.1016/j.cognition.2003.11.002>
- Embleton, K., Haroon, H., Morris, D., Lambon-Ralph, M. A., & Parker, G. (2010). Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Human Brain Mapping*, 31(10), 1570-1587.
- Federmeier, K. D., & Laszlo, S. (2009). Time for Meaning: Electrophysiology Provides Insights into the Dynamics of Representation and Processing in Semantic Memory. In H. R. Brian (Ed.), *The Psychology of Learning and Motivation* (Vol. 51, pp. 1-44). San Diego, CA: Academic Press.
- Fell, J., Klaver, P., Lehnertz, K., Grunwald, T., Schaller, C., Elger, C. E., & Fernandez, G. (2001). Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nat Neurosci*, 4(12), 1259-1264.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1), 1-47. doi: <http://dx.doi.org/10.1093/cercor/1.1.1>
- Fiez, J. A. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. *Human Brain Mapping*, 5(2), 79-83. doi: [http://dx.doi.org/10.1002/\(SICI\)1097-0193\(1997\)5:2<79::AID-HBM1>3.0.CO;2-J](http://dx.doi.org/10.1002/(SICI)1097-0193(1997)5:2<79::AID-HBM1>3.0.CO;2-J)
- Fischer, S. E., Wickline, S. A., & Lorenz, C. H. (1999). Novel real-time R-wave detection algorithm based on the vectorcardiogram for accurate gated magnetic resonance acquisitions. *Magnetic Resonance in Medicine*, 42(2), 361-370. doi: [http://dx.doi.org/10.1002/\(sici\)1522-2594\(199908\)42:2<361::aid-mrm18>3.0.co;2-9](http://dx.doi.org/10.1002/(sici)1522-2594(199908)42:2<361::aid-mrm18>3.0.co;2-9)
- Frahm, J., Merboldt, K.-D., & Hänicke, W. (1988). Direct FLASH MR imaging of magnetic field inhomogeneities by gradient compensation. *Magnetic Resonance in Medicine*, 6(4), 474-480. doi: <http://dx.doi.org/10.1002/mrm.1910060412>
- Freyer, F., Becker, R., Anami, K., Curio, G., Villringer, A., & Ritter, P. (2009). Ultrahigh-frequency EEG during fMRI: Pushing the limits of imaging-artifact correction. *NeuroImage*, 48(1), 94-108. doi: <http://dx.doi.org/10.1016/j.neuroimage.2009.06.022>
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78-84.
- Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, 91(4), 1357-1392. doi: <http://dx.doi.org/10.1152/physrev.00006.2011>
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7), 2458-2463. doi: <http://dx.doi.org/10.1073/pnas.0509389103>
- Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *NeuroImage*, 20, S8-S17.
- Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory Language Comprehension: An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information. *Brain and Language*, 74(2), 289-300. doi: <http://dx.doi.org/10.1006/brln.2000.2313>
- Friston, K. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, 2(1-2), 56-78.
- Friston, K. (2003). Functional Integration. In R. S. J. Frackowiak, K. Friston, C. Firth, R. Dolan, C. J. Price, S. Zeki, J. Ashburner & W. Penny (Eds.), *Human Brain Function* (2nd Edition ed., pp. 971-999). London, UK: Academic Press.
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, 13(3), 361-377.

- Galton, C., Patterson, K., Graham, K., Lambon-Ralph, M. A., Williams, G., Antoun, N., . . . Hodges, J. R. (2001). Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. *Neurology*, *57*(2), 216-225.
- Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123-144. doi: [http://dx.doi.org/10.1016/S0926-6410\(02\)00244-6](http://dx.doi.org/10.1016/S0926-6410(02)00244-6)
- Geschwind, N. (1970). Organization of Language and Brain. *Science*, *170*(3961), 940-944.
- Glover, G. H., & Law, C. S. (2001). Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. *Magnetic Resonance in Medicine*, *46*(3), 515-522. doi: <http://dx.doi.org/10.1002/mrm.1222>
- Gough, P. M., Nobre, A. C., & Devlin, J. T. (2005). Dissociating Linguistic Processes in the Left Inferior Frontal Cortex with Transcranial Magnetic Stimulation. *The Journal of Neuroscience*, *25*(35), 8010-8016. doi: <http://dx.doi.org/10.1523/jneurosci.2307-05.2005>
- Griffith, T. D., Rees, A., & Green, G. G. R. (1999). Disorders of human complex sound processing. *Neurocase: The Neural Basis of Cognition*, *5*(5), 365-378.
- Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*(10), 685-694.
- Hagoort, P. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (Vol. 4th Ed, pp. 819-836). Boston, MA: MIT press.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439 - 483.
- Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. *Science*, *304*(5669), 438-441. doi: <http://dx.doi.org/10.1126/science.1095455>
- Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 801-811. doi: 10.1098/rstb.2007.2089
- Halgren, E., Wang, C., Schomer, D. L., Knake, S., Marinkovic, K., Wu, J., & Ulbert, I. (2006). Processing stages underlying word recognition in the anteroventral temporal lobe. *NeuroImage*, *30*(4), 1401-1413. doi: <http://dx.doi.org/10.1016/j.neuroimage.2005.10.053>
- Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, *12*(2), 140-149. doi: <http://dx.doi.org/10.1093/cercor/12.2.140>
- Hart, J., & Gordon, B. (1990). Delineation of single-word semantic comprehension deficits in aphasia, with anatomical correlation. *Annals of Neurology*, *27*(3), 226-231. doi: <http://dx.doi.org/10.1002/ana.410270303>
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*, 131-138.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1-2), 67-99.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393-402.
- Hodges, J. R., Graham, N., & Patterson, K. (1995). Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory*, *3*(3-4), 463-495. doi: <http://dx.doi.org/10.1080/09658219508253161>
- Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic Dementia: Progressive Fluent Aphasia with Temporal Lobe Atrophy. *Brain*, *115*(6), 1783-1806. doi: <http://dx.doi.org/10.1093/brain/115.6.1783>
- Holcomb, P. J., & Neville, H. J. (1990). Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-related Brain Potentials. *Language and Cognitive Processes*, *5*(4), 281-312. doi: 10.1080/01690969008407065

