# Investigating the motor-sensory learning of foreign speech

**Anna Jane Simmonds**

2012

Imperial College London
Cognitive Neuroimaging Group,
MRC Clinical Sciences Centre,
Computational, Cognitive and Clinical Neuroimaging Laboratory

This thesis is submitted for the degree of
Doctor of Philosophy

# Acknowledgements

Firstly, I should like to thank the Medical Research Council for funding my PhD research and all the volunteers who participated in the studies presented in this thesis.

I have been privileged to have two exceptional supervisors, Professor Richard Wise and Dr Rob Leech, both of whom have provided ongoing inspiration and guidance throughout the course of my PhD. I could not have asked for better mentors and I am very thankful to them both.

I am grateful for the help given by my colleagues in the Computational, Cognitive and Clinical Neuroimaging Laboratory (C3NL) and the radiographers at the Robert Steiner MRI Unit, Hammersmith Hospital.

Particular thanks are due to Catherine Collins and Ozlem Redjep for their assistance with the study presented in Chapter 3 of this thesis, to Dr Paul Iverson and his colleagues in the Speech, Hearing and Phonetic Sciences department at University College London for their help with the study presented in Chapter 5, and to the Traumatic Brain Injury group at C3NL for the resting state data presented in Chapter 6.

Finally, I thank my family for their constant encouragement. Without their support I would never had made the leap from teacher to neuroscientist and would still be teaching foreign speech, rather than researching the motor-sensory learning of it.

# Declaration of publications

The data presented in Results Chapter 4 of this thesis have been published in the Journal of Neurophysiology (Simmonds et al., 2011a). Much of the Introduction to this thesis has been published in Frontiers in Psychology (Simmonds et al., 2011b). The data presented in Results Chapters 5 and 6 have been submitted for publication.

Simmonds, A.J., Wise, R.J.S., Dhanjal, N.S., and Leech, R. (2011a). A comparison of sensory-motor activity during speech in first and second languages. J Neurophysiol *106*, 470-478.

Simmonds, A.J., Wise, R.J.S., and Leech, R. (2011b). Two tongues, one brain: imaging bilingual speech production. Frontiers in Psychology *2*: *10.3389/fpsyg.2011.00166*.

# Declaration of originality

The study design, participant recruitment, MRI scanning and behavioural analysis for the data included in Chapter 3 were conducted with the help of two MSc students, Catherine Collins and Ozlem Redjep, in the Computational, Cognitive and Clinical Neuroimaging Laboratory (C3NL). The stimuli for Chapter 5 were developed and recorded by Dr Paul Iverson and his colleagues from the Speech, Hearing and Phonetic Sciences department at University College London. They also conducted the speech ratings. The 'resting state' dataset used in Study 3 of Chapter 6 of this thesis was collected by the Traumatic Brain Injury Group within C3NL.

All other work contained in this thesis is my own and conforms to the rules and guidelines set out for PhD theses by Imperial College London.

# Abbreviations

| | |
|---|---|
| AFP | anterior forebrain pathway |
| ANOVA | analysis of variance |
| aPut | anterior putamen |
| antPT | anterior planum temporale |
| $B_0$ | longitudinal static magnetic field in the MRI scanner |
| BET | Brain Extraction Tool |
| BOLD | Blood-oxygen level dependent |
| BVAT | Bilingual Verbal Ability Tests |
| CBF | cerebral blood flow |
| COPE | contrast of parameter estimate |
| CPE | certificate of proficiency in English |
| CSF | cerebrospinal fluid |
| dcaud | dorsal caudate |
| deoxyHb | deoxygenated haemoglobin |
| DIVA | directions into velocities of articulators |
| DLM | dorsal lateral nucleus of anterior thalamus |
| DMN | default mode network |
| DVD | developmental verbal dyspraxia |
| EPI | echo-planar image |
| EV | explanatory variable |
| FEAT | fMRI Expert Analysis Tool |
| FLAME | FMRIB's Local Analysis of Mixed Effects |
| FLIRT | FMRIB's Linear Image Registration Tool |
| fMRI | functional magnetic resonance imaging |
| FMRIB | Oxford Centre for Functional Magnetic Resonance Imaging of the Brain |
| fOP | frontal operculum |
| FOXP2 | forkhead box P2 |
| FSL | FMRIB's Software Library |
| FWHM | full width at half maximum |
| GLM | General Linear Model |

| | |
|---|---|
| HRF | haemodynamic response function |
| HVC | high vocal centre |
| ICA | Independent Component Analysis |
| IFG | inferior frontal gyrus |
| L1 | native language |
| L2 | second language |
| LMAN | lateral magnocellular nucleus of the anterior nidopallium |
| McFLIRT | Motion Correction FMRIB's Linear Registration Tool |
| MNI | Montreal Neurological Institute |
| MRI | magnetic resonance imaging |
| NMR | nuclear magnetic resonance |
| oxyHb | oxygenated haemoglobin |
| PDP | posterior descending pathway |
| PE | parameter estimate |
| PET | positron emission tomography |
| PLA | primary language acquisition |
| PMC | premotor cortex |
| pre-SMA | pre-supplementary motor area |
| pSTG | posterior superior temporal gyrus |
| PT | planum temporale |
| PO | parietal operculum |
| postPT | posterior planum temporale |
| RA | robust nucleus of acropallium |
| RF | radio frequency |
| ROI | region of interest |
| SLA | secondary language acquisition |
| SLI | specific language impairment |
| SMA | supplementary motor area |
| SNR | signal-to-noise ratio |
| Spt | Sylvian parietal temporal |
| STP | superior temporal plane |
| STS | superior temporal sulcus |
| T1 | T1-weighted structural MRI scan |
| T2 | T2-weighted structural MRI scan |

| | |
|---|---|
| TE | echo time |
| TPJ | temporo-parietal junction |
| TR | repetition time |
| vAPL | ventral anterior parietal lobe |
| vcaud | ventral caudate |
| WM | white matter |
| X | area X within songbird medial striatum |

# Abstract

This thesis presents an investigation of bilingualism as a motor learning skill, with success ultimately measured in terms of strength of a foreign accent, in contrast to the many studies of bilingualism in terms of linguistic competence. My research used functional magnetic resonance (fMRI) imaging to investigate feedforward (motor) and feedback (auditory and somatosensory) systems involved in the production of foreign speech and how these systems are modulated by proficiency levels. I investigated the function of the frontal operculum and the temporo-parietal junction (TPJ) – planum temporale (posterior auditory association cortex) and parietal operculum (somatosensory association cortex) – during speech. The frontal operculum, strongly lateralised to the left, has been associated with speech since Broca performed his classic post mortem lesion-deficit analysis. Interest in the TPJ has arisen because of recent publications proposing the posterior half of the left planum temporale (± adjacent parietal operculum) as a 'sensorimotor interface' for speech production. My research compared activity within the frontal operculum and the TPJ during overt and covert speech. A second fMRI study examined retrospective proficiency based on existing language skills in people with English as a foreign language who were scanned during speech production in their native language and in English. A third fMRI study manipulated proficiency by training monolingual native English participants in the production of foreign speech sounds, with scanning pre- and post-training. This allowed measures of changes in activity (indicating rapid plasticity) following a short period of behavioural training in articulating novel foreign speech sounds. Training effects were observed predominantly in the striatum, and further analyses indicated that striatal activity in vocal learning is modulated by proficiency.

# Table of contents

**List of figures** …… **16**

**List of tables** ……..…**19**

**1 Introduction** ..... **20**

*1.1 Speech production* ..... *21*

1.1.1 Linguistic aspects of speech production ..... 21

1.1.2 Motor-sensory aspects of speech production ..... 22

1.1.3 Self-monitoring in L1 and L2 ..... 23

*1.2 The representation of multiple languages in the bilingual brain* ..... *25*

*1.3 Previous research on bilingualism: linguistic and cognitive aspects* ..... *26*

1.3.1 Differences between first and second language acquisition ..... 27

1.3.2 Age of acquisition and proficiency ..... 30

1.3.3 Cognitive differences in adult L2 learners ..... 34

*1.4 Why has previous bilingualism research largely ignored motor-sensory aspects of learning?* ..... *35*

1.4.1 Interlanguage phonology ..... 35

1.4.2 Difficulty assessing speech ..... 36

1.4.3 Scan artefacts of speech production ..... 37

*1.5 Evidence for motor-sensory contributions in bilingualism* ..... *38*

1.5.1 Cortical contributions to motor-sensory control in bilingualism ..... 38

1.5.2 Subcortical contributions to motor-sensory control in bilingualism ..... 40

1.5.3 Adaptations to auditory and somatosensory feedback during L1 production ..... 43

1.5.4 Increased activity in response to sensory feedback during L2 production ..... 45

*1.6 Potential links between bilingual articulation and comprehension* ..... *46*

1.6.1 Language input and its impact on speech output ..... 46

1.6.2 Speech output and its impact on language input ..... 47

*1.7 Parallels between human speech learning and songbird song acquisition* ..... *47*

*1.8 Multiple networks involved in L1 and L2 motor-sensory control of speech production* ..... *51*

*1.9 Implications of taking a motor-sensory approach to bilingual speech production* ..... *52*
1.9.1 Pedagogic implications: adults acquiring native-like levels of production ..... 52
1.9.2 Babbling adults? ..... 53
1.9.3 The role of literacy in L1 and in L2 ..... 53
1.9.4 Clinical implications ..... 54
*1.10 Benefits of researching neural bases of bilingualism* ..... *55*
*1.11 Main aims and hypotheses of this thesis* ..... *56*
1.11.1 Main aims ..... 56
1.11.2 Main hypotheses ..... 57
*1.12 Thesis overview* ..... *57*
1.12.1 Chapter 2 ..... 57
1.12.2 Chapter 3 ..... 57
1.12.3 Chapter 4 ..... 58
1.12.4 Chapter 5 ..... 59
1.12.5 Chapter 6 ..... 60
1.12.6 Chapter 7: Thesis summary and future directions ..... 60
**2 Methods ..... 61**
*2.1 Introduction to magnetic resonance imaging* ..... *61*
2.1.1 MRI principle summary ..... 61
2.1.2 Nuclear Magnetic Resonance (NMR) ..... 61
2.1.3 The magnetic fields $B_0$ and $B_1$ ..... 62
2.1.4 Shimming coil ..... 63
2.1.5 Radiofrequency (RF) coils ..... 63
2.1.6 X Y Z gradient coils ..... 64
2.1.7 MRI parameters ..... 64
2.1.8 T1-weighted images ..... 65
2.1.9 T2-weighted images ..... 65
2.1.10 T2*-weighted images ..... 66
*2.2 Functional MRI* ..... *66*
2.2.1 Blood-oxygen level dependent fMRI ..... 66
2.2.2 The haemodynamic response ..... 67

2.2.3 Echo-planar imaging for fast fMRI acquisition ..... 67
2.2.4 Model-based fMRI analysis ..... 67
2.2.5 'Sparse' scanning ..... 68
2.2.6 Data preprocessing ..... 69
2.2.6.1 Brain extraction ..... 70
2.2.6.2 Temporal filtering ..... 70
2.2.6.3 Motion correction ..... 70
2.2.6.4 Spatial smoothing ..... 71
*2.3 Statistical analysis ..... 71*
2.3.1 The General Linear Model ..... 72
2.3.1.1 The dependent variable ..... 72
2.3.1.2 The independent variable ..... 72
2.3.1.3 Parameters ..... 72
2.3.1.4 Error ..... 72
2.3.1.5 The GLM Matrix ..... 72
2.3.2 First-level analysis ..... 73
2.3.3 Fixed-effects analysis (intermediate level) ..... 74
2.3.4 Mixed-effects analysis (higher level) ..... 74
2.3.5 Thresholding ..... 75
*2.4 Registration ..... 75*
*2.5 Region of interest analysis ..... 77*
*2.6 Scanning protocol ..... 78*
2.6.1 Scanner parameters ..... 78
2.6.2 Structural T1 ..... 78
2.6.3 Functional MRI ..... 78
**3 Using overt and covert speech to investigate motor-sensory speech control ..... 79**
*3.1 Aims and hypothesis ..... 82*
*3.2 Materials and methods ..... 82*
3.2.1 Participants ..... 82
3.2.2 Experimental conditions ..... 82
3.2.3 Behavioural performance ..... 85
3.2.4 Data acquisition ..... 85

3.2.5 Data analysis ..... 85
3.2.6 Region of Interest analysis ..... 85
*3.3 Results ..... 88*
3.3.1 Behavioural performance ..... 88
3.3.2 Functional MRI analysis ..... 88
3.3.2.1 Main effect of Overt/Covert ..... 89
3.3.2.2 Main effect of Speaking/Counting ..... 90
3.3.2.3 Overt Speaking > Rest ..... 91
3.3.2.4 Covert Speaking > Rest ..... 91
3.3.2.5 Overt Counting > Rest ..... 91
3.3.2.6 Covert Count > Rest ..... 91
3.3.2.7 Summary of the comparison of conditions with Rest ..... 94
3.3.2.8 Direct comparisons between conditions ..... 95
3.3.3 ROI analysis ..... 97
*3.4 Discussion ..... 99*
**4 A comparison of motor-sensory activity during speech in first and second languages ..... 102**
*4.1 Aim and hypotheses ..... 103*
*4.2 Materials and methods ..... 103*
4.2.1 Participants ..... 103
4.2.2 Behavioural assessment of language proficiency ..... 104
4.2.3 Functional MRI experimental tasks ..... 104
4.2.4 Materials ..... 105
4.2.5 Speech Rate ..... 107
4.2.6 Regions of interest ..... 107
*4.3 Results ..... 108*
4.3.1 Language proficiency ..... 108
4.3.2 Functional localisers ..... 109
4.3.3 Functional MRI whole-brain analyses of speech conditions ..... 110
4.3.3.1 Speech > Rest ..... 111
4.3.3.2 L2 > L1 ..... 112
4.3.4 Region of interest analyses ..... 113
4.3.5 Different L1s ..... 115

4.3.6 Co-varying proficiency behavioural measures ..... 117
4.3.7 Speech rate ..... 117
*4.4 Discussion ..... 119*

**5 Learning to produce non-native speech sounds ..... 125**

*5.1 Aim and hypotheses ..... 126*
*5.2 Materials and methods ..... 126*
5.2.1 Participants ..... 126
5.2.2 Materials ..... 127
5.2.3 Training ..... 129
5.2.3.1 Phoneme discrimination testing ..... 129
5.2.3.2 Functional MRI experimental tasks ..... 129
5.2.3.3 Filtering out CSF and WM ..... 131
5.2.3.4 Regions of interest ..... 131
5.2.3.5 Additional ROI analyses ..... 131
*5.3 Behavioural results ..... 132*
5.3.1 Phoneme discrimination ..... 132
5.3.2 Behavioural performance ..... 132
5.3.3 Speech ratings ..... 133
5.3.4 Speech rating scores ..... 134
5.3.5 Trained and untrained sounds ..... 135
5.3.6 Coefficients of variation ..... 137
5.3.7 Signal-to-Noise ratios for the speech scores ..... 139
*5.4 Whole-brain results ..... 140*
5.4.1 ANOVA: Language (Native and Non-native) and Session (pre- and post-training) for the Listening task ..... 140
5.4.1.1 Contrasts of Non-native Listening with Rest and Native Listening with Rest – Session 1 ..... 142
5.4.1.2 Direct contrasts of Native Listening with Non-native Listening – Sessions 1 and 2 ..... 144
5.4.1.3 Listening ROI analyses ..... 145
5.4.2 ANOVA: Language (Native, Non-native) x Session (Session 1 and Session 2) for the Repeating task ..... 147

5.4.2.1 Contrasts of Non-Native Repeating with Rest and Native Repeating with Rest – Both Sessions ..... 148
5.4.2.2 Direct contrasts of Native Repeating with Non-native Repeating – Sessions 1 and 2 ..... 149
5.4.2.3 ROIs from a whole-brain Language x Session interaction for the Repeating task ..... 151
5.4.3 Interactions for each Non-native language separately with Native ..... 154
5.4.3.1 Cortical region of interest analyses: Repeating ..... 156
5.4.4 Whole-brain results with speech score ..... 157
5.4.4.1 Non-native Repeating > Rest ..... 157
5.4.4.2 Whole-brain results with separate speech scores ..... 157
*5.5 Discussion ..... 159*

**6 Parallel competing systems in the control of speech ..... 165**

*6.1 Aim and hypothesis ..... 165*
*6.2 Methods ..... 166*
6.2.1 Participants ..... 166
6.2.2 Functional MRI acquisition and analysis ..... 166
6.2.3 Functional MRI experimental procedures ..... 166
6.2.4 Data analysis ..... 166
6.2.5 Defining the left perisylvian language area ..... 166
6.2.6 Decomposing the left perisylvian language area into discrete signals .. 168
*6.3 Results ..... 169*
6.3.1 Study 1 ..... 169
6.3.2 Study 2 ..... 175
6.3.3 Study 3 ..... 176
6.3.4 Comparison between studies ..... 177
6.3.5 Study 1: comparing across sessions ..... 180
6.3.6 Study 1: comparing different levels of performance ..... 181
6.3.7 Study 1: Listening ..... 182
*6.4 Discussion ..... 183*

**7 Thesis summary and future directions ..... 188**

*7.1 Summary of results and main findings ..... 188*
*7.2 Future directions ..... 192*

*7.3 Conclusion ..... 196*

**8 References ..... 197**

**9 Appendix ..... 223**

*Appendix 1, Cluster table for the ANOVA in Chapter 3 ..... 223*

*Appendix 2, Cluster table for the speech conditions > Rest in Chapter 3..... 223*

*Appendix 3, Cluster table for the direct contrasts in Chapter3 ..... 224*

*Appendix 4, Cluster table for the functional localisers in Chapter 4 ..... 225*

*Appendix 5, Cluster table for the speech conditions in Chapter 4 ..... 225*

*Appendix 6, Example page of articulatory anatomy from the training materials used in Chapter 5 ..... 226*

*Appendix 7, Example page of spectrograms from the training materials used in Chapter 5 ..... 227*

*Appendix 8, Speech rating correlations for the individual sounds in Chapter 5..... 228*

*Appendix 9, Speech scores for the individual sounds in Chapter 5..... 229*

*Appendix 10, Cluster table for the Listening ANOVA in Chapter 5..... 230*

*Appendix 11, Cluster table for the Listening > Rest conditions in Chapter 5..... 231*

*Appendix 12, Cluster table for the Listening contrasts in Chapter 5..... 231*

*Appendix 13, Cluster table for the Repeating ANOVA in Chapter 5..... 232*

*Appendix 14, Cluster table for the Repeating > Rest conditions in Chapter 5..... 232*

*Appendix 15, Cluster table for the Repeating contrasts in Chapter 5..... 233*

*Appendix 16, Cluster table for the separate Repeating ANOVAs in Chapter 5..... 234*

*Appendix 17, Cluster table for the whole-brain results with speech score improvements for Non-native Repeating and each Language separately, in Chapter 5 ..... 235*

# List of figures

*Figure 1.1, Universal and language-specific stages of speech development* ..... 29

*Figure 1.2, Schematic of cortical motor-sensory regions involved in speech production that are the focus of this thesis* ..... 39

*Figure 1.3, The major regions within the basal ganglia* ..... 41

*Figure 1.4, The song system of songbirds* ..... 50

*Figure 2.1, Properties of protons with spin outside and in the magnetic field* ..... 63

*Figure 2.2, T1- and T2-weighted images* ..... 65

*Figure 2.3, A Schematic description of continuous and "sparse" imaging* ..... 69

*Figure 2.4, Schematic outline of fMRI data analysis* ..... 73

*Figure 2.5, Individual T1-weighted images and a standard brain template* ..... 77

*Figure 3.1, Regions of interest* ..... 87

*Figure 3.2, Main effect of Overt/Covert* ..... 89

*Figure 3.3, Main effect of Speaking/Counting* ..... 90

*Figure 3.4, Whole-brain results for three speech tasks, each against Rest (legend overleaf)* ..... 92

*Figure 3.5, Three speech tasks, each against rest* ..... 94

*Figure 3.6, Overt and Covert direct contrasts* ..... 96

*Figure 3.7, ROI analyses for the four experimental conditions* ..... 98

*Figure 4.1, Examples of stimuli* ..... 105

*Figure 4.2, Functional localiser results* ..... 110

*Figure 4.3, Speech > Rest* ..... 111

*Figure 4.4, L2 > L1* ..... 112

*Figure 4.5, ROI results* ..... 114

*Figure 4.6, L2 > L1 signal change for different L1s in ROIs (legend overleaf)* ..... 116

*Figure 4.7, Speech rates for L1 and L2 and a Speech-rate x Language ANOVA* ... 118

*Figure 5.1, Stimuli presentation for the Listening and Repeating conditions* ..... *130*

*Figure 5.2, Phoneme discrimination results, all languages* ..... *132*

*Figure 5.3, Rater correlations, all languages* ..... *134*

*Figure 5.4, Speech scores from all languages in both Sessions* ..... *136*

*Figure 5.5, Coefficients of variation* ..... *138*

*Figure 5.6, Signal-to-noise ratios for the speech scores in both Sessions* ..... *139*

*Figure 5.7, Listening ANOVA* ..... *141*

*Figure 5.8, Contrasts of Non-native Listening with Rest and Native Listening with Rest, Session 1* ..... *143*

*Figure 5.9, Native > Non-native Listening Session 1 and Non-native > Native Listening Session 2* ..... *145*

*Figure 5.10, ROI analyses of Native and Non-native Listening in both Sessions* .... *146*

*Figure 5.11, Repeating ANOVA* ..... *147*

*Figure 5.12, Non-native Repeating > Rest and Native Repeating > Rest* ..... *148*

*Figure 5.13, Non-native Repeating > Native Repeating, and Native Repeating > Non-native Repeating in both Sessions* ..... *150*

*Figure 5.14, Repeating Language x Session interaction and ROIs* ..... *151*

*Figure 5.15, ROI mean time series for Native and Non-native Repeating* ..... *152*

*Figure 5.16, ROI activity across all runs* ..... *153*

*Figure 5.17, Interactions for each Non-native language separately with Native* ..... *155*

*Figure 5.18, ROI percent signal change for Repeating* ..... *156*

*Figure 5.19, Non-native Repeating, with speech score* ..... *157*

*Figure 5.20, Repeating individual languages, with speech score* ..... *158*

*Figure 6.1, ROI masks* ..... *167*

*Figure 6.2, Methods used for finding sub-regions within the left perisylvian language area and finding whole-brain functional connectivity with these sub-regions* ..... *169*

*Figure 6.3, Whole-brain connectivity maps* ..... *171*

*Figure 6.4, Activity for the temporal and parietal signals in response to different speech production tasks* ..... *173*

*Figure 6.5, Default mode network component* ..... *174*

*Figure 6.6, Whole-brain connectivity maps for the two components for the three datasets* ..... *176*

*Figure 6.7, Qualitative similarity of the whole-brain functional connectivity patterns for the temporal (pSTG) component* ..... *178*

*Figure 6.8, Qualitative similarity of the whole-brain functional connectivity patterns for the parietal (vAPL) component* ..... *179*

*Figure 6.9, ROI activity for Repeating all languages for both components in Session 2* ..... *180*

*Figure 6.10, Activity for different levels of performance for both components in both Sessions* ..... *181*

*Figure 6.11, Listening for both components in both Sessions* ..... *182*

*Figure 7.1, A schematic diagram of cortical motor-sensory systems involved in speech production* ..... *191*

*Figure 7.2, Masks of the five components from the ICA, shown on sagittal slices of the left hemisphere* ..... *194*

*Figure 7.3, Connectivity patterns and task responses for components from a Broca's area mask (legend overleaf)* ..... *194*

# List of tables

*Table 3.1, Stimuli list with values for linguistic variables ..... 84*

*Table 3.2, Number of excluded trials from participants who made incorrect responses during the fMRI speech tasks..... 88*

*Table 3.3, Significant t-tests following significant ROI ANOVAs..... 98*

*Table 4.1, Details of stimuli used from the International Picture Naming Project.... 106*

*Table 4.2, Behavioural proficiency scores ..... 109*

*Table 5.1, List of non-native words and native non-words in each Session ..... 128*

*Table 5.2, Number of missed repeat trials..... 133*

# 1 Introduction

This thesis examines the response of motor-sensory neural systems during speech production in native (L1) and non-native (L2) languages. The research investigated changes in the neural systems as a result of proficiency, using behavioural methods and functional neuroimaging of healthy volunteers.

For adults, speaking is a highly automatic skill; in a native language, speakers do not have to concentrate on how they move their articulators (the lips, tongue and jaw), they just speak. However, despite fully mastering one language in childhood, adults find it difficult to speak a second language, even after years of training. L2 is spoken with an accent that clearly marks the speaker as non-native and even after years of practice there is little prospect of achieving native-like proficiency (Clarke and Garrett, 2004; Flege, 1995). In contrast, children can acquire two or even more languages at the same time and are able to master them to approximately equal proficiency, with little or no foreign accent. The criteria for 'being good at speaking' are numerous, and even in the target language there are regional variations; yet it is possible to use functional imaging to investigate language proficiency in terms of the changes in neural network function that underlie changes in behavioural measures.

Despite the interest in imaging the neural systems supporting the execution and sensory monitoring of speech, this has only recently been extended into bilingualism and the effects of speaking a second language with an accent. Investigating why it is so much more difficult to learn a foreign language as an adult and what can be done to make the learning process more successful will help improve the effectiveness of foreign language teaching and learning. There are also implications for language acquisition in general. As well as the interest in understanding the differences between native speech production and speaking a foreign language with an accent, the studies presented in this thesis offer the potential for insights into the compensatory mechanisms that may be engaged in patients with motor speech disorders, such as dysarthria, resulting from both vascular and neurodegenerative diseases.

The bilingual participants included in Chapter 4 of this thesis were successive bilinguals, meaning that L2 was learnt after L1 was already established. Although these bilingual speakers had reached high levels of linguistic proficiency in L2 (English for these participants), the persistence of a foreign accent showed that it was clearly still a non-native language. Simultaneous bilinguals, who acquired L1 and L2 concurrently in childhood and have balanced proficiency and native accents in both languages, may not demonstrate the motor and sensory differences that I present in this thesis. Therefore, those speakers were not included in the studies here, and a separate study will be required to investigate motor-sensory processing in balanced bilinguals. Young learners who acquire a second language in childhood and go on to reach native proficiency sometimes stop using their L1 and only use L2, in effect becoming monolingual again, but with a different language. This requires different processing skills than those bilingual speakers who master a second language while maintaining their first (Snow, 2002), so, again, these speakers were not investigated here. Chapter 4 of this thesis focuses on bilingual speakers who learnt L2 after L1 had been established, and maintained L1 whilst acquiring L2, and Chapter 5 investigates the motor-sensory learning of foreign speech by monolingual speakers of English.

In the first section of this Chapter I briefly review previous research on bilingualism, to place the motor-sensory aspects of speaking in L2 in context. I then consider how learning to speak L2 requires the retuning of the neural circuits involved in motor control of articulation, and the effects on auditory and somatosensory feedback systems. I also discuss why the system rarely becomes so finely and accurately tuned that L2 can be spoken without an accent. Finally, I consider the pedagogical and clinical implications of taking a motor-sensory perspective on speech production.

## 1.1 Speech production

### 1.1.1 Linguistic aspects of speech production

A range of linguistic stages are involved in speech production. Speech is the final expression of concepts and emotions, translated through linguistic pathways that involve lexical, syntactic, phonological and phonetic stages (Levelt, 1989), as well as prosody. These stages have been defined and refined over many decades using

behavioural measures in normal participants (Levelt et al., 1999) and lesion-deficit analyses of patients with focal lesions (Shallice, 1988). More recently the links between linguistics and brain anatomy and function have become intensively studied in normal participants, first with positron emission tomography (PET) and then with functional magnetic resonance imaging (fMRI). Such functional neuroimaging studies have enabled large meta-analyses that have considered native speech comprehension in terms of semantics, syntax and phonology (Vigneau et al., 2006); (Binder et al., 2009), and native single word speech production in more precise terms of lemma retrieval and selection, phonological code retrieval, syllabification and final motor output (Indefrey and Levelt, 2004).

### 1.1.2 Motor-sensory aspects of speech production

The linguistic studies mentioned above have been followed by investigations of the motor and sensory representations and processes that control speech production, with studies designed to investigate speech-related breath control (Loucks et al., 2007), laryngeal function (Simonyan et al., 2009) and articulatory movements (Sörös et al., 2010). Speech production is a complex motor-sensory process, involving rapid sequential motor movements, with transitions occurring over tens of milliseconds. It relies on the integration of feedforward motor and feedback sensory signals, with online self-monitoring guiding rapid modification of motor commands to the larynx, pharynx and articulators, requiring coordination of up to 100 muscles (Ackermann and Riecker, 2010). This integration allows the maintenance of intelligible speech, even under adverse speaking conditions. It depends on motor (frontal), auditory (temporal) and somatosensory (parietal) cortex, as well as the insular cortices, subcortical nuclei and cerebellum (Guenther, 2006; Ventura et al., 2009) (Golfinopoulos et al., 2010).

When a speaker plans to say something, they initiate a motor plan. This sends commands to the articulators, instructing them how to move to produce the speech sound. Feedforward signals from premotor and motor cortex, as well as modulatory contributions from many other regions, control air flow through the larynx, tension of the vocal folds and the movements of the articulators (Gracco and Lofqvist, 1994). The final production of the intended acoustic signal is carried out by shaping of the vocal tract (Nieto-Castanon et al., 2005). Concurrent with these motor processes,

efference copies of the intended motor plan are fed forward to auditory and somatosensory cortices. These cortices, in superior temporal and anterior parietal regions, respectively, receive the auditory and somatosensory information associated with speaking, delayed by several milliseconds after motor execution, to provide retrospective online monitoring of the sensory consequences of the utterance. If there is a mismatch between the intended and actual sensory consequences, subsequent rapid motor adjustments can be made to maintain fluent speech (Golfinopoulos et al., 2010; Guenther et al., 2006). Feedback monitoring may also involve other regions. For example, the involvement of insular cortex in the McGurk effect (Bamiou et al., 2003) suggests that it plays a role in multi-modal processing, such as the integration of auditory and visual inputs during speech perception, and mid-insular cortex may also contribute to the integration of auditory and somatosensory feedback during speech production (Dhanjal et al., 2008).

When speakers use their native language, both somatosensory and auditory feedback systems show some suppression of activation during speech production (Aliu et al., 2008; Paus et al., 1996). Although auditory association cortex is active during overt speech, this activity is less than when listening to the voice of another (Ventura et al., 2009; Wise et al., 1999). This reduction in activity reflects cortical modulation by a parallel copy of the motor command, which could originate from premotor cortex, primary motor cortex or both, sent to sensory cortical regions processing articulatory feedback signals. This allows more efficient feedback-appropriate sensory processing (Eliades and Wang, 2008), permitting the rapid detection and correction of articulatory errors (Golfinopoulos et al., 2010; Guenther et al., 2006).

### 1.1.3 Self-monitoring in L1 and L2

For both adults and children, self-monitoring is involved in speech production, when the expected sensory signal is compared with the actual (Golfinopoulos et al., 2010; Guenther et al., 2006). Speech sounds are defined primarily by their acoustic structure, and so self-monitoring is normally considered to depend on the auditory domain (Directions Into Velocities of Articulators (DIVA) model - (Bohland and Guenther, 2006). However motor movements also generate somatosensory feedback, and recently it has been shown that whereas some speakers rely on

auditory feedback information, others seem to rely more on somatosensory feedback (Lametti et al., 2012). Thus, feedback systems monitor the sensory consequences of a speech motor command (based on the shape, height, position or protrusion of the articulators), as well as how the speech sounds, and compare these to the expected consequences of that command (Golfinopoulos et al., 2010). A prediction of sensory consequences of a movement is made, based on the efference copy of the motor command, and once the movement has been made its consequences are compared with the sensory prediction. The sensory feedback that occurs in response to any skilled self-produced movement is generally the same as expected feedback (Blakemore et al., 2001). The DIVA computational model is based on the premise that speech acquisition relies on connections between dual-sensory (auditory and somatosensory) feedback and feedforward motor control (Golfinopoulos et al., 2010). Improved proficiency in speech production is also linked to an ability to distinguish different speech sounds (Nasir and Ostry, 2009), and the ability to distinguish between the acoustic properties of the different phonemes of one's culture-specific language (L1) in infancy contributes to vocal learning. If two phonemes are not represented in L1, but are in L2, then a native speaker of L1 who acquires L2 later in life may not distinguish between these two phonemes when producing them; the classic example being the difficulties native speakers of Japanese have with the English phonemes /l/ and /r/. Therefore, imperfect auditory 'templates' result in imperfect speech production, and the quality of speech in L2 depends on a strong interplay between motor and sensory processes. The focus of my thesis is on the neural adaptations in motor and sensory systems as normal participants speak in a non-native language in comparison with when they speak in their native tongue.

Much of the work presented here, particularly the region of interest analyses, investigates the temporo-parietal junction. Interest in the function of this region – the planum temporale (the posterior auditory association cortex in the supratemporal plane) and parietal operculum (the location of somatosensory association cortex) - during speech has become a major topic in the speech and language literature (Hickok et al., 2009). Predictive feedforward and post-articulatory sensory feedback signals are, it is proposed, integrated in this region, although this assertion has largely been based on L1 speech production. The left planum temporale and adjacent parietal operculum has been proposed as a 'sensorimotor interface' (area

Spt) for speech production (Buchsbaum and D'Esposito, 2008; Hickok et al., 2003; Hickok et al., 2009; Simmonds et al., 2011a). The studies presented in this thesis investigate this motor-sensory interface in native and non-native speech production, and its response to the learning of foreign speech sounds.

## 1.2 The representation of multiple languages in the bilingual brain

Opposing views on the neural organisation and representation of multiple languages in bilinguals exist. One key hypothesis is that, in a bilingual brain, different languages recruit different neural networks. This idea was first proposed in 1867 when Scoresby-Jackson claimed that L1 located to Broca's area and other languages spread forward from there (Ahlsen, 2006). Bilingual aphasia research has provided indirect evidence in support of this view, with some studies reporting decreased performance in only one language, with the other language(s) remaining unaffected (Abutalebi et al., 2001), as have reports of single language impairments following neurosurgery (van den Noort et al., 2006). Kim and colleagues (1997) and Kovelman and colleagues (2008) report imaging studies that suggest separate systems for L1 and L2; the latter group proposing that the linguistic properties of a language, in particular morphosyntactic differences, may affect the organisation of that language in a bilingual brain. Kim and colleagues (1997) also found separate activation patterns for first and second languages, but instead of discussing linguistic properties of individual languages, they argue that the age of acquisition of the second language leads to different representations in frontal-lobe language areas. They found that early bilinguals used similar areas for both languages, whereas late bilinguals had spatially separated language representations. The majority of imaging studies, however, provide evidence for a single language network for L1 and L2 (Briellman et al., 2004; Chee et al., 1999; Hernandez et al., 2000; Illes et al., 1999). Callan and colleagues (2004) and Chee and colleagues (2003) also suggested overlapping networks for both languages, with a more extensive system for L2.

A further hypothesis regarding the neural representations of different languages in a bilingual brain goes beyond a simple same/different dichotomy, and proposes that bilinguals develop separate centres for different language functions implicated in bilingualism, such as language selection, switching and translation (Potzl, 1930, in

(Ahlsen, 2006)). However, this has been rejected by Jakobson (1941) and Paradis (2004), who claim that a bilingual's ability to select and switch languages is no different than a monolingual's ability to select the appropriate language register for various social situations (in Ahlsen, (2006)).

Van den Noort and colleagues (2006) suggest two reasons that might explain these differing findings, namely the range of experimental paradigms used in various studies and the type of participants included. Other factors that may also affect results include levels of proficiency, age of acquisition and levels of exposure to all languages concerned (Abutalebi et al., 2001). As they explain, it is widely accepted that children acquire a second language more easily than adults. Second language vocabulary can be thoroughly acquired by adults, although phonological and morphological aspects of the second language are more problematic for adults than children. These points are discussed in more detail throughout this introduction.

## 1.3 Previous research on bilingualism: linguistic and cognitive aspects

Articulation interacts with syntactic, lexical, phonological and phonetic constraints and previous functional imaging studies of bilingualism have largely concentrated on these aspects and much less on the motor and sensory aspects. Here I consider why the control of articulation during L2 production has received so little interest in the functional neuroimaging literature in the past.