- Holle, H., Obleser, J., Rueschemeyer, S.-A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, 49(1), 875-884. doi: <http://dx.doi.org/10.1016/j.neuroimage.2009.08.058>
- Huang-Hellinger, F. R., Breiter, H. C., McCormack, G., Cohen, M. S., Kwong, K. K., Sutton, J. P., . . . Rosen, B. R. (1995). Simultaneous functional magnetic resonance imaging and electrophysiological recording. *Human Brain Mapping*, 3(1), 13-23. doi: <http://dx.doi.org/10.1002/hbm.460030103>
- Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension. *Journal of Cognitive Neuroscience*, 18(4), 665-679. doi: <http://dx.doi.org/10.1162/jocn.2006.18.4.665>
- Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *NeuroImage*, 36(3), 924-932. doi: <http://dx.doi.org/10.1016/j.neuroimage.2007.03.059>
- Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping*, 26(2), 128-138. doi: <http://dx.doi.org/10.1002/hbm.20148>
- Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, 10(11), 480-486.
- Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic Perception and the Temporal Cortex. *NeuroImage*, 15(4), 733-746. doi: <http://dx.doi.org/10.1006/nimg.2001.1027>
- Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. *Brain*, 129(8), 2132-2147. doi: <http://dx.doi.org/10.1093/brain/awl153>
- Jezzard, P., Matthews, P. M., & Smith, S. M. (2001). *Functional MRI: an introduction to methods*. Oxford: Oxford University Press.
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization Training Results in Shape- and Category-Selective Human Neural Plasticity. *Neuron*, 53(6), 891-903.
- Jobard, G., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *NeuroImage*, 34(2), 784-800. doi: <http://dx.doi.org/10.1016/j.neuroimage.2006.06.067>
- Jones, R. A., Schirmer, T., Lipinski, B., Elbel, G. K., & Auer, D. P. (1998). Signal undershoots following visual stimulation: A comparison of gradient and spin-echo BOLD sequences. *Magnetic Resonance in Medicine*, 40(1), 112-118. doi: <http://dx.doi.org/10.1002/mrm.1910400116>
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99, 122-149.
- Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253-260. doi: [http://dx.doi.org/10.1016/S0093-934X\(03\)00335-3](http://dx.doi.org/10.1016/S0093-934X(03)00335-3)
- Kennan, R. P., Zhong, J. H., & Gore, J. C. (1994). Intravascular susceptibility contrast mechanisms in tissues. *Magnetic Resonance in Medicine*, 31, 9-21.
- Kircher, T., Sass, K., Sachs, O., & Krach, S. (2009a). Priming words with pictures: Neural correlates of semantic associations in a cross-modal priming task using fMRI. *Human Brain Mapping*, 30(12), 4116-4128. doi: <http://dx.doi.org/10.1002/hbm.20833>
- Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., . . . Green, A. (2009b). Neural interaction of speech and gesture: Differential activations of metaphoric co-verbal gestures. *Neuropsychologia*, 47(1), 169-179. doi: <http://dx.doi.org/10.1016/j.neuropsychologia.2008.08.009>
- Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., . . . Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3), 786-802. doi: <http://dx.doi.org/10.1016/j.neuroimage.2008.12.037>

- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2–3), 169-195. doi: [http://dx.doi.org/10.1016/S0165-0173\(98\)00056-3](http://dx.doi.org/10.1016/S0165-0173(98)00056-3)
- Klimesch, W., Doppelmayr, M., Stadler, W., Pöllhuber, D., Sauseng, P., & Röhme, D. (2001). Episodic retrieval is reflected by a process specific increase in human electroencephalographic theta activity. *Neuroscience Letters*, 302(1), 49-52. doi: [http://dx.doi.org/10.1016/S0304-3940\(01\)01656-1](http://dx.doi.org/10.1016/S0304-3940(01)01656-1)
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463-470.
- Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, 62(1), 621-647. doi: <http://dx.doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205. doi: <http://dx.doi.org/10.1126/science.7350657>
- Kutas, M., Neville, H., & Holcomb, P. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening, and signing. *Electroencephalography and Clinical Neurophysiology, Supplement*, 39, 325-330.
- Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles, M. G. Jennings & M. G. H. Coles (Eds.), *Advances in Psychophysiology* (pp. 139-187). Greenwich, CT: JAI.
- Lambon-Ralph, M. A., McClelland, J. L., Patterson, K., Galton, C., & Hodges, J. R. (2001). No Right to Speak? The Relationship between Object Naming and Semantic Impairment: Neuropsychological Evidence and a Computational Model. *Journal of Cognitive Neuroscience*, 13(3), 341-356. doi: <http://dx.doi.org/10.1162/08989290151137395>
- Lambon-Ralph, M. A., Sage, K., Jones, R. W., & Mayberry, E. J. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences*, 107(6), 2717-2722. doi: <http://dx.doi.org/10.1073/pnas.0907307107>
- Lambon Ralph, M. A., Graham, K. S., Patterson, K., & Hodges, J. R. (1999). Is a Picture Worth a Thousand Words? Evidence from Concept Definitions by Patients with Semantic Dementia. *Brain and Language*, 70(3), 309-335. doi: <http://dx.doi.org/10.1006/brln.1999.2143>
- Lambon Ralph, M. A., & Patterson, K. (2008). Generalization and Differentiation in Semantic Memory. *ANNALS OF THE NEW YORK ACADEMY OF SCIENCES*, 1124(1), 61-76. doi: <http://dx.doi.org/10.1196/annals.1440.006>
- Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual Knowledge Is Underpinned by the Temporal Pole Bilaterally: Convergent Evidence from rTMS. *Cerebral Cortex*, 19(4), 832-838. doi: <http://dx.doi.org/10.1093/cercor/bhn131>
- Laszlo, S., & Federmeier, K. D. (2008). Minding the PS, queues, and PXQs: Uniformity of semantic processing across multiple stimulus types. *Psychophysiology*, 45(3), 458-466. doi: 10.1111/j.1469-8986.2007.00636.x
- Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology*, 48(2), 176-186. doi: <http://dx.doi.org/10.1111/j.1469-8986.2010.01058.x>
- Leaver, A. M., & Rauschecker, J. P. (2010). Cortical Representation of Natural Complex Sounds: Effects of Acoustic Features and Auditory Object Category. *The Journal of Neuroscience*, 30(22), 7604-7612. doi: <http://dx.doi.org/10.1523/jneurosci.0296-10.2010>
- Lebedev, M. A., Messinger, A., Kralik, J. D., & Wise, S. P. (2004). Representation of Attended Versus Remembered Locations in Prefrontal Cortex. *PLoS Biol*, 2(11), e365. doi: <http://dx.doi.org/10.1371/journal.pbio.0020365>
- Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., & Price, C. J. (2008). The cortical dynamics of intelligible speech. *The Journal of Neuroscience*, 28(49), 13209-13215. doi: <http://dx.doi.org/10.1523/JNEUROSCI.2903-08.2008>