Learning a new language is normally understood as adaptations of the neural systems involved in linguistic processing; that is, those involved in acquiring a new syntax and lexicon (Francis, 1999; Kovacs and Mehler, 2009; Kovelman et al., 2008), and those components of domain-general cognitive control systems involved in translation and switching between languages (Abutalebi and Green, 2007; Costa and Santesteban, 2004; Crinion et al., 2006; Hernandez et al., 2000; Price et al., 1999; Rodriguez-Fornells et al., 2002). Age and order of acquisition and/or proficiency in the different languages, as well as the way in which new languages are learnt, and the modality (signed/spoken) are all considered to play a role in how L2, L3, etc. are represented in the brain relative to L1, and the control processes that operate on these representations (Vaid and Hull, 2002). The work presented in this thesis is

novel as it was designed to determine the motor-sensory aspects of bilingual speech production, and not the linguistic or cognitive processing. Bilingual fMRI studies of both production and comprehension at the word-, sentence- and discourse-level, as well as inflectional morphology have been reviewed by Indefrey (2006). Control mechanisms in bilingual language production have been reviewed in the article by Abutalebi and Green (2008). Abutalebi and Green (2008) specify a model that integrates separable neural systems responsible for distinct aspects of cognitive control involved in bilingual speech production. These systems include the prefrontal cortex (updating the language, inhibition of the language not in use and error correction), the anterior cingulate cortex (attention, conflict monitoring and error detection), the basal ganglia (language selection) and the inferior parietal lobule (maintenance of representations and working memory). Abutalebi and colleagues (2009) have also shown how this model can be applied to bilingual aphasic research, using dynamic causal modelling, in combination with behavioural and imaging data. This cognitive control model is the predominant neurocognitive model that exists in the bilingual literature, and the model of speech production I discuss in Chapter 7 of this thesis complements this cognitive control model.

### 1.3.1 Differences between first and second language acquisition

From a theoretical perspective, the stages by which first and second languages are acquired differ, especially when the latter begins once the former has already been established. In this vein, acquisition and learning can be defined as two separate processes (Krashen and Terrell, 1983). Acquisition can be thought of as an implicit process that enables the speaker to develop functional skills without theoretical knowledge. In contrast, learning can be thought of as knowing about the language, a more explicit and conscious process. In the late 1970s, a distinction between first and second language acquisition was proposed by Lamendella (1977), who defined primary language acquisition (PLA) as a child's acquisition of one or more languages, before the age of 5 years. Secondary language acquisition (SLA) after the age of 5 years was defined as both explicit formal foreign language learning and the natural acquisition of another language (acquired without formal instruction, but through frequent exposure and use). Lamendella suggested that PLA is largely dependent on innate neural systems, whose plasticity declines after a critical period, after which different neural systems are recruited when learning a language. Thus, the

hypothesis is that there are two "hierarchies" of language processing: communicative and cognitive. PLA and natural SLA use the communicative hierarchy, and formal SLA uses the cognitive hierarchy (with the implication that this is more dependent on top-down frontal executive control systems). Functional imaging research into bilingualism has investigated both these hierarchies, even if not explicitly. These hierarchies cope more or less well with different stages of linguistic processing; for example, the vocabulary of L2 can be thoroughly acquired by adults but its phonological and morphological aspects are less problematic for children.

For a child, language acquisition begins with speech perception. Initially, the speech-perception skills of infants are language-general, which offers the potential to acquire any human language to which the child is exposed. The phonetic repertoire of a language is based on both the individual consonant and vowel sounds and the permissible combinations of these sounds in the creation of words and phrases (Jacquemot et al., 2003). A child acquiring their first language goes through various stages of linguistic development that correlate with brain volume increases during the first years (Sakai, 2005) and occur within a set time frame for that particular language (Kovacs and Mehler, 2009). Before an infant is six months old, their perceptual system has become tuned to the phonetic repertoire of their native language, and making distinctions between non-native phonemes may be more difficult (Kuhl, 2004) (see Figure 1.1).

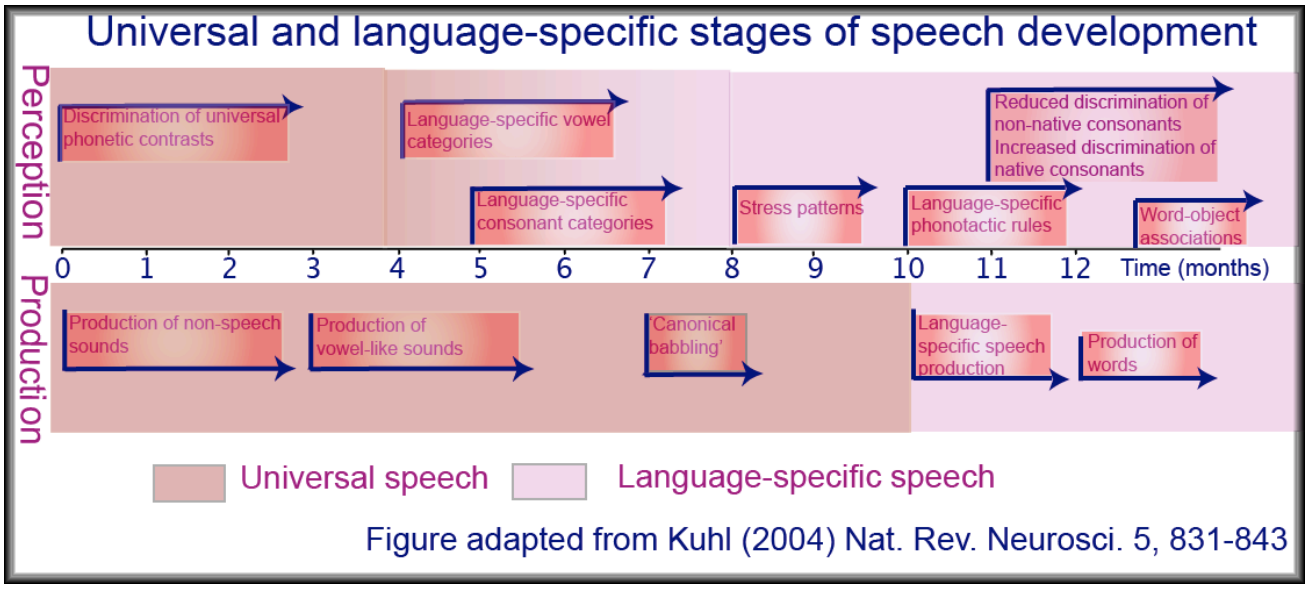



***Figure 1.1, Universal and language-specific stages of speech development***

*Universal and language-specific stages of development for speech-perception and speech-production in typically developing human infants from birth to one year. Published in Simmonds et al. (2011b).*

This early stage is apparent as a 'silent period', during which the infant listens to language without attempts to produce speech sounds. Before any meaningful utterances are made, children first practise moving their articulators and produce sounds in the form of babbling. These meaningless linguistic sounds enable the development of motor-sensory integration, and by listening to and feeling the response to sounds, children develop speech production skills. Acquisition of language occurs by matching speech sounds to articulatory gestures required to produce those sounds (Buchsbaum et al., 2005). By the time the child is nine months old its babbling is becoming language-specific, and a skilled child developmental psychologist can tell the difference between the babblings of children from different language cultures. This is the earliest evidence of language-specific motor-sensory processing. By one year, babbling turns to speech, beginning with single words, followed by short phrases and then sentences. Although linguistic errors are frequent at this stage, the accent of the child is clearly that of a native speaker. In contrast, those who learn a second language as an older child or an adult speak with an accent that clearly marks them as a non-native speaker.

These observations indicate some form of ‘critical periods’ in language development, and after these periods the cortical plasticity of the motor and auditory systems become more limited. The formation of representations of foreign speech sounds may be less accurate later in development, so that the learner may not be able to perceive and represent the auditory nuances of certain vowels and consonants that are readily detected by a native speaker. Sounds in L2 that do not exist or are ‘phonologically ungrammatical’ in L1 are often assimilated into the closest acceptable form in the native language (Dupoux et al., 1999).

The acquisition of L2 production is quite different from that for L1. There is no babbling stage, and the learners acquire new words with the explicit knowledge about their meaning. There is also a strong tendency to translate a word in L2 into its corresponding word in L1 when listening (Thierry and Wu, 2007; Wu and Thierry, 2010), and both L2 and L1 phonological representations are retrieved during covert word production in L2 (Wu and Thierry, 2011). Second language learners may also ‘fossilise’ at a relatively early stage, when the speaker feels that they have developed L2 sufficiently for their purpose, reflecting the more explicit nature of L2 acquisition and an acceptance of what is enough to ‘get by’. This means that they may not strive to tune their auditory cortex to generate accurate, long-term representations of L2 syllables, words and phrases, as spoken by a native speaker. L2 consonant or vowel strings that are not permissible in the L1 may be corrected through insertion of an ‘epenthetic’ vowel, or through substitution of a sound (Jacquemot et al., 2003). Therefore, the online perceptual monitoring of what they are producing is never sufficiently fine-tuned to drive improvements in the long-term motor representations of articulation. The same factors that are used to investigate linguistic processing differences between two languages, such as the age of acquisition, the level of proficiency reached, the amount of exposure to the second language and the degree of language use in everyday life (Abutalebi et al., 2001), also apply to the differences in motor and sensory representations for the two languages.

### 1.3.2 Age of acquisition and proficiency

Second language acquisition research has undergone many changes since it was first established. In the 1950s there was a shift from considering language acquisition as learning parrot-fashion to viewing it as a cognitive process, by which a limited

resource, i.e. vocabulary, can be combined and used to express an unlimited range of concepts. The two most common claims for late bilingual learners relate to the actual age of acquisition and the level of proficiency attained. The debate about age of acquisition began in 1967 when Lenneberg first put forward his “critical period hypothesis”, which proposed a specific time period in which language acquisition can occur, with only a language learnt before puberty being mastered to native proficiency (Lenneberg, 1967). The concept of this ‘critical period’ was based on the idea that certain unspecified ‘electro-chemical’ changes in the brain had reached maturity by the age of 10-12 years, after which implicit language acquisition can no longer occur (Lenneberg, 1967). Lateralisation of cognitive functions has also been proposed as support for critical periods, as hemispheric specialisation is established by early puberty. These proposals carry the inherent suggestion of loss of neocortical plasticity with maturation, meaning that with development language acquisition becomes increasingly difficult.

The critical period hypothesis was considered in more detail in the 1970s when Krashen re-examined Lenneberg’s data and concluded that whilst the critical period exists, it ends much sooner than Lenneberg suggested. Krashen proposed that the process of first language acquisition is complete by the age of 5 years and a second language learnt after that period would not reach native-like proficiency (in Danesi (1994).

The motor and sensory aspects of bilingual speech production are clearly susceptible to some sort of ‘critical period’, in that it is much easier for younger children to learn a second language. I have already discussed the tuning of auditory cortex to language-specific speech sounds during the first year of life, and this must impact on the phonological competence accompanying the acquisition of L2, both in its perception and production (Scovel, 1969). Therefore, certain linguistic skills (vocabulary and grammar) are not as susceptible to age of acquisition limitations. However, it has been suggested that certain aspects of grammar acquisition are also less plastic than vocabulary acquisition; while semantic development uses associative learning mechanisms that are adaptable to later L2 learners, syntactic development uses a computational mechanism that is less plastic (Neville and Bavelier, 2001; Scherag et al., 2004). Scovel (1988) redefined the critical period hypothesis to apply to a specific

age (before puberty), a specific neurobiological change (lateralisation) and a specific linguistic skill (the ability to sound like a native speaker). He claimed that native-like pronunciation is not possible for adult second language learners. There are, of course, individual differences in levels of proficiency reached by later L2 learners, even with similar amounts and types of language training. It has previously been shown that there are brain structural differences between individuals that partly predict their ability at perceiving non-native speech sounds (Golestani et al., 2007).

The notion of critical periods for motor control (as involved in control of the articulators) can apply to fields outside language. It has been suggested that for experts in fields requiring great muscular dexterity (music, dance, skating, etc.), the acquisition process began in childhood, and the same could be said of language experts (Archibald, 1988). Penfield and Roberts (1959) also suggested that speech is not the only skill that is better acquired in childhood; it is also true for talents as diverse as piano and violin playing and skiing. With any general adult skill acquisition, not just foreign language learning, performance is variable, with different strategies for and types of learning, and this is entirely expected (Bley-Vroman, 1990). The critical period hypothesis, especially as applied to speech production, might have less to do with cognition and more to do with fine motor control. The articulation of human speech uses the many muscles that control breathing, the larynx and the articulators, and so it could be that the reason adults have difficulty in mastering L2 speech is due to declining motor dexterity. This would explain why the mastery of the cognitive aspects of language, such as syntax and vocabulary, remains possible for adult learners, whereas control of pronunciation reaches a level of proficiency below that of native speakers. Pronunciation is the only ‘physical’ part of language with complex neuromuscular demands (gestures and handwriting use simple movements compared with speech production), and correct pronunciation is strongly dependent on sensory feedback of how and where the articulators are moving, with specific timings and sequences (Scovel, 1988). Other language aspects are ‘cognitive’ or ‘perceptual’, rather than ‘physical’ (Scovel, 1988). Scovel suggested that it is self-evident that the motor expression of language would be most affected by the loss of neural plasticity that is hypothesised to occur with age. Even those who dispute the ‘critical period’ hypothesis in general often accept that pronunciation may be one aspect in which the hypothesis could be valid (Walsh and Diller, 1978). This view is

also supported by Long (1990), who has proposed that the age of a learner affects phonological attainment, with supra-segmental phonology being possible up the age of six years but the ability to acquire segmental phonology ending soon after that. However, Flege (1981) argues against this view, and suggests instead that accents arise not as a result of loss of plasticity but rather incorrect use of acoustic models of L2, due to interference from L1 (in Scovel, 1988). Under this psychomotor view, it could be argued that accents can be overcome if learners adapt their phonetic model of L2 phonemes to be less affected by L1 phonemes. However, Flege does point out that even if an L2 speaker could adapt their pronunciation to be more closely matched to the L2 phonology, there would still be differences between their pronunciation and that of a speaker for whom the L2 is a native language, even if these differences are only detectible using highly detailed acoustic analyses. He also claims that such phonetic learning of the L2 would affect pronunciation in the learner's L1. Other arguments against Scovel's biological constraints include the suggestion that accents can arise as a result of an adult L2 learner's attitude and lack of motivation or discipline (Taylor, 1974), sociocultural expectations of language learning (Hill, 1970), and 'cognitive maturation', which causes adults to learn L2 differently from children (Dulay et al., 1982).

There are occasional exceptions to a 'critical period' as a hard biological constraint to later acquisition of a second language. A few adult learners of a second language manage to do so without an accent; and, in contrast, there are young learners who have slight accents (Flynn and Manuel, 1991). Abilities in foreign speech sound learning and articulation vary according to the individual, and previous work has shown that these individual differences correlate with brain structural differences in left insula/prefrontal cortex, left temporal cortex and bilateral parietal cortices (Golestani et al., 2006; Golestani et al., 2007). Salisbury (1962) and Sorensen (1967) also reported adult learners who reached native-like proficiency in several languages, in contexts where multilingualism is necessary, such as in New Guinea and the Northwest Amazon. The social context of the language learner affects the level of proficiency attained. In contexts where it is necessary to speak as a native, the speaker will continue to progress, rather than fossilising at the level of adequate communication, albeit with a non-native accent. To understand the neural basis of

these exceptions requires research into the motor and sensory control of bilingual speech production.

In addition to age of acquisition, it has been suggested that language networks are affected by proficiency level. Abutalebi and colleagues (2008) demonstrated that activity in prefrontal regions reduces as the level of proficiency increases. Using a picture-naming task with German-French bilinguals, they found that when using the less proficient language, activation in the left caudate and anterior cingulate cortex was more extensive. One interpretation is that the processes required to produce language become more automatic, requiring less domain-general executive control, as the language becomes more familiar; in the jargon, the production becomes less goal-directed and more habitual. A number of studies have shown greater activation for lower proficiency (less proficient bilinguals compared with more proficient bilinguals, or the less proficiently spoken language of a bilingual compared with their more fluent language) (Chee et al., 2001; Golestani et al., 2006; Stein et al., 2009). Several studies have also shown greater activation in prefrontal regions for tasks that require greater 'top-down' processing (Frith et al., 1991; Raichle et al., 1994). These studies have also shown that as a task becomes more automated and the processing is more 'bottom-up', frontal activity decreases and activity in more posterior regions increases. The networks involved in bilingual speech production are more extensive than those concerned with linguistic processing; the increased activation for less proficient language production likely relates to cognitive processing as well. In line with this, processing a less fluent language can be considered more effortful and top-down, whereas a more fluent language is more automatic and bottom-up. This fits with Green's (2003) 'convergence' hypothesis, which states that convergence is possible because networks adapt; as proficiency in L2 increases, the representation of L2 and its processing profile converge with those of native speakers of that language. Qualitative differences between native and L2 speakers disappear as proficiency increases.

### 1.3.3 Cognitive differences in adult L2 learners

It has also been argued, however, that not only do adults have the ability to acquire a second language as proficiently as young learners, they can be even more successful language learners. In some respects adults can be considered to have

improved language-learning capabilities (Walsh and Diller, 1978). They can do better at certain cognitive levels, such as those involving grammatical and semantic complexity, as the neural systems responsible for these processes develop with age (Schleppegrell, 1987). Older learners also have the advantage of a well-established first language, and they have the ability to integrate L2 with what they already know about L1. Their cognitive systems are more highly developed than those of young learners, enabling them to make higher-order associations and generalisations (Schleppegrell, 1987). It has been shown that the skill of phonetic learning is stable within individuals and has structural correlates (Golestani et al., 2007). These cognitive differences in bilingualism are still being widely researched by a number of research groups.

## 1.4 Why has previous bilingualism research largely ignored motor-sensory aspects of learning?

It is usually obvious when a late bilingual is speaking their late-acquired L2, because of their accent. Research into second language acquisition has considered the degree of accent, but not in terms of motor-sensory control. In this section of the introduction I consider why functional imaging has largely ignored motor-sensory aspects of second language learning and how it can build upon data provided by linguistics research.

### 1.4.1 Interlanguage phonology

A common analysis technique for L2 acquisition research is Contrastive Analysis, comparing two languages by investigating L1 and describing how it is different or similar to L2. The 1960s gave rise to the new notion of interlanguage phonology, in which comparisons are not only made between the target L2 and the L1, but also between the target L2 and variations of L2 that the speaker develops as proficiency increases. Learners have their own interlanguage phonologies, based on temporary rules that they develop throughout the learning process. Although research into phonology has been largely neglected, particularly by the functional imaging community, Alario and colleagues (2010) have recently investigated contrastive phonology in L1 and L2. They found that late bilinguals, but not early bilinguals, were sensitive to non-target syllable frequency. They interpret this by suggesting that syllable representations differ for the two groups of bilinguals. Early bilinguals are

proposed to have independent syllable representations, whereas late bilinguals use the same representation for the two languages. For late bilinguals the syllable representation for L2 is based on earlier L1 experience and consequently L2 representations rely on non-native L1-like patterns. It has been suggested that the failure of non-native speakers to accurately produce L2 speech sounds may be a problem of phonetic implementation (articulation), rather than one of phonological encoding (auditory discrimination of speech sounds). It is apparent that a learner's interlanguage phonology results from adapting the motor-sensory system during the course of acquiring L2. These aspects of non-native speech production could be investigated with functional neuroimaging techniques, especially using more sensitive analyses, such as multivariate pattern analysis (Raizada et al., 2010).

The acquisition of phonetic and articulatory skills must depend as much on sensory as motor processing. The study of these lower-level processes may not be best served by investigating highly constrained speech production tasks. For example, single word tasks, such as naming, reading aloud or repeating, do not reflect well what occurs during self-generated propositional speech. This may explain why earlier studies of motor differences in the production of L1 and L2 only demonstrated altered activity in the basal ganglia (Frenck-Mestre et al., 2005; Klein et al., 1994). Understanding the motor-sensory aspects of speech production, either in L1 or L2, is more sensitively studied with participants producing whole phrases and sentences, rather than stimulus-led single word production, such as naming or repeating.

### 1.4.2 Difficulty assessing speech

Another reason why the motor-sensory aspects of bilingual speech production are overlooked could relate to the difficulty in comparing these lower-level processes. Tasks such as picture naming, direct translation or grammaticality judgments (all of which are frequently used in bilingual imaging studies) are easily assessed in an objective manner. Answers given by the participants are either correct or incorrect, enabling comparison across conditions (i.e. native or non-native languages) only for trials in which the task was successfully completed. In the case of studies investigating cognitive control, responses with longer reaction times or containing errors can be investigated. These trial-by-trial analyses are made possible by the development of event-related fMRI. However, with studies of overt propositional

speech involving sentences, assessment is more subjective. The way in which one judges speaking proficiency is difficult to define. Whilst a native speaker may be able to listen and deem someone to be a ‘good’ speaker, they often find it more difficult to explain why. The notion of fluency can be used, which is being able to communicate a message effectively, in real time, without undue hesitation or delay, as speech is a ‘real-time’ phenomenon (Bygate, 1987). Therefore, although the correct use of vocabulary and syntax, spoken at a normal conversational rate, is central to language proficiency, the same sentence produced by two different proficient bilinguals may vary widely in accent. The problem is to come up with a reliable measure of ‘accent’. Perhaps the most sensitive indicator is a rater scale (Southwood and Flege, 1999), in which a rater assigns a value to their judgement of a speech utterance, with one end of the scale reflecting a strong native accent and the other end a strong non-native accent. Other approaches, such as analysis of speech spectrograms (Arslan and Hansen, 1997) or possibly electropalatograms, may trade improved accuracy in assessing specific motor movements and consequences with a more limited ability to generalise to all the dimensions that characterise human articulation.

### 1.4.3 Scan artefacts of speech production

Motion-induced scan artefacts can be categorised as two types: direct (resulting from head or jaw movements) and indirect signal changes (variations in the magnetic field) (Gracco et al., 2005). Motion artefacts can both mask genuine signal changes due to neural activity and generate apparent ‘activity’ time-locked to speech production. These artefacts arise in fMRI studies of overt speech production, encouraging many researchers to investigate covert speech production instead. A number of functional neuroimaging studies have reported activity within left temporo-parietal cortex, in a region termed Spt, during the covert production of both speech and non-speech vocalisations (Hickok et al., 2009; Hickok and Poeppel, 2000; Pa and Hickok, 2008), and related this to a motor-sensory interface for speech production. However, the investigation of imagined movements is only a partial substitute for the investigation of actual movements, and the results obtained with covert speech need to be directly contrasted with overt speech to fully understand speech-related activity in auditory and somatosensory association cortex. Studies making comparisons between covert and overt speech are discussed in more detail in the introduction to Chapter 3, which

was designed to identify common and different systems involved in both types of speech production. The assumption made when using covert speech is that 'virtual' interactions between sensory and motor speech-related systems become active from the planning of an utterance, even in the absence of actual production.

There are now well-established techniques ("sparse" scanning) to acquire functional data with fMRI while minimising movement- and respiratory-related artefact during overt speech at the level of both single words and sentences (Gracco et al., 2005; Hall et al., 1999). In addition, "sparse" fMRI can be used to minimise auditory masking, as the overt speech is produced during silent periods during which functional images are not being acquired. This is possible because the signal in fMRI, the haemodynamic response function (HRF), which relies on changes in blood flow in response to net regional synaptic activity (the blood oxygen-level dependent – BOLD – signal), extends over many seconds. More detail about sparse scanning is given in Chapter 2 of this thesis, section 2.2.5 (p68).

## 1.5 Evidence for motor-sensory contributions in bilingualism

In speech production models the cognitive planning stages are often described in great detail, whereas articulation is listed as a simple motor output (Indefrey and Levelt, 2004). Models that do provide detail of phonological encoding and articulation, and can encompass developmental changes in size of the articulators, have been developed for monolingual speech production (Guenther et al., 2006; Golfinopoulos et al., 2010). It has been proposed that the bilingual speaker is more than the sum of two monolingual speakers (Abutalebi and Green, 2007) and consequently, monolingual speech production models are not necessarily sufficient to explain bilingual speech production.

### 1.5.1 Cortical contributions to motor-sensory control in bilingualism

There is convincing evidence that the motor system generates internal representations of speech sounds (Wilson and Iacoboni, 2006). In a native language, these representations match auditory input received. The oral motor movements necessary for producing native speech sounds are highly over-learned and automatic, integrating feedforward motor and feedback auditory and somatosensory information. However, in a foreign language, auditory and somatosensory input does

not match the internal representations. Ludlow and Cikaoja (1998) propose that, for a fluent speaker, the internal representation of speech produced by oneself is highly similar to the perception of speech produced by other native speakers. In contrast, when learning a foreign language, perception and internal representations likely match less well. The oral motor movements necessary for non-native phonemes are unfamiliar and require greater engagement of motor-sensory neural feedback systems. Speakers need to become aware of the differences between their internal representation of speech and the perception of speakers of that foreign language. It is then necessary to map that new perception onto their own internal representation, in order to be able to produce accurate speech sounds in the foreign language (Ludlow and Cikaoja, 1998). In this thesis I suggest that auditory feedback is crucial to this process of modifying the speech motor-control system. When learning and speaking a foreign language, it is also hypothesised that online modifications to existing articulatory-acoustic relationships are necessary to produce accurate speech sounds. The persistent accent in late learners of a second language is likely to be the result of a failure to achieve the same proficiency in integrating the motor feedforward and sensory feedback control of articulation that is achieved when speaking in their native tongue. Figure 1.2 shows a schematic diagram of the motor-sensory systems involved in speech production that are the focus of this thesis.

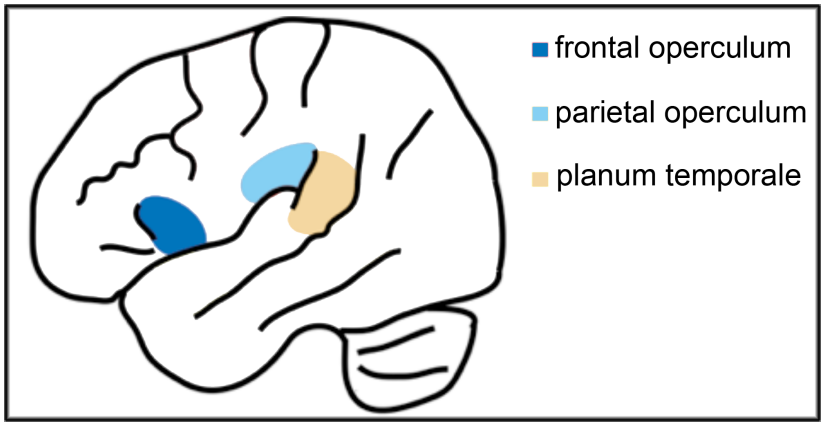



***Figure 1.2, Schematic of cortical motor-sensory regions involved in speech production that are the focus of this thesis***

*A schematic diagram displaying the frontal operculum (dark blue), secondary somatosensory cortex (pale blue) and planum temporale (beige). Adapted from Simmonds et al. (2011b).*

The ease with which a non-native speech sound can be produced ("producibility") affects brain networks activated by perception of this foreign sound. Wilson and Iacoboni (2006) found that listening to non-native phonemes with lower producibility (i.e. those that are difficult to produce) resulted in greater activation in auditory areas. Listening to non-native phonemes also led to higher activation in premotor cortex than native phonemes. The motor system is necessarily involved in speech production and is therefore also affected by producibility. The lower the producibility, the greater the mismatch between the existing internal representation and the incoming sensory signal. Oral motor movements necessary for non-native phoneme production are unfamiliar, whereas those necessary for producing native speech sounds are highly over-learned and automatic (Moser et al., 2009). Moser and colleagues (2009) demonstrated that the production of unfamiliar speech sounds resulted in greater extent and intensity in the BOLD signal, compared to the production of familiar speech sounds. Increased activity in motor speech networks may directly reflect the lack of familiarity with the motor commands necessary to produce the target. Moser and colleagues discuss their results with particular reference to the anterior insula and adjacent frontal operculum, and their roles in the formation and sequencing of articulatory gestures for novel native and non-native speech sounds embedded in non-words. It has been suggested that the insula is involved in allocating auditory attention and is activated more strongly for processing unfamiliar rather than familiar auditory stimuli (Bamiou et al., 2003; Seeley et al., 2007). However, the contribution of anterior insular cortex may be quite different, and relate to a domain-general system, which includes anterior cingulate cortex, known as the saliency network, which exerts control over many forms of non-reflexive motor response (Seeley et al., 2007).

### 1.5.2 Subcortical contributions to motor-sensory control in bilingualism

In addition to the cortical regions discussed above, the motor-sensory learning of foreign speech also relies on subcortical areas. Figure 1.3 presents a schematic diagram of the key regions within the basal ganglia, from which region of interest masks were created to investigate the striatal effects of learning foreign speech presented in Chapter 5.

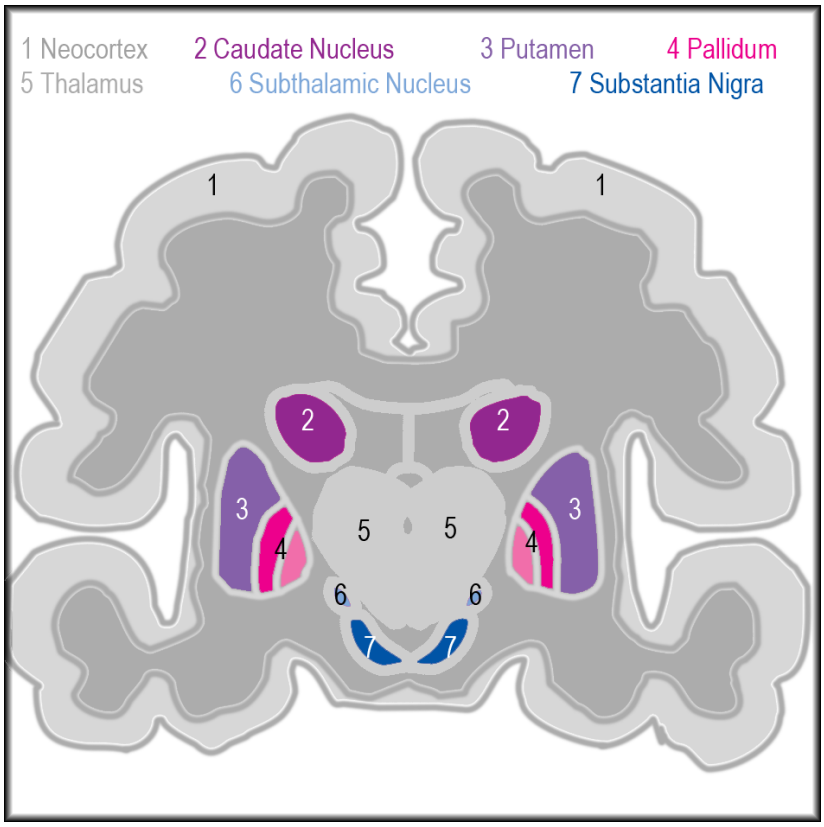



***Figure 1.3, The major regions within the basal ganglia***

*The basal ganglia is made up of the striatum (shown in purple (caudate nucleus (2) and putamen (3)), the pallidum (shown in pink (4), with inner and outer segments, which have different connectivity), the subthalamic nucleus (6) and substantia nigra (7). Also included in this figure are the neocortex (1) and the thalamus (5). Figure adapted from Graybiel* (2000).

It has been shown, using diffusion tensor imaging (DTI) to reveal white matter tracts in vivo, that the component of the putamen that forms part of the 'motor loop' is connected to both primary sensory and motor areas and to medial premotor cortex in the posterior part of the SMA (Lehéricy et al., 2004), areas active during overt speech production. In addition, Booth and colleagues (2007) suggested that the cerebellum and basal ganglia may be recruited in the modulation of articulatory or phonological output representations, as demonstrated by the use of a rhyming task with fMRI. With L2 production, consideration has to be given to the effects on the recorded BOLD signal resulting from the production of novel or partially trained sequences of motor commands, and their effects on feedback signals that are only partially attuned to the sound and somatosensations of L2. These signals will be further confounded by an

increased number of speech errors, whether or not the speaker is consciously aware of these errors, and if so whether attempts at self-correction are or are not initiated. This contrasts with the highly automatic processing of native speech.

Klein and colleagues (1994) found activation in the left putamen when participants repeated words in L2, which they attribute to the role of left basal ganglia in articulation, particularly the precise motor timing of speech output, which is less automatic and more 'effortful' in a language acquired later in life. This study was followed up in 1995 with three lexical search tasks (rhyme generation, synonym generation and translation), with word repetition as a control task (Klein et al., 1995). They found similar areas of activation for both within- and across-language searches, i.e. there were no significant differences related to whether the task used phonological or semantic cues or whether it used L1 or L2. Contrasting L1 translation with L1 repetition resulted in increased activation in the left putamen, as did the contrasts of L1 translation – L1 rhyming, L2 synonyms – L1 synonyms and L2 translation – L1 translation. Klein and colleagues (1999) extended their studies from English-French bilinguals to English and Mandarin Chinese bilinguals, using highly proficient speakers who had learnt L2 during adolescence. Using a noun-verb single word generation task, they demonstrated that common cortical areas were activated in fluent bilinguals who use both languages in daily life. Therefore, it appeared that similar brain regions are active even when the languages are typologically distant, such as English and Mandarin Chinese, and when L2 is acquired later in life. This is in line with their previous finding of similar brain regions for word repetition in L1 and in L2 (Klein et al., 1994; 1995). However, only the earlier studies (1994; 1995) demonstrated activity that could be attributed to motor control, and no increased activity in the left putamen associated with L2 was observed in their more recent study (1999), despite the fact that the L2 Mandarin Chinese was heavily accented. One suggestion put forward by the authors to explain this disparity between studies was that the latter study required mono- or bisyllabic production, whereas in the earlier study responses were mostly bi- or multisyllabic. Further studies are required in order to investigate the effect of syllable counts on brain regions involved in articulation, particularly the basal ganglia.

Frenck-Mestre and colleagues (2005) also found increased activity in the left putamen for bilinguals who had learnt L2 after the age of 12, compared to early bilinguals, in a reading aloud task, both at word- and sentence-level. Otherwise activity in cortical, subcortical and cerebellar regions was identical for both L1 and L2 in both groups. The authors suggest that learning to produce new articulatory patterns necessary for speaking an L2 requires adaptation of existing neural networks, rather than recruitment of new networks. Of course, relating activity in the basal ganglia to particular processes is problematic, as different regions of the striatum are connected to widely distributed cortical areas that subserve very different functions: motor, cognitive and emotional.

Aglioti and colleagues (1996) discuss the first neurolinguistically assessed case of bilingual subcortical aphasia and found that, due to a left capsulo-putaminal lesion, the patient had a speech deficit in their L1, with the much less practiced L2 being relatively spared. The patient spontaneously spoke only L2, and when L1 speech was elicited, it was non-fluent, slow and characterised by a low voice. They also report that the patient spoke L2 with a foreign accent, which they attributed to left basal ganglia pathology. The authors argued that the left basal ganglia is involved in implicit verbal memory and lesions here tend to affect the most automatic language of a bilingual. Despite the rarity of this dissociation, it has also been reported by Gelb (1937, in Paradis (1983)), who described a patient who could no longer speak L1 (German) but could speak L2 (Latin), which he had studied formally as an adult. Speedie and colleagues (1993) reported a bilingual patient with a basal ganglia lesion, but on the right, and while propositional speech was unaffected, automatic (non-propositional) speech was impaired, in both L1 and in L2 (in Aglioti et al., 1996).

The role of the basal ganglia in native and non-native speech is investigated in Chapter 5 of this thesis, in terms of comparing across languages and also how activity changes with learning.

### 1.5.3 Adaptations to auditory and somatosensory feedback during L1 production

Congenitally deaf speakers have demonstrated that somatosensory feedback plays an important role in language acquisition, and it has been shown that somatosensory

information constitutes a principal component of the speech target, independently of the acoustic information (Tremblay et al., 2003). In that study, participants received altered somatosensory feedback due to mechanical jaw perturbations, although auditory feedback remained unaltered. Even when speech acoustics were unaltered by the jaw perturbation, participants adapted their jaw movements when producing speech, indicating that the somatosensory target in speech is monitored independently of the acoustics. Nasir and Ostry (2006) provided evidence for the central role of somatosensory feedback in speech production by using a robotic device to alter jaw movements during speech. Speech acoustics were unaffected, demonstrating a dissociation between the influences of auditory and somatosensory feedback on speech production. Houde and Jordan (1998) investigated how altering auditory feedback affects speech motor control, relating their study to previous work in which limb motor control systems adapt to changes in visual feedback. Using native English speakers, they demonstrated that the control system involved in vowel production adapts to altered auditory feedback. Similarly, using native Mandarin speakers, Jones and Munhall (2005) found that in response to altered auditory frequency feedback, speakers automatically adjusted the pitch of their speech. This suggests that the motor control of vocal pitch requires continuous mapping between the laryngeal motor system and the vocal output and that this mapping relies on auditory feedback. A further study, by Tourville and colleagues (2008), who altered auditory feedback during single word reading in L1, demonstrated an increase in activity in posterior auditory association cortex (including planum temporale) and in the parietal operculum (second-order somatosensory cortex). The participants had no awareness of this alteration, and yet they automatically altered speech production as a motor compensation to counter the auditory perturbation.

Adaptations to motor output based on sensory feedback can be described in two ways. Unexpected perturbations to auditory or somatosensory feedback result in rapid compensatory motor changes, whereas a more constant change results in remapping sensory signals to motor output (Shiller et al., 2009). Sensory input from speech is compared with predicted sensory consequences in order to further control production, and speech adaptation involves both input and output processes simultaneously (Shiller et al, 2009). This comparison between predicted and actual movement and their sensory consequences, as well as any mismatch signals, are

likely sent via climbing fibres that project from the inferior olive to the cerebellum (Blakemore et al., 2001). These projections are not always carried out effectively, and this can lead to impaired speech. For example, in stutterers, one proposal is that the speech-motor plan is not initiated accurately and as a result, the perceptual prediction of the speech sound gets sent to the auditory system repeatedly and is used as an inhibitory signal. Speech production is impaired, with the initial sound of a word often being repeated (Brown et al., 2005). Stuttering results in cerebellar over-activation, which may be due not only to motor over-activity but also as a response to an action-consequence mismatch, which is consistent with cerebellar activity as the "discrepancy signal" (Brown et al., 2005). Stuttering also results in over-activity in the midbrain and under-activity of cortical motor and premotor regions involved in articulation (Watkins et al., 2008).

### 1.5.4 Increased activity in response to sensory feedback during L2 production

When speakers use L1 it has been shown that there is a paradoxical suppression of neural activity ('sensory gating') in second-order somatosensory association cortex in the parietal operculum (Dhanjal et al., 2008). Although auditory association cortex is active during overt speech, this activity is less than when listening to the voice of another (Ventura et al., 2009; Wise et al., 1999), providing further evidence of the importance of internal sensory feedback. This suppression of activity may indicate the efficiency of online sensory-motor monitoring during L1 speech production, which is highly automatic. In contrast, L2 speech production is less automatic and may result in increased activity in response to sensory feedback. It has also been shown that processes such as articulation and post-articulatory monitoring result in greater activation for bilinguals than monolinguals (Parker Jones et al., 2012).

Chapter 4 of this thesis shows that, by considering self-generated, fluent, sentential speech, a more extensive picture of the distributed neural systems involved in non-native speech production is observed (Simmonds et al., 2011a). In the study presented in Chapter 4 I used fMRI to compare native and non-native propositional speech production with the specific aim of revealing the changes in motor-sensory control when switching from native speech production to speech in a second language. Participants were instructed to give definitions of visually presented

pictures in either their native language (L1) or English (L2 for all participants). Rest was included as the baseline condition. I predicted and observed that altered feedback processing in the non-native language resulted in increased activity in sensory regions, particularly in second-order somatosensory cortex. This network involves both auditory and somatosensory areas not previously revealed by previous functional imaging studies of bilingualism. My results demonstrate that learning to speak a second language after the normal period of 'innate' language acquisition (i.e. an L2 that is acquired after L1 has already been established) has functional consequences on cortex involved in auditory and somatosensory feedback control of articulation (Simmonds et al., 2011a). Demonstrating sensory feedback in bilingualism was made possible by using a task dependent on propositional speech production, rather than single word production. A prospective training study on novel speech sounds is required to understand the relative motor-sensory contributions from feedforward motor systems, sensory predictions and sensory feedback. This is examined in Chapter 5 of this thesis.

## 1.6 Potential links between bilingual articulation and comprehension

In this section I examine whether perception and production of phonological features can be dissociated or whether production abilities depend on accurate perception of the target phonological distinctions. I also discuss how neuroimaging data of native and non-native speech production can be used to inform current theoretical models of bilingual language processing. Although it is widely accepted that speech production results in interactions between the auditory system and the articulatory motor system, these interactions resulting from speech perception are still being debated (Möttönen and Watkins, 2011).

### 1.6.1 Language input and its impact on speech output

Flege and Hillenbrand (1984) investigated limits on phonetic accuracy of adult L2 learners. They suggest that a non-native accent is in part due to phonological and phonetic differences between the speaker's L1 and L2. They cite Weinreich's (1953) hypothesis that a non-native learner makes substitutions for phones or phonemes when there are similarities between them in L1 and L2. For example, the phone /s/ sounds similar in French and English but the place of articulation is different in the

two languages (dental in French and alveolar in English). A native-English adult learner of French perceives the acoustic similarity between the native and non-native /s/ and does not adapt production, pronouncing /s/ in French as an alveolar phone (Flege and Hillenbrand, 1984). Using the /r-l/ contrast varying in frequency of the second and third formants in a study with English, Japanese and German speakers, Iverson and colleagues (2003) showed that speakers of different languages attend to different dimensions of the perceptual input, even when the same stimulus is used. There are many other examples of perceptual biases with respect to how speech sounds are heard, which likely influence L2 production. This is often reported as a function of the L1 phonetic repertoire, for example in Kuhl and colleagues' (2008) review of phonetic perception development models and Hickok and colleagues' (2011) review of auditory input affecting language output.

### 1.6.2 Speech output and its impact on language input

Lenneberg (1962, in Krashen (1982)) presented the case of a boy with severe congenital dysarthria. However, the boy was able to understand spoken English perfectly. Lenneberg claimed that the boy had acquired "competence" in the language, without ever speaking it himself. In Krashen's view (1982), producing speech output does play a role in language acquisition, but only indirectly. The benefit of speaking is not that it improves language acquisition itself, but rather that the acquisition and use of fluent speech encourages dialogue with others. Thus, speaking increases being spoken to, the quantity of which is matched by the quality, as native speakers use more natural language to learners they deem to be at a higher level. Native speech is often modified to include "foreigner talk", i.e. simplifying the language to make it accessible to a non-native speaker. A second language learner who does not attempt to speak much, who makes lots of mistakes, has a strong non-native accent and speaks hesitantly, will often be spoken to in a more simple version of the target language than a speaker who appears fluent.

## 1.7 Parallels between human speech learning and songbird song acquisition

Humans are highly skilled auditory and vocal learners. Although many animal species are auditory learners, and through experience can readily distinguish environmental sounds, most are not vocal learners. The vocalisations they utter are

innate and remain unmodified throughout life. For example, the zebra finch is an age-limited learner that learns its song as a juvenile and this song remains constant throughout life (Eda-Fujiwara et al., 2012). However, a few species, namely songbirds, parrots and hummingbirds, are skilled vocal learners and continue developing their imitation of vocalisations, even in adulthood. These vocal learners therefore provide an insight into speech learning, which in this thesis is further extended to motor-sensory learning of foreign speech.

Over a decade ago a review article drew parallels between song acquisition in songbird hatchlings and the development of speech production in human children (Doupe and Kuhl, 1999). Since then there have been a number of review articles attempting to identify anatomical homologues in the brains of vocal learning birds with human brain regions involved in speech production. The attempt to draw these parallels has been advanced by the revision of the nomenclature of regions within the brains of birds, consequent upon greater understanding of their structural organisation. The ability to perform detailed experiments on birds to define the regions and pathways for song perception and production and song learning has stimulated a number of review articles that have directly related the knowledge gained from such animal studies to human neuroanatomy.

Both vocal learning in humans and song learning in songbirds require a period of auditory-guided motor learning during a sensitive period of development. Songbirds acquire their song before a specific stage in development, which mirrors the difficulty adult humans have in speaking a foreign language without an accent. As well as similarities in the timing of learning, human speech learning and birdsong acquisition have parallels at the behavioural, neural and genetic levels. A particular common neural substrate involves the forkhead box P2 *(FOXP2)* gene.

Of particular relevance to this thesis, specifically the striatal aspects of vocal learning presented in Chapter 5, is the role of the anterior forebrain pathway (AFP) in songbirds. This pathway is critical for learning, but after crystallisation of their song, a lesion of the AFP does not interfere with continuing song production. In contrast, song acquisition is severely impaired by a lesion within the AFP. A key component of the songbird AFP is area X. The AFP is the homologue of the pallial-striatal-pallidal-

thalamic-pallial loops in the mammalian brain, although in birds the striatal-like and pallidal-like neurons lie within one structure, area X.

Figure 1.4 shows a schematic diagram of the song system of songbirds. The posterior descending pathway (PDP) is used for the production of learned song, whereas the AFP is involved in the acquisition of the pattern (Nottebohm, 2005). The LMAN is not necessary for the production of adult song, only learning in juvenile song. When LMAN neurons are silent, or absent, the HVC – RA pathway produces an accurate, established pattern. When LMAN – RA neurons are firing, there is much more variability in the song. This variability is needed to reach accurate imitation of a pattern. HVC – RA neurons are continually produced in adulthood and are therefore replaceable. HVC – Area X neurons are produced in the embryo or in the initial days post-hatching and are not replaceable. The brain's capacity to learn is often explained in terms of synapses, such as synaptic number, efficacy or plasticity. However, the neuronal replacement in songbirds indicates that learning may be enabled by new neurons, which are themselves transient and replaceable, which replace older neurons that may be preventing further learning (Nottebohm, 2002). This could be due to the dendritic and synaptic changes necessary for learning being less possible in older neurons and the new neurons reinstating the plasticity for learning (Nottebohm, 2005). Another explanation could be that in certain types of learning it is the neuron itself, rather than the synapse, that is the unit of learning. The demonstration of neuronal replacement in song learning in songbirds furthers our understanding of the acquisition and maintenance of complex skills.

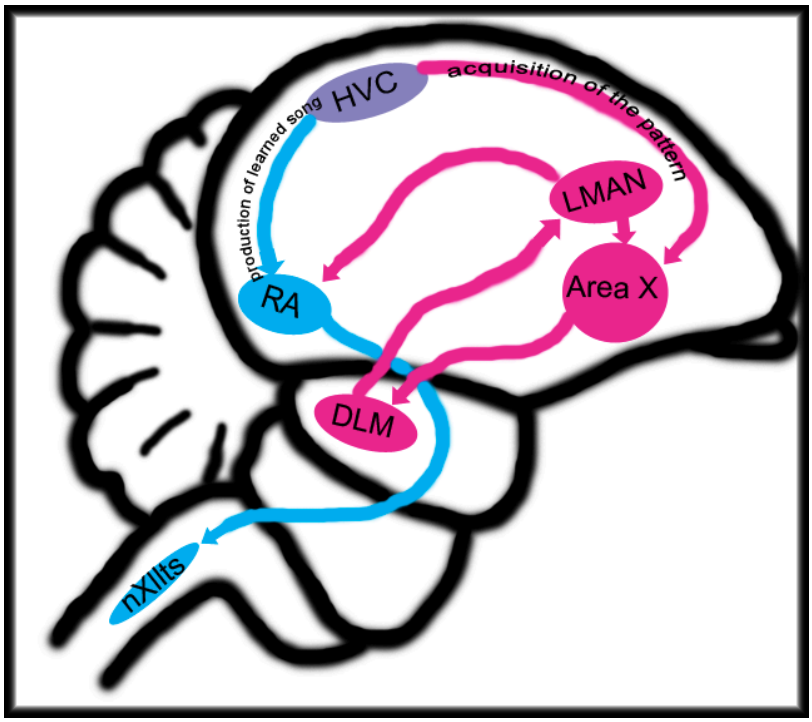



***Figure 1.4, The song system of songbirds***

*The HVC (shown in purple) projects to the two pathways, the PDP (shown in blue) and the AFP (shown in pink). Information from the HVC ultimately projects to the nXIIts, which then projects to the vocal muscles. The HVC projects to the RA directly, through the PDP and indirectly through the AFP, via Area X, the DLM and LMAN. Figure adapted from Nottebohm (2005).*

The AFP comparison hypothesis proposes that the comparison between the actual vocalisation produced, judged by auditory feedback, and the adult song template, involves the AFP (Bolhuis et al., 2010). However, it has been shown that AFP activity is not affected by auditory input when a bird is singing, which suggests that auditory feedback monitoring recruits regions outside this pathway (Leonardo, 2004). The 'error-correction hypothesis' for the AFP proposes that the basal ganglia calculate an 'error' signal by comparing the intended vocalisation with actual performance (Bolhuis et al., 2010). At early stages of development a young bird produces variable attempts at the adult song, but with practice and feedback the adult song template can be accurately matched by the learner. Such variability is necessary for reinforcement-based trial-and-error learning, as the learning process requires exploration of a range of action sequences, evaluation of performance with each and modifications to behaviour that result in improved performance (Ölveczky et al.,

2005). It has been shown that variability in the neural activity in the AFP correlates with performance variability (Kao et al., 2005). Variability in young birds' song is reduced following transient inactivation of the lateral magnocellular nucleus of the nidopallium (LMAN), which is the target for basal ganglia outflow (Ölveczky et al., 2005). LMAN inactivation reduced variability in song performance, and this reduction was most pronounced for birds at an early stage of song development, which show the greatest song variability. Research into song learning supports the view that trial-and-error learning, which depends on behavioural variability in the early stages, depends on the basal ganglia (Doya and Sejnowski, 1995). Similar support has also been shown in a rodent study, which investigated variability of neuronal firing in the basal ganglia throughout non-vocal motor learning. Neuronal firing was highly variable at the initial stage of learning, but following training became more consistent (Barnes et al., 2005). The variable firing during learning is considered to represent 'neural exploration', whereas the stable firing after learning reflects 'neural exploitation'. The striatal results in vocal learning presented in Chapter 5 are consistent with this view. In Session 1, when all sounds were unfamiliar, performance was highly variable and basal ganglia activity was also high. In Session 2, post-training, performance was less variable, and basal ganglia activity reduced.

## 1.8 Multiple networks involved in L1 and L2 motor-sensory control of speech production

As discussed earlier in the introduction, modern neuroimaging techniques have advanced our understanding of the distributed anatomy of speech production, beyond that inferred from clinic-pathological correlations. However, much remains unknown about functional interactions between anatomically distinct components of this speech production network. One reason for this is the need to separate spatially overlapping competing neural signals supporting diverse cortical functions. It is unlikely that there is a close one-to-one correspondence between individual brain regions and the many processes established by psycholinguistic research, such as word retrieval, semantics, syntax and phonology. This thesis focuses on the anatomical-functional correspondence for articulation accompanied by sensory (auditory and somatosensory) feedback. Chapter 6 of this thesis identifies distinct signals from posterior left perisylvian 'language' cortex that make functionally

different contributions to speech, depending on the type of speech production required.

## 1.9 Implications of taking a motor-sensory approach to bilingual speech production

### 1.9.1 Pedagogic implications: adults acquiring native-like levels of production

Whether native-like pronunciation should be the ultimate goal of foreign language learners has been debated. There are instances of non-native speakers who use their L2 more to communicate with other non-native speakers than native speakers, so for them, a near-native accent is not necessary (McArthur, 2002). It has also been suggested that speakers may wish to maintain their non-native accent in order to keep their cultural identity (Avery and Ehrlich, 1992). In the case of the English language, there are many varieties and no single agreed upon pronunciation (Dauer, 2005). However, as reported by Derwing (2003), the majority of foreign language learners aim for perfect native pronunciation and this thesis focuses on the foreign language learning with success measured by native-like pronunciation.

Taking a more motor-sensory approach to understanding bilingualism implies that, in line with L1 acquisition, adult second language learners might benefit from a mute period - a period of intense auditory exposure to L2 before attempting to produce the sounds. This may prove beneficial in enabling the learner to hear (and thus produce) subtly different phonetic features, new phoneme distinctions and unfamiliar sequences of stress patterns. One possibility is that an artificially induced mute period may protect the learner from using first language phonological categories to represent the L2 system, thus enabling higher levels of production performance and avoiding L1 transfer or interference.

Neufeld (1979) showed that, with the right method of instruction, adults can acquire native-like pronunciation (in Archibald (1998)). Students were trained on pronunciation of certain sounds from Inuktitut, a language to which they had not been exposed previously. The learning process involved a lot of time listening to the language, with no attempt at producing the sounds. They were later instructed to

produce the sounds and native speakers subsequently rated their speech attempts, with much of the speech deemed to be native-like. Neufeld claimed that the silent period at the beginning helped the students to accurately produce the language later. Removing students' own attempts allowed perception to remain more plastic, such that the L2 acoustic template is heard accurately before erroneous phonetic utterances in L2 become crystallised. Producing the sound too early, and therefore incorrectly, would have influenced this acoustic template and thus hindered their production. Bley-Vroman (1990) makes the comparison between an adult language learner listening to a foreign language before attempting to speak it and an adult learning to play poker watching a poker game before playing. This can be considered a useful strategy for learning a complex task, rather than a consequence of an innate language, or poker acquisition faculty (Bley-Vroman, 1990).

### 1.9.2 Babbling adults?

As well as the benefits adult second language learners might gain from a mute period, there is the possibility that a babbling phase might also improve non-native speech learnt later in life. By imitating the target speech sounds in isolation, before attempting to produce them in word form, adult learners might develop more accurate efference copies of the motor commands required for the production of the sounds, allowing more efficient feedback for the monitoring and correction of articulatory errors.

### 1.9.3 The role of literacy in L1 and in L2

Skilled readers are very familiar with the written form of their native language, and automatically decode a grapheme by producing a phonological representation of the sound (Snow, 2002) (although when reading text skilled readers progress to recognise the whole word form and read at an appropriate speed). For example, the letter "p" is highly familiar to a literate adult native speaker of English and each time they read that letter, they associate it with the typical English phonological representation, which is aspirated and has a relatively long voice-onset time. If a native English speaker begins learning Spanish, it is likely that they would transfer the typical phonological representation used in spoken English and therefore pronounce the Spanish "p" with an English accent, rather than the non-aspirated bilabial stop with a short voice-onset time that a native Spanish speaker would

produce. It would seem possible therefore, that learning a language without being exposed to the written form would lead to more accurate pronunciation. In line with this, it could also be expected that learning a second language with a different orthography to one's native language would also result in more accurate foreign language pronunciation. Learning orally has similar advantages to listening before speaking. Representations of L2 speech sounds are more plastic and less influenced by automated activation of native representations that might be triggered by reading a letter.

### 1.9.4 Clinical implications

Motor speech disorders arise due to impaired motor and feedback systems. Comparing speech production in a foreign language with that in the native language provides insights into how feedforward and feedback systems operate, which could help explain what goes wrong in the case of motor speech disorders, such as dysarthria. Dysarthria is characterised by problems with pronunciation, making speech sound slurred. Focal lesions resulting in dysarthria can be very differently located: the cerebral cortex, white matter, basal ganglia, thalamus, cerebellum and brain stem, anywhere that might compromise the motor-sensory pathways that control the final outflow from motor neurons to the articulatory muscles. Patients show slow and uncoordinated movements of the muscles required for speech (Sellars et al., 2005), possibly with impaired prosody (Ackermann and Hertrich, 2000). Although communication disorders after stroke are often the result of aphasia and associated with problems with word understanding and word retrieval, in dysarthric patients the predominant disorder is with the motor aspects of speech. Although natural recovery can occur, so that dysarthria, although still present, is so mild as not to impair intelligibility, in a proportion of patients intelligibility is so compromised that alternative communication systems need to be used.

Greater understanding of how brain processes underlying speech production adapt to non-native speech may help develop more effective therapy techniques to be developed in order to improve dysarthric speech. Similar to previous research on bilingualism, most research on intervention in communication disorders following stroke has focused on impairment of linguistic and semantic representations and

processes. In contrast, the motor-sensory control of speech production has received much less attention.

## 1.10 Benefits of researching neural bases of bilingualism

Investigating adult foreign language learning complements child language development and together this research leads to greater understanding of the fundamental principles of language and cognition (Bley-Vroman, 1990). Theories of second language acquisition can inform our understanding of how the brain works. Conversely, understanding how the brain can adapt to multiple languages could provide an empirical basis to constrain cognitive and linguistic theories of second language acquisition (Danesi, 1994). For the learning of second languages to become more effective, the teaching of such languages needs to become more "brain-compatible". There is no clear evidence for domain-specific foreign language learning in adults and as such, it would be more beneficial to incorporate broader findings from more domain-general cognition (Bley-Vroman, 1990). This Chapter has discussed the largely overlooked motor-sensory aspects of bilingual speech production and argued that investigating these aspects will lead to a more comprehensive understanding of bilingual speech and ways in which the teaching of this can become more effective. Chapters 3 - 6 of this thesis present my investigation of the motor-sensory learning of foreign speech.

## 1.11 Main aims and hypotheses of this thesis

### 1.11.1 Main aims

The literature that I have discussed in the introduction to this thesis provides the context for the studies presented here. Despite the extensive research into the motor-sensory control of speech in L1, its response to L2 is not known. In this thesis, I present fMRI studies on healthy volunteers that use a variety of speech production tasks involving native and non-native languages, at different levels of familiarity. The aims of this thesis were fourfold:

- The first was to investigate the use of overt and covert speech production tasks to examine motor-sensory control in both propositional and non-propositional speech.

- The second aim of this thesis was to investigate the response of motor-sensory control systems to speaking a non-native language, by using late learners of L2 with high proficiency.

- The third aim was to investigate the role of language proficiency on speech motor-sensory control, by taking monolingual speakers and training them in the production of non-native speech and scanning them both before and after a short period of training.

- The fourth aim was to investigate the distinct signals from posterior left perisylvian 'language' cortex, in response to different types of speech production, using native and non-native languages, both before and after training, as well as comparing between overt and covert speech production.

### 1.11.2 Main hypotheses

The studies presented in this thesis were designed to test four general hypotheses related to the motor-sensory learning of foreign speech:

- First, that overt speech engages a more distributed system than covert speech and is therefore the only appropriate method to use for investigation of motor-sensory control.

- Second, that speaking in a non-native language, relative to native, requires greater engagement of motor-sensory control systems.

- Third, that proficiency in non-native speech production leads to reduction of activity in these motor-sensory control regions, both cortical and subcortical.

- Fourth, that posterior left perisylvian 'language' cortex can be separated into distinct signals that may make functionally different contributions to speech, depending on the type of speech production required, and that show differences in response following training in foreign speech.

The specific hypotheses for each of the four studies presented in this thesis are stated in the following section.

## 1.12 Thesis overview

### 1.12.1 Chapter 2

In Chapter 2, I introduce the neuroimaging techniques and analytical tools that I used in the studies presented in Chapters 3 - 6.

### 1.12.2 Chapter 3

In Chapter 3 I present a study that investigated activity in response to both overt and covert speech, using propositional sentential speech and non-propositional speech (counting). Previous literature suggests that during covert speech, i.e. sub-vocal inner speech, interactions between sensory and motor speech-related systems will still be active. I designed an experiment using fMRI that would directly compare brain

activity in overt and covert speech production tasks, using both propositional and non-propositional speech. I had the assistance of two MSc students to acquire the data, and they helped me with some of the analyses. In this study, I proposed that only overt speech, with auditory feedback from the sound of one's own voice, and somatosensory feedback from speech-related movement, is a reliable method for investigating motor-sensory processes in speech production. I demonstrated that overt, not covert, speech fully engages the motor-sensory speech production systems. In addition I showed that the pattern of activity for overt propositional speech reflects the sum of overt non-propositional speech and covert propositional speech. In line with the findings from this study, the other studies presented in this thesis all used overt speech production.

### 1.12.3 Chapter 4

In Chapter 4 I present a study that investigated motor-sensory integration in native and non-native speech production. This study was based on the assumption that to speak in a foreign language means to speak with an accent, and this is likely to be the result of a failure to achieve the same proficiency in integrating the motor feedforward and sensory feedback control of L2 articulation that is achieved when speaking in the native language. In this study I proposed that production of L2 would place increased demands on the motor-sensory control system. To test this hypothesis I used fMRI to investigate motor-sensory integration, measuring the haemodynamic response in motor and sensory (auditory and somatosensory) cortices as the participants produced connected overt speech using L1 and L2. Participants had a variety of L1s and all had English as L2. Speech rate was recorded during the scanning session, to ensure that this was not a confound when analysing the imaging data. This work demonstrated that within the parietal operculum, increased activity expressed a mismatch between the efference copy of motor commands to the articulators, the intended auditory target (L2 without an accent), and actual somatosensory and auditory feedback. This study paved the way for the investigation of the motor-sensory control of speech production as participants learn foreign speech sounds, which is presented in Chapter 5.

### 1.12.4 Chapter 5

In Chapter 5 I present a study investigating the role of proficiency in the motor-sensory control of speech production. The main hypothesis for this study was based on the results presented in Chapter 4, that producing non-native speech sounds requires greater engagement of motor-sensory neural feedback systems, and that familiarity with non-native speech leads to reductions in motor-sensory activity, specifically in the frontal operculum and in the temporo-parietal junction. I demonstrated that non-native word, relative to native non-word, repetition resulted in increased activity in widespread bilateral cortical and subcortical regions. In addition I showed that before training, during anticipatory listening (that is, listening and preparing to repeat), there was activity in a bilateral fronto-parietal system that was greater for native non-words than non-native words. This difference in activity within this bilateral system reduced after training on the pronunciation of the non-native words. This was an unexpected result, and I have interpreted these functional changes in terms of processes associated with working memory and attention. This same fronto-parietal network was also observed during overt repetition trials on both native non-words and non-native words, but activity was not modulated by training.

Of greater interest were the changes in basal ganglia activity during the repetition trials. The avian striatum is known to be involved in song learning in songbirds, resulting in the hypothesis that the human striatum must be involved in vocal learning for speech, and Chapter 5 presents data demonstrating the role of the striatum in vocal learning in the adult human brain. An interaction between Language and Session – that is, the effect of training with non-native words – was observed predominantly in the left and right striatum, with peaks in both the caudate nucleus and anterior putamen. Further analyses confirmed Language x Session interactions for German and Mandarin but not for Spanish. The behavioural data demonstrated learning for all languages, but Spanish was rated the least accurate, both before and after training. This indicates that striatal activity in vocal learning is modulated by proficiency. Further, the profile of striatal activity across the duration of the pre-training Session indicated an effect of early learning, occurring within minutes.

### 1.12.5 Chapter 6

In Chapter 6 I took three separate fMRI datasets (two using speech production, from Chapters 3 and 5 of this thesis, and one when participants were scanned while they were at 'rest'). In each I decomposed the neural activity from the left posterior perisylvian speech region into discrete signals. This decomposition identified two overlapping signals, one centred on the posterior superior temporal gyrus (pSTG) and the other on the adjacent ventral anterior parietal lobe (vAPL). The connectivity of the pSTG was with bilateral superior temporal and inferior frontal regions, whereas that of the vAPL was with other parietal regions. In addition, I demonstrated that the patterns of evoked activity for the two signals dissociate in response to different speech demands.

### 1.12.6 Chapter 7: Thesis summary and future directions

In Chapter 7 I discuss the main results of this thesis and the limitations of the studies presented here, which could be addressed with future studies. One aspect relates to the methodological limits that could be enhanced through the use of higher-resolution scanning and physiological monitoring. Based on the findings from Chapter 5, I suggest a possible study using a clinical population to further investigate striatal involvement of vocal learning. Specifically, I propose a study of patients with Parkinson's disease at various stages of reduced midbrain dopamine innervation of the striatum. Investigating vocal learning in patient groups with striatal dysfunction would provide further insight into the role of the striatum in vocal learning and the specific contributions of separate striatal structures. Finally, I discuss one way the connectivity study presented in Chapter 6 could be extended by focusing on a different region of interest and I propose that Broca's area could be used as a mask for investigating whole-brain connectivity. I present some preliminary data showing the connectivity maps and the responses to native and non-native listening and repeating for components within Broca's area.

# 2 Methods

In this Chapter I discuss the methods and analytical techniques used in the studies included in this thesis. I introduce the general principles of magnetic resonance imaging (MRI). I then present the methods I have used to process and analyse the functional and structural imaging data. In the final section of this Chapter I explain the scanning protocol and scanning parameters used in these studies.

## 2.1 Introduction to magnetic resonance imaging

### 2.1.1 MRI principle summary

MRI is an imaging technique that produces images of biological tissue when it is placed in a strong magnetic field. These images are produced using a series of changing magnetic gradients and pulse sequences that create fluctuating electromagnetic fields. The MRI scanner uses a range of pulse sequence types to detect different tissue properties. It is therefore possible to create images that differentiate between grey and white matter, fluid and tissue or high and low proton density. MRI is based on the physics underlying nuclear magnetic resonance (NMR), which is explained briefly in the following section. The different layers of magnetic coils in the scanner are also each explained in turn, namely the static magnetic field coil ($B_0$), the shimming coils, the radiofrequency coils and the X, Y, Z gradient coils.

### 2.1.2 Nuclear Magnetic Resonance (NMR)

Nuclear Magnetic Resonance is based on the spin property of protons in the nucleus of an atom. Protons usually spin in random directions but align when placed in a magnetic field, and this difference in spin property allows image acquisition through MRI scanning, which detects the change. The spin of a proton causes it to produce a small magnetic field, known as the magnetic moment, and the particle responds as if it were a magnet with north and south poles. The majority of human tissue is water, which is high in hydrogen, and the nucleus of a hydrogen atom produces a significant magnetic moment. MRI scanners are therefore usually configured to detect hydrogen, because of its evident magnetic moment, and prevalence in the human body.

### 2.1.3 The magnetic fields $B_0$ and $B_1$

The direction of spin is usually random but when a proton with a non-zero spin, such as hydrogen, is placed in a strong magnetic field ($B_0$), the particle aligns itself either parallel or anti-parallel to the direction of the magnetic field (the Z plane, from the head to toe of a body placed in a scanner). The parallel alignment is known as a low energy configuration and anti-parallel alignment as high energy configuration. The higher the $B_0$, the more particles align in parallel to the magnetic field (low energy state). Figure 2.1 shows the properties of protons with spin in a normal and a magnetic field. At rest, the axes of the spin align with the static magnetic field. To create a measurable signal, it is necessary to excite the protons. After excitation the axes of the spin precess about the magnetic field lines, with a motion similar to the gyration of a spinning top around its main axis.

$B_0$ is a static magnetic field and the radiofrequency (RF) transmitter coil generates a dynamic magnetic field, known as $B_1$. These two magnetic fields, $B_0$ and $B_1$, are at right angles to each other. Once the RF transmitter coil generates the $B_1$ magnetic field, the hydrogen protons begin to precess about the magnetic field line, and when the $B_1$ magnetic field is switched off, these protons gradually return to their original orientation. As these protons go back to where they were, known as 'relaxing', they emit a signal that is detected by the RF receiver coil. The strength of the magnetic field determines the frequency of the proton spin precession. Changing the local spin precession frequency enables identification of proton spin locations, and this is carried out by X, Y and Z gradient coils, explained in more detail in a later section.

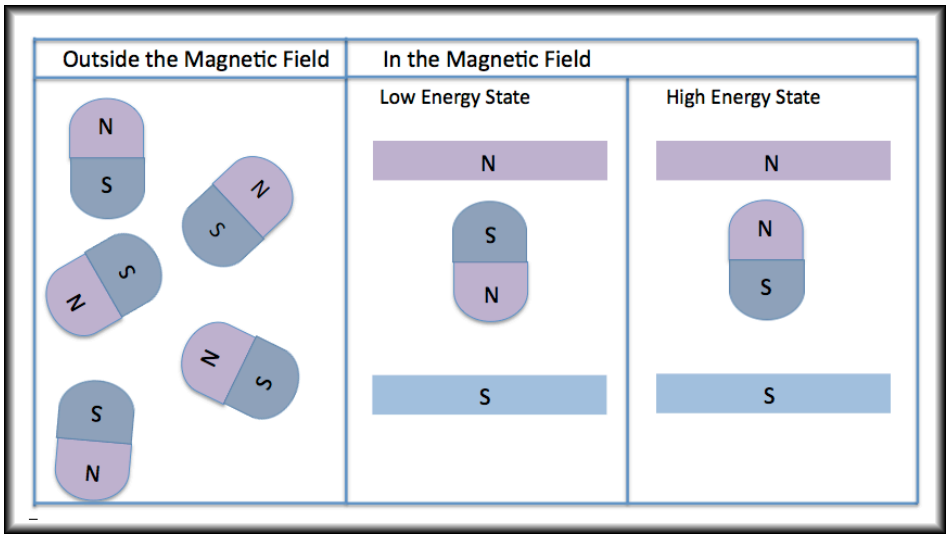



***Figure 2.1, Properties of protons with spin outside and in the magnetic field***

*The spin property of protons is random outside a magnetic field, but causes them to align with or against the direction of the magnetic field they are placed in. The low energy state reflects hydrogen protons in alignment along the direction of the main magnetic field (a NorthSouth-NorthSouth configuration). Following a radiofrequency pulse orthogonal to the main magnetic direction, protons move to the higher energy state, in an anti-parallel configuration to the main magnetic field (a NorthNorth-SouthSouth configuration).*

### 2.1.4 Shimming coil

Magnetic fields, especially when there is a body inside them, are not homogenous. However, MRI needs a homogenous magnetic field, otherwise the imaging is susceptible to artefacts such as image distortions and signal drop-out. The shimming coil improves magnetic field homogeneity by producing an additional magnetic field that is added to the scanner's overall magnetic field. Shimming is done before the scan starts and the settings remain constant throughout the scan.

### 2.1.5 Radiofrequency (RF) coils

RF coils are used to both transmit and receive an RF pulse. Transmitting an RF pulse excites the spins and receiving the pulse measures the signal. RF coils are aligned perpendicular to the $B_0$ field and transmit RF energy at the pulsating proton frequency, allowing the tissue protons to absorb this energy. After the RF pulse from the coil, the RF energy is re-emitted by the tissue, and the RF coils then receive this

signal. In the studies presented in this thesis, transmission and reception were combined in a single “transmit-receive” coil. The RF pulse converts the protons from low energy state to high energy state and as the RF energy is re-emitted by the protons, they return to the low energy state. The time taken for the protons to return to the low energy state is known as T1 relaxation time.

### 2.1.6 X Y Z gradient coils

The RF coil detects the presence of hydrogen atoms as a function of time but does not provide any spatial information. The location of spins is identified by their frequency response, detected by gradient coils. Differences in precession frequency can be created along the directions of the magnetic gradients, with alterations to the static field strength. There are three gradient directions: X is horizontal (left to right), Y is vertical and Z is horizontal (head to toe) and each gradient coil generates a magnetic field that increases in strength along one spatial direction. For example, along the Z-plane, the $B_0$ field is stronger at one end of the bore of the magnet and weaker at the other end. In the centre of the bore, the gradient field is zero, meaning that the $B_0$ field is its normal strength. The use of gradient coils enables three-dimensional reconstruction of images by determining frequency measurements of the magnetic resonance (MR) signal at different spatial locations.

### 2.1.7 MRI parameters

Variations in MR signal received by the RF coil depend on the time taken for proton spins to return to the low energy state (relaxation) and whether relaxation is complete before the next RF transmission pulse. The two main factors that alter the timing of MR image acquisition are the repetition time (TR) and the echo time (TE). The TR is the time interval between each excitation pulse and the TE is the time interval between excitation and relaxation. Adjusting these two parameters allows collection of different signal intensities using T1, T2 or T2* relaxation, which are each explained in the following sections. As precession moves back towards the $B_0$, energy in the longitudinal plane (z-plane) increases, which is referred to as T1 recovery, and energy in the transverse plane (x-y plane) decreases as the spins rapidly lose their phase coherence, which is referred to as T2 decay.

### 2.1.8 T1-weighted images

For most anatomical brain images T1-weighting is the usual contrast. T1 relaxation time is the time taken for the excited spin to return to the low energy state, aligned along the longitudinal z-plane, the direction of the main magnetic field. In this type of image the relative signal intensity of voxels is dependent upon the T1 value of the tissue. Voxels with a short T1 value are bright and those with long T1 values are dark, resulting in a T1-weighted image in which fluid is black, grey matter is dark grey and white matter is light grey. Figure 2.2A shows an example of a T1-weighted image.

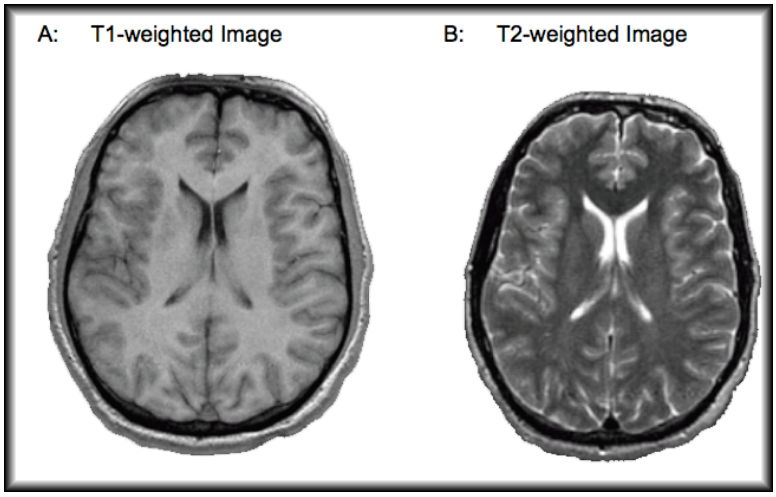



***Figure 2.2, T1- and T2-weighted images***

*Panel A shows a T1-weighted image in which fluid is black, grey matter is dark grey and white matter is light grey. Panel B shows a T2-weighted image in which fluid is bright, grey matter is light grey and white matter is dark.*

### 2.1.9 T2-weighted images

The T2 signal varies for different tissue types, with protons that have different spin and magnetic fields. The temporary and random 'spin-spin' interactions between these protons result in a potential loss of signal, known as T2 relaxation. Compared to T1 relaxation, T2 relaxation is very rapid. For a T2-weighted image, the T1 contrast needs to be minimal and the TR must be long so that recovery is almost complete in all tissue types. Whereas T1 governs the rate of longitudinal relaxation (the z-plane), T2 governs the rate of transverse relaxation (the x-y plane). The appearance of a T2-weighted image is opposite to a T1-weighted image, so fluid is bright, grey matter is light grey and white matter is dark. T2-weighted images are

particularly useful in clinical structural scans as the T2-contrast reveals pathological conditions, such as tumours, more clearly. Figure 2.2B shows an example of a T2-weighted image.