- Liebenthal, E., Ellingson, M. L., Spanaki, M. V., Prieto, T. E., Ropella, K. M., & Binder, J. R. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*, 19(4), 1395-1404. doi: [http://dx.doi.org/10.1016/S1053-8119\(03\)00228-3](http://dx.doi.org/10.1016/S1053-8119(03)00228-3)
- Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron*, 62(2), 281-290. doi: <http://dx.doi.org/10.1016/j.neuron.2009.02.025>
- LoCasto, P. C., Krebs-Noble, D., Gullapalli, R. P., & Burton, M. W. (2004). An fMRI Investigation of Speech and Tone Segmentation. *Journal of Cognitive Neuroscience*, 16(9), 1612-1624. doi: <http://dx.doi.org/10.1162/0898929042568433>
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150-157.
- Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD Signal. *Annual Review of Physiology*, 66(1), 735-769. doi: <http://dx.doi.org/10.1146/annurev.physiol.66.082602.092845>
- Lowe, M. J., Lurito, J. T., Mathews, V. P., Phillips, M. D., & Hutchins, G. D. (2000). Quantitative Comparison of Functional Contrast From BOLD-Weighted Spin-Echo and Gradient-Echo Echoplanar Imaging at 1.5 Tesla and H215O PET in the Whole Brain. *Journal of Cerebral Blood Flow & Metabolism*, 20(9), 1331-1340.
- Luders, H., Lesser, R. P., Hahn, J., Dinner, D. S., Morris, H., Resor, S., & Harrison, M. (1986). Basal temporal language area demonstrated by electrical stimulation. *Neurology*, 36(4), 505-510.
- Luders, H., Lesser, R. P., Hahn, J., Dinner, D. S., Morris, H. H., Wyllie, E., & Godoy, J. (1991). Basal temporal language area. *Brain*, 114(2), 743-754. doi: <http://dx.doi.org/10.1093/brain/114.2.743>
- Luzzi, S., Snowden, J. S., Neary, D., Coccia, M., Provinciali, L., & Lambon Ralph, M. A. (2007). Distinct patterns of olfactory impairment in Alzheimer's disease, semantic dementia, frontotemporal dementia, and corticobasal degeneration. *Neuropsychologia*, 45(8), 1823-1831. doi: <http://dx.doi.org/10.1016/j.neuropsychologia.2006.12.008>
- Mandelkow, H., Halder, P., Boesiger, P., & Brandeis, D. (2006). Synchronization facilitates removal of MRI artefacts from concurrent EEG recordings and increases usable bandwidth. *NeuroImage*, 32(3), 1120-1126. doi: <http://dx.doi.org/10.1016/j.neuroimage.2006.04.231>
- Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal Dynamics of Modality-Specific and Supramodal Word Processing. *Neuron*, 38(3), 487-497.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177-190. doi: <http://dx.doi.org/10.1016/j.jneumeth.2007.03.024>
- Marshall, J. F. (1984). Brain Function: Neural Adaptations and Recovery From Injury. *Annual Review of Psychology*, 35(1), 277-308. doi: <http://dx.doi.org/10.1146/annurev.ps.35.020184.001425>
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1-2), 71-102.
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, 11(2), 194-201.
- Matsuda, T., Matsuura, M., Ohkubo, T., Ohkubo, H., Atsumi, Y., Tamaki, M., . . . Kojima, T. (2002). Influence of arousal level for functional magnetic resonance imaging (fMRI) study: Simultaneous recording of fMRI and electroencephalogram. *Psychiatry and Clinical Neurosciences*, 56(3), 289-290. doi: <http://dx.doi.org/10.1046/j.1440-1819.2002.01016.x>
- Matsumoto, R., Imamura, H., Inouchi, M., Nakagawa, T., Yokoyama, Y., Matsushashi, M., . . . Ikeda, A. (2011). Left anterior temporal cortex actively engages in speech

- perception: A direct cortical stimulation study. *Neuropsychologia*, 49(5), 1350-1354. doi: <http://dx.doi.org/10.1016/j.neuropsychologia.2011.01.023>
- McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, 4(4), 310-322.
- McGettigan, C., Evans, S., Rosen, S., Agnew, Z. K., Shah, P., & Scott, S. K. (2012). An Application of Univariate and Multivariate Approaches in fMRI to Quantifying the Hemispheric Lateralization of Acoustic and Linguistic Processes. *Journal of Cognitive Neuroscience*, 24(3), 636-652. doi: http://dx.doi.org/10.1162/jocn_a.00161
- Menon, V., Ford, J. M., Lim, K. O., Glover, G. H., & Pfefferbaum, A. (1997). Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection. *NeuroReport*, 8(14), 3029-3037.
- Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999). EEG and MEG: forward solutions for inverse methods. *IEEE Transactions on Biomedical Engineering*, 46(3), 245-259.
- Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., & Malach, R. (2005). Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science*, 309(5736), 951-954. doi: <http://dx.doi.org/10.1126/science.1110913>
- Mullinger, K., Brookes, M., Stevenson, C., Morgan, P., & Bowtell, R. (2008a). Exploring the feasibility of simultaneous electroencephalography/functional magnetic resonance imaging at 7 T. *Magnetic Resonance Imaging*, 26(7), 968-977. doi: <http://dx.doi.org/10.1016/j.mri.2008.02.014>
- Mullinger, K., Debener, S., Coxon, R., & Bowtell, R. (2008b). Effects of simultaneous EEG recording on MRI data quality at 1.5, 3 and 7 tesla. *International Journal of Psychophysiology*, 67(3), 178-188. doi: <http://dx.doi.org/10.1016/j.ijpsycho.2007.06.008>
- Mullinger, K., Morgan, P., & Bowtell, R. (2008c). Improved artifact correction for combined electroencephalography/functional MRI by means of synchronization and use of vectorcardiogram recordings. *Journal of Magnetic Resonance Imaging*, 27(3), 607-616. doi: <http://dx.doi.org/10.1002/jmri.21277>
- Mullinger, K., Yan, W., & Bowtell, R. (2011). Reducing the gradient artefact in simultaneous EEG-fMRI by adjusting the subject's axial position. *NeuroImage*, 54(3), 1942-1950. doi: <http://dx.doi.org/10.1016/j.neuroimage.2010.09.079>
- Mummary, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. S. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *The Journal of the Acoustical Society of America*, 106(1), 449-457.
- Mummary, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. J., & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: Relationship between temporal lobe atrophy and semantic memory. *Annals of Neurology*, 47(1), 36-45. doi: [http://dx.doi.org/10.1002/1531-8249\(200001\)47:1<36::aid-ana8>3.0.co;2-I](http://dx.doi.org/10.1002/1531-8249(200001)47:1<36::aid-ana8>3.0.co;2-I)
- Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432-434.
- Nagel, I. E., Schumacher, E. H., Goebel, R., & D'Esposito, M. (2008). Functional MRI investigation of verbal selection mechanisms in lateral prefrontal cortex. *NeuroImage*, 43(4), 801-807. doi: <http://dx.doi.org/10.1016/j.neuroimage.2008.07.017>
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a Left-lateralized Response Specific to Intelligible Speech Using fMRI. *Cerebral Cortex*, 13(12), 1362-1368. doi: <http://dx.doi.org/10.1093/cercor/bhg083>
- Nestor, P. J., Fryer, T. D., & Hodges, J. R. (2006). Declarative memory impairments in Alzheimer's disease and semantic dementia. *NeuroImage*, 30(3), 1010-1020. doi: <http://dx.doi.org/10.1016/j.neuroimage.2005.10.008>
- Neville, H., Nicol, J. L., Barsis, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically Based Sentence Processing Classes: Evidence from Event-Related Brain Potentials. *Journal of Cognitive Neuroscience*, 3(2), 151-165. doi: <http://dx.doi.org/10.1162/jocn.1991.3.2.151>