### 2.1.10 T2*-weighted images

Similar to a T2-weighted image, T2*-weighted acquisition requires a long TR and medium TE. The pulse sequence uses magnetic field gradients to generate the signal echo. T2 signal assumes a homogeneous $B_0$ field but as explained in section 2.1.4 of this Chapter (p63), the magnetic field is actually inhomogeneous, due to the presence of a human body in the scanner. T2* signal reflects the sum of all these effects, including the spin-spin interactions and the combined effects of inhomogeneities in $B_0$. T2* relaxation is even quicker than T2 relaxation, which makes it particularly useful for functional imaging. T2*-weighted images are sensitive to the amount of deoxygenated haemoglobin present in the blood, and this is explained more in the following section.

## 2.2 Functional MRI

Functional MRI allows investigation of changes in neural activity in response to particular cognitive tasks and depends on the use of energy in neuronal synaptic activity. Such activity in a particular brain region leads to a metabolic response, in the form of glucose and oxygen consumption, and a haemodynamic response. The latter is evident in an increase in cerebral blood flow (CBF), blood volume and blood oxygenation. The following section describes the physiological basis of the functional MRI signal.

### 2.2.1 Blood-oxygen level dependent fMRI

Functional MRI measures the BOLD response (blood-oxygen level dependent imaging), which relies on the difference in magnetisation of oxy- and deoxyhaemoglobin. Neuronal activity leads to an increase in oxygenated blood due to the increased oxygen consumption and increased CBF induced by neuronal firing. The amount of oxygenated blood produced is greater than is required for any given neural activity (Fox and Raichle, 1986), so an increase in local oxygenated haemoglobin (oxyHb) is evident. OxyHb is diamagnetic and has no magnetic properties and the deoxygenated haemoglobin (deoxyHb), which results from the

transfer of oxygen to the neurons involved in a particular neuronal activity, is paramagnetic. Due to the difference in magnetic susceptibility, the scanner magnet can therefore detect the change in proportion of oxy- and deoxyhaemoglobin.

### 2.2.2 The haemodynamic response

The timecourse of the change in the local ratio of oxyHb to haemoglobin is known as the haemodynamic response function (HRF). BOLD-fMRI has a delayed temporal sensitivity relative to the underlying neural activity. This delay is modelled when analysing fMRI data in order to allow identification of changes in the BOLD contrast linked to specific cognitive processes used in the fMRI task. For the analysis of the studies presented in this thesis, I used a canonical HRF (from FSL), which is an assumed profile of the vascular response and the sum of local neuronal activity. The haemodynamic response function (HRF) models the timing of the neurovascular response and its effect on the BOLD signal. The timecourse of neural activity is rapid, whereas the vascular response is considerably slower. There is an initial dip, followed by an increase in BOLD, with the peak reached approximately 6 seconds after stimulus onset.

### 2.2.3 Echo-planar imaging for fast fMRI acquisition

Acquiring images of brain function requires a fast sequence that acquires images at approximately the same rate as the task demands. The most common sequence used in functional imaging is echo-planar imaging (EPI). This is a T2*-weighted sequence and emits one RF pulse from a transmitter coil before introducing rapidly changing magnetic field gradients. The MR signal is recorded from an entire image slice at a time. The high speed of this type of image acquisition results in lower spatial resolution than for a conventional structural MRI scan and an increase in artefact and distortion susceptibility. For the analysis of EPI images, a T1 image is also required, so that the EPI image can be registered onto an image with better resolution, which is explained in more detail in section 2.4 of this Chapter.

### 2.2.4 Model-based fMRI analysis

The main aim of model-based fMRI experiments is to identify brain regions (measured in voxels) that respond to the change in brain state throughout the data acquisition. Typically brain ‘activity’, measured with the BOLD signal, is compared

between two conditions that differ in a contrast of interest. The basic principle of functional MRI is that, by comparing the differences in BOLD responses to two tasks that only differ in one specific parameter, it is possible to identify brain regions specific to a given task. This approach is known as the subtractive method and is useful in fMRI, which cannot give an absolute measurement, but rather a measure of activity in one condition relative to another. FMRI demonstrates which regions are active during any given task, but such a result does not suggest that a region is essential for carrying out that task. To determine which brain regions are essential for any given task, evidence is required from lesion studies, either real in the form of focal brain pathology, or 'virtual' and temporarily-induced with transcranial magnetic stimulation. For the analyses presented in this thesis I used FEAT (FMRI Expert Analysis Tool) version 5.98, which is a part of FSL (FMRIB's Software Library; www.fmrib.ox.ac.uk/fsl) (Smith et al., 2004).

### 2.2.5 'Sparse' scanning

In speech production studies, overt speaking results in motion-related signal changes that confound the measured activity. These confounds can be direct or indirect signal changes independent of any neuronal-related signal change (Birn et al., 1999). Head movement, related to jaw movements involved in speech, is the direct source of signal artefact (Birn et al., 1999). Motion close to, but outside the field of view, leads to variations in the magnetic field and image warping, resulting in indirect signal artefacts (Birn et al., 1998). These magnetic field distortions cause voxels to shift in the phase-encoding direction, and the measured signal does not accurately reflect signal changes due to neural activation. To avoid these artefacts in the studies presented in this thesis, I used 'sparse' temporal sampling. Figure 2.3 shows a schematic description of continuous and sparse imaging, with the predicted BOLD responses to both auditory stimuli and noise from the scanner. Sparse scanning uses a silent period in between volume acquisition, which reduces the negative influence of scanner noise on the recorded signal. This technique allows for presentation of auditory stimuli without interference from background scanner noise (Hall et al., 1999) and minimises speech movement-related artefact (Gracco et al., 2005). In this type of scanning the pauses in data acquisition are approximately 8-12 seconds, followed by 2-3 seconds of a single volume acquisition. The BOLD response reaches

its peak in auditory cortex at between 8-10 seconds, coinciding with the data acquisition for that trial. This technique is particularly useful in the types of studies I present in this thesis, due to the involvement of auditory processing. In Chapter 5 there was an auditory stimulus that participants needed to attend to and in all experiments (Chapters 3 - 6) participants were required to speak overtly, and they heard themselves speak. Auditory feedback from the speech production would be hindered by noise from the scanner as this also activates the auditory system, reduces response to non-scanner sounds and masks the auditory stimuli presented to participants (Petkov et al., 2009).

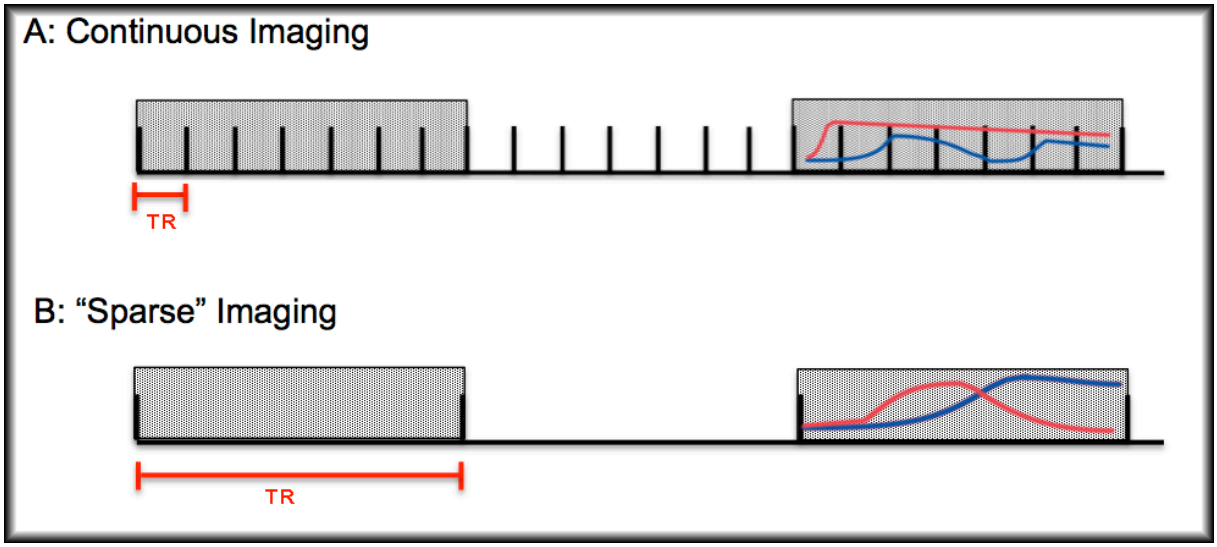



***Figure 2.3, A Schematic description of continuous and "sparse" imaging***

*A: Continuous Imaging; B: Sparse Imaging. The shaded boxes represent the periods of stimulation and short vertical lines indicate acquisition of a single data volume. Shown in the right-hand shaded box are the predicted BOLD responses to the scanner noise (in red) and to a sound stimulus (in blue). Adapted from Hall et al., (1999) and Petkov et al., (2009).*

### 2.2.6 Data preprocessing

The preprocessing stages performed on the analyses presented in this thesis are explained in the following section, which covers brain extraction, temporal filtering, motion correction and spatial smoothing. FMRI data analysis also requires registration to a standard anatomical template and this is explained in section 2.4 of this Chapter (p75).

#### *2.2.6.1 Brain extraction*

The T1-weighted structural image contains brain and non-brain tissue, so prior to statistical analysis non-brain tissue must be removed. For this I used the Brain Extraction Tool (BET), which uses a surface model approach to remove all non-brain parts of an image (Smith, 2002). A tessellated mesh of triangles is used as the surface model, which is then fitted to the brain surface in the T1-weighted image. Brain extraction was performed on the EPI images as part of the automated processes in FEAT.

#### *2.2.6.2 Temporal filtering*

This increases the signal-to-noise ratio by removing noise from the EPI data. Minor instabilities in the scanning hardware can cause slow changes in the overall signal intensity and this type of low frequency noise can be removed by high-pass filtering. In my analyses I used a temporal high-pass filtering to correct for baseline drifts in the signal. This filtering was carried out as part of an automated process in the FEAT analyses.

#### *2.2.6.3 Motion correction*

For accurate analysis the anatomical location of a single voxel must remain constant throughout the length of the scan. A typical functional imaging scan usually lasts approximately an hour, during which time it is impossible for participants to keep their head completely still. Head movements can reduce the signal-to-noise ratio as well as lead to artefacts that manifest as apparent ‘activations’ and any changes due to head motion are likely to be much larger than any changes due to brain activity, so it is important to account for them. I used FSL’s McFLIRT (Motion Correction FMRIB’s Linear Registration Tool) for the motion correction in my analyses. McFLIRT aligns all the images to a selected reference volume (in my analyses, it was the middle volume). In the output for an individual analysis, FEAT reports the amount of movement in each direction required to align each volume. FEAT produces a summary of relative and absolute motion, which displays motion artefacts. The motion parameters identified by FEAT are also entered in the design matrix to model any changes in signal intensity correlating with head motion. In addition, I used the FSL tool motion outliers, which identifies timepoints with a high amount of residual intensity change from the motion corrected data. A confound matrix of outliers is

created and this is included in the FEAT analysis. This is more beneficial than simply removing volumes with high levels of motion as it does not require any adjustments to the rest of the model with regard to timing, and it correctly accounts for signal changes on either side of the excluded timepoint, as well as correctly adjusting the degrees of freedom.

#### *2.2.6.4 Spatial smoothing*

Spatial smoothing improves the signal-to-noise ratio by filtering out high spatial frequency noise. Signal from a biologically plausible source is assumed to take the form of a spatially smooth area of activity, of approximately 5 to 8 mm in diameter. Spatial smoothing uses the matched filter theorem, which states that the optimal signal is gained by using a size of filter equivalent to the size of the expected signal. It works by averaging the data in each voxel with the data in surrounding voxels and subsequently smoothing the data. This also helps to account for individual variation in the distribution of gyri, sulci and cytoarchitectonic borders across participants. The size of the filter determines the shape of the function that is used to smooth the data across voxels. Spatial smoothing convolves the data with a Gaussian kernel, which has a normal distribution curve, with a full width at half maximum (FWHM) of the expected cluster size. In the whole-brain analyses presented in this thesis I used a 5mm FWHM Gaussian kernel for Chapters 3 and 5 and 8mm FWHM in Chapter 4. There are some circumstances when it is preferable to avoid directly smoothing and I discuss this further in Chapters 3 – 5 when I present my region of interest analyses.

## 2.3 Statistical analysis

In fMRI data, the stimulus effect is very small, compared to the background noise and can be as little as a 1% change in the overall BOLD signal. However, through the use of statistical software packages, such as FSL, true physiological changes can reliably be identified. FSL uses a multi-level approach to analysing fMRI data. At each level the General Linear Model (GLM) is used to produce summary statistics that are then passed on to the next level (Beckmann et al., 2003). Typically, the first level analyses each run for each scan, the intermediate level uses a fixed effects analysis to combine runs within participants and the higher level uses random effects analysis to average data across participants or compare between groups. I explain the General

Linear Model and the different levels of analysis in more detail in the following sections.

### 2.3.1 The General Linear Model

The General Linear Model ($y = x\beta + e$) describes a response (y), such as a voxel's BOLD response, in terms of all its contributing factors ($x\beta$) in a linear combination (x is the design matrix; β is the matrix of the parameter estimates for each EV), while also accounting for the contribution of error (e).

#### *2.3.1.1 The dependent variable*

The dependent variable, y, describes the BOLD response at each time point from a single voxel, as measured by the fMRI scan.

#### *2.3.1.2 The independent variable*

Also known as the predictor, this considers all available knowledge about experimentally controlled factors and potential confounds.

#### *2.3.1.3 Parameters*

Also known as regression coefficients or beta weights, these quantify how much each specific predictor (x) independently influences the dependent variable (y). The parameter (β) chosen for a model should minimise the error by reducing the amount of variance in y that is left unexplained.

#### *2.3.1.4 Error*

This is the variance in the data (y) that is not explained by the linear combination of predictors (x).

#### *2.3.1.5 The GLM Matrix*

As samples of a response (y) are taken many times in functional imaging, this equation actually represents a matrix. Each predictor (x) has an expected signal time course, which contributes to the dependent variable y. Figure 2.4 shows a schematic outline of how the GLM is used to explain the fMRI data.

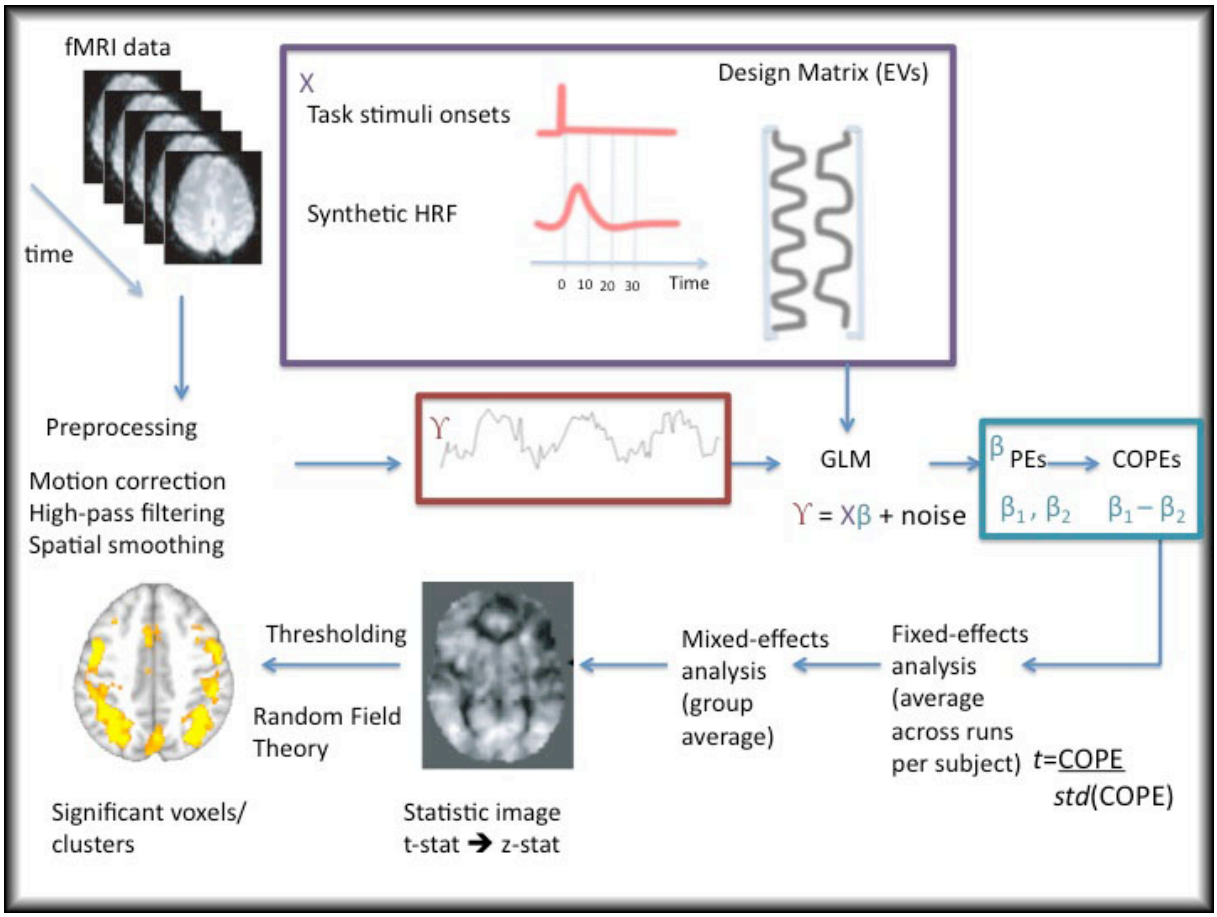



***Figure 2.4, Schematic outline of fMRI data analysis***

*The General Linear Model (GLM) explains the data (Y) by taking the timecourse of each experimental condition, convolved with a synthetic HRF (X). Parameter estimates (β) reflect how well each EV (X) fits the data (Y) at each voxel. Contrasts of parameter estimates (COPEs) are generated by subtracting one PE from another. COPEs are converted into t-statistic images by dividing the COPE by its standard error. The t-statistic is then transformed into a z-statistic image, which is then thresholded using Gaussian Random Field-based cluster or voxel inference. Fixed-effects analysis is used for averaging across runs within participants, and mixed-effects analysis is used for higher-level multi-participant group averages.*

### 2.3.2 First-level analysis

At the first level, the GLM is used for voxelwise time series analysis done separately for each run for each individual. The GLM is a model of the expected BOLD responses given the specific timings of the various experimental conditions. The design matrix models all of the experimental conditions as 'explanatory variables' (EVs), which describe the time course of each condition, convolved with a

stereotypical HRF. The design matrix is therefore a description of what would be expected if there were an effect of any condition. This is then compared to the real data in order to evaluate if any significant effects exist. For the analyses presented in this thesis I used FSL's one-column format, with a binary timecourse for each EV, with 1 indicating that that timepoint was the specific condition and 0 to indicate that it was not. Each EV in the design matrix results in a parameter estimate (PE) image, which reflects how strongly the EV fits the data at each voxel. This estimates both the signal and the noise (the effect and the variance). The standard error is derived from the residual noise after fitting the model, and to convert a PE into a t-statistic image, the PE is divided by its standard error. This t-statistic image is then transformed into a z-statistic image. To compare brain activation across two conditions, contrasts of parameter estimates (COPE) are generated by subtracting one PE from another and calculating the combined standard error. This contrast performs a t-test at each voxel of the image, testing whether the difference between parameter estimates for two experimental conditions is significantly different from zero.

### 2.3.3 Fixed-effects analysis (intermediate level)

With the exception of the study presented in Chapter 3, which only had one run, in my studies participants performed more than one run of the same task. These runs were analysed separately at the first-level and then combined at the intermediate level. This level uses a within-participant fixed-effects analysis. The fixed-effects model assumes that for a given participant there is no difference between the variances of different runs. This type of analysis explores significant effects from the specific participants included and indicates that the same result would be found if the study were to be repeated on the same group of participants. It does not make any inferences about the wider population.

### 2.3.4 Mixed-effects analysis (higher level)

Mixed-effects analysis considers the variance of each participant and accounts for between-participant variance and the possibility of 'outliers' (unusually large or small effects). This type of analysis can be used to make inferences at the population-level by indicating that the same result would be found if the study were to be repeated in a different group of participants with the same demographic profile. Comparisons between different groups can also be made with this type of analysis. At the higher

level of all the analyses presented in this thesis I used mixed-effects in order to be able to make generalisations to the wider population. I used FLAME (FMRIB's Local Analysis of Mixed Effects) for this (Beckmann et al., 2003; Woolrich et al., 2004).

### 2.3.5 Thresholding

The statistical tests described above result in a z-statistic image, with one z-score for each of the many thousand voxels in the brain. With such a large amount of data points, even with a standard statistical threshold of $p < 0.01$ ($z > 2.3$), there may be up to 2000 voxels that appear significant by chance (for example, 1% of 200,000). Subsequently the risk of falsely rejecting the null hypothesis (type II error) is very high and it is therefore necessary to correct for multiple comparisons. If all the voxels were considered to be independent, then a Bonferroni correction, a traditional multiple comparison correction method, could be used. However, as there are so many data points to consider, and they are not truly independent but display spatial correlation, this type of correction is often too conservative for applying to fMRI data (Nichols and Hayasaka, 2003). In a typical fMRI data set, the z-scores of any one voxel are highly correlated with the values of neighbouring voxels, so a more appropriate method of correction is one that accounts for the number of possible independent observations in regions of activity that are biologically plausible, i.e. from the spatially smoothed data.

In FSL, cluster-based correction for multiple comparisons is carried out using the Gaussian Random Field Theory. As explained earlier, the data analysis produces an image showing a z-statistic for each voxel. The threshold (typically, as in the analyses I present in this thesis) of $z > 2.3$ is applied, so that voxels with a z-statistic of lower than 2.3 are set to 0, which allows identification of contiguous clusters. The original z-statistic image is then masked using these significant clusters. Inference is then based on cluster size. The specific threshold I used in the studies presented in this thesis used Gaussian Random Field-based cluster inference with a standard height threshold of $z > 2.3$ and a cluster significance threshold of $p < 0.05$.

## 2.4 Registration

There is considerable variability in the size and shape of individual brains and so it is important to ensure that the same brain regions are being compared across

participants before making any inferences about which regions are involved in any given task. To do this, each image needs to be put into an identical ‘standard space’. This is also important due to the poor spatial resolution of fMRI data. In FSL registration is carried out by FMRIB’s Linear Image Registration Tool (FLIRT) (Jenkinson et al., 2002). There are two stages to this. First, EPI functional data is registered to the T1-weighted structural image (after brain-extraction) as this has much higher resolution. This registration process uses six degrees of freedom transformations (translation and rotations in each of the three dimensions). During this process each slice from the data set is adjusted so that it is aligned with a target slice, by using transformations in six directions so that all slices match up. This target slice can be any slice and in the analyses I present in this thesis, I used the middle volume. The second stage registers the high-resolution structural image to a standard brain template, using 12 degrees of freedom transformations. In my analyses I used the MNI-152 (Montreal Neurological Institute) template, which is based on the T1 MRI scans of 152 normal participants. The six additional affine transformations used in this stage of registration correct for the differences in the size and shape of brains across participants. Figure 2.5 shows the variation across individual T1-weighted images and the difference between those images and the standard brain template.

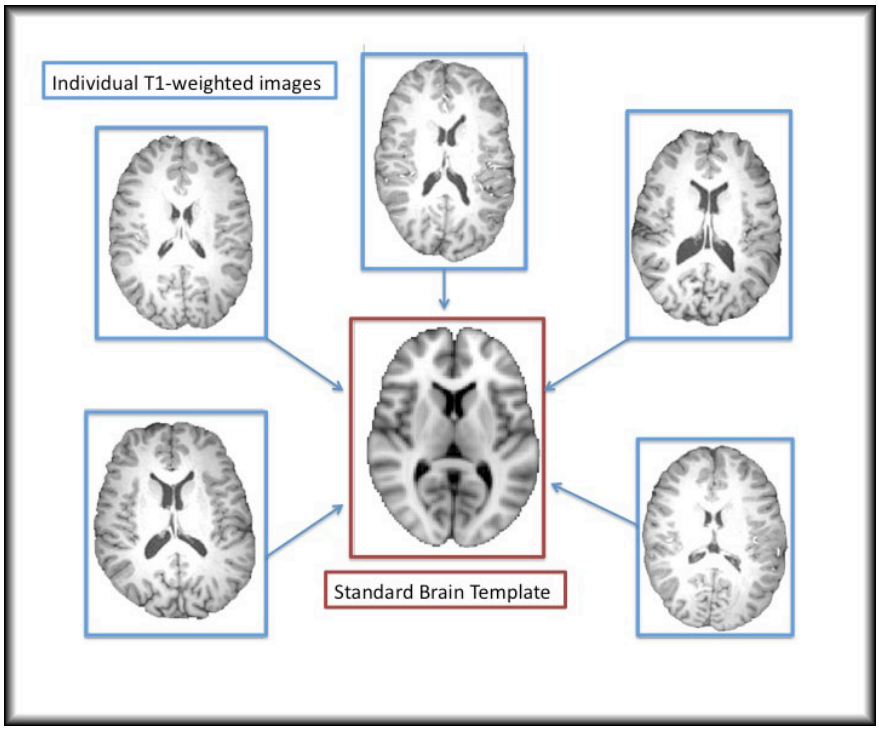



***Figure 2.5, Individual T1-weighted images and a standard brain template***

*Individual T1-weighted images show considerable variation across participants. In order to combine data from a number of participants and reveal significant results at the group level, it is necessary to register each participant's image to a standard brain template.*

## 2.5 Region of interest analysis

When there is a clear hypothesis about the involvement of a specific brain region in a particular task, statistical power can be increased by using a region of interest (ROI) analysis. This type of analysis may increase the signal-to-noise ratio and reduce the problem of multiple comparisons. ROIs can be defined from significant activations revealed by higher-level analyses from independent data, but in the studies presented in this thesis, I used anatomical boundaries. The FSL tool Featquery examines the FEAT results and extracts the mean percentage BOLD signal change within an ROI.

The neurovascular coupling has not been shown to be uniform throughout the brain, so it is not always appropriate to compare main effects between ROIs. Therefore,

ROI analyses typically compare the effect size within one region across different tasks, which is the technique used in the studies presented here.

## 2.6 Scanning protocol

### 2.6.1 Scanner parameters

Throughout the work presented in this thesis, the scanning parameters were kept constant. MRI data were obtained using a Philips (Best, The Netherlands) Intera 3.0 Tesla MRI scanner using Nova Dual gradients, a phased array head coil and sensitivity encoding with an undersampling factor of 2.

### 2.6.2 Structural T1

High-resolution images (T1-weighted) were acquired with the following acquisition parameters: matrix size 208 x 208; slice thickness = 1.2mm, 0.94mm x 0.94mm in plane resolution, 150 slices; TR = 9.6 ms; TE = 4.5 ms; flip angle 8°.

### 2.6.3 Functional MRI

MRI data were obtained using a Philips Intera 3.0 Tesla scanner, using dual gradients, a phased array head coil, and sensitivity encoding with an undersampling factor of 2. Functional MR images were obtained using a T2*-weighted, gradient-echo, echoplanar imaging (EPI) sequence with whole-brain coverage (TR = 8s, acquisition time = 2s, giving 6s for the participants to speak during silence; echo time = 30ms; flip angle, 90 ˚). Thirty-two axial slices with a slice thickness of 3.25mm and an interslice gap of 0.75mm were acquired in ascending order (resolution, 2.19 x 2.19 x 4.00mm; field of view, 280 x 224 x 128 mm). Quadratic shim gradients were used to correct for magnetic field inhomogeneities within the brain. T1-weighted images were also acquired for structural reference. Stimuli were presented visually using E-Prime software (Psychology Software Tools) run on an IFIS-SA system (In Vivo Corporation). Sounds were delivered through MR-compatible headphones and speech was recorded using a fibre-optic noise-cancelling microphone.

Participants in the studies presented in this thesis had no history of neurological problems or hearing loss. The studies were approved by the local research ethics committee and all participants gave informed written consent.

# 3 Using overt and covert speech to investigate motor-sensory speech control

Investigating speech production with fMRI has proved to be a challenge. This Chapter addresses the broad question of how to investigate speech production, to provide the context within which to study native and non-native speech. The main difficulties relate to artefacts generated by movements of the articulators and changes in respiration when participants produce speech. These artefacts occur whether the study is limited to syllable repetition or extended to investigate more natural connected speech. In order to avoid artefacts as much as possible, much fMRI research on speech production has used covert rather than overt speech. A series of covert speech production studies by Hickok and colleagues have resulted in defining a region for auditory-motor integration (the 'sensorimotor interface') for speech production. These authors locate, with considerable anatomical precision, this interface in left posterior planum temporale, sometimes extending up into the parietal operculum (Buchsbaum et al., 2005; Hickok et al., 2003; Hickok et al., 2009; Pa and Hickok, 2008), and have termed it 'area Spt'. It is considered to be a homologue of a subportion of non-human primate auditory area Tpt (Pandya and Sanides, 1973), which Galaburda and Sanides (1980) locate caudal to the external parakoniocortex and extending towards the temporo-parietal junction. The assumption is that area Spt is active in the absence of actual motor or sensory activity, perhaps as the result of pre-articulatory feedforward signals from premotor areas to sensory regions that respond to post-articulatory sensory feedback. However, it remains an open question whether it is possible to study speech production without actual vocal production. The study presented in this Chapter was designed to compare activity in response to both overt and covert speech production, to identify common systems involved in both types of speech production and also to reveal functional components that are not present when speaking covertly. Two different types of speech task were used in this study, both propositional and non-propositional speech (specifically, sentence production and counting).

Previous studies have directly compared overt and covert speech, but with conflicting results. Similar patterns of activation for the two types of speech have been reported

(Palmer et al., 2001; Yetkin et al., 1995), as have different activation patterns (Barch et al., 1999; Huang et al., 2001; Shuster and Lemieux, 2005). These inconsistent results may be partly explained by the very small numbers of participants included in these studies (ten or fewer), which will have resulted in the results being less reliable. These previous studies also used continuous image acquisition, rather than sparse sampling, even for overt speech production. Kemeny and colleagues (2005) have used continuous acquisition for overt speech production, to compare EPI (BOLD) and arterial spin labelling, and they suggest that artefacts arising from using the BOLD technique to investigate speech production degrade activity in perisylvian cortex and increase 'activation' in temporopolar/frontobasal regions. As head movement was not particular large in their study they suggest that these artefacts arise as a consequence of the susceptibility effect generated by tongue and jaw movements and airflow changes that cause inhomogeneities in the magnetic field. The false-positive temporopolar activations and the degraded inferior frontal gyrus (IFG) signal led the authors to conclude that continuous overt speech production cannot be accurately investigated using BOLD imaging. The present study used the BOLD technique, but also employed sparse imaging to avoid these artefacts.

Another reason for the discrepancy across these previous studies could be due to the specific tasks involved. One task used in the comparison of overt and covert speech has been the Stroop test (Barch et al., 1999), which emphasises domain-general cognitive functions, such as selective attention and inhibition of prepotent responses, rather than speech production itself. Word-stem completion, i.e. generating a word in response to a three-letter cue, has also been used for comparing the two types of speech production (Palmer et al., 2001; Rosen et al., 2000), as has word generation following a letter cue (Lurito et al., 2000; Yetkin et al., 1995). These tasks only require participants to produce single words, which again does not fully engage the speech production system. Further, like the Stroop task, they are dependent on frontal executive systems that are not necessarily engaged during normal propositional speech; and they were originally developed by neuropsychologists to probe problem solving rather than language. In order to investigate speech production in detail, it is preferable to use more ecologically-valid speech tasks. The results will then relate more closely to normal daily discourse.

Therefore, in this study, tasks were specifically designed to avoid single word responses.

The overt and covert conditions included in this study required participants to use both propositional and non-propositional speech. These conditions ensured that the speech produced would be similar across the two types of speech production (overt and covert), in terms of the rate of word production. Overt and covert speech clearly differ with regard to output to the motor units controlling the articulators, larynx and respiratory muscles, and sensory (auditory and somatosensory) feedback.

## 3.1 Aims and hypothesis

This study was a precursor to investigating non-native speech production, and was designed to answer two broad questions related to imaging speech production, prior to subsequent studies, presented in later Chapters of this thesis, that focus on non-native speech production.

The aim was to investigate whether covert speech provides accurate insight into the processes supporting normal overt speech, other than activity associated with feedforward motor and feedback sensory activity; and whether intention (covert speech) activates a 'sensorimotor interface' in the absence of actual motor and sensory activity. The hypothesis was that overt speech engages a more distributed system than covert speech and is therefore the only appropriate method to use for investigation of motor-sensory control. The aim of this study was achieved by comparing activity in premotor, motor and sensory regions during both overt and covert speech, with ROI analyses in the frontal and parietal opercula and the plana temporale.

## 3.2 Materials and methods

### 3.2.1 Participants

Seventeen right-handed native speakers of English (eight female) took part in this study, with an age range of 21 years, 10 months – 61 years, 3 months (mean 28 years, 3 months).

### 3.2.2 Experimental conditions

There were four language tasks (propositional and non-propositional speech with two response types, overt and covert speech) and a rest baseline (referred to as Overt Speaking, Overt Counting, Covert Speaking, Covert Counting and Rest). The propositional speech tasks required participants to describe nouns, which were selected using the Medical Research Council psycholinguistic database (Wilson, 1998). All had high values for concreteness and imagability, although frequency was variable. Fifty nouns were selected from the database and then randomly assigned to either the overt or covert speaking conditions. There were no significant differences between mean values for concreteness, imagability or frequency between the words

assigned to the overt condition and those assigned to the covert condition. The list of nouns and their psycholinguistic values are presented in Table 3.1. Non-propositional speech was tested with a counting task, counting upwards from one for the duration of the trial at a rate of ~one per second. Stimuli were displayed on an MRI compatible screen for 7.5 seconds, and participants were instructed to start speaking as soon as the stimuli appeared on the screen. The end of the task was indicated by a fixation cross. All tasks were preceded by an image that indicated whether the following task was to be performed overtly or covertly. The rest condition consisted of a series of Xs displayed on the screen, and no response from the participant was required. There were 25 trials of each speech condition and 20 rest trials, presented in a pseudo-randomised order.

| Stimuli | Concreteness | Imagability | Frequency |
|---|---|---|---|
| | | *Overt Trials* | |
| aunt | 564 | 567 | 22 |
| band | 590 | 579 | 53 |
| branch | 583 | 548 | 33 |
| bush | 585 | 549 | 14 |
| camp | 571 | 588 | 75 |
| cheek | 565 | 561 | 20 |
| earth | 580 | 580 | 15 |
| frog | 619 | 617 | 1 |
| glove | 607 | 596 | 9 |
| heart | 605 | 617 | 173 |
| lamb | 633 | 614 | 7 |
| lamp | 615 | 575 | 18 |
| lunch | 552 | 602 | 33 |
| nurse | 588 | 617 | 17 |
| page | 571 | 555 | 66 |
| plug | 558 | 583 | 23 |
| pool | 573 | 577 | 111 |
| road | 583 | 609 | 197 |
| rope | 608 | 596 | 15 |
| sauce | 576 | 569 | 20 |
| shed | 611 | 602 | 11 |
| straw | 603 | 568 | 15 |
| thumb | 638 | 599 | 10 |
| tool | 570 | 538 | 40 |
| worm | 611 | 578 | 4 |
| | | *Covert Trials* | |
| blade | 584 | 568 | 13 |
| bowl | 575 | 579 | 23 |
| cake | 624 | 624 | 13 |
| deer | 631 | 624 | 13 |
| doll | 588 | 565 | 10 |
| flame | 582 | 598 | 17 |
| fork | 592 | 598 | 14 |
| fruit | 612 | 587 | 35 |
| grape | 611 | 591 | 3 |
| hill | 588 | 607 | 72 |
| knee | 593 | 597 | 35 |
| mouse | 624 | 615 | 10 |
| pond | 623 | 599 | 25 |
| sheet | 608 | 594 | 45 |
| shirt | 616 | 612 | 27 |
| shop | 549 | 561 | 63 |
| snake | 621 | 627 | 44 |
| sock | 581 | 553 | 4 |
| song | 514 | 578 | 70 |
| steak | 646 | 647 | 10 |
| stew | 603 | 587 | 5 |
| stove | 591 | 592 | 15 |
| tent | 608 | 593 | 20 |
| wood | 606 | 577 | 55 |
| wool | 608 | 586 | 10 |

***Table 3.1, Stimuli list with values for linguistic variables***

*There were no significant differences ($p > 0.2$) between the mean values of the two word lists (overt and covert) for concreteness, imagability or frequency. Mean values (with standard deviation) were: Overt Concreteness 590.36 (23.37), Covert Concreteness 599.12 (27.37), Overt Imagability 583.36 (23.22), Covert Imagability 594.36 (22.29), Overt Frequency 40.08 (50.05) and Covert Frequency 26.04 (20.90).*

### 3.2.3 Behavioural performance

Audio recordings were taken of the participants' vocalisations in order to ensure that the participants responded appropriately with regard to overt speech and to ensure that they had not spoken overtly during a covert trial. Rare trials in which the participants failed to respond during an overt trial, or spoke overtly when the task should have been carried out covertly, were excluded from analyses.

### 3.2.4 Data acquisition

The fMRI acquisition parameters were the same as described in Chapter 2.6 of this thesis, with one run of 120 volumes and a TR of 10 seconds.

### 3.2.5 Data analysis

The initial whole-brain analysis was a 2 (Overt and Covert) x 2 (Speaking and Counting) ANOVA. Subsequent to this, I contrasted directly between each condition and the common baseline condition of Rest. Finally, I performed direct contrasts between the various speech-related conditions.

### 3.2.6 Region of Interest analysis

Four ROIs were defined on an individual basis, separately for each hemisphere, using Freesurfer's autosegmentation, based on gyral and sulcal landmarks (detailed below) (Figure 3.1). Three ROIs were around the temporo-parietal junction (anterior and posterior planum temporale, and parietal operculum) and the fourth was the frontal operculum. There were three reasons for selecting the three regions of interest around the temporo-parietal junction. First, it has been proposed based on a computational model of speech production that these regions integrate predictive feedforward and post-articulatory sensory (auditory and somatosensory) feedback signals during articulation (Golfinopoulos et al., 2010; Guenther, 2006). Second, these regions, with a particular emphasis on the left posterior planum temporale, are considered the sensorimotor interface for speech production, area Spt (Hickok et al., 2003; Hickok et al., 2009; Pa and Hickok, 2008). Third, although posterior auditory and somatosensory association cortices are closely adjacent and anatomically variable (Ono et al., 1990; Penhune et al., 1996), it is possible to separate these regions anatomically using probabilistic anatomical atlases. The frontal operculum

was selected to investigate changes in the premotor responses to different types of speech.

The frontal operculum (fOp) and the planum temporale (PT) were labelled using Freesurfer's automatic parcellation. The anterior and posterior halves of the planum temporale (antPT and postPT, respectively) were defined on an average brain surface and then applied to individual brains using Freesurfer. This decision was made in light of the functional heterogeneity of the PT, with the anterior half displaying a more canonical auditory response and the posterior half corresponding to both auditory-motor integration and responding to acoustic stimulation produced by the human vocal tract (Pa and Hickok, 2008). In the absence of a defined parietal operculum (PO) within Freesurfer, I employed the boundaries by Eickhoff and colleagues (2006). In brief, the parietal operculum comprises dorsal cortex within the posterior lateral sulcus, with the anterior border defined by the post-central sulcus and the medial border by the circular sulcus of the insula. The cortical surface was reconstructed from each participant's high-resolution T1 scan using Freesurfer (Dale et al., 1999), and the ROIs were then automatically defined for each individual's reconstructed cortical surface. This approach has been shown to be comparable in accuracy to manual labelling of brain regions (Fischl et al., 2002). Mean effect sizes for overt and covert propositional and non-propositional speech conditions, relative to rest, were calculated for each individual. For this analysis the functional data were not spatially smoothed before averaging, in order to avoid any possibility of blurring of activation across the Sylvian fissure between temporal and parietal lobes.

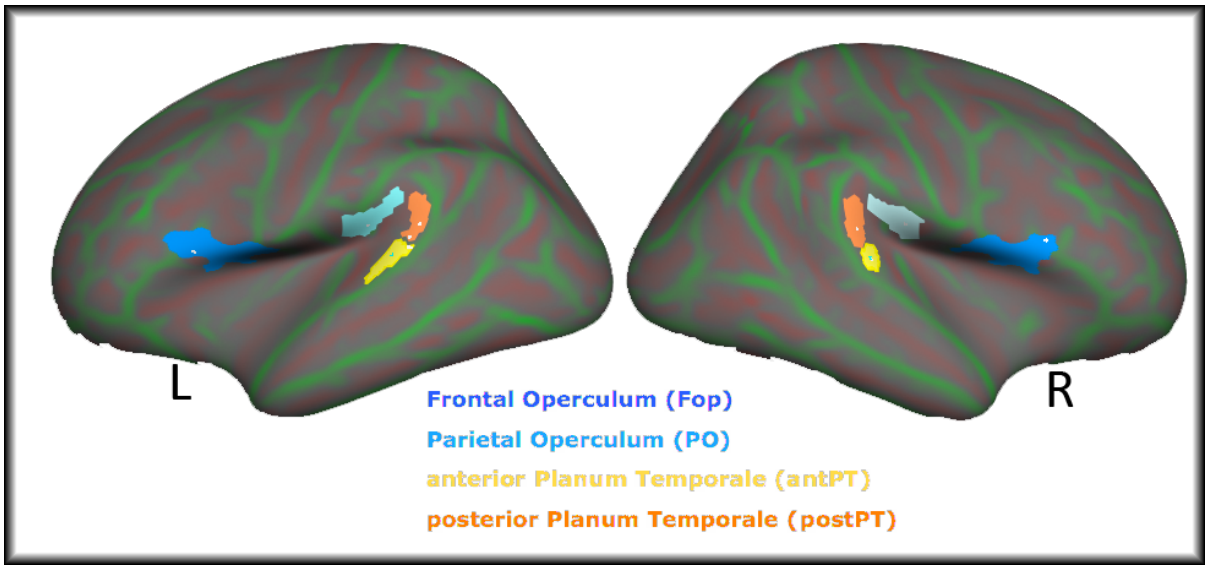



***Figure 3.1, Regions of interest***

*The masks used for the ROI analyses are shown in both hemispheres on an inflated Freesurfer standard brain image. The frontal operculum is shown in dark blue, parietal operculum in light blue, anterior planum temporale in yellow and posterior planum temporale in orange. These masks were used for analyses in Chapters 3, 4 and 5.*

## 3.3 Results

### 3.3.1 Behavioural performance

Technical problems meant that audio recordings were not available for one of the participants, who was then excluded from analysis. The 16 participants with audio recordings generally responded accurately. Seven participants made incorrect responses during the speech tasks, and these trials were excluded from fMRI analyses. The specific conditions for each incorrect trial are shown in Table 3.2.

| Participant | Overt Speaking | Covert Speaking | Overt Counting | CovertCounting |
|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 0 |
| 7 | 0 | 0 | 0 | 1 |
| 10 | 4 | 0 | 4 | 0 |
| 12 | 2 | 0 | 0 | 0 |
| 13 | 0 | 0 | 1 | 1 |
| 14 | 0 | 1 | 0 | 0 |
| 16 | 2 | 2 | 0 | 0 |

***Table 3.2, Number of excluded trials from participants who made incorrect responses during the fMRI speech tasks***

*The number of trials excluded for each speech condition (out of 25) for the seven participants who responded incorrectly during the fMRI speech tasks.*

### 3.3.2 Functional MRI analysis

The whole-brain 2 (Overt and Covert) x 2 (Speaking and Counting) ANOVA revealed a main effect of Overt/Covert and a main effect of Speaking/Counting, but no significant interaction. A table reporting the main cluster peaks for the main effects is shown in Appendix 1 (p223 of this thesis).

### *3.3.2.1 Main effect of Overt/Covert*

The main effect of Overt/Covert was observed in medial premotor cortex (the supplementary motor area, SMA), primary sensorimotor and auditory cortices, thalami, putamen and paravermal cerebellum (Figure 3.2). The pattern of activity was symmetrically distributed between both cerebral and cerebellar hemispheres.

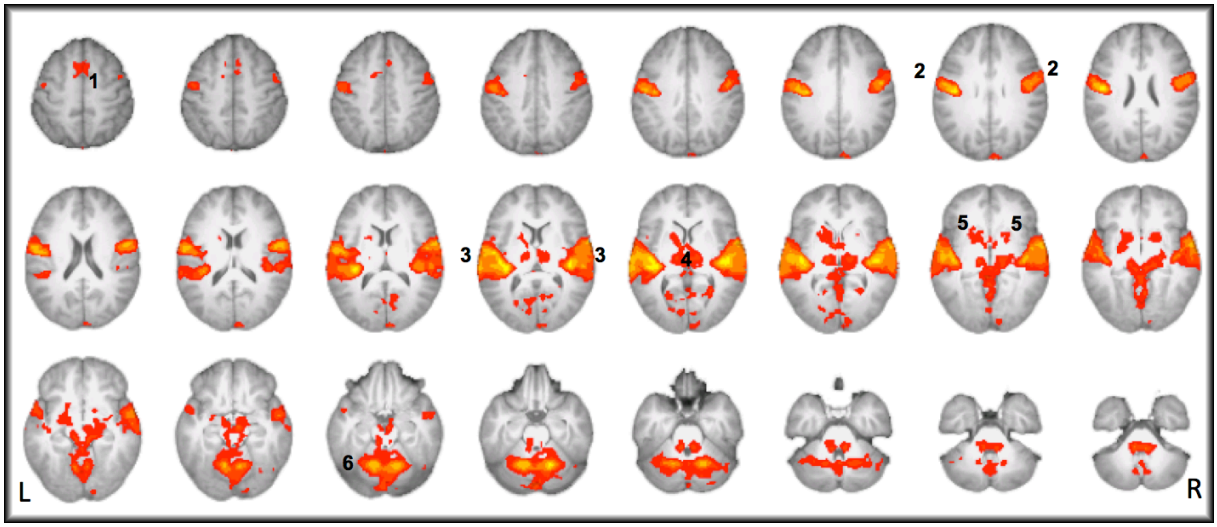



***Figure 3.2, Main effect of Overt/Covert***

*Axial slices are shown in 4mm decrements; from left to right, the top panel shows slices 50 – 22, middle panel slices 18 – -10 and lower panel slices -14 – -42. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Bilateral activity is shown in: (1) medial premotor cortex, (2) primary sensorimotor cortex, (3) primary and association auditory cortex, (4) left and right thalamus, (5) basal ganglia and (6) paravermal cerebellum. Results are displayed on a standard brain template (MNI152) and left is shown on the left.*

### *3.3.2.2 Main effect of Speaking/Counting*

The main effect of Speaking/Counting was observed in left inferior frontal gyrus extending to the superior frontal gyrus, as well as pre-supplementary motor area (pre-SMA), paracingulate cortex, left inferolateral temporal cortex, bilateral thalami, caudate and putamen, and parvermal cerebellum (Figure 3.3). What appears to be activity in the left lateral cerebellum is likely to be activity smoothed down from the adjacent inferior temporal cortex.

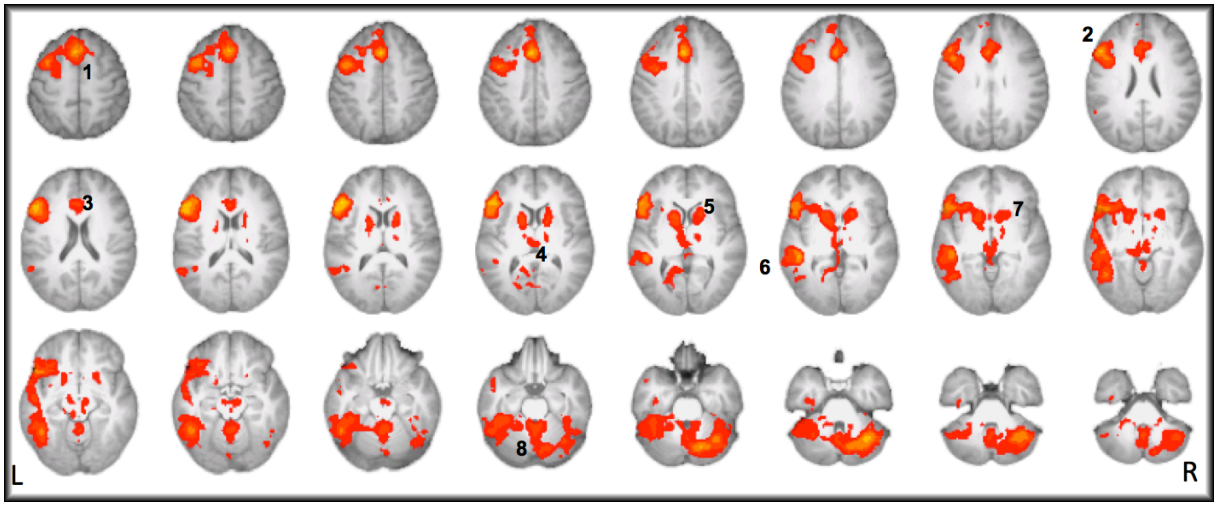



***Figure 3.3, Main effect of Speaking/Counting***

*Axial slices are shown in 4mm decrements; from left to right, the top panel shows slices 50 – 22, middle panel slices 18 – -10 and lower panel slices -14 – -42. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Activity is shown in: (1) pre-SMA, extending in lower planes into anterior cingulate cortex, (2) left inferior frontal gyrus, (3) paracingulate cortex, (4) bilateral thalamus, (5) bilateral caudate nucleus, (6) left inferolateral temporal cortex, (7) bilateral putamen and (8) cerebellum. Results are displayed on a standard brain template (MNI152) and left is shown on the left.*

#### *3.3.2.3 Overt Speaking > Rest*

Overt Speaking produced bilateral activation: in the supplementary motor area, extending into anterior cingulate and paracingulate cortices; primary sensorimotor cortices; secondary somatosensory cortices within the parietal opercula; basal ganglia; thalami; the superior temporal gyri; and the cerebellum (Figure 3.4A). In addition, there was left-lateralised activity in the inferior and middle frontal gyri, frontal operculum, and posterior inferior temporal gyrus. A table reporting the main cluster peaks for these contrasts is shown in Appendix 2 (p223 of this thesis).

#### *3.3.2.4 Covert Speaking > Rest*

Covert Speaking returned an activation pattern that appeared to be similar to but less extensive and more left-lateralised than overt propositional speech (Figure 3.4B). There was activity in the pre-SMA, extending into cingulate and paracingulate cortices, left inferior frontal gyrus, left middle temporal gyrus and right cerebellum.

#### *3.3.2.5 Overt Counting > Rest*

Overt Counting also resulted in an activation pattern with some regions that appeared common to those observed during overt propositional speech, but were less extensive (Figure 3.4C). At the statistical threshold employed, there was bilateral activation in primary sensorimotor and auditory cortices, secondary somatosensory cortices in the parietal opercula, and in paravermal cerebellum.

#### *3.3.2.6 Covert Count > Rest*

Covert Counting, relative to Rest, did not result in any activity for the given cluster correction of $z > 2.3$, a finding that replicates Kawashima and colleagues' (2000) silent counting condition.

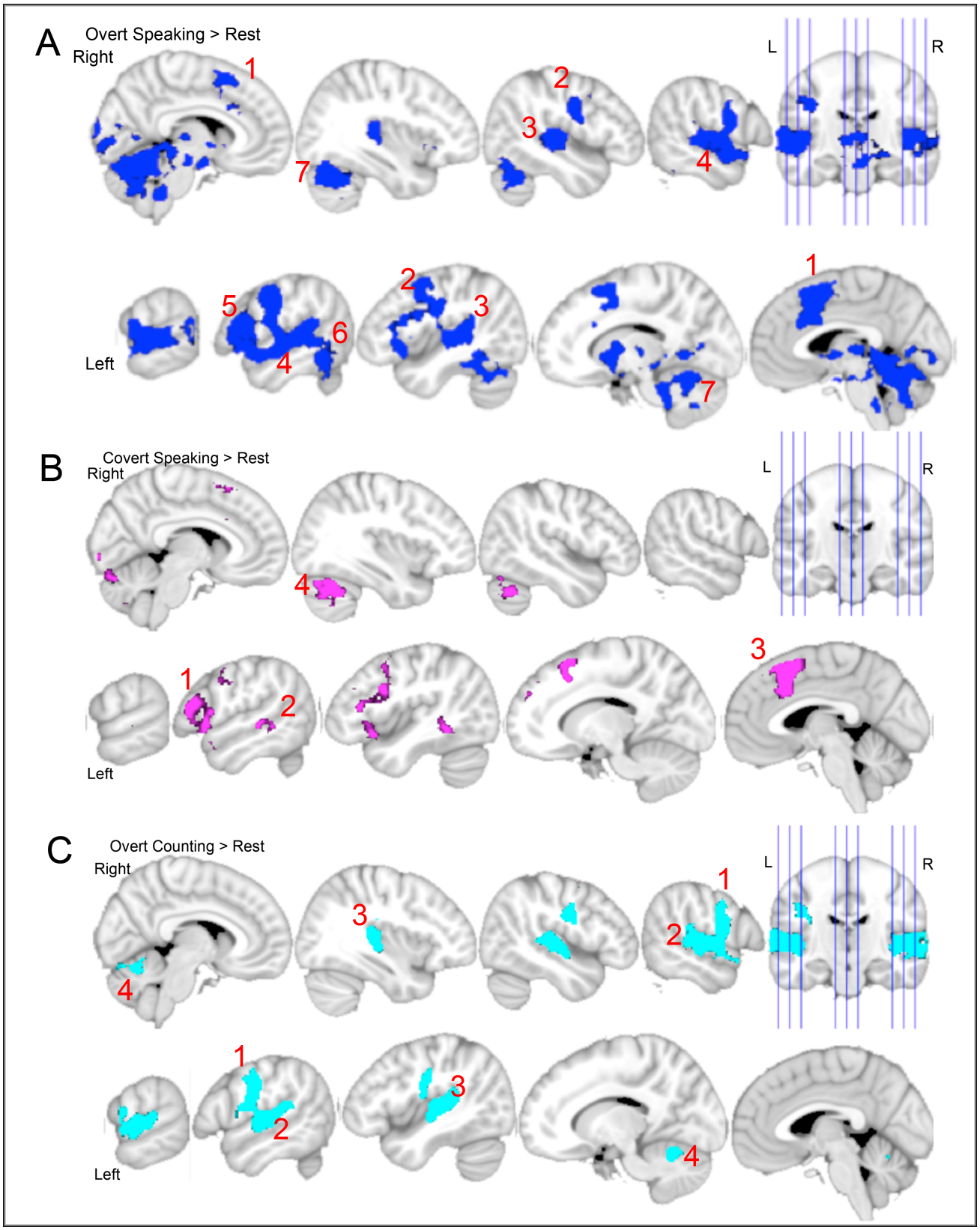



***Figure 3.4, Whole-brain results for three speech tasks, each against Rest (legend overleaf)***

***Figure 3.4, Whole-brain results for three speech tasks, each against Rest (Figure on previous page)***

*A: Overt Speaking > Rest (dark blue) revealed activity in 1) the SMA, extending into anterior cingulate and paracingulate cortices, 2) bilateral primary sensorimotor cortices, 3) bilateral secondary somatosensory cortices, 4) bilateral superior temporal gyri, 5) left inferior frontal gyrus, 6) left posterior inferior temporal gyrus and 7) the cerebellum. B: Covert Speaking > Rest (pink) revealed activity in 1) left inferior frontal gyrus, 2) left middle temporal gyrus, 3) left pre-SMA, extending into anterior cingulate and paracingulate cortices and 4) the right cerebellum. C: Overt Counting > Rest (turquoise) revealed activity in 1) bilateral primary sensorimotor cortices, 2) bilateral primary auditory cortices, 3) bilateral secondary somatosensory cortices and 4) bilateral paravermal cerebellum. In each section sagittal slices are shown; the top panel shows slices from the right hemisphere, (from left to right: x =7, 37, 47, 57 and the orthogonal coronal slice); the bottom panel shows slices from the left hemisphere (from left to right: x = -63, -53, -43, -13, -3). Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Results are displayed on a standard brain template (MNI152).*

### *3.3.2.7 Summary of the comparison of conditions with Rest*

On visual inspection of the statistical maps of the contrast of the various speech-related conditions with Rest, Overt Speaking activated the voxels that were also active in the combined distribution of voxels observed with Covert Speaking and Overt Counting (Figure 3.5).

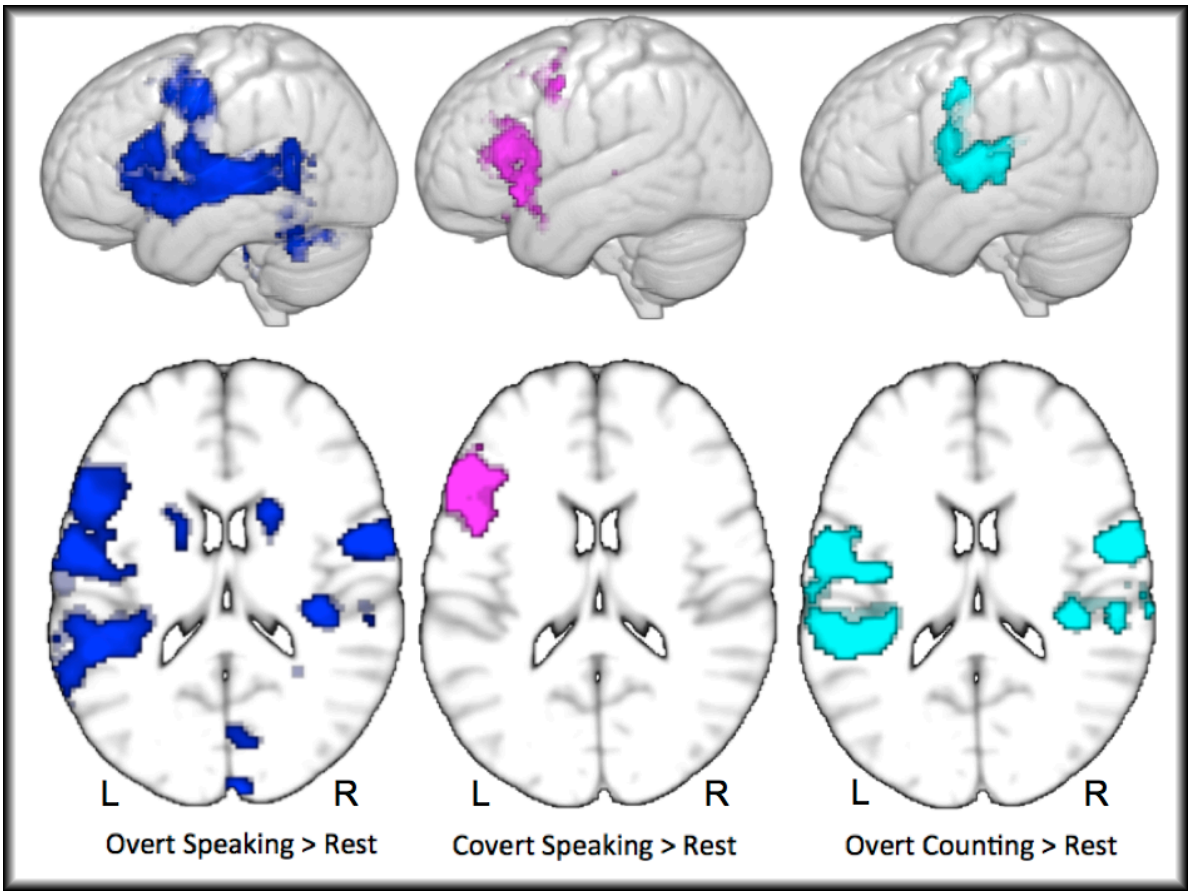



***Figure 3.5, Three speech tasks, each against rest***

*The results for three speech tasks, each against rest, are shown: Overt Speaking (dark blue), Covert Speaking (pink) and Overt Counting (turquoise). The sum of Covert Speaking and Overt Counting is closely matched to Overt Speaking. The top panel shows activity presented on a left hemisphere render of a standard brain template and the bottom panel shows an axial slice, z = 16. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05.*

#### *3.3.2.8 Direct comparisons between conditions*

The differences in the activation patterns across conditions, with Rest as the common baseline condition, presented in section 3.3.2.7 of this Chapter (p94), were based on visual inspection of the thresholded images. In this section, the four experimental conditions were directly compared. The direct contrast of Overt against Covert, for Speaking only, resulted in bilateral activity in the SMA, primary sensorimotor cortices, primary and association auditory cortices in the superior temporal gyri and the paravermal cerebellum. In addition there was signal in the left caudate nucleus but no other activity in the basal ganglia at the statistical threshold employed (Figure 3.6A). The same contrast for Counting revealed activity in a very similar distribution (Figure 3.6B).

The reverse contrast of Covert against Overt did not show activation at the threshold used, for either Speaking or Counting.

Directly contrasting Speaking with Counting, both for Overt (Figure 3.6C) and Covert speech (Figure 3.6D), demonstrated activity in the left inferior frontal gyrus and the basal ganglia. The Covert contrast showed additional activity in the anterior cingulate and the cerebellum (Figure 3.6D).

The opposite contrast of Counting against Speaking revealed activity in the posterior cingulate and right parietal operculum for both Overt (Figure 3.6E) and Covert (Figure 3.6F), with additional activity in the precuneus and right angular gyrus for Overt.

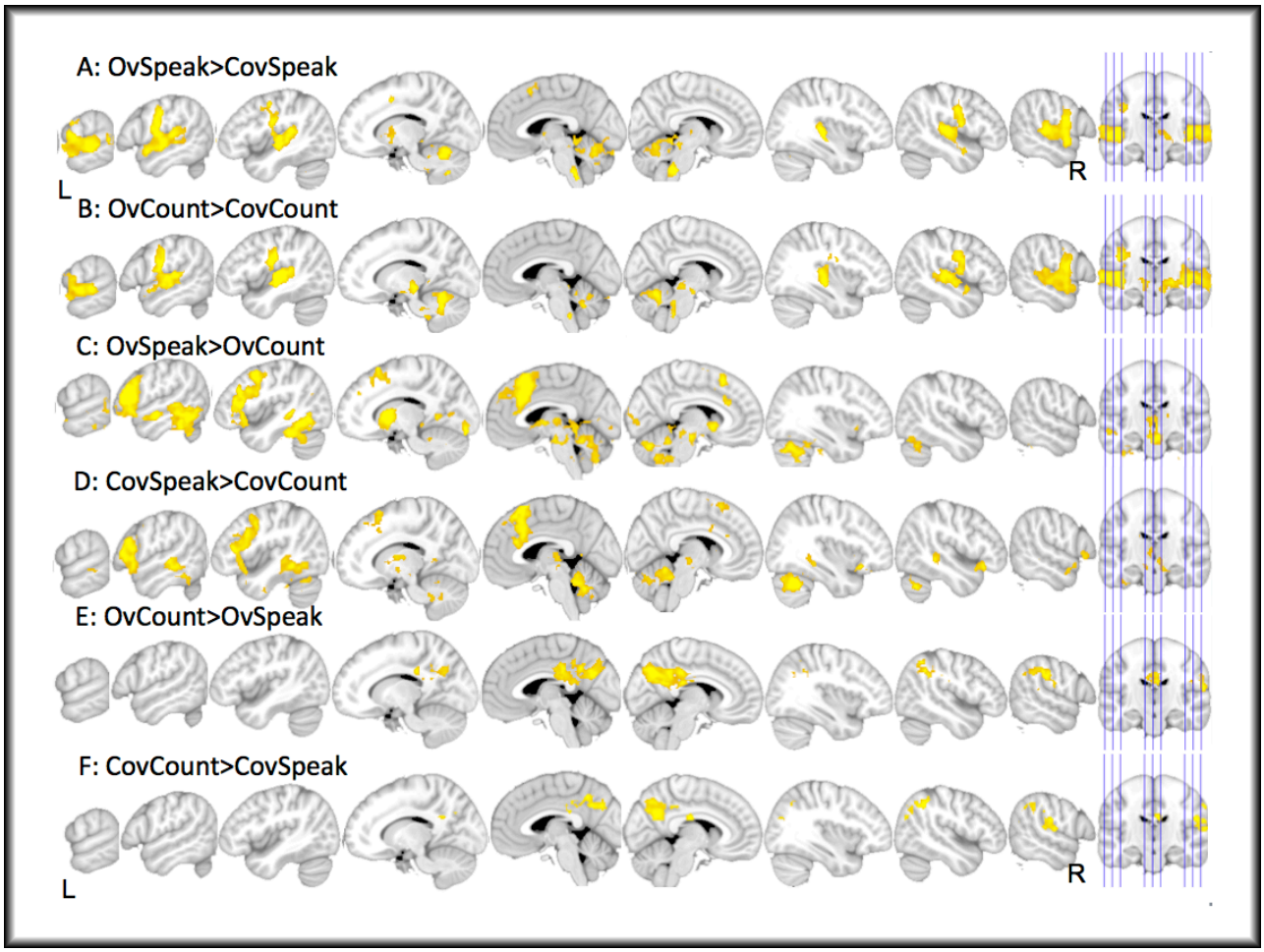



***Figure 3.6, Overt and Covert direct contrasts***

*A: Overt Speaking > Covert Speaking; B: Overt Counting > Covert Counting; C: Overt Speaking > Overt Counting; D: Covert Speaking > Covert Counting; E: Overt Counting > Overt Speaking; F: Covert Counting > Covert Speaking. For the six contrasts, sagittal slices are shown, from left to right x = -63, -53, -43, -13, -3, 7, 37, 47, 57 and the orthogonal coronal slice. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Results are displayed on a standard brain template (MNI152). A table reporting the main cluster peaks for these contrasts is shown in* Appendix 3 *(p224 of this thesis).*

### 3.3.3 ROI analysis

Activity in the theoretically motivated ROIs was measured for the four experimental conditions. The posterior ROIs allowed investigation of activity related to sensory feedback during speech production and the anterior ROI investigated the premotor control of articulation (Golfinopoulos et al., 2010). Therefore, these ROIs investigated the core motor-sensory aspects of different types of speech production, rather than cognitive or linguistic processing. They were entered into eight separate 2 (Overt and Covert) x 2 (Speaking and Counting) ANOVAs. In view of the multiple ANOVAs, the threshold for statistical threshold was set at $p < 0.01$. There was a significant main effect of Task (Speaking or Counting) in the left frontal operculum, (Speaking > Counting, $F$ (1,16) = 23.575, $p < 0.0005$) (Figure 3.7A). There was a main effect of type (Overt or Covert) in the left and right anterior PT (left Overt > Covert, $F$ (1,16) = 19.086, $p < 0.00005$; right Overt > Covert, $F$ (1,16) = 9.833, $p < 0.006$ (Figure 3.7). There was no task by type interaction in either hemisphere. Post-hoc, two-tailed paired t-tests, revealed significant differences in the left frontal operculum and left and right anterior planum temporale. These results are listed in Table 3.3. As is evident from Figure 3.6, neither of the posterior plana temporale demonstrated significant differences in activity between the speech conditions. Further, regional activity during these conditions was no greater than during Rest. In the adjacent parietal opercula, activity during Covert Speaking was significantly below Rest. Even in the anterior PT, activity was only present for the Overt conditions. Therefore, there was no evidence from this study that could confirm the extensive functional imaging literature on area Spt.

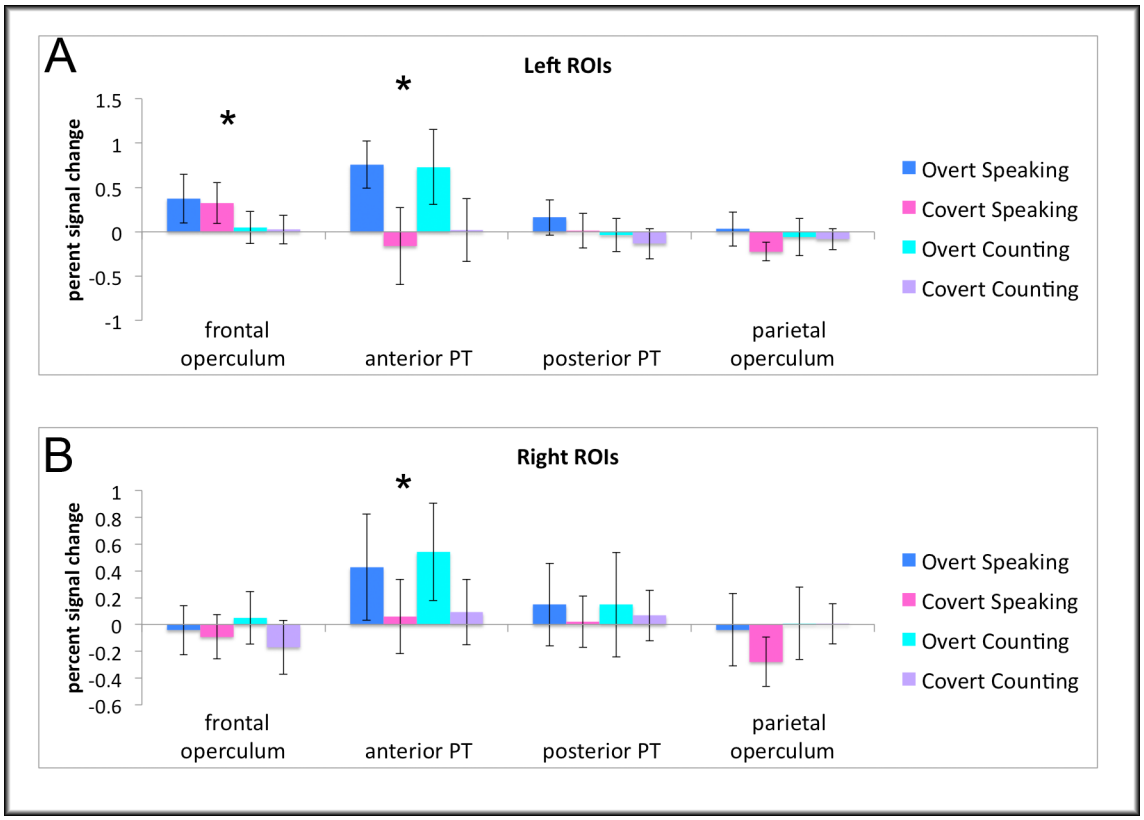



***Figure 3.7, ROI analyses for the four experimental conditions***

*A: Left ROIs, B: Right ROIs. Mean effect sizes for each of the speaking tasks contrasted with rest are shown: Overt Speaking (dark blue), Covert Speaking (pink), Overt Counting (turquoise) and Covert Counting (lilac). Error bars represent 95% confidence intervals. * significant interaction (Overt/Covert) x (Speaking/Counting).*

| **ROI** | **Contrast (mean)** | ***t*** | ***df*** | ***p*** |
|---|---|---|---|---|
| Left FOp | Overt Speaking (0.352) > Overt Counting (0.063) | 2.535 | 16 | 0.02 |
| | Covert Speaking (0.298) > Covert Counting (0.025) | 4.08 | 16 | 0.001 |
| Left antPT | Overt Speaking (0.752) > Covert Speaking (-0.009) | 5.145 | 16 | <0.001 |
| | Overt Counting (0.769) > Covert Counting (0.130) | 2.871 | 16 | 0.01 |
| Right antPT | Covert Counting (0.092) > Overt Counting (0.543) | -2.263 | 16 | 0.02 |

***Table 3.3, Significant t-tests following significant ROI ANOVAs***

## 3.4 Discussion

This study investigated propositional (sentence-level) speech production using fMRI and contrasted this condition with non-propositional speech (counting), with both covert and overt performance of these two speech tasks. The results for Overt Speaking relative to Rest and Overt Counting relative to Rest were broadly in accord with previous studies using position emission tomography (Awad et al., 2007; Blank et al., 2002), a technique that is less susceptible to motion and susceptibility artefacts than fMRI (Devlin et al., 2000; Scott and Wise, 2003). The striking differences were greater activity in left dorsolateral prefrontal cortex and in the left posterior inferolateral temporal cortex compared with the previous PET studies. This may relate to the task demands - noun definition in the current study compared to free narrative speech in response to a prompt question in the previous PET studies. Noun definition rather than free narrative speech was used because of data acquisition with sparse sampling, which required the participants to produce short time-limited sentences, maximum 7.5 seconds long. This was easy for the participants to perform when providing short 'dictionary' definitions to concrete nouns, but clearly an impossible requirement if unconstrained narrative production was the task. Activity associated with Covert (both Speaking and Counting) was essentially the same as during Overt, without activity that can be attributed to the motor-sensory control of articulation; namely, medial frontal activity (pre-SMA extending into anterior paracingulate and cingulate cortices) and the left dorsolateral prefrontal and posterior inferolateral temporal regions. The activity associated with Overt relative to Covert Speaking was symmetrically distributed between the cerebral and cerebellar hemispheres, in regions that can be attributed to lower levels of processing controlling the motor execution, with accompanying sensory monitoring, of speech production. The left-lateralised frontal and temporal activity is clearly associated with higher levels of language processing, and it is evident that this was also associated with a crossed response in the right lateral cerebellar cortex. The functional role of the cerebellum in higher order speech processes is unknown, although anatomical loops between prefrontal cortex and the cerebellum have been established (Ramnani and Miall, 2001).