- Newman, S. D., Just, M. A., Keller, T. A., Roth, J., & Carpenter, P. A. (2003). Differential effects of syntactic and semantic processing on the subregions of Broca's area. *Cognitive Brain Research*, 16(2), 297-307. doi: [http://dx.doi.org/10.1016/S0926-6410\(02\)00285-9](http://dx.doi.org/10.1016/S0926-6410(02)00285-9)
- Niazy, R. K., Beckmann, C. F., Lannetti, G. D., Brady, J. M., & Smith, S. M. (2005). Removal of fMRI environment artifacts from EEG data using optimal basis sets. *NeuroImage*, 28(3), 720-737. doi: <http://dx.doi.org/10.1016/j.neuroimage.2005.06.067>
- Noonan, K. A., Jefferies, E., Corbett, F., & Lambon Ralph, M. A. (2009). Elucidating the Nature of Deregulated Semantic Cognition in Semantic Aphasia: Evidence for the Roles of Prefrontal and Temporo-parietal Cortices. *Journal of Cognitive Neuroscience*, 22(7), 1597-1613. doi: <http://dx.doi.org/10.1162/jocn.2009.21289>
- Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond Inferior Prefrontal Involvement in Semantic Control: Evidence for the Additional Contribution of Dorsal Angular Gyrus and Posterior Middle Temporal Cortex. *Journal of Cognitive Neuroscience*, 25(11), 1824-1850. doi: 10.1162/jocn_a_00442
- Noppeney, U., Phillips, J., & Price, C. J. (2004). The neural areas that control the retrieval and selection of semantics. *Neuropsychologia*, 42(9), 1269-1280. doi: <http://dx.doi.org/10.1016/j.neuropsychologia.2003.12.014>
- Norris, D. G., Zysset, S., Mildner, T., & Wiggins, C. J. (2002). An Investigation of the Value of Spin-Echo-Based fMRI Using a Stroop Color-Word Matching Task and EPI at 3 T. *NeuroImage*, 15(3), 719-726.
- Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG* (2nd Ed. ed.). New York: Oxford University Press.
- Obleser, J., & Kotz, S. A. (2010). Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. *Cerebral Cortex*, 20(3), 633-640. doi: <http://dx.doi.org/10.1093/cercor/bhp128>
- Obleser, J., Lahiri, A., & Eulitz, C. (2003). Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *NeuroImage*, 20(3), 1839-1847. doi: <http://dx.doi.org/10.1016/j.neuroimage.2003.07.019>
- Obleser, J., Scott, S. K., & Eulitz, C. (2006). Now You Hear It, Now You Don't: Transient Traces of Consonants and their Nonspeech Analogues in the Human Brain. *Cerebral Cortex*, 16(8), 1069-1076. doi: <http://dx.doi.org/10.1093/cercor/bhj047>
- Obleser, J., Wise, R. J. S., Alex Dresner, M., & Scott, S. K. (2007a). Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *J. Neurosci.*, 27(9), 2283-2289. doi: <http://dx.doi.org/10.1523/jneurosci.4663-06.2007>
- Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007b). Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI. *Cerebral Cortex*, 17(10), 2251-2257. doi: <http://dx.doi.org/10.1093/cercor/bhl133>
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., . . . Hickok, G. (2010). Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. *Cerebral Cortex*, 20(10), 2486-2495. doi: <http://dx.doi.org/10.1093/cercor/bhp318>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia*, 9(1), 97-113.
- Oostendorp, T. F., & van Oosterom, A. (1989). Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. *IEEE Transactions on Biomedical Engineering*, 36(3), 382-391. doi: <http://dx.doi.org/10.1109/10.19859>
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*. doi: <http://dx.doi.org/10.1155/2011/156869>
- Ordidge, R. J., Gorell, J. M., Deniau, J. C., Knight, R. A., & Helpert, J. A. (1994). Assessment of relative brain iron concentrations using T2-weighted and T2*-weighted MRI at 3 Tesla. *Magnetic Resonance in Medicine*, 32(3), 335-341. doi: <http://dx.doi.org/10.1002/mrm.1910320309>

- Orfanidou, E., Marslen-Wilson, W. D., & Davis, M. H. (2006). Neural Response Suppression Predicts Repetition Priming of Spoken Words and Pseudowords. *Journal of Cognitive Neuroscience*, 18(8), 1237-1252. doi: <http://dx.doi.org/10.1162/jocn.2006.18.8.1237>
- Parker, G. J. M., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A. M., Ciccarelli, O., & Lambon-Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *NeuroImage*, 24(3), 656-666.
- Parkes, L. M., Schwarzbach, J. V., Bouts, A. A., Deckers, R. h. R., Pullens, P., Kerskens, C. M., & Norris, D. G. (2005). Quantifying the spatial resolution of the gradient echo and spin echo BOLD response at 3 Tesla. *Magnetic Resonance in Medicine*, 54(6), 1465-1472. doi: <http://dx.doi.org/10.1002/mrm.20712>
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976-987.
- Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., & Davis, M. H. (2010a). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *NeuroImage*, 52(4), 1410-1419. doi: <http://dx.doi.org/10.1016/j.neuroimage.2010.05.015>
- Peelle, J. E., Johnsrude, I. S., & Davis, M. H. (2010b). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, 4. doi: <http://dx.doi.org/10.3389/fnhum.2010.00051>
- Pei, X., Leuthardt, E. C., Gaona, C. M., Brunner, P., Wolpaw, J. R., & Schalk, G. (2011). Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *NeuroImage*, 54(4), 2960-2972. doi: <http://dx.doi.org/10.1016/j.neuroimage.2010.10.029>
- Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74(3), 374-388.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., . . . Roberts, T. (2000). Auditory Cortex Accesses Phonological Categories: An MEG Mismatch Study. *Journal of Cognitive Neuroscience*, 12(6), 1038-1055. doi: <http://dx.doi.org/10.1162/08989290051137567>
- Pinard, M., Chertkow, H., Black, S., & Peretz, I. (2002). A Case Study of Pure Word Deafness: Modularity in Auditory Processing? *Neurocase*, 8(1-2), 40-55. doi: <http://dx.doi.org/10.1093/neucas/8.1.40>
- Piwonica-Worms, K. E., Omar, R., Hailstone, J. C., & Warren, J. D. (2010). Flavour processing in semantic dementia. *Cortex*, 46(6), 761-768. doi: <http://dx.doi.org/10.1016/j.cortex.2009.07.002>
- Pobric, G., Jefferies, E., & Lambon-Ralph, M. A. (2007). Anterior temporal lobes mediate semantic representation: Mimicking semantic dementia by using rTMS in normal participants. *Proceedings of the National Academy of Sciences*, 104(50), 20137-20141. doi: <http://dx.doi.org/10.1073/pnas.0707383104>
- Pobric, G., Jefferies, E., & Lambon-Ralph, M. A. (2010). Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia*, 48(5), 1336-1342.
- Poeppel, D. (1996). A Critical Review of PET Studies of Phonological Processing. *Brain and Language*, 55(3), 317-351.
- Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1999). Functional Specialization for Semantic and Phonological Processing in the Left Inferior Prefrontal Cortex. *NeuroImage*, 10(1), 15-35. doi: <http://dx.doi.org/10.1006/nimg.1999.0441>
- Poser, B. A., & Norris, D. (2007). Fast spin echo sequences for BOLD functional MRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 20(1), 11-17. doi: <http://dx.doi.org/10.1007/s10334-006-0063-x>
- Poser, B. A., & Norris, D. G. (2009). Investigating the benefits of multi-echo EPI for fMRI at 7 T. *NeuroImage*, 45(4), 1162-1172. doi: <http://dx.doi.org/10.1016/j.neuroimage.2009.01.007>

- Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, 55(6), 1227-1235. doi: <http://dx.doi.org/10.1002/mrm.20900>
- Posse, S., Wiese, S., Gembris, D., Mathiak, K., Kessler, C., Grosse-Ruyken, M.-L., . . . Kiselev, V. G. (1999). Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic Resonance in Medicine*, 42(1), 87-97. doi: [http://dx.doi.org/10.1002/\(sici\)1522-2594\(199907\)42:1<87::aid-mrm13>3.0.co;2-o](http://dx.doi.org/10.1002/(sici)1522-2594(199907)42:1<87::aid-mrm13>3.0.co;2-o)
- Prat, C. S., & Just, M. A. (2008). Brain bases of individual differences in cognition. *Psychological Science Agenda*, 22.5.
- Prat, C. S., & Just, M. A. (2011). Exploring the Neural Dynamics Underpinning Individual Differences in Sentence Comprehension. *Cerebral Cortex*, 21(8), 1747-1760. doi: <http://dx.doi.org/10.1093/cercor/bhq241>
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *ANNALS OF THE NEW YORK ACADEMY OF SCIENCES*, 1191(1), 62-88. doi: <http://dx.doi.org/10.1111/j.1749-6632.2010.05444.x>
- Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, 62(2), 816-847. doi: <http://dx.doi.org/10.1016/j.neuroimage.2012.04.062>
- Price, C. J., Moore, C., Humphreys, G. W., & Wise, R. J. S. (1997). Segregating semantic from phonological processes during reading. *Journal of Cognitive Neuroscience*, 9(6), 727-733. doi: <http://dx.doi.org/10.1162/jocn.1997.9.6.727>
- Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5), 952-962. doi: [http://dx.doi.org/10.1002/\(sici\)1522-2594\(199911\)42:5<952::aid-mrm16>3.0.co;2-s](http://dx.doi.org/10.1002/(sici)1522-2594(199911)42:5<952::aid-mrm16>3.0.co;2-s)
- Pulvermüller, F. (2007). Brain processes of word recognition as revealed by neurophysiological imaging. In G. Gaskell (Ed.), *Handbook of Psycholinguistics* (pp. 119-140): Oxford University Press.
- Pulvermüller, F., Assadollahi, R., & Elbert, T. (2001). Neuromagnetic evidence for early semantic access in word recognition. *European Journal of Neuroscience*, 13(1), 201-205.
- Pulvermüller, F., Cooper-Pye, E., Dine, C., Hauk, O., Nestor, P. J., & Patterson, K. (2010). The Word Processing Deficit in Semantic Dementia: All Categories Are Equal, but Some Categories Are More Equal than Others. *Journal of Cognitive Neuroscience*, 22(9), 2027-2041. doi: <http://dx.doi.org/10.1162/jocn.2009.21339>
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11(5), 351-360.
- Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Electrocortical distinction of vocabulary types. *Electroencephalography and Clinical Neurophysiology*, 94(5), 357-370.
- Pulvermüller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and Verbs in the Intact Brain: Evidence from Event-related Potentials and High-frequency Cortical Responses. *Cerebral Cortex*, 9(5), 497-506. doi: <http://dx.doi.org/10.1093/cercor/9.5.497>
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Reviews Neuroscience*, 12(6), 718-724.
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), 11800-11806.
- Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207), 111-114. doi: <http://dx.doi.org/10.1126/science.7701330>
- Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and Cognition*, 69(2), 269-278. doi: <http://dx.doi.org/10.1016/j.bandc.2008.08.007>

- Robinson, S., Windischberger, C., Rauscher, A., & Moser, E. (2004). Optimized 3 T EPI of the amygdalae. *NeuroImage*, 22(1), 203-210. doi: <http://dx.doi.org/10.1016/j.neuroimage.2003.12.048>
- Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The Neural Mechanisms of Speech Comprehension: fMRI studies of Semantic Ambiguity. *Cerebral Cortex*, 15(8), 1261-1269. doi: <http://dx.doi.org/10.1093/cercor/bhi009>
- Rogers, T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K., & Price, C. J. (2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. *Cognitive, Affective, & Behavioral Neuroscience*, 6(3), 201-213. doi: <http://dx.doi.org/10.3758/CABN.6.3.201>
- Rogers, T. T., Lambon-Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, 111(1), 205-235.
- Rolls, E. T. (1991). Neural organization of higher visual functions. *Current Opinion in Neurobiology*, 1(2), 274-278. doi: [http://dx.doi.org/10.1016/0959-4388\(91\)90090-T](http://dx.doi.org/10.1016/0959-4388(91)90090-T)
- Romanski, L., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci*, 2(12), 1131-1136.
- Romanski, L. M., & Averbach, B. B. (2009). The Primate Cortical Auditory System and Neural Representation of Conspecific Vocalizations. *Annual Review of Neuroscience*, 32(1), 315-346. doi: <http://dx.doi.org/10.1146/annurev.neuro.051508.135431>
- Romanski, L. M., Averbach, B. B., & Diltz, M. (2005). Neural Representation of Vocalizations in the Primate Ventrolateral Prefrontal Cortex. *Journal of Neurophysiology*, 93(2), 734-747. doi: <http://dx.doi.org/10.1152/jn.00675.2004>
- Rugg, M. (1983). The relationship between evoked potentials and lateral asymmetries of processing. In A. Gaillard & W. Ritter (Eds.), *Tutorials in event related potential research: Endogenous components*. Amsterdam.
- Rugg, M. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, 18(4), 367-379. doi: <http://dx.doi.org/10.3758/bf03197126>
- Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M.-S., . . . Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, 105(46), 18035-18040. doi: <http://dx.doi.org/10.1073/pnas.0805234105>
- Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Küpper, H., Kellmeyer, P., . . . Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *NeuroImage*, 49(4), 3187-3197. doi: <http://dx.doi.org/10.1016/j.neuroimage.2009.11.009>
- Scarff, C. J., Reynolds, A., Goodyear, B. G., Ponton, C. W., Dort, J. C., & Eggermont, J. J. (2004). Simultaneous 3-T fMRI and high-density recording of human auditory evoked potentials. *NeuroImage*, 23(3), 1129-1142. doi: <http://dx.doi.org/10.1016/j.neuroimage.2004.07.035>
- Schmidt, C. F., Boesiger, P., & Ishai, A. (2005). Comparison of fMRI activation as measured with gradient- and spin-echo EPI during visual perception. *NeuroImage*, 26(3), 852-859.
- Schmithorst, V. J., Dardzinski, B. J., & Holland, S. K. (2001). Simultaneous correction of ghost and geometric distortion artifacts in EPI using a multiecho reference scan. *IEEE Transactions on Medical Imaging*, 20(6), 535-539. doi: <http://dx.doi.org/10.1109/42.929619>
- Schwarzbauer, C., Mildner, T., Heinke, W., Brett, M., & Deichmann, R. (2010). Dual echo EPI - The method of choice for fMRI in the presence of magnetic field inhomogeneities? *NeuroImage*, 49(1), 316-326.
- Schwarzbauer, C., & Porter, D. A. (2010). Single shot partial dual echo (SPADE) EPI—an efficient acquisition scheme for reducing susceptibility artefacts in fMRI.