Overt Counting resulted in activity in speech motor regions, such as those involved in controlling movement of the articulators and larynx. Activity in premotor, motor, sensory and paravermal cerebellar regions was observed, which can be reasonably attributed to the motor-sensory control of articulation. There was no activity in the higher-order cortices observed during Speaking. This was to be expected, as counting only requires word retrieval from one lexical category, with little if any conceptual processing. Covert Counting did not show significant activity above Rest at the statistical threshold used in the analyses. Although cognitively these two conditions are different, it is now very well established that the Rest state results in widespread activity in the so-called default mode network (Raichle et al., 2001). This study was perhaps not sufficiently powered to reflect any low-level activity associated with Covert Counting, over and above the distributed activity associated with the rest condition.

Medial frontal activity in pre-SMA and anterior cingulate/paracingulate cortex, evident in both Overt and Covert Speaking conditions, have previously been shown to be involved in word generation (Crosson et al., 1999) and sequencing of syllables and phonemes (Ghosh et al., 2008). The anterior cingulate cortex has also been implicated in performance monitoring and error detection (Carter et al., 1998). Noun definition, as used in the present study, requires selection and retrieval of items from semantic memory, which will have involved these domain-general executive processes.

With regard to the motor-sensory ROI analyses in this study, Covert Speaking only showed activity in the left frontal operculum. The ROIs at the temporo-parietal junction (anterior and posterior planum temporale and parietal operculum) were not active for Covert Speaking. This region has been proposed to be a ‘sensorimotor interface’, with ‘virtual’ activation during covert speech. However, the present study only revealed activity here (and then only in the anterior PT) for the two Overt conditions. The paucity of activity at the temporo-parietal junction in the present study does not accord with the studies of Hickok and colleagues. However, an absence of activity does not imply an absence of involvement. Multivariate analyses may allow functional connections between regions to be identified in the absence of increased

activity evident on univariate analyses of contrasts between conditions. These issues are discussed further in Chapter 6.

In summary, this study demonstrated that the motor-sensory speech production systems are only evident with overt, not covert, speech, if the analyses are based on subtractive contrasts between conditions. The pattern of activity for Overt Speaking reflects the sum of Overt Counting and Covert Speaking. Having established the broad systems involved in speech production, the subsequent studies presented in this thesis focus more specifically on non-native speech and how these systems are modulated by proficiency in a second language. These subsequent studies all employed overt speech to investigate motor-sensory aspects of native and non-native speech.

# 4 A comparison of motor-sensory activity during speech in first and second languages

To speak in a foreign language, particularly when it is acquired in adulthood, means to speak with an accent (Flege, 1995). This is likely to be the result of a failure to achieve the same proficiency in integrating the motor feedforward and sensory feedback control of articulation that is achieved when speaking in the native language. This study investigated modulations of activity within speech motor-sensory systems in response to native (L1) and non-native (L2) speech production using fMRI. Learning a second language involves learning less familiar patterns of motor control, which requires modifications to the existing speech-motor control system and to existing articulator-acoustic relationships. This study investigated whether sensory feedback is central to this modification process, which enables production of accurate speech sounds in a non-native language. It is proposed that, during speech production, rapid neural pathways integrate feedforward discharges from premotor or motor cortex encoding articulatory gestures with signals from auditory and somatosensory feedback. These match expectation with outcome to monitor online for articulatory errors (Golfinopoulos et al., 2010; Guenther et al., 2006; Ventura et al., 2009).

When speakers use L1, it has been shown that there is paradoxical suppression of neural activity ('sensory gating') in sensory association cortex (temporo-parietal cortex), both somatosensory and auditory (Dhanjal et al., 2008); (Ventura et al., 2009). The junction of temporal and parietal cortex within the posterior end of the lateral sulcus has been proposed as the site for sensorimotor integration during articulation (Hickok et al., 2003; Hickok et al., 2009). Modulation of the response within this region would be expected when participants speak a second language (L2) with an accent, compared to when they speak their native tongue (L1). This study investigated the hypothesis that this heteromodal sensory region does respond differently during production of L2 compared to production of L1; in particular, that responses in posterior auditory and somatosensory association cortices increase during production of L2. This study has been published in the Journal of Neurophysiology (Simmonds et al., 2011a).

## 4.1 Aim and hypotheses

The aim of this study was to investigate the response of motor-sensory control systems to speaking a non-native language in late learners of L2 with high proficiency. The prediction was that L2 speech production would engage a more extended system than L1. This hypothesis proposed that production of L2 would place altered demands on sensory (auditory and somatosensory) control. Specifically, I predicted an increased response in planum temporale and parietal operculum, the locations of auditory and somatosensory association cortices respectively, when speaking in L2 relative to L1.

## 4.2 Materials and methods

### 4.2.1 Participants

Eighteen highly proficient non-native English speakers (9 female; mean age 27 years, 4 months; range 20 years, 11 months – 35 years, 10 months) participated. All participants had learned English as a second language (mean age of starting to use English: 12 years, 1 month; range 5-21 years). They all had considerable exposure to L2 as they were all living in England (mean time in England 1 year, 10 months; range 4 months – 6 years). I deliberately chose highly proficient speakers, living in England and working using English. These participants had all passed English language proficiency examinations for UK university entrance or successful visa applications. Therefore they were able to speak at a normal conversational rate in both L1 and L2. I excluded very early bilinguals who would be dominant in both languages as, although the amount of speech produced in the two languages would be matched, the hypothesis was that altered sensory-motor activity would be observed in a late-acquired language. I also excluded less proficient bilinguals or those who did not speak English every day, as the amount of speech produced in the two languages would be biased towards the use of L1. As I was not examining linguistic or cognitive systems and was not expecting effects to be language-specific, I deliberately chose participants from different language backgrounds, with the expectation that the results would apply to any L1 group. There was therefore a range of first languages, from both European and non-Indo-European language groups: two Greek, two Italian, one German, one Dutch, three French, three Spanish, two Polish, one Russian, two Chinese (Mandarin), and one Indonesian.

### 4.2.2 Behavioural assessment of language proficiency

Proficiency in English was assessed using standardised language tests. There were three measures for spoken English: age of learning English, amount of English used daily and the speaking component of the Certificate of Proficiency in English (CPE) exam, from the University of Cambridge ESOL Examinations. There were two other scores, one a measure of knowledge of English vocabulary (the picture-naming component of the Bilingual Verbal Ability Tests (BVAT) (Muñoz and Singleton, 2011), and the other a measure of proficiency at reading, in which participants were recorded reading a text aloud, to judge reading accuracy.

### 4.2.3 Functional MRI experimental tasks

In the fMRI scanning session, participants produced overt propositional speech in both their native language (L1) and English (L2). Rest was included as the baseline condition. During speech trials, participants were instructed to give definitions of visually presented pictures. For example, when seeing a picture of an apple, the response might be, “This is a round fruit, it can be red or green, it grows on a tree, it tastes nice”. Table 4.1 shows the stimuli used and their properties. Each picture appeared twice, once for each language and country flags indicated the language to be used (L1 or L2). The pictures and the requirement for speaking in L1 and in L2 in response to a specific picture were switched between runs but kept constant across participants. Participants performed blocks of 6-12 trials in one language, followed by 3 rest trials and then 6-12 trials in the second language, to avoid the confound of repeated language switching, which would have occurred with single-trial randomisation. As highly proficient speakers had been deliberately chosen for this study, all participants were able to complete the tasks.

At the beginning of each scanning session, somatosensory and auditory functional localisers were also conducted (TR = 5 seconds, 96 volumes). The first condition involved silently moving the tongue from the floor of the mouth to the upper ridge of the hard palate, with the jaw open and still (Dhanjal et al., 2008). The second condition involved listening to band-pass filtered (either 0-1kHz or 1-2kHz) amplitude-modulated (either 8 or 32 Hz) white noise to activate widely throughout the supratemporal plane (Warren and Griffiths, 2003). In addition, there was a tongue condition with no motor output, obtained by directing a puff of air onto the tongue

(Sereno and Huang, 2006). However, this was confounded by co-activation of auditory and motor areas, due to some participants moving their tongue and lips in response to the air-puff, as well as some participants reporting hearing the air-puff being applied. Therefore, only the first two conditions were used in data analysis.

### 4.2.4 Materials

The picture stimuli consisted of 60 black and white line drawings of objects, from the International Picture Naming Project (University of California, San Diego). The stimuli had been normed and matched for a range of linguistic and visual variables (see Table 4.1). They were ranked by naming latency (measured by speed of response in a picture naming task) and grouped into objects with short and long naming latencies. There were 30 pictures for each category (30 Short Latency and 30 Long Latency) and each picture appeared twice, once for L1 and once for L2. The 30 stimuli for the rest condition consisted of black and white random figures not depicting any clear object, to provide a high-level baseline. Figure 4.1 shows examples of stimuli used for the experimental and rest conditions.

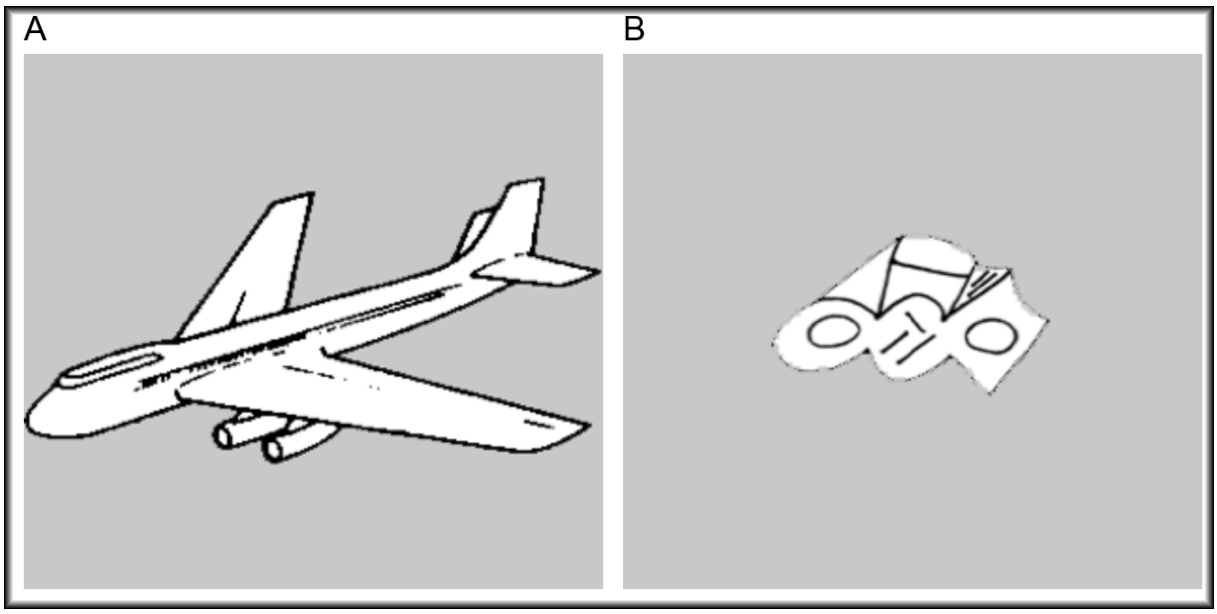



***Figure 4.1, Examples of stimuli***

*A: An example stimulus for the experimental condition, during which participants were instructed to define the object in the picture. Each picture appeared twice, once for L1 and once for L2. B: An example stimulus for the rest condition, during which participants were instructed not to speak.*

| Stimuli | Naming Latency | Familiarity | Syllables | Characters | Log Frequency CELEX | AOA | Visual Complexity |
|---|---|---|---|---|---|---|---|
| | | | *Short Naming Latency* | | | | |
| book | 656 | 1 | 1 | 4 | 6.075 | 1 | 8619 |
| eye | 700 | 0.98 | 1 | 3 | 6.261 | 1 | 9104 |
| balloon | 702 | 1 | 2 | 7 | 1.946 | 1 | 8015 |
| dog | 702 | 1 | 1 | 3 | 4.754 | 1 | 12012 |
| pencil | 702 | 1 | 2 | 6 | 2.996 | 2 | 7899 |
| bell | 703 | 1 | 1 | 4 | 3.332 | 3 | 11109 |
| sock | 712 | 1 | 1 | 4 | 2.944 | 1 | 8316 |
| camera | 725 | 1 | 3 | 6 | 3.611 | 2 | 16408 |
| turtle | 734 | 1 | 2 | 6 | 1.609 | 1 | 14768 |
| house | 745 | 0.98 | 1 | 5 | 6.409 | 1 | 18069 |
| frog | 751 | 1 | 1 | 4 | 2.303 | 1 | 14773 |
| sun | 762 | 1 | 1 | 3 | 5.03 | 1 | 18102 |
| cat | 766 | 0.96 | 1 | 3 | 4.22 | 1 | 9894 |
| finger | 775 | 0.98 | 2 | 6 | 4.82 | 1 | 5370 |
| spoon | 777 | 1 | 1 | 5 | 2.773 | 1 | 7344 |
| airplane | 778 | 0.7 | 2 | 8 | 1.946 | 1 | 16810 |
| ring | 785 | 1 | 1 | 4 | 1.386 | 3 | 7652 |
| television | 786 | 0.61 | 2 | 2 | 0 | 1 | 18950 |
| helicopter | 793 | 1 | 4 | 10 | 2.833 | 2 | 18241 |
| tree | 796 | 1 | 1 | 4 | 5.257 | 1 | 26074 |
| clown | 804 | 1 | 1 | 5 | 1.609 | 2 | 21244 |
| moon | 804 | 1 | 1 | 4 | 4.094 | 1 | 3730 |
| carrot | 806 | 1 | 2 | 6 | 2.197 | 1 | 13201 |
| banana | 808 | 1 | 3 | 6 | 2.197 | 1 | 8767 |
| apple | 810 | 1 | 2 | 5 | 3.434 | 1 | 8241 |
| knife | 816 | 1 | 1 | 5 | 3.807 | 2 | 8773 |
| broom | 821 | 1 | 1 | 5 | 2.197 | 1 | 11261 |
| glove | 848 | 1 | 1 | 5 | 2.996 | 3 | 11509 |
| leaf | 848 | 1 | 1 | 4 | 4.407 | 3 | 26600 |
| table | 852 | 0.98 | 2 | 5 | 5.464 | 1 | 12010 |
| | | | *Long Naming Latency* | | | | |
| fire | 854 | 0.96 | 2 | 4 | 5.094 | 3 | 52543 |
| kangaroo | 856 | 1 | 3 | 8 | 1.386 | 3 | 14555 |
| web | 869 | 0.68 | 3 | 9 | 0 | 3 | 14705 |
| alligator | 881 | 0.9 | 4 | 9 | 1.099 | 2 | 14874 |
| palmtree | 908 | 0.86 | 2 | 8 | 0 | 3 | 18577 |
| parrot | 910 | 0.79 | 2 | 6 | 1.609 | 3 | 18115 |
| gorilla | 944 | 0.7 | 3 | 7 | 1.386 | 3 | 17084 |
| piggybank | 965 | 0.94 | 3 | 9 | 0 | 3 | 24489 |
| sink | 984 | 0.96 | 1 | 4 | 2.773 | 1 | 26560 |
| rhinoceros | 998 | 0.77 | 4 | 10 | 1.099 | 3 | 18320 |
| stairs | 1011 | 0.74 | 1 | 6 | 3.807 | 1 | 27602 |
| swan | 1049 | 0.74 | 1 | 4 | 2.079 | 3 | 12465 |
| violin | 1051 | 0.82 | 3 | 6 | 1.946 | 3 | 8571 |
| trumpet | 1053 | 0.69 | 2 | 7 | 2.197 | 3 | 13615 |
| nest | 1059 | 0.73 | 1 | 4 | 2.89 | 3 | 12296 |
| saxophone | 1061 | 0.81 | 3 | 9 | 0.693 | 3 | 8795 |
| panda | 1071 | 0.38 | 2 | 5 | 0.693 | 3 | 29117 |
| teapot | 1085 | 0.44 | 2 | 6 | 1.609 | 3 | 17625 |
| ironingboard | 1105 | 0.9 | 4 | 12 | 0 | 3 | 12848 |
| sweater | 1122 | 0.55 | 2 | 7 | 2.773 | 1 | 11622 |
| moose | 1158 | 0.76 | 1 | 5 | 0.693 | 2 | 23330 |
| dresser | 1163 | 0.48 | 2 | 7 | 1.792 | 3 | 21173 |
| canoe | 1164 | 0.62 | 2 | 5 | 1.946 | 3 | 27029 |
| lawnmower | 1166 | 0.96 | 3 | 9 | 0 | 2 | 18238 |
| vase | 1171 | 0.94 | 1 | 4 | 2.079 | 3 | 20221 |
| leopard | 1194 | 0.54 | 2 | 7 | 2.197 | 3 | 23203 |
| stethoscope | 1209 | 0.93 | 3 | 11 | 0.693 | 3 | 13841 |
| tractor | 1216 | 0.87 | 2 | 7 | 2.485 | 2 | 9518 |
| grasshopper | 1234 | 0.67 | 3 | 11 | 1.386 | 3 | 13119 |
| paperclip | 1262 | 0.81 | 3 | 9 | 0 | 3 | 21555 |

***Table 4.1, Details of stimuli used from the International Picture Naming Project***

*The mean naming latency for words in the short latency group was 768.9 (standard deviation 51.21) and for the long latency group 1059.1 (standard deviation 121.72). The naming latencies for the two groups were significantly different ($p < 0.00001$).*

### 4.2.5 Speech Rate

Online speech output was recorded and syllables were counted from recordings of each speech trial. Due to technical failure, speech output from six participants was not recorded and they were not included in subsequent analyses of speech rate. To remove noise from the sound of the scanner, Praat (Boersma and Weenink, 2007) was used to break the sound up into sections that only consisted of speech, and not scanner sounds.

Syllable counts were calculated in Praat (De Jong and Wempe, 2009), which enabled automatic detection of syllable nuclei, so that speech rate could be measured without manual transcription. This was especially important for this study, as the native languages of participants varied widely. The script considers all peaks above a certain threshold as possible syllables then discards peaks that are not followed and preceded by dips of at least 2dB in intensity. Trials were then split into high and low speech rate, based on the median for each individual, and entered into a 2 x 2 ANOVA with 2 factors: language (L1 and L2) and speech rate (high and low).

### 4.2.6 Regions of interest

In this Chapter I used two converging methodologies to define the theoretically motivated ROIs, one sensitive to anatomical variability, as in Chapter 3 (see Figure 3.1, p87), and the other sensitive to functional individual variability. The anatomically-defined parietal operculum and adjacent planum temporale ROIs were defined on an individual basis using gyral and sulcal landmarks, using the technique described in Chapter 3. The increased motor demands when articulating in the less familiar L2 were also investigated, with ROIs placed in the left and right frontal operculum. The placing of these ROIs was made with reference to the probabilistic atlas. For these anatomically defined ROIs, the functional data were not spatially smoothed before averaging, in order to avoid any possibility of blurring of activation across the Sylvian fissure between temporal and parietal lobes. Mean effect sizes for L1 and L2 speech conditions, relative to rest, were calculated for each individual within each ROI.

To complement this anatomical analysis, ROIs sensitive to individual functional variability were also created. First, ROIs were defined using a combination of probabilistic anatomical masks from the FSL structural atlas (Harvard-Oxford Cortical

Structural Atlas) and masks from individual functional localiser results. Analyses were run using independent anatomically constrained, functionally defined masks for each participant. Each participant's individual motor tongue somatosensory and auditory functional localisers were multiplied with either a parietal operculum (somatosensory) or planum temporale (auditory) anatomical mask. These ROIs were then investigated in each hemisphere separately.

Five participants were excluded from further analyses using functionally defined regions of interest, due to absence of significant activity for the functional localisers within those regions. One participant was excluded from analyses using individual anatomical regions of interest, due to a technical error.

## 4.3 Results

### 4.3.1 Language proficiency

All behavioural measures showed a high level of English proficiency for all participants and confirmed that all participants used English daily for work. The CPE speaking assessment demonstrated that all participants showed 'fluent, spontaneous expression in clear, well-structured speech' (see Table 4.2). There were two other scores, one a measure of knowledge of English vocabulary (BVAT difference between native language and English mean 10.5 (+/- 1.3 s.e.m.) and the other a measure of proficiency at reading (correct words out of 209 mean 202.5 (+/- 1.0 s.e.m.). These measures were included as regressors in the analyses of the imaging data. Table 4.2 shows scores from the behavioural assessments of spoken language proficiency, demonstrating that all participants were highly proficient and a relatively homogeneous group.

| Subject | Native Language (L1) | Age at Time of Scanning, yr (mo) | Age Started English, yr | English Experience, yr | Use of English* (maximum score, 21) | CPE speaking (Grade**) |
|---|---|---|---|---|---|---|
| 1 | Greek | 24 (10) | 12 | 12 | 18 | C1+ |
| 2 | Italian | 25 (2) | 7 | 18 | 14 | C2 |
| 3 | German | 25 (5) | 11 | 14 | 18 | nd |
| 4 | Dutch | 32 (7) | 11 | 21 | 12 | C2 |
| 5 | French | 24 (2) | 14 | 10 | 21 | nd |
| 6 | French | 26 (7) | 12 | 14 | 14 | nd |
| 7 | Spanish | 27 (7) | 20 | 7 | 11 | C1 |
| 8 | Chinese | 26 (10) | 11 | 15 | 13 | nd |
| 9 | Chinese | 26 (4) | 12 | 14 | 13 | C1+ |
| 10 | Spanish | 26 (5) | 8 | 18 | 17 | B2+ |
| 11 | Indonesian | 23 (6) | 17 | 6 | 21 | B2+ |
| 12 | Polish | 31 (6) | 13 | 18 | 13 | C1 |
| 13 | Russian | 26 (8) | 11 | 15 | 17 | B2+ |
| 14 | Polish | 25 (11) | 12 | 13 | 18 | C1+ |
| 15 | Spanish | 32 (11) | 5 | 27 | 15 | C1 |
| 16 | Italian | 28 (5) | 13 | 15 | 19 | B2+ |
| 17 | Greek | 20 (11) | 9 | 11 | 9 | C1 |
| 18 | French | 35 (10) | 21 | 14 | 12 | nd |

***Table 4.2, Behavioural proficiency scores***

** Participants were asked to describe their current use of English on a sliding scale from 1 to 7 with native language only as 1 and English only as 7, for three categories: at work, at home and other places. A score of 21 would show that the participant only used English and a score of 3 means that they only use their native language.*
*** There are 11 grades in the Cambridge ESOL Examinations Assessments scales for the Certificate of Proficiency in English (CPE) (from high to low: A1, A1+, A2, A2+, B1, B1+, B2, B2+, C1, C1+, C2). Grades used in the present study are from the speaking component only, not the CPE assessment as a whole.*
*nd Data not available for these participants.*

### 4.3.2 Functional localisers

Due to individual anatomical variability, it is difficult to separate parietal operculum from planum temporale. As a multi-sensory area of cross-modal integration, these regions are activated by speech, from both auditory and somatosensory contributions. Functional localisers were run for individual participants, to identify auditory and motor areas within this multi-sensory region. Listening to band-pass filtered amplitude-modulated noise resulted in increased activation bilaterally in the superior temporal gyrus, which is involved in auditory processing (Chen et al., 2009)

(Figure 4.2A). Silently moving the tongue resulted in bilateral motor cortex activation, as expected (Figure 4.2B).

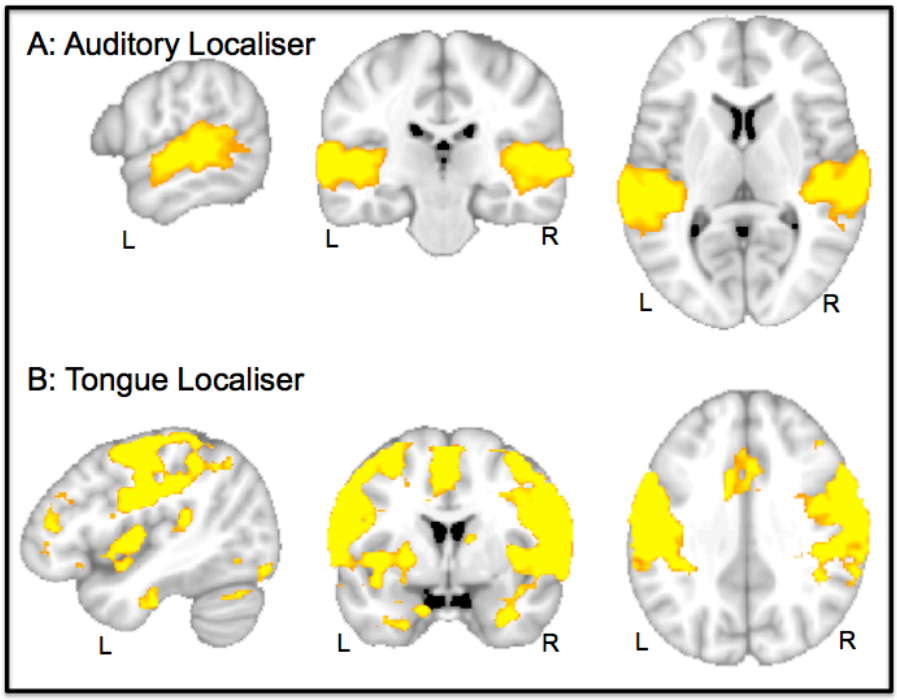



***Figure 4.2, Functional localiser results***

*Activation patterns for the two localisers, shown on a sagittal, coronal and axial slice for each. A: auditory (x = -57, y = -26, z = 7) and B: tongue (x = -43, y = 1, z = 34). Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Results are displayed on a standard brain template (MNI152) and left is shown on the left. A table reporting main cluster peaks for these contrasts is shown in* Appendix 4 *(p225 of this thesis).*

### 4.3.3 Functional MRI whole-brain analyses of speech conditions

This study was designed to investigate the motor-sensory differences in speaking a late acquired second language compared to speaking in a native language. The initial whole-brain analysis contrasted speech (in both languages) and the non-speech baseline, and L2 was then directly contrasted with L1. ROI analyses were conducted with both functionally and anatomically-defined regions. I then present analyses investigating the effects of different L1s and behavioural proficiency. The final part of this results section reports the effects of speech rate and I present a whole-brain 2 (Language: L1 and L2) x 2 (High and Low speech rate) ANOVA. No effects of type of word (short or long naming latency) were found; therefore all results

are reported collapsed across word type. A table reporting main cluster peaks for these contrasts is shown in Appendix 5 (p225 of this thesis).

#### *4.3.3.1 Speech > Rest*

Speech (L1 + L2) contrasted with a non-speech 'rest' baseline condition, demonstrated activity in a distribution that accorded with previous studies of propositional speech production (Blank et al., 2002; Dhanjal, et al. 2008). These included bilateral primary somatosensory and auditory cortical regions, the left frontal operculum and the midline vermal cerebellum. At a more lenient statistical threshold ($p < 0.01$, uncorrected), activity in midline premotor cortex (PMC), both the supplementary motor area (SMA) proper and the pre-SMA, lateral PMC, the bodies of both caudate nuclei, left and right basal ganglia and both thalami were evident (Figure 4.3).

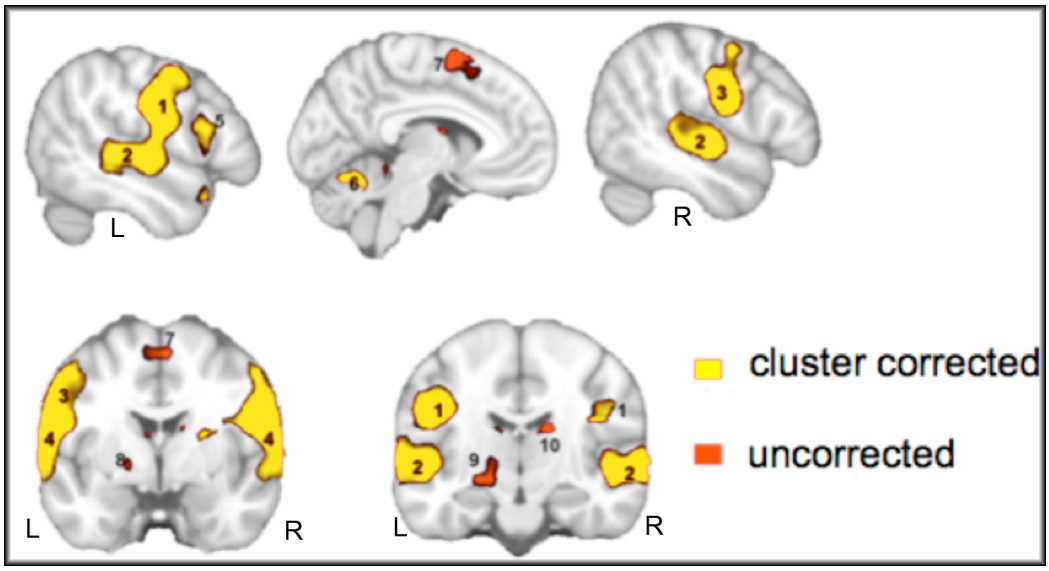



***Figure 4.3, Speech > Rest***

*Speech (L1+L2) > Rest, cluster corrected, z < 2.3, p < 0.05 (yellow), uncorrected (red). Cluster-corrected activity (p < 0.05, orange) in bilateral primary somatosensory (1), primary auditory (2), primary motor (3), secondary somatosensory (4), left frontal operculum (5) and midline vermal cerebellum (6). Additional uncorrected activity (red) is shown in midline premotor cortex (7), pallidum (8), thalamus (9) and caudate (10). Top panel shows three sagittal slices, x = -52, -6, 52. Bottom panel shows two coronal slices, y = -2, -16. Results are displayed on a standard brain template (MNI152) and left is shown on the left.*

#### *4.3.3.2 L2 > L1*

Directly contrasting production of L2 with L1 demonstrated greater activity in a limited number of cerebral and cerebellar cortical regions (threshold $p < 0.05$, corrected): left frontal operculum, left anterior superior temporal gyrus, left lateral PMC, medial PMC (both the SMA and pre-SMA), left temporo-parietal cortex (both parietal operculum and planum temporale) and midline vermal cerebellum (Figure 4.4). In addition, there was significantly greater activity in a number of bilateral subcortical regions: the basal ganglia and the lateral thalami. The increased cognitive demands of producing L2 during picture description were reflected in increased activity in anterior medial prefrontal cortex.

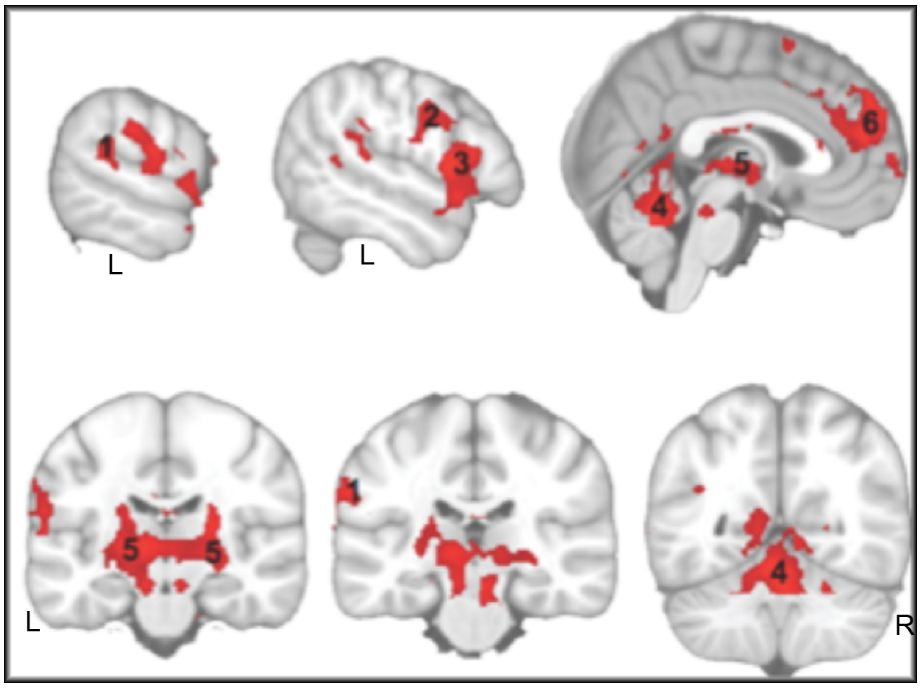



***Figure 4.4, L2 > L1***

*Sagittal (top) and coronal (bottom) slices showing increased activity in motor feedforward and sensory feedback systems for L2. Greater activity is shown in left temporo-parietal cortex (1), pre-motor cortex (2), left frontal operculum (3), midline vermal cerebellum (4), thalami (5) and anterior medial prefrontal cortex (6). Top panel shows three sagittal slices, x = -61, -52, 3. Bottom panel shows three coronal slices, y = -18, -25, -56. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Results are displayed on a standard brain template (MNI152) and left is shown on the left.*

### 4.3.4 Region of interest analyses

Averaging activity during production of L1 and L2 within the ROIs confirmed and extended the impression from the whole-brain analyses. Within the functionally-defined planum temporale, both speech production conditions resulted in increased activity relative to rest, although activity was marginally less during production of L1 relative to L2 (Figure 4.5Aii). These effects were present in both hemispheres although the overall magnitude was greater on the left. In both left and right functionally-defined parietal opercula there was significant activation in response to production of L2, relative to both the production of L1 and to rest, whereas activity during production of L1 was no different from rest (Figure 4.5Aii). The whole brain analysis at the statistical threshold chosen ($p < 0.05$, cluster corrected) had only shown the difference in activity on the left. However, to validly assess statistical lateralisation differences in activation, ROI analyses must be used (Jernigan et al., 2003). Using ROIs demonstrated no significant lateralisation of the neural response in these regions, suggesting that if anything there was only a mild bias in lateralisation.

Using anatomically defined ROIs for the left posterior planum temporale and left parietal operculum, a broadly similar pattern of results was observed, with L2 showing significantly greater activation than L1 in both regions (Figure 4.5Bi). However, in contrast to the whole-brain analyses and the functionally-defined ROI analyses, the anatomical ROIs demonstrated no significant activation for L1 speech production versus rest. Within the frontal operculum, there was significantly greater activity for L2 than L1 in the left hemisphere ($p < 0.01$) and no significant differences in the right hemisphere (Figure 4.5Bii).

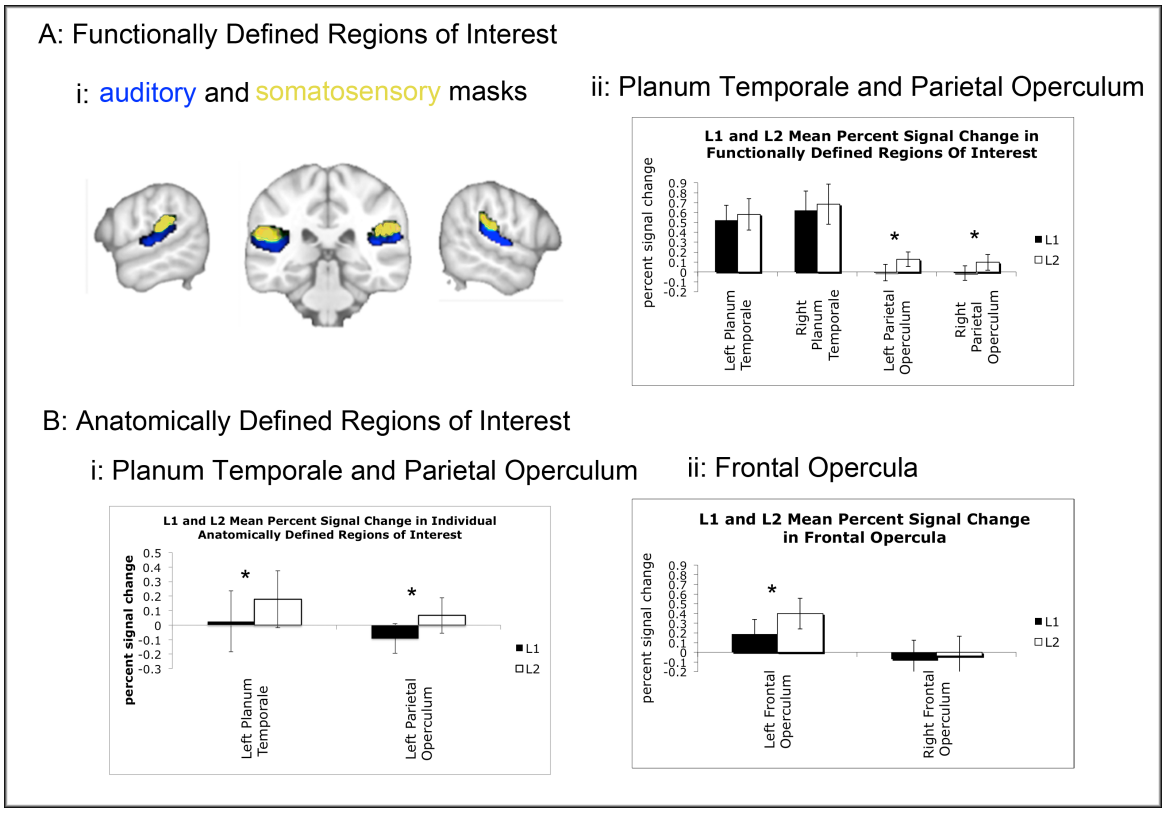



***Figure 4.5, ROI results***

*A) Functionally Defined ROIs. Ai: auditory (blue) and somatosensory (yellow) masks displayed on a standard brain template (MNI152). From left to right the figure shows a sagittal slice, x = -58, a coronal slice, y = -32 and a sagittal slice, x = 58. Aii: Percent signal changes for L1 and L2 in the functionally defined ROIs on smoothed data, where there was overlap of the ROIs. Within the auditory ROI, both L1 and L2 were active relative to baseline, with L1 marginally suppressed relative to L2. Within the somatosensory ROI, L1 showed negative activation (not significantly different from the non-speech baseline), whereas L2 showed significant positive activation. L2 was significantly greater than L1, in both hemispheres (left $p < 0.05$, right $p < 0.001$). B) Anatomically Defined ROIs. Bi: Percent signal changes for L1 and L2 in the lateral posterior planum temporale and parietal operculum ROIs on unsmoothed data, where overlap had been avoided. Within both regions, L2 showed significantly greater activation relative to L1 (planum temporale $p < 0.001$, parietal operculum $p < 0.05$). Bii: Percent signal changes for L1 and L2 in the frontal opercula ROIs, defined with a probabilistic structural atlas (L2 > L1, $p < 0.01$ on the left). Error bars represent standard error. * significant language differences.*

The functionally-defined regions of interest were also investigated in each hemisphere separately. In the planum temporale, the difference between L1 and L2 in the left hemisphere was significant (L1 = 0.521, L2 = 0.582, t (16) = -3.158, $p < 0.01$) but not significant in the right. The parietal operculum also showed a significant difference between L1 and L2 in the left hemisphere (L1 = -0.006, L2 = 0.128, t (16) = -2.905, $p < 0.01$) but not in the right.

Mean effect sizes for L1 and L2 speech conditions, relative to rest, were entered into a 2 (hemisphere) x 2 (language) ANOVA, performed separately for the functionally-defined masked analyses. With the planum temporale mask, there was a main effect of Language (F (1, 16) = 6.619; $p < 0.05$) and no significant effect of hemisphere and no significant interaction. With the parietal operculum mask, there was a main effect of Language (F (1, 16) = 4.916; $p < 0.05$) and a main effect of Hemisphere (F (1,16) = 13.832; $p < 0.01$). There was no significant interaction.

### 4.3.5 Different L1s

Different native languages had no demonstrable influence on differences in articulating in L1 relative to the common L2 of English. Therefore, I did not find evidence that the motor-sensory consequences of speaking in L2 are dependent on L1, or are related to, for example, differences in stress patterns between languages (Figure 4.6). However, a much larger fMRI study may reflect subtle neural differences when a common L2 is spoken by a group of speakers with different L1s.

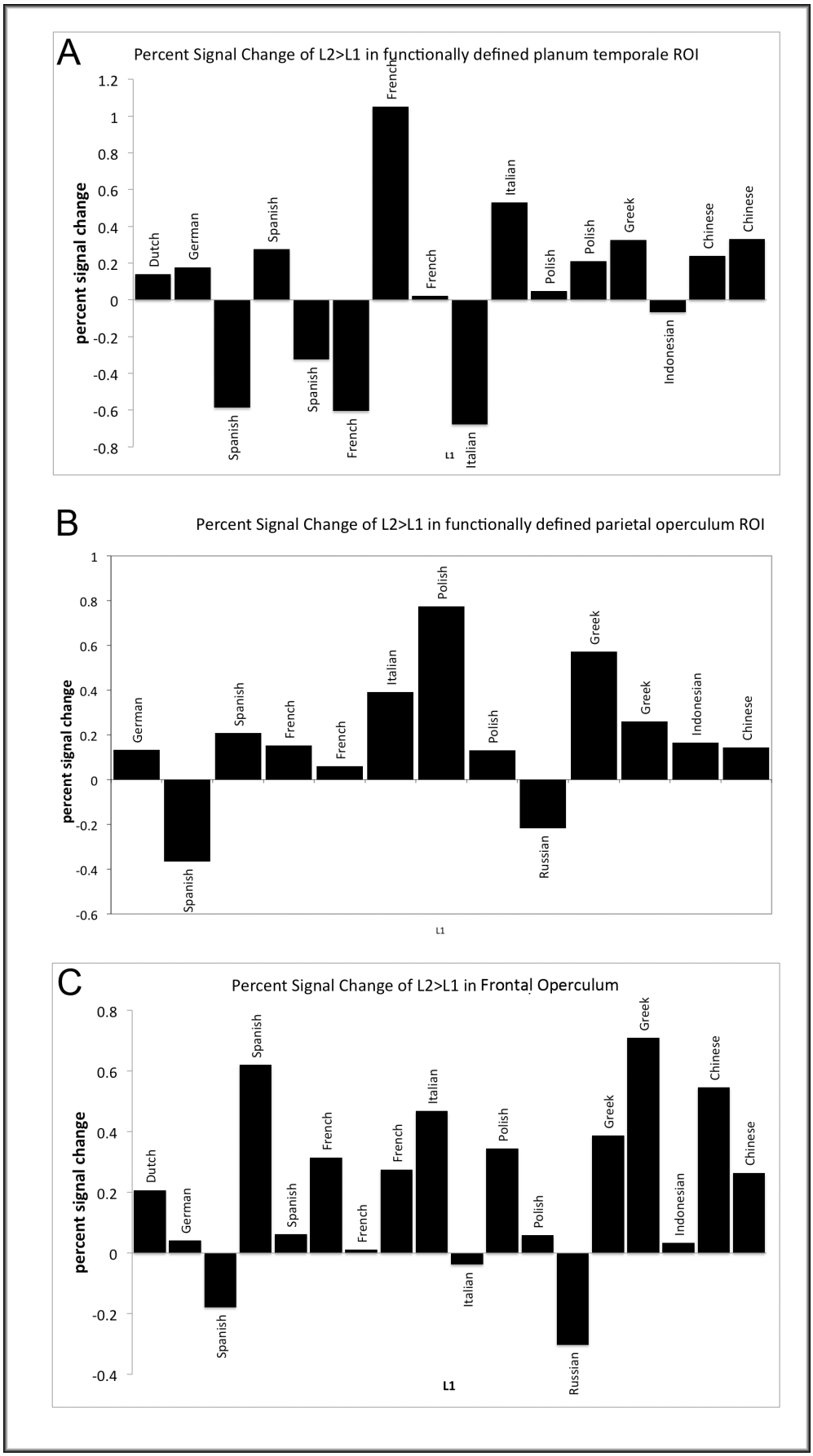



***Figure 4.6, L2 > L1 signal change for different L1s in ROIs (legend overleaf)***

***Figure 4.6, L2 > L1 signal change for different L1s in ROIs (Figure on previous page)***

*The difference between L2 and L1 for each participant within the ROIs was broadly similar for different native languages, even across language groups. Plots are shown for the left ROIs in three regions. A: Functionally defined planum temporale ROI. For two participants (1 Greek and 1 Russian) there was no signal in this ROI for either language. B: Functionally defined parietal operculum ROI. For five participants (1 Dutch, 1 Spanish, 1 French, 1 Italian and 1 Chinese) there was no signal in this ROI for either language. C: Anatomically defined frontal operculum ROI.*

### 4.3.6 Co-varying proficiency behavioural measures

In addition to considering group differences in the neuroimaging analyses, all measures of English proficiency were included as covariates in ROI analyses. None of these covariates showed any significant correlation with activity in the data from the ROIs, although this was not a primary aim of the current study.

### 4.3.7 Speech rate

One potential confound with using volitional propositional speech is that the number of utterances produced is unconstrained and can differ between languages. Analyses of the recordings of participants' speech demonstrated approximately equivalent numbers of syllables spoken in L1 relative to L2 (mean speech rate in syllables for L1 = 8.2; L2 = 7.8; t (21) = 1.437, $p$ = 0.165, paired t-test). There were no significant differences in speech rate across the two runs (mean difference of 0.1 syllables). Figure 4.7 shows the online rate of speaking in L1 and L2 at the time of scanning, and although there was considerable inter-individual variability of speech rate across both scanning runs, this plot shows that within individuals, rates were very closely matched between L1 and L2.

In addition, a subsequent whole-brain ANOVA on a subset of participants, testing for effects of speech rate and L1/L2 language activation, revealed no significant effects of speech rate or interaction with language (even at a liberal statistical threshold), while revealing a significant main effect of language in cortical motor-sensory feedback systems, in particular, temporo-parietal cortex, see Figure 4.7. It should be stressed that although inter-individual speech rate was variable, intra-individual rate

was relatively constant. Therefore, this analysis will not have demonstrated the specific effect of speech rate. This specific effect has been investigated previously by a study design which varies speech rate systematically within each individual, such as by varying the rate of repetition (Wise et al., 1999). Speech rate was not intended as a marker of proficiency; this was independently measured using more sensitive out-of-scanner assessments.

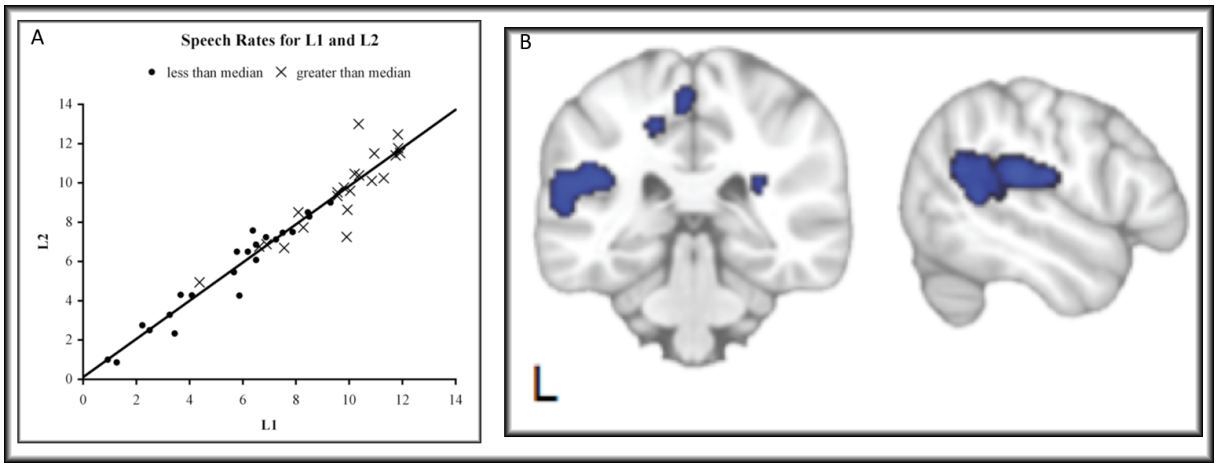



***Figure 4.7, Speech rates for L1 and L2 and a Speech-rate x Language ANOVA***

*A: Speech rate measured in syllables produced per trial for L1 and L2. The circles represent trials with lower speech rates (less than the median) and the crosses represent trials with higher speech rates (the median or greater). B: Results from the ANOVA testing for effects of speech rate and L1/L2 language activation revealed no significant effects of speech rate or interaction with language. A significant main effect of language was revealed in cortical motor-sensory feedback systems, in particular the parietal operculum. Coronal (y -34) and sagittal (x -51) slices showing the main effect of language are displayed. Results are displayed on a standard brain template (MNI152) and left is shown on the left.*

## 4.4 Discussion

The focus of this study was to investigate modulations of activity within speech motor-sensory systems in response to first and second language speech production. Previous work has shown that speech production in a native language is associated with reduced activation, compared to rest, in the parietal operculum, specifically secondary somatosensory association cortex. This study was designed to investigate whether this is also true for speech production in a foreign language. To ensure that speech produced in the scanner was fluent and well matched across L1 and L2, despite the two being unbalanced, I selected highly proficient late bilinguals. They had all acquired L2 after early childhood and despite high proficiency, spoke with a non-native accent.

The specific hypothesis was that speaking in L2 (with an accent), relative to L1 (without an accent), would be associated with increased activity in both posterior sensory (planum temporale and parietal operculum) and anterior premotor regions (frontal operculum). Different approaches to the analysis (including whole-brain voxelwise analysis as well as regions of interest defined both on individual anatomy and using individual functional localisers) provided converging evidence of increased activation for L2 relative to L1 within temporo-parietal cortex and left frontal operculum. Given previous work investigating native speech production, as well as work delineating the more basic sensory properties and connectivity of these regions in humans and non-human primates, temporo-parietal cortex is highly likely to be involved in heteromodal auditory and somatosensory feedback control.

There are now a number of functional neuroimaging studies that have demonstrated the role of the temporo-parietal junction in both the overt and covert production of human speech and non-speech vocalisations (Hickok et al., 2009; Hickok and Poeppel, 2000; Pa and Hickok, 2008; Wise et al., 2001). In the group analysis on the whole-brain and using individual auditory functional localisers on images normalised into standard anatomical space, the posterior superior temporal plane (STP) was significantly active during both L1 and L2 speech production relative to rest, as illustrated in Figure 4.3; an analysis informed by, and compatible with, the earlier studies. The discrepancy between the findings from this approach to the analysis and

the earlier studies is that there was no functional asymmetry between the hemispheres. Therefore, it would appear that the approach adopted in this study has not isolated precisely the same area, known as cortical area Spt and defined as the posterior part of the planum temporale and, less certainly, the adjacent parietal operculum (Pa and Hickok, 2008), which is described as a sensory-motor integration area for speech. Area Spt, it is proposed, is strongly left-lateralised, and connected, functionally and anatomically, with the left frontal operculum (Hickok and Poeppel, 2007). Clearly, the studies by Hickok and colleagues did not study direct interactions between feedforward and feedback vocalisation pathways. As the ventral premotor cortex is left-lateralised for speech production and is activated by covert speech when motor cortex is not (Basho et al., 2007), data from Hickok and colleagues indicates the existence of a left-lateralised 'pre-articulatory sensory' area that is active during covert vocalisation when auditory and somatosensory association areas remain inactive. It has been proposed that left ventral premotor cortex and Spt are strongly connected, both anatomically and functionally (Hickok and Poeppel 2007). The present study confirmed the left-lateralisation of frontal opercular activity even if the speech-related activity in the posterior planum temporale is bilateral, but with the additional observation, apparent from the whole-brain contrast of L2 with L1 production shown in Figure 4.4, that this left frontal opercular activity is significantly greater during L2 production. With the data from the ROIs with parietal operculum, using the individual tongue functional localisers on images normalised into standard anatomical space, again there was no functional asymmetry; but unlike in the posterior STP, activity was significantly greater for L2 production, with activity no greater for L1 production than in the rest condition.

Despite one of the functional localisers resulting in co-activation in other regions, the two remaining localisers provided appropriate masks for region of interest analyses. The functionally defined region of interest analyses demonstrated a significant difference between L1 and L2 in auditory and in somatosensory regions. When looking at hemispheres separately, there was a significant difference between L1 and L2 in the auditory region in the left hemisphere, but not in the right. The somatosensory region also showed a significant difference between L1 and L2 in the left hemisphere but not in the right. This suggests that differences in sensory areas between native and non-native languages follow language lateralisation.

A further, and anatomically more refined, ROI analysis was then performed on individual unsmoothed images and using anatomical boundaries available in Freesurfer (the posterior planum temporale) and from the publication of Eickhoff and colleagues (2006) (the parietal operculum). This demonstrated that the posterior extent of the left planum temporale was only active during L2 production, relative to L1, with activity during L1 production no different from that during the rest condition. Directly adjacent to this part of the planum temporale, on the dorsal surface of the lateral sulcus, activity in the parietal operculum was also significantly active during L2 relative to L1 production, with activity during L1 production less, albeit not significantly, than during the rest condition. Although Galaburda and Sanides (1980) have argued, on cytoarchitectonic evidence, that auditory cortex extends over the human parietal operculum, this region has been associated with somatosensory processing of external stimuli in functional imaging studies (reviewed in a meta-analysis by Eickhoff and colleagues, 2006).

Evidence to support the claim that the observed difference between L1 and L2 is due to sensory-motor changes between first and second languages comes from the precise anatomical locations of this change. ROIs were placed on the planum temporale and parietal operculum using objective anatomical criteria on unsmoothed data from each participant (thereby minimising inaccurate parcellation of cortical regions that may accompany ROIs placed on normalised smoothed images). The increased activity in temporo-parietal cortex when speaking in L2 is, by its very location, related to sensory feedback; this was corroborated further by the activation differences between L1 and L2 in regions defined by the non-linguistic functional localisers. In addition to sensory-motor differences between L1 and L2, there were inevitably also linguistic or cognitive processing during speech production. Such factors (e.g., phonological processing differences, differences in attentional control or translation influences from L1 into L2) plausibly explain the pattern of activation seen in a number of the regions reported, such as in inferior and superior frontal regions and others. However, the regions analysed with ROIs, the planum temporale the parietal operculum, are sensory areas, and are not implicated in cognitive or linguistic processing.

In this Chapter I have demonstrated that activity, compatible with sensory-motor integration, in posterior planum temporale and adjacent parietal operculum was only evident during L2 production; with activity during both L1 and L2 production confined to a more anterior part of the posterior STP. One conclusion from the profiles of activity observed in parietal cortex is that production of L1, which is highly automatic with few motor errors, requires little online somatosensory monitoring, and this is restricted to part of the posterior STP, and not evident at all within the second-order somatosensory association cortex of the parietal operculum. In contrast, production of L2, less automatic and more prone to error, engages much closer sensory monitoring of any discrepancies between the predicted and actual sensory outcome of movements, and this is reflected in the activity distributed more widely throughout the planum temporale and within the parietal operculum.

A more plausible interpretation, in keeping with animal studies investigating auditory feedback, is that feedback monitoring, both auditory and somatosensory, is always involved during speech production, even when there is no measurable fMRI signal. In non-human primate auditory cortex, single cell recordings have revealed that the majority of neurons within auditory cortex are suppressed during vocalisation, but a subset of these suppressed neurons remain sensitive to auditory feedback, and increase their firing rate in response to disturbances in this feedback (Eliades and Wang, 2008). By analogy, the absence of signal in secondary somatosensory cortex during native speech may reflect increased feedback efficiency on the part of a subset of neurons. Feedback monitoring in L1 engages a subset of sensory association neurons distributed within the posterior STP and the parietal operculum, with suppression of many other neurons not engaged by the monitoring of speech production. The net result, detectable at the resolution of fMRI, could result in no measurable increase in BOLD response or even a deactivation relative to a rest condition. In the present study, there was a non-significant suppression of activity during L1 production relative to rest in the parietal operculum, and in the earlier study by Dhanjal and colleagues (2008) this suppression reached significance. The conclusion from this interpretation is that a reduced or absent signal in secondary sensory cortex during native speech indicates increased feedback efficiency by a limited number of neurons, tuned by experience from an early age. Viewed in this way, the increase in signal in the temporo-parietal junction in non-native speech

indicates feedback processing that has never become as optimally efficient as native language acquisition during early childhood. The consequence would be a less reliable efference copy of articulator motor commands sent to or processed in sensory cortex, and a consequent inability to appropriately suppress neurons involved in feedback. These two explanations are, of course, not mutually exclusive, and the signal I observed in sensory cortices may represent a combination of the two effects.

What is not established in this study is the sensory modality that is influencing temporo-parietal activity. Heteromodal processing of sensory information occurs early in 'unimodal' sensory association cortices, and the merging of auditory and somatosensory information in the posterior superior temporal plane and parietal operculum does not divide along strict unimodal sensory boundaries (Smiley et al., 2007).

Although the discussion here has emphasised the role of the parietal operculum, based on the original hypothesis, other regions that were apparent in the contrast of L2 > L1, such as the lateral thalamic nuclei and the superior cerebellum, are included in the distributed network envisaged in the computational model of speech production devised by Golfinopoulos and colleagues (2010). Therefore, the differential processing within temporo-parietal cortex for the production of L1 and L2 is only one component process across a distributed cortical, subcortical and cerebellar network. However, it is the one sensory region that exhibited the considerable plasticity within the adult brain that accompanied acquisition of L2. Midline cerebellar cortex has also been proposed as a component of the pathway for feedforward commands for speech production (Guenther et al., 2006). I observed increased activity in response to L2 relative to L1 within these cerebellar regions. This mirrors the response observed in the study of Tourville and colleagues (2008). As such, increased midline cerebellar activation provides converging evidence for motor-sensory processing differences between L1 and L2.

Tourville and colleagues (2008) showed sudden unexpected perturbations of auditory feedback during single word reading in L1 result in an event-related increase in activity in posterior auditory association cortex (including planum temporale) and in

parietal operculum. This was associated with an online compensatory alteration in articulation to counter the perceived perturbation, despite the speakers' lack of awareness of the mismatch between production and auditory feedback. This automatic compensation offers an explanation for why, in the present study, the shift in activity within temporo-parietal cortex accompanying the change from L1 to L2 was not sensitive to measures of L2 proficiency and daily use. This study was performed on highly educated participants, working in England and continually using English in their professional lives. This deliberate choice of a homogeneous population explains why there were unlikely to be imaging-behavioural correlations. The choice of population was to demonstrate altered function in temporo-parietal cortex even though the participants were proficient and practiced in L2.

This Chapter has demonstrated motor-sensory differences when participants speak in their non-native versus native languages, revealing bilingualism as a motor-sensory skill. The following Chapter presents a study that investigated changes in this system over time (with a serial fMRI study) as participants acquire a second language through training. This Chapter also demonstrated the advantage of using individual anatomically defined ROIs rather than functional localisers, particularly a non-speech auditory localiser, such as the amplitude-modulated noise used here, and therefore in the following Chapter, only anatomical ROIs were included.

# 5 Learning to produce non-native speech sounds

The previous Chapter demonstrated that regions involved in integrating motor feedforward signals with sensory feedback signals are more active during non-native than native speech production. This study used a prospective training fMRI paradigm to explore the functional importance of this increased motor-sensory activation for producing non-native speech sounds, and the plasticity within this system with a short period of training. Participants were scanned after training on production of novel speech sounds using previously unfamiliar phonemes. Manipulating prospective proficiency allows for greater experimental control, to provide insights into how proficiency and experience drive the organisation of somatosensory feedback systems.

Repeating verbal stimuli is considered to involve three different processing pathways, semantic, lexical and phonological (Klein et al., 2006). The present study had no semantic or lexical component as participants were not taught the meaning of the words and were instructed to focus solely on articulation and would therefore only recruit the phonological pathway. Both the native and non-native conditions involved novel stimuli and activated the “phonological loop” for storing an unfamiliar sequence of phonemes (Baddeley, 1986; Gathercole, 2006). Although the native stimuli consisted of familiar phonemes, the combination sequence was novel as they were non-words, and the non-native stimuli consisted of unfamiliar phonemes and sequences.

The development of human speech requires vocal imitation in infants, using auditory-guided motor learning in the ‘babbling’ phase. Songbirds go through similar stages of vocal development, beginning with a ‘subsong’ phase before the adult song can be accurately imitated, and it is proposed that these similarities between development of human speech and birdsong can be observed at behavioural, neural and genetic levels (Bolhuis et al., 2010). Both humans and songbirds require auditory input in order to develop their own vocalisations. Without exposure to speech or songs from adults, production of accurate vocalisations is not possible. Both speech and birdsong production also require control of the motor system, matching vocal production to an auditory template, error detection and modification to vocal signals

where necessary. Vocal production requires constant monitoring using auditory feedback throughout learning for both songbirds and human infants. Monitoring continues after the learning process but reliance on auditory feedback is less important for adults than young learners (Bolhuis et al., 2010). This training study aimed to re-create the 'babbling' phase in adult speakers, by introducing them to single words and asking them to listen to them multiple times and to practise saying them, without understanding what the words meant or being able to use the words to construct sentences. It was therefore expected that auditory feedback would be much more important for the production of the non-native words, but not for the native non-words which, although unfamiliar as a whole utterance, consisted of familiar phonemes. The results from this study are interpreted in line with findings from the birdsong literature, which describes song learning in great detail and has many parallels with human speech learning.

## 5.1 Aim and hypotheses

The aim of this study was to examine how activation patterns for speech production develop throughout learning, judged by proficiency level. The hypotheses were that producing non-native speech sounds would result in greater dependency on the function of motor-sensory systems, and that activity in part of this network would decline with practice.

## 5.2 Materials and methods

### 5.2.1 Participants

Twenty-two monolingual native speakers of English (10 female) participated. The average age was 26 years, 8 months (range 19 years, 11 months to 40 years, 5 months). Speech recordings were lost from one participant, who was then excluded from the fMRI analyses as it was not possible to determine whether the tasks had been carried out appropriately, nor was any measure of improvement possible. There were technical difficulties in one of the three runs for three separate participants, so those runs were also excluded from the analyses. In total, data were analysed from three runs for 18 participants and from two runs for 3 participants.

### 5.2.2 Materials

For the non-native condition the stimuli came from three different languages, based on four speech sounds per language, and were produced in collaboration with a phonetician from University College London, Dr Paul Iverson. He and his colleagues developed and recorded the stimuli, which I then edited for use in the study. These speech sounds were specifically chosen to manipulate a different place and manner of articulation in each language. The words were real bisyllabic words, matched for word length. The native stimuli consisted of bisyllabic non-words, also matched for word length. Native speakers of each language, one male and one female, produced the stimuli. Audio-visual recordings were made of the stimuli for inclusion in the training materials, but audio only versions were used in the scanning Sessions. There were four different sounds from each non-native language. Of the four sounds for each language, the pronunciation of two was trained and two was untrained. For Mandarin, the words involved four different tones (Trained: t1 and t4; Untrained: t2 and t3); for Spanish, four different consonants were used (Trained: j and y; Untrained: d and r); and for German, four different vowel sounds (Trained: u and ü; Untrained: o and ö). Table 5.1 lists the non-native words for each sound for each language.

**Session 1**

| To be Trained | | Not to be Trained | |
|---|---|---|---|
| **Spanish** | | | |
| **j** | **y** | **r** | **d** |
| jaba | yema | rato | data |
| jade | yeso | remo | dato |
| jaque | yodo | reno | dedo |
| jota | yuca | robo | dote |
| jugo | yudo | roca | duda |
| **German** | | | |
| **uu** | **ü** | **ö** | **oo** |
| Bube | Düne | Böhme | doofe |
| Fusse | Düse | höbe | lose |
| Kufe | fühle | höhle | Mode |
| Lupe | Güte | löte | pose |
| nute | mühte | töne | Sohne |
| **Mandarin** | | | |
| **1** | **4** | **2** | **3** |
| chu1 fa1 | bu4 cha1 | ji2 cha1 | di3 ya1 |
| di1 wa1 | da4 ba1 | ji2 ta1 | ju3 fa1 |
| pi1 fa1 | da4 ma1 | tu2 sha1 | Ma3 la1 |
| sha1 la1 | fu4 fa1 | tu2 ya1 | Nu3 wa1 |
| ta1 sha1 | la4 ma1 | yi2 ma1 | wu3 ba1 |
| 出發 | 不差 | 級差 | 抵押 |
| 低窪 | 大巴 | 吉他 | 舉發 |
| 批發 | 大媽 | 屠殺 | 馬拉 |
| 沙拉 | 復發 | 塗鴉 | 女媧 |
| 他殺 | 辣媽 | 姨媽 | 五八 |

**Nonwords**

beakie
beavna
beetoo
pablee
padoo
pazlay
tonni
tossi
tovaie
tozko

**Session 2**

| Trained | | Untrained | |
|---|---|---|---|
| **Spanish** | | | |
| **j** | **y** | **r** | **d** |
| gema | yace | reno | daga |
| giro | yate | rezo | dama |
| jaba | yodo | robo | dato |
| jota | yuca | roca | dedo |
| jugo | yudo | roto | dote |
| **German** | | | |
| **uu** | **ü** | **ö** | **oo** |
| Kufe | Bühne | Böhme | Hose |
| Lupe | Düne | löte | kohle |
| Musse | fühle | schöne | lose |
| mute | mühte | töne | Mode |
| nute | wüsche | wöge | pose |
| **Mandarin** | | | |
| **1** | **4** | **2** | **3** |
| chu1 fa1 | da4 ba1 | bo2 fa1 | da3 fa1 |
| di1 wa1 | da4 ma1 | ji2 cha1 | Gu3 ba1 |
| ku1 ma1 | fu4 fa1 | ji2 ta1 | ju3 fa1 |
| ta1 ma1 | ke4 la1 | ni2 ba1 | Ma3 la1 |
| ta1 sha1 | li4 ta1 | tu2 ya1 | Nu3 wa1 |
| 出發 | 大巴 | 勃發 | 打發 |
| 低窪 | 大媽 | 級差 | 古巴 |
| 姑媽 | 復發 | 吉他 | 舉發 |
| 他媽 | 克拉 | 泥巴 | 馬拉 |
| 他殺 | 利他 | 塗鴉 | 女媧 |

**Nonwords**

beavna
beeso
beetoo
bepoe
paffi
panni
pazlay
tockee
tonni
tozko

***Table 5.1, List of non-native words and native non-words in each Session***

*Non-native To-be-trained and Trained words are shown in pink boxes, Non-native Not-to-be-trained and Untrained words are shown in lilac boxes. Native non-words are in beige boxes.*

### 5.2.3 Training

Participants underwent a week of training, one hour per day, with different exercises each day. Training materials were developed specifically for the purposes of this study. All training materials were provided on a CD and participants were given a booklet of exercises and a blank CD for recording their production of the speech sounds. The first part of the training introduced participants to the main articulators involved in speech production, using interactive exercises and multiple-choice tests (an example page is shown in Appendix 6, p226 of this thesis). The next three sessions looked at each of the speech sound groups (Spanish, German and Mandarin) in turn, with the two speech sounds for each language. The non-native speech sounds were also presented using interactive exercises, including listening to the sounds, learning the main articulators used in each sound, and watching video-clips of a native speaker producing the sound. Participants were instructed to record their repetition attempts and were given guidance on creating spectrograms of their speech and comparing those to spectrograms from the native speakers. An example of these spectrograms is shown in Appendix 7 (p227 of this thesis). The final training session put all the non-native speech sounds together, with many repetitions of each. The training focused on producing the sounds multiple times. A listen-and-repeat paradigm was used in training, with participants recording their attempts to confirm that they complied with the training.

#### *5.2.3.1 Phoneme discrimination testing*

With such unfamiliar phonemes as used in this study, it is possible that participants would not be able to discriminate between different sounds and might produce one attempt for a variety of phonemes. To ensure that participants were able to discriminate between the phonemes, I tested their perception of the different sounds used in the study. If participants were able to distinguish differences in the sounds when perceiving them, production errors are more likely to relate to incorrect motor planning, rather than trying to match to an imperfect auditory template.

#### *5.2.3.2 Functional MRI experimental tasks*

In both fMRI scanning Sessions participants listened to and repeated Native and Non-native words. Rest was included as the baseline condition. During speech trials participants saw two white circles on the screen. They were instructed that if the two

circles remained white, it was a listening trial and they were not required to speak, and the participants then heard two more words containing the same sound. If the second circle turned black, that was the cue to repeat the word, and the participants heard no more than the initial word. Hearing three stimuli or hearing and then repeating one stimulus allowed for same-length trials for these two conditions. In each run (three in each of the two scanning Sessions), there were 20 repeat trials for each of the non-native language groups, 6 listen trials for each non-native language group, 10 native repeat trials, 2 native listen trials and 15 rest trials. Figure 5.1 shows the stimuli presentation for the Listening and Repeating trials.

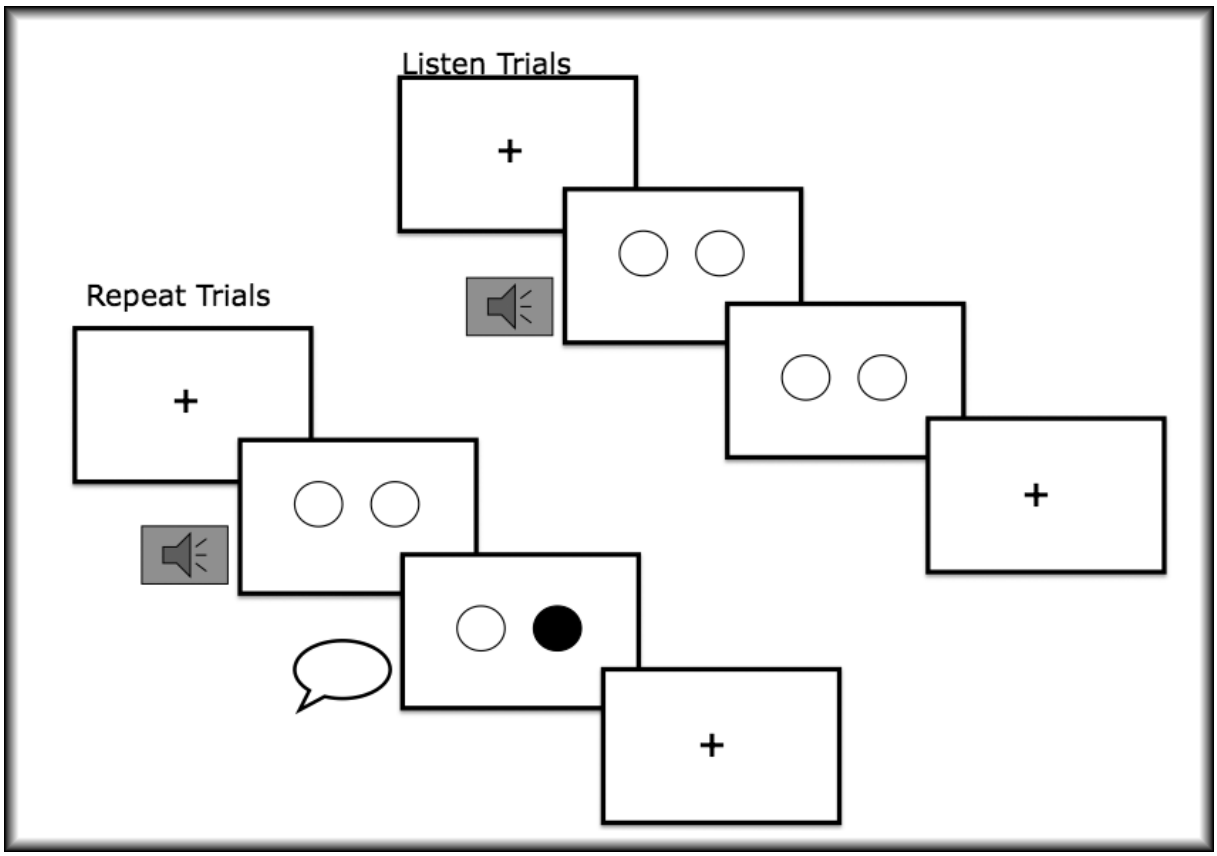



***Figure 5.1, Stimuli presentation for the Listening and Repeating conditions***

*One important point to note from the stimuli presentation design is that when the participants first heard the stimuli, they did not know whether the trial was for a listen or repeat condition. The visual cue for 'go' was the second circle turning black. For 'no-go' the cue was hearing the second word, followed subsequently by a third with no visual change in the circles.*

#### *5.2.3.3 Filtering out CSF and WM*

As well as removing variance associated with the six motion variables and the motion outliers on an individual basis, I also wanted to ensure that white matter (WM) and cerebro-spinal fluid (CSF) would not affect the results. Therefore I also removed variance associated with the timecourses of WM and CSF from the whole brain functional data using ordinary least squares linear regression. To calculate the time course for the white matter and CSF, a 3mm-radius sphere was created based on the coordinates MNI, -26, -22, 28 and MNI, 2, 10, 8 respectively, and the mean timecourse across the sphere calculated.

#### *5.2.3.4 Regions of interest*

Four ROIs (frontal operculum, parietal operculum, anterior planum temporale and posterior planum temporale) were defined on an individual basis using gyral and sulcal landmarks, separately for each hemisphere, using the methods described in Chapter 3 (see Figure 3.1). The eight ROIs were then automatically defined for each individual's reconstructed cortical surface. Mean effect sizes for native and non-native speech conditions, relative to rest, were calculated for each individual. For this analysis the functional data were not spatially smoothed before averaging, in order to avoid any possibility of blurring of activation across the Sylvian fissure between temporal and parietal lobes.