- NeuroImage*, 49(3), 2234-2237. doi:
<http://dx.doi.org/10.1016/j.neuroimage.2009.10.059>
- Scott, S. K., Blank, C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400-2406. doi:
<http://dx.doi.org/10.1093/brain/123.12.2400>
- Scott, S. K., Leff, A. P., & Wise, R. J. S. (2003). Going beyond the information given: a neural system supporting semantic interpretation. *NeuroImage*, 19(3), 870-876. doi: [http://dx.doi.org/10.1016/S1053-8119\(03\)00083-1](http://dx.doi.org/10.1016/S1053-8119(03)00083-1)
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech---A positron emission tomography study. *The Journal of the Acoustical Society of America*, 120(2), 1075-1083.
- Sereno, S. C., Cameron, C. B., & O'Donnell, P. J. (2003). Context Effects in Word Recognition: Evidence for Early Interactive Processing. *Psychological Science*, 14(4), 328-333.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science*, 270(5234), 303-304. doi: <http://dx.doi.org/10.1126/science.270.5234.303>
- Sharp, D. J., Scott, S. K., & Wise, R. J. S. (2004). Retrieving meaning after temporal lobe infarction: The role of the basal language area. *Annals of Neurology*, 56(6), 836-846. doi: <http://dx.doi.org/10.1002/ana.20294>
- Simmons, W. K., & Martin, A. (2009). The anterior temporal lobes and the functional architecture of semantic memory. *Journal of the International Neuropsychological Society*, 15(05), 645-649.
- Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two Neurocognitive Mechanisms of Semantic Integration during the Comprehension of Visual Real-world Events. *Journal of Cognitive Neuroscience*, 20(11), 2037-2057. doi: 10.1162/jocn.2008.20143
- Smith, M., & Halgren, E. (1987). Event-related potentials during lexical decision: effects of repetition, word frequency, pronounceability, and concreteness. In R. Johnson, J. Rohrbaugh & R. Parasuraman (Eds.), *Current Trends in Event-Related Potential Research* (pp. 417-421). New York: Elsevier.
- Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., . . . Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage*, 54(2), 875-891. doi: <http://dx.doi.org/10.1016/j.neuroimage.2010.08.063>
- Speck, O., & Hennig, J. (1998). Functional Imaging by I0- and T2* -parameter mapping using multi-image EPI. *Magnetic Resonance in Medicine*, 40(2), 243-248. doi: <http://dx.doi.org/10.1002/mrm.1910400210>
- Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2006). Converging Language Streams in the Human Temporal Lobe. *The Journal of Neuroscience*, 26(28), 7328-7336. doi: <http://dx.doi.org/10.1523/jneurosci.0559-06.2006>
- Stark, C. E. L., & Squire, L. R. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences*, 98(22), 12760-12766. doi: <http://dx.doi.org/10.1073/pnas.221462998>
- Stenger, V. A., Boada, F. E., & Noll, D. C. (2000). Three-dimensional tailored RF pulses for the reduction of susceptibility artifacts in T*2-weighted functional MRI. *Magnetic Resonance in Medicine*, 44(4), 525-531. doi: [http://dx.doi.org/10.1002/1522-2594\(200010\)44:4<525::aid-mrm5>3.0.co;2-l](http://dx.doi.org/10.1002/1522-2594(200010)44:4<525::aid-mrm5>3.0.co;2-l)
- Stevens, K. N. (1983). Acoustic properties used for the identification of speech sounds. *ANNALS OF THE NEW YORK ACADEMY OF SCIENCES*, 405(1), 2-17. doi: <http://dx.doi.org/10.1111/j.1749-6632.1983.tb31613.x>
- Stroman, P. W., Krause, V., Frankenstein, U. N., Malisza, K. L., & Tomanek, B. (2001). Spin-echo versus gradient-echo fMRI with short echo times. *Magnetic Resonance Imaging*, 19(6), 827-831.

- Stuss, D. T., Sarazin, F. F., Leech, E. E., & Picton, T. W. (1983). Event-related potentials during naming and mental rotation. *Electroencephalography and Clinical Neurophysiology*, *56*(2), 133-146.
- Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, *150*(1-2), 225-244. doi: [http://dx.doi.org/10.1016/S0378-5955\(00\)00203-3](http://dx.doi.org/10.1016/S0378-5955(00)00203-3)
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792-14797.
- Thulborn, K. R., Chang, S. Y., Shen, G. X., & Voyvodic, J. T. (1997). High-resolution echo-planar fMRI of human visual cortex at 3.0 tesla. *NMR in Biomedicine*, *10*(4-5), 183-190. doi: [http://dx.doi.org/10.1002/\(sici\)1099-1492\(199706/08\)10:4/5<183::aid-nbm469>3.0.co;2-w](http://dx.doi.org/10.1002/(sici)1099-1492(199706/08)10:4/5<183::aid-nbm469>3.0.co;2-w)
- Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional Specialization in Rhesus Monkey Auditory Cortex. *Science*, *292*(5515), 290-293. doi: <http://dx.doi.org/10.1126/science.1058911>
- Tranel, D., Grabowski, T. J., Lyon, J., & Damasio, H. (2005). Naming the Same Entities from Visual or from Auditory Stimulation Engages Similar Regions of Left Inferotemporal Cortices. *Journal of Cognitive Neuroscience*, *17*(8), 1293-1305. doi: <http://dx.doi.org/10.1162/0898929055002508>
- Tranel, D., Kemmerer, D., Adolphs, R., Damasio, H., & Damasio, A. R. (2003). Neural correlates of conceptual knowledge for actions. *Cognitive Neuropsychology*, *20*(3-6), 409-432. doi: <http://dx.doi.org/10.1080/02643290244000248>
- Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*(1), 1-15. doi: <http://dx.doi.org/10.1016/j.bandl.2010.03.008>
- van Petten, C., & Rieffelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*(4), 485-508. doi: [http://dx.doi.org/10.1016/0028-3932\(94\)00133-A](http://dx.doi.org/10.1016/0028-3932(94)00133-A)
- Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*(9), 867-880.
- Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The Response of Left Temporal Cortex to Sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550-560. doi: <http://dx.doi.org/10.1162/08989290260045800>
- Vandenberghe, R., Price, C. J., Wise, R., Josephs, O., & Frackowiak, R. S. J. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature*, *383*(6597), 254-256.
- Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, *2*(4), 229-239.
- Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., . . . Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing?: Insights from a meta-analysis. *NeuroImage*, *54*(1), 577-593.
- Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., . . . Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414-1432.
- Visser, M., Embleton, K. V., Jefferies, E., Parker, G. J., & Lambon-Ralph, M. A. (2010). The inferior, anterior temporal lobes and semantic memory clarified: Novel evidence from distortion-corrected fMRI. *Neuropsychologia*, *48*(6), 1689-1696.
- Visser, M., Jefferies, E., & Lambon-Ralph, M. A. (2009). Semantic Processing in the Anterior Temporal Lobes: A Meta-analysis of the Functional Neuroimaging Literature. *Journal of Cognitive Neuroscience*, *22*(6), 1083-1094. doi: <http://dx.doi.org/10.1162/jocn.2009.21309>

- Visser, M., & Lambon-Ralph, M. A. (2011). Differential Contributions of Bilateral Ventral Anterior Temporal Lobe and Left Anterior Superior Temporal Gyrus to Semantic Processes. *Journal of Cognitive Neuroscience*, 23(10), 3121-3131. doi: http://dx.doi.org/10.1162/jocn_a_00007
- Volz, S., Hattungen, E., Preibisch, C., Gasser, T., & Deichmann, R. (2009). Reduction of susceptibility-induced signal losses in multi-gradient-echo images: Application to improved visualization of the subthalamic nucleus. *NeuroImage*, 45(4), 1135-1143.
- Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of Sounds in the Auditory Stream: Event-Related fMRI Evidence for Differential Activation to Speech and Nonspeech. *Journal of Cognitive Neuroscience*, 13(7), 994-1005. doi: <http://dx.doi.org/10.1162/089892901753165890>
- Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering Meaning: Left Prefrontal Cortex Guides Controlled Semantic Retrieval. *Neuron*, 31(2), 329-338.
- Weiskopf, N., Hutton, C., Josephs, O., & Deichmann, R. (2006). Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3T and 1.5T. *NeuroImage*, 33(2), 493-504.
- Weiss, S., & Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language*, 85(2), 325-343. doi: [http://dx.doi.org/10.1016/S0093-934X\(03\)00067-1](http://dx.doi.org/10.1016/S0093-934X(03)00067-1)
- Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2011). The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cerebral Cortex*, 21(5), 1066-1075. doi: <http://dx.doi.org/10.1093/cercor/bhq180>
- Wise, R. J. S., Scott, S. K., Blank, C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain*, 124(1), 83-95. doi: <http://dx.doi.org/10.1093/brain/124.1.83>
- Yan, W., Mullinger, K., Brookes, M., & Bowtell, R. (2009). Understanding gradient artefacts in simultaneous EEG/fMRI. *NeuroImage*, 46(2), 459-471. doi: <http://dx.doi.org/10.1016/j.neuroimage.2009.01.029>
- Yan, W., Mullinger, K., Geirsdottir, G., & Bowtell, R. (2010). Physical modeling of pulse artefact sources in simultaneous EEG/fMRI. *Human Brain Mapping*, 31(4), 604-620. doi: <http://dx.doi.org/10.1002/hbm.20891>
- Yang, Q. X., Dardzinski, B. J., Li, S., Eslinger, P. J., & Smith, M. B. (1997). Multi-gradient echo with susceptibility inhomogeneity compensation (MGESIC): Demonstration of fMRI in the olfactory cortex at 3.0 T. *Magnetic Resonance in Medicine*, 37(3), 331-335. doi: <http://dx.doi.org/10.1002/mrm.1910370304>
- Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is 'where' in the human auditory cortex? *Nat Neurosci*, 5(9), 905-909.

Appendices

Appendix 1: Word/Pseudoword stimuli

High imageability		Low imageability	
Word	Pseudoword	Word	Pseudoword
Ball	Lall	Base	Hase
Beach	Weach	Beat	Leat
Bed	Ked	Bit	Mit
Big	Mig	Born	Dorn
Call	Nall	Choose	Cheel
Church	Curch	Cost	Bost
Cup	Mup	Cut	Wut
Dead	Mead	Deal	Geal
Deep	Feep	Death	Peath
Face	Dace	Fear	Mear
Fight	Gight	Fell	Rell
Fish	Bish	Fine	Bine
Gas	Ras	Gate	Pate
Hall	Rall	Half	Falf
Hard	Fard	Harm	Marm
Heart	Reart	Hope	Lope
King	Bing	Kill	Rill
Laugh	Taugh	Lack	Dack
Light	Dight	Life	Bife
Love	Pove	Lord	Mord
Male	Lale	Main	Tain
March	Barch	Mean	Rean
Meal	Beal	Miss	Biss
Moon	Shoon	Mood	Sood
Mouth	Touth	Move	Cove
Night	Pight	Need	Keed
Nose	Cose	None	Rone
Page	Nage	Pain	Nain
Park	Hark	Part	Wart
Poor	Roor	Push	Fush
Rain	Lain	Race	Nace
Red	Ped	Reach	Heach
Rich	Mich	Rest	Hest
Roof	Foof	Rule	Bule
Shop	Wop	Shape	Lape
Size	Hize	Side	Fide
Tall	Chall	Talk	Shalk
Teeth	Meeth	Team	Ceam
Wall	Sall	Warm	Tarm
Win	Hin	Wish	Pish

Appendix 2: Semantic/Letter categorisation stimuli

Words			
	Probes		Target
train	hovercraft	aeroplane	bus
taxi	boat	bicycle	coach
car	truck	helicopter	van
cello	banjo	harp	trumpet
keyboard	saxophone	trombone	violin
clarinet	harpsichord	organ	recorder
envelope	pencil	ruler	notebook
bikini	shirt	blouse	jacket
dress	shoes	jumper	tie
hat	skirt	mittens	socks
gloves	stockings	shawl	boots
knife	plate	fork	cup
bowl	spoon	mug	saucer
whisky	rum	lager	wine
beer	scotch	brandy	gin
screwdriver	spade	nail	paintbrush
rake	saw	scissors	pliers
brick	tile	mortar	plaster
chair	futon	shelf	bed
couch	desk	cabinet	wardrobe
cooker	fridge	dishwasher	freezer
microwave	television	stereo	toaster
radio	printer	computer	telephone
gun	sword	dagger	pistol
quiche	chips	pizza	pasta
sausage	hotdog	burger	pasty
sandwich	pudding	pie	flan
bag	purse	basket	suitcase
comb	soap	deodorant	toothbrush
candle	torch	lamp	bulb
bus	hovercraft	aeroplane	harp
bicycle	coach	taxi	pudding
truck	helicopter	car	cider
motorbike	tractor	lorry	shirt
trumpet	banjo	cello	lamp
trombone	saxophone	keyboard	beer
guitar	tuba	cymbal	newspaper
notebook	envelope	pencil	boat
blouse	jacket	bikini	spoon
dress	tie	shoes	bat
gloves	boots	shawl	bread
saucer	bowl	mug	trousers
pan	kettle	pot	violin
whisky	wine	rum	purse
gin	scotch	brandy	comb
saw	scissors	pliers	piano
cement	plaster	mortar	yoghurt
sofa	bench	bookcase	ruler

bed	chair	shelf	paperclip
cabinet	desk	wardrobe	chisel
sink	bath	shower	spade
dishwasher	fridge	freezer	rake
toaster	microwave	television	gun
skittles	marbles	ball	cup
dagger	pistol	sword	burger
sausage	pasty	hotdog	jumper
basket	bag	suitcase	ferry
soap	deodorant	toothbrush	lager
torch	bulb	candle	futon
ceiling	roof	floor	train

Letters

	Probes		Target
aaaaa	aaaaaaaaa	aaaaaaaaa	aaa
bbbb	bbbb	bbbbbbb	bbbbb
ccc	cccc	ccccccccc	ccc
eeee	eeee	eeee	eeeeeee
ffffff	ffffff	ffffff	fffff
hhhhhhh	hhhhhhhhh	hhhhh	hhhhhhh
kkkkkkk	kkkkk	kkkkk	kkkkkkk
lllll	llll	lllll	lllll
mmmmm	mmmm	mmmmmm	mmm
nnn	nnnn	nnnnnnn	nnnn
oooo	oooooooo	oooo	oooo
rrrr	rrrr	rrrr	rrr
ssss	ssss	sss	sssss
uuuuu	uuu	uuuuu	uuu
vvvv	vvvvv	vvvvv	vvv
xxxxxxxxx	xxxxx	xxxx	xxxxxxxxx
zzzz	zzz	zzzzzzz	zzzzz
aaaaa	aaaa	aaaaa	aaaaaaa
cccc	cccc	cccc	ccc
dddd	ddd	dddddd	ddddddd
ffff	ffff	fffffff	fffff
hhhhhhh	hhhhhhhhh	hhhhh	hhhhhh
iiii	iiiiii	iiiiiii	iiiiiii
lllll	llll	lllll	lllll
nnnn	nnnn	nnnn	nnnn
oooooo	ooooo	ooooo	oooo
rrrrrr	rrrrrr	rrrr	rrrr
sss	sssss	sssss	sssssss
ttt	ttttt	ttttttt	ttttttt
uuuuu	uuuu	uuuu	uuu
aaa	aaaaaaaaa	aaaaaaaaa	rrrrrrr
bbbb	bbb	bbbbbb	ttt
ccc	cccc	ccccccccc	kkkkkkk
ddddddd	dddddd	ddd	aaaa
eeeeee	eeee	eee	vvvv
ffff	fffffff	ffffff	iiiiii
iii	iiii	iiii	ffff
kkkkkkk	kkkkk	kkkk	eeeeeee

lllll	llll	lllll	rrrrrr
mmm	mmmmm	mmmmmm	ttttt
ooooo	ooooooooo	ooooo	uuuuu
sssss	sssss	sss	aaaa
ttt	ttttt	tttt	ssss
uuuu	uuu	uuuuu	dddddd
vvv	vvvvv	vvvvv	fff
zzzzz	zzz	zzzzzzz	kkkkk
aaaaaaa	aaaa	aaaaa	lllll
bbbbbbb	bbbb	bbbbbb	zzzz
ccc	cccc	cccc	xxxx
ddddddd	dddd	dddddd	cccc
eeeeeee	eeee	eeeeee	vvvv
ffffff	fffff	fffffff	bbbbbb
hhhhhhh	hhhhhhhhhhh	hhhhh	mmmmmmmm
kkkkkkkk	kkk	kkkk	eeeeeee
lllll	llll	lllll	rrrrrr
ooooo	oooooo	oooooo	uuuuuuuu
sssssss	sssss	sssss	aaa
tttttttt	ttttt	ttttttt	ssss
uuuu	uuuuu	uuuu	dddddd
vvvvvv	vvvv	vvvvv	fff

Appendix 3: Auditory sentence stimuli

The driver starts the engine
The bag bumps on the ground
The boy did a handstand
The cat sits on the bed
The lorry carried fruit
The rain came down
The ladders near the door
The ball went into the goal
The boy knew the game
The police chased the car
The school finished early
The dog played with the stick
The farmer keeps the ball
The lady wore a coat
The milk came in a bottle
The man cleaned his shoes
The boys running away
The rooms getting cold
The girl kicked the table
The wife helped her husband
The girl lost her doll
The cooks making a cake
The child grabs the toy
The matches lie on the shelf
The parks near the road
The cook cuts some onions
The light went out
The paint dripped on the ground
The mother stirs the tea
The lady goes to the shop
The sun melted the snow
The fathers coming home
The lorry drove up the road
The broom stood in the corner
The woman tidied her house
The children dropped the bag
The dog came back

The bus went early
The father forgot the bread
The boy forgot his book
The family bought a house
The jug stood on the shelf
The ball broke the window
The boy hurried to school
The man's painting a sign
The lady packed her bag
The trains moving fast
The child drank some milk
The car hit the wall
The cleaner used a broom
The dog jumped on the chair
The daughter laid the table
The chicken laid some eggs
The dogs go for a walk
The lady stayed for tea
The driver waits by the corner
The mother tied the string
The cow gave some milk
The boy got into bed
The father writes a letter
The girl's washing her hair
The children helped the milkman
The tree fell on the house
The fruit came in a box
The husband brings some flowers
The mouse found the cheese
The lorry climbed the hill
The dog sleeps in a basket
The girl plays with the baby
The children washed the plates
The match fell on the floor
The shop closed for lunch
The children waved at the train
The girl held a mirror
The flower stands in a pot