#### *5.2.3.5 Additional ROI analyses*

Following a 2 x 2 ANOVA looking at Language (Native and Non-native) and Session (Pre- and Post-training), the strongest activation was observed in the basal ganglia. I subsequently carried out ROI analyses using anatomical putamen and caudate masks from the Harvard-Oxford probabilistic atlas. A single voxel was selected from within these masks, in each hemisphere separately, for the anterior putamen (aPut) (left -22, 8, -2; right 18, 8, -2), the dorsal caudate (dcaud) (left -14, 12, 12; right 14, 12, 12) and the ventral caudate (vcaud) (left -14, 18, 4; right 14, 18, 4). A 5mm sphere was then created around those six voxels. The spheres were then constrained by the interaction functional result to create the six ROI masks. Separate masks were created for each participant by transforming these standard space masks into each participant's native space, to allow for use in first level, individual

analyses. These analyses were to determine the direction of significant changes, rather than for statistical analysis.

## 5.3 Behavioural results

### 5.3.1 Phoneme discrimination

The mean percentage correct for the phoneme discrimination task was 91.4 (range 72.0-98.0). The mean scores for Mandarin and German were identical to each other and almost at ceiling (mean 93.3, range 58.8-100). Mean scores were lower for Spanish but still well above chance (mean 87.5, range 68.7-100), (Figure 5.2). The most common type of error was misidentifying a matching pair as a mismatch and this most frequently happened for the Spanish phonemes. Discrimination of unpaired phonemes led to most errors being made for Mandarin phonemes (t1 and t2, t1 and t3, and t1 and t4). For the German phonemes errors were made in discriminating between o and ö, ö and u, ö and ü, u and ü). Figure 5.2 presents the results for the phoneme discrimination task.

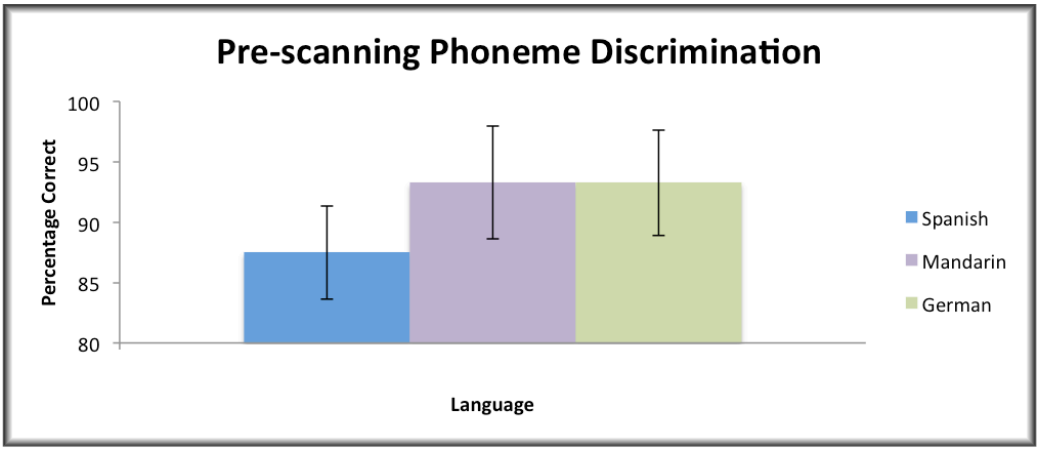



***Figure 5.2, Phoneme discrimination results, all languages***

*Accuracy rates were almost at ceiling for Mandarin and German, although lower for Spanish. Spanish is shown in blue, Mandarin in lilac and German in green. Error bars display 95% confidence intervals.*

### 5.3.2 Behavioural performance

Performance during the scanning Sessions was very high and most participants performed the tasks correctly. Five participants occasionally failed to produce a word during a repeat trial and these volumes were excluded from fMRI analyses. Table 5.2

lists the number of missed repeat trials for each language. There were no occasions when a participant attempted to repeat in a listening trial.

| Participant | Spanish | Mandarin | German | Native |
|---|---|---|---|---|
| 3 | 2 | 0 | 1 | 0 |
| 8 | 3 | 2 | 5 | 0 |
| 13 | 13 | 9 | 3 | 0 |
| 21 | 4 | 4 | 2 | 6 |
| 22 | 1 | 0 | 0 | 0 |

***Table 5.2, Number of missed repeat trials***

*Performance within the scanning Sessions was generally very accurate; only five participants missed trials and even then, only a small number of trials had to be excluded.*

### 5.3.3 Speech ratings

Speech recordings from participants' scanning Sessions were rated by two native speakers of each of the three languages, colleagues of Dr Paul Iverson. I edited the in-scanner recordings, with each word as a separate sound file, presented randomly across participants and Sessions, which the raters then listened to and judged. A 7-point scale was used to judge the degree of native-like performance (1 = non-native-like, very strong foreign accent, 7 = native-like, no foreign accent). Each utterance from each participant was given a score from each of the two native raters. For Spanish, the inter-rater correlation was 0.483, $p < 0.001$; for Mandarin 0.594, $p < 0.001$, and for German 0.543, $p < 0.001$. The correlations differed for the specific phonemes included in the study (Figure 5.3). For Spanish, the correlations for individual sounds were as follows: Trained j = 0.366 and y = 0.247; Untrained d = 0.364 and r = 0.535; for Mandarin, Trained t1 = 0.553 and t4 = 0.566; Untrained t2 = 0.641 and t3 = 0.461; and for German, Trained u = 0.582 and ü = 0.538; Untrained o = 0.418 and ö = 0.629 (see Appendix 8, p228 of this thesis, for the density plots of rating correlations for each sound individually). Using the rating scores from the two native raters for each language, I then calculated the mean score for each utterance for each participant and ran statistical analyses to investigate behavioural performance in both Sessions.

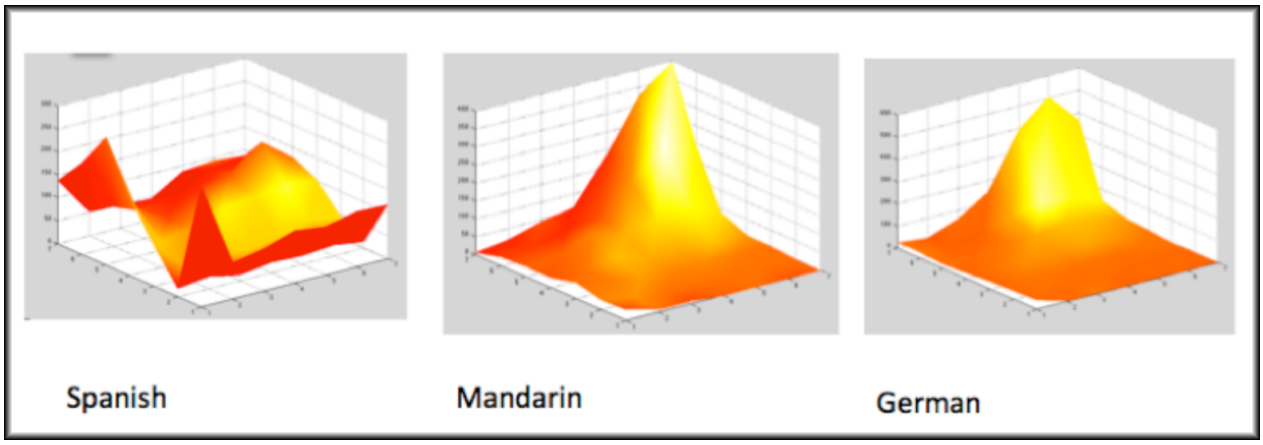



***Figure 5.3, Rater correlations, all languages***

*Density plots showing speech rating correlations for each of the three languages (Eilers and Goeman, 2004). The x-y domain shows the scores from the two raters (using a 1-7 scale) and the z-plane shows the number of observations at each datapoint. Density is reflected with a red-yellow intensity, with yellow areas revealing the highest correlations.*

### 5.3.4 Speech rating scores

Performance in both Sessions varied for the three different speech sound groups. Highest scores were for the German words, closely followed by the Mandarin words, with Spanish words scoring the least. This pattern was the same in both Sessions. Two-tailed paired t-tests revealed significant differences between the mean scores for each language: Mandarin > Spanish (Mandarin = 5.34, Spanish = 3.94, $t(20) = -10.716$, $p < 0.0001$) and German > Spanish (German = 5.53, $t(20) = -10.314$, $p < 0.0001$. The difference between Mandarin and German was not significant. In Session 2, Mandarin > Spanish (Mandarin = 5.55, Spanish = 4.17, $t(20) = -9.537$, $p < 0.0001$) and German > Spanish (German = 5.67, $t(20) = -8.760$, $p < 0.0001$). Averaging across all sounds in each language group, there were significant improvements for all languages, although Spanish and Mandarin improved more than German. The hypothesis for the study predicted a behavioural improvement and therefore one-tailed paired t-tests were used to analyses the difference across Sessions: Spanish $t(20) = -2.734$, $p < 0.01$; Mandarin $t(20) = -2.066$, $p < 0.05$ and German $t(20) = -1.717$, $p < 0.05$. There was wide variability across participants but all improved on at least one of the speech sound groups; some improved on two, some

on all three. Spanish ‘r’ was by far the lowest scoring phoneme, both for Session 1 and Session 2. The order of score from highest to lowest across Sessions only changed for Mandarin, with the largest improvement for t4. The largest improvements across all Sessions were for the Spanish consonant ‘j’, Mandarin tone ‘t4’ and German vowel ‘u’. Performance on the German vowel ‘ü’ decreased after training, but it remained the best pronounced of all speech sounds across both Sessions. Figure 5.4A shows the average scores for the three different language groups, in both Sessions. Appendix 9 (p229 of this thesis) shows the scores in both Sessions for each speech sound individually.

### 5.3.5 Trained and untrained sounds

One-tailed paired t-tests revealed significant improvements for the trained sounds for Spanish (Session 1 mean = 4.60, Session 2 mean = 4.91, $t(20) = -2.895$, $p < 0.005$) and Mandarin (Session 1 mean = 5.58, Session 2 mean = 6.01, $t(20) = -3.950$, $p < 0.0005$) (Figure 5.4B). The trained sounds for German did not show a significant improvement overall because performance on one of the sounds (‘ü’) decreased, although it remained the highest scoring German sound (see above). There were no significant improvements for the untrained sounds (Figure 5.4C).

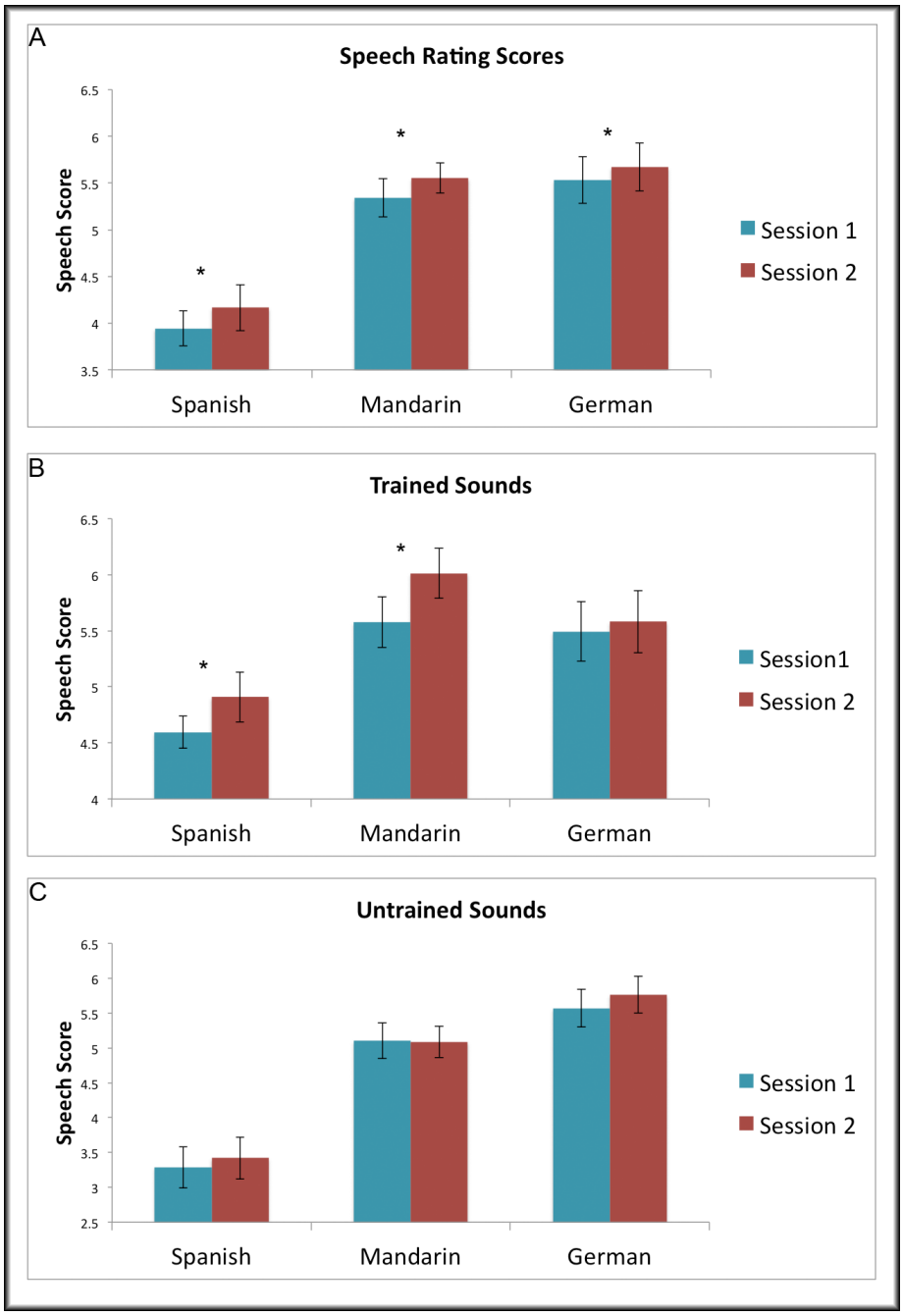



***Figure 5.4, Speech scores from all languages in both Sessions***

*Mean scores from Session 1 are shown in green and from Session 2 in red. Error bars display 95% confidence intervals. A: Average scores for the three speech sound groups in both Sessions. One-tailed paired t-tests revealed significant improvements across Session for all languages ($p < 0.05$). B: Scores for the Trained sounds only for each of the three languages. One-tailed paired t-tests revealed significant improvements for the Trained sounds for Spanish ($p < 0.005$) and Mandarin ($p < 0.0005$). C: Scores for the Untrained sounds only. There were no significant differences across the Sessions for the Untrained sounds.*

### 5.3.6 Coefficients of variation

As well as the score in each Session, I also calculated the coefficients of variation (standard deviation divided by the mean), to determine how variable the speech production attempts were in each Session. Averaging across all sounds, two-tailed paired t-tests revealed significant reductions across Session for Spanish ($p < 0.05$) and German ($p < 0.01$) (Figure 5.5A). For the Trained sounds only, two-tailed paired t-tests revealed significant reductions for the trained sounds for Mandarin ($p < 0.01$) and German ($p < 0.05$) (Figure 5.5B) and for the Untrained sounds only, two-tailed paired t-tests revealed significant reductions for the untrained sounds in German ($p < 0.05$) (Figure 5.5C).

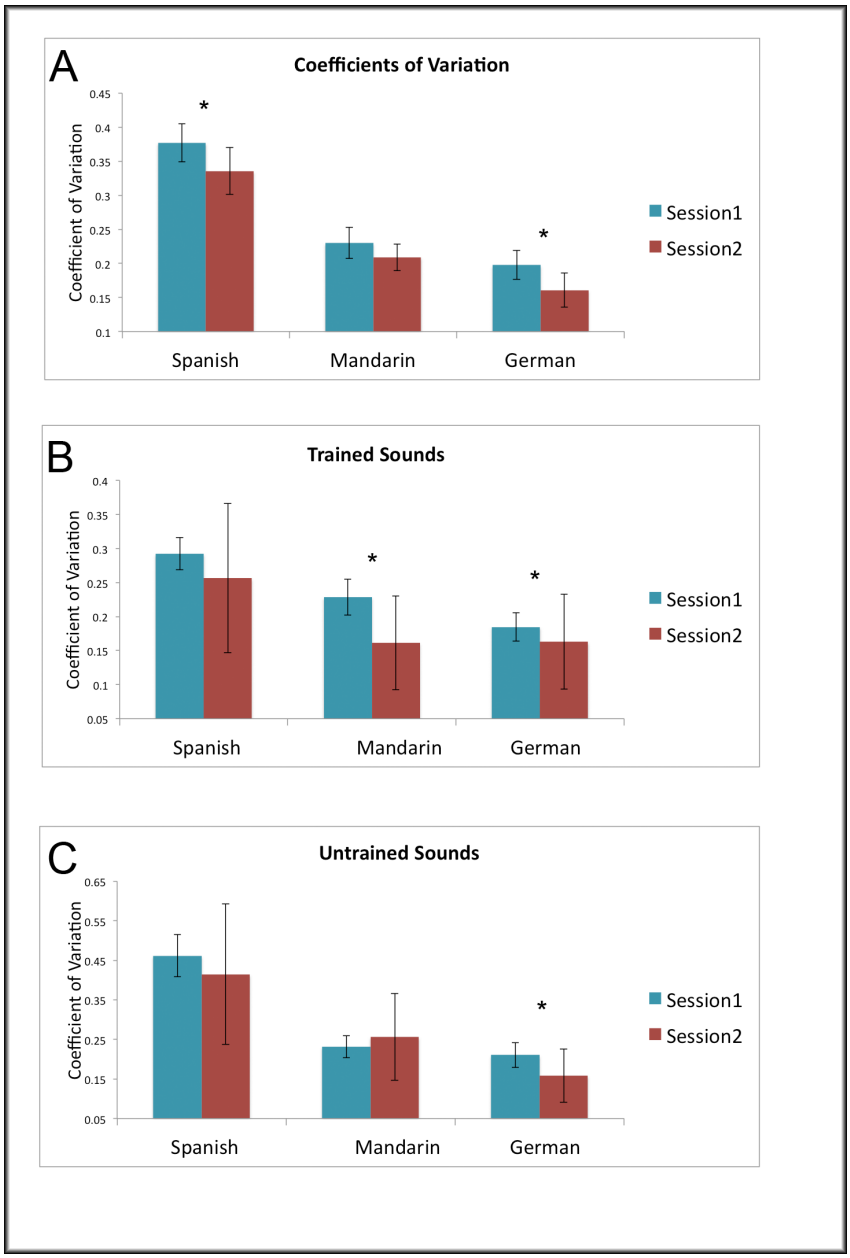



***Figure 5.5, Coefficients of variation***

*Coefficients of variation from Session 1 are shown in green and from Session 2 in red. Error bars display 95% confidence intervals. A: Average coefficients of variation for the three speech sound groups in both Sessions. Two-tailed paired t-tests revealed significant reductions across Session for Spanish ($p < 0.05$) and German ($p < 0.01$). B: Coefficients of variation for the Trained sounds only for each of the three languages. Two-tailed paired t-tests revealed significant reductions for the Trained sounds for Mandarin ($p < 0.01$) and German ($p < 0.05$). C: Coefficients of variation for the Untrained sounds only. Two-tailed paired t-tests revealed significant reductions for the Untrained sounds in German ($p < 0.05$).*

### 5.3.7 Signal-to-Noise ratios for the speech scores

The signal–to-noise ratio (SNR) was calculated by dividing the mean by the standard deviation. Figure 5.6A displays the SNR plotted against the mean scores for Session 1 and Figure 5.6B displays the same for Session 2. Whilst the relationship between the two measures is clear in both Sessions, the curve is more pronounced in Session 2, indicating a higher SNR for conditions with higher mean scores. The standard deviations are also shown, indicating highest variability in both Sessions for Spanish, which had the lowest rated mean score, and lowest variability for German, which was rated the highest mean score. In Session 1 the variability was not significantly different between the 'To-be-trained' sounds and the 'Not-to-be-trained' sounds, whereas in Session 2, the Untrained sounds had higher variability than Trained sounds for both Spanish and Mandarin (Figure 5.6C and D).

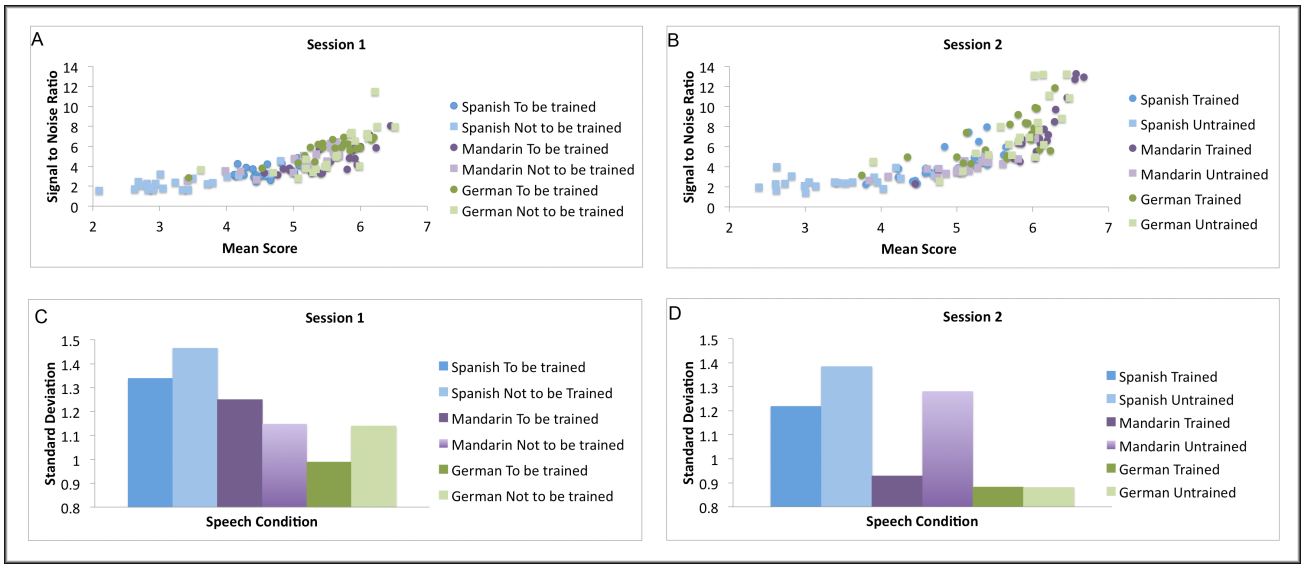



***Figure 5.6, Signal-to-noise ratios for the speech scores in both Sessions***

*The signal to noise ratios plotted against the mean scores are shown for Session 1 (A) and Session 2 (B), with standard deviations for Session 1 (C) and Session 2 (D). Darker shades reflect the To-be-trained sounds in Session 1 and the Trained sounds in Session 2; lighter shades reflect the Not-to-be-trained sounds in Session 1 and the Untrained sounds in Session 2; Spanish (blue), Mandarin (purple) and German (green).*

## 5.4 Whole-brain results

This study was designed to investigate training effects of producing non-native sounds. A 2 (Language: Native and Non-native) x 2 (Session: Pre- and Post-training) factorial design was used. I also included a listening task, in order to investigate whether training on production had an effect on listening, so a 2 x 2 ANOVA was also performed for the listening trials. The difference in cognitive demands between the two tasks (Listening and Repeating) was not the focus of this study and therefore the tasks were analysed separately. Further, the sensory inputs (both visual and auditory) were different between the Listening and Repeating trials, as were the executive demands ('no-go' and 'go' for Listening and Repeating, respectively). Therefore, I did not perform a 2 x 2 x 2 ANOVA. Following the presentation of the 2 x 2 ANOVA for each task, I also present post-hoc contrasts of Language and Session, as well as ROI analyses. The final part of this results section presents decreases in activity for each individual language and the correlations between signal decreases and behavioural improvements. No effects of type of word (Trained or Untrained) were found; therefore all results are reported collapsed across word type.

### 5.4.1 ANOVA: Language (Native and Non-native) and Session (pre- and post-training) for the Listening task

The main effect of Language was observed in bilateral inferior frontal gyri (Figure 5.7A). The main effect of Session was observed in the right middle temporal gyrus, left and right intra-parietal sulcus, left secondary somatosensory cortex, right primary somatosensory cortex, right middle frontal gyrus and left premotor cortex (Figure 5.7B). The interaction was observed in the right intra-parietal sulcus and right middle frontal gyrus (Figure 5.7C).

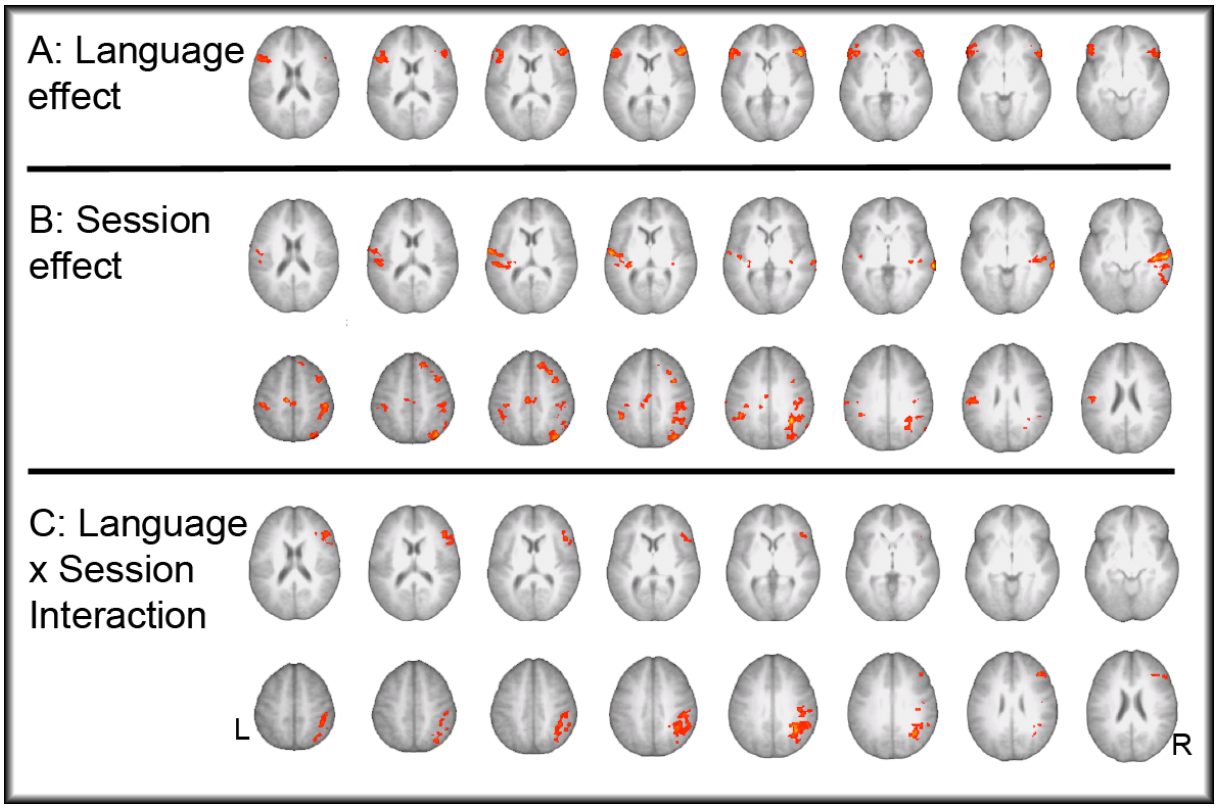



***Figure 5.7, Listening ANOVA***

*A: Main effect of Language, B: Main effect of Session, C: Language x Session Interaction. Results are displayed on a standard brain template (MNI152) and axial slices are shown in neurological orientation, with left on the left. Slices are shown from in 4mm decrements from 18 to -10 for A and the top panels of B and C. Lower panels of B and C show slices in 4mm decrements from 50 to 22. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. A table reporting main cluster peaks for these contrasts is shown in* Appendix 10 *(p230 of this thesis).*

Subsequent to the 2 x 2 ANOVA, direct contrasts between conditions were analysed, commencing with contrasts of each Language versus Rest, and then direct contrasts between Native and Non-Native in Sessions 1 and 2. As will become apparent, these thresholded direct contrasts provided more information than the 2 x 2 ANOVA, and avoided Type 1 errors.

#### *5.4.1.1 Contrasts of Non-native Listening with Rest and Native Listening with Rest – Session 1*

In Session 1, for Non-native Listening relative to the Rest baseline, extensive cerebellar and cerebral systems, symmetrically distributed between the hemispheres were observed (Figure 5.8). There was the expected activity in bilateral primary and association auditory cortex in response to the auditory stimuli, and in bilateral visual cortex in response to the prompting visual stimuli. In addition there was bilateral cerebellar hemisphere activity and bilateral lateral and medial (SMA) premotor activity. This activity probably relates to preparation to repeat; as at the time of hearing the first non-native word the participants did not know whether they were required to repeat this stimulus or, in contrast, listen only to a further two non-native words. In addition, there was bilateral activity in frontal opercular/anterior insular cortex and within the intraparietal sulci. I can speculate that these high-order networks relate to the 'no-go' response to hearing the second word, attention to the auditory and visual stimuli, and working memory. These are incidental findings from the subtraction against the low-level baseline of Rest and will not be discussed, or speculated upon, further. The contrast of Native Listening with Rest showed an identical distribution of activity.

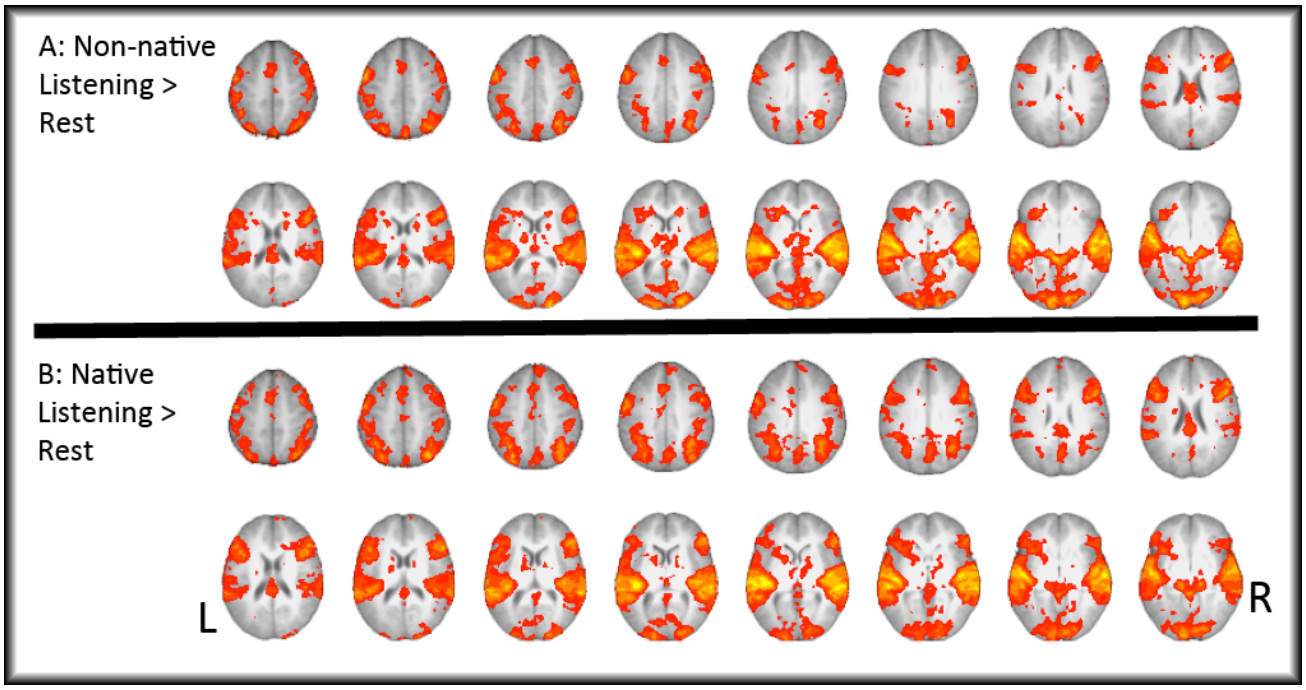



***Figure 5.8, Contrasts of Non-native Listening with Rest and Native Listening with Rest, Session 1***

*A: Non-native Listening > Rest; B: Native Listening > Rest. Axial slices are shown in 4mm decrements; top panel shows slices 50 – 22 and lower panel slices 18 – -10. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Results are displayed on a standard brain template (MNI152) and left is shown on the left. A table reporting main cluster peaks for these contrasts is shown in* Appendix 11 *(p231 of this thesis).*

#### *5.4.1.2 Direct contrasts of Native Listening with Non-native Listening – Sessions 1 and 2*

The direct contrasts between the two Listening conditions were more revealing. At Session 1, there was activity observed above threshold in the contrast of Native with Non-native, but not in the reverse contrast. Native > Non-native Listening showed high-order bilateral fronto-parietal activity that was symmetrically distributed between the two cerebral hemispheres (Figure 5.9A). The right-lateralised activity was observed in the Language x Session interaction shown in Figure 5.7C. The direct contrast reveals that this was, in fact, a bilateral system. The symmetry of distributed activity argues against a Type 2 error for the left-lateralised system. When the Language x Session interaction was run at a lower threshold, this bilateral system was clearly evident. Nevertheless, these bilateral fronto-parietal networks probably relate to greater attention and working memory during Session 1 for Native non-words, and will not be discussed further.

Of much greater interest, in view of the hypotheses that run through this thesis, were the contrasts at Session 2. For this Session, there was no activity observed above threshold in the contrast of Native with Non-native. However, there was significant activity in the reverse contrast, symmetrically distributed between the hemispheres, in posterior auditory cortices (including the plana temporale) and the parietal opercula (PO) (Figure 5.9B). This distributed activity was also clearly evident at the lower threshold for the Language x Session interaction. Its symmetrical distribution between the hemispheres strongly argues for this being signal and not noise. The implication from this result is that there was enhanced activity in <u>bilateral</u> posterior supratemporal plane (STP) and adjacent PO after participants have undergone a period of vocal learning on Non-native words. The alternative is that activity in these regions became less for Native non-words across Sessions. These possibilities were resolved by the ROI analyses, which determined the relationships between activities for Native and Non-native and Sessions 1 and 2.

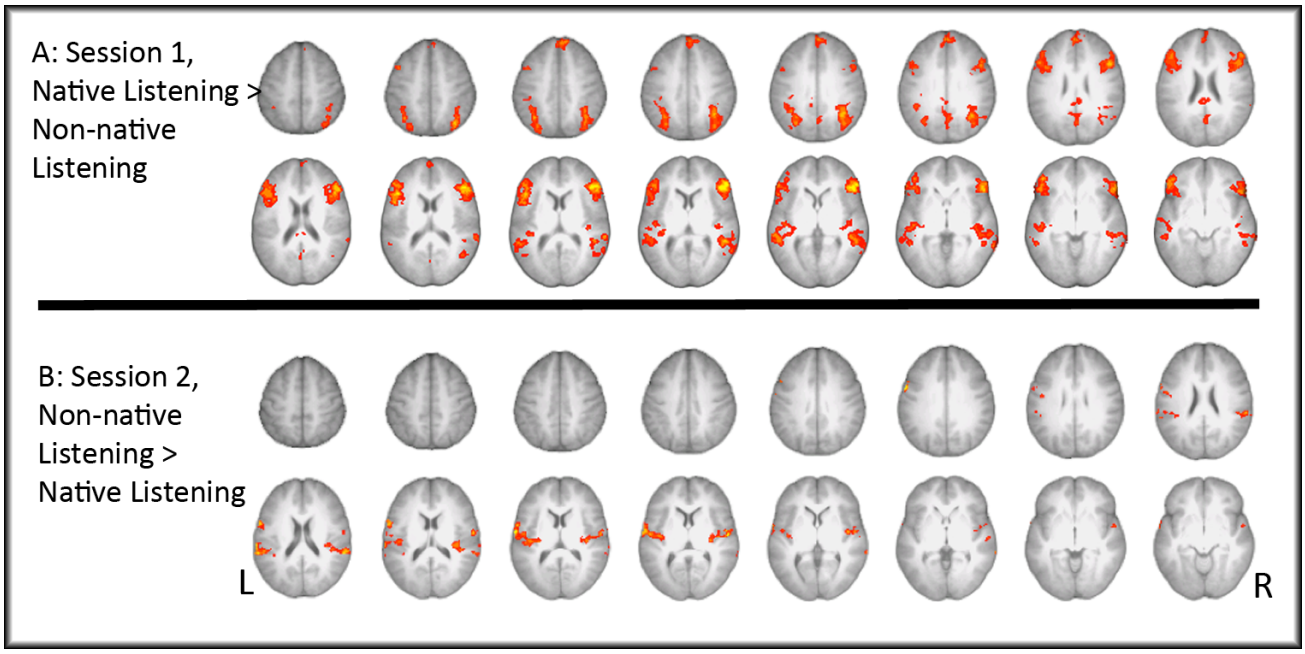



***Figure 5.9, Native > Non-native Listening Session 1 and Non-native > Native Listening Session 2***

*A: Session 1 Native Listening > Non-native Listening; B: Session 2 Non-native Listening > Native Listening. Axial slices are shown in 4mm decrements; top panel shows slices 50 – 22 and lower panel slices 18 – -10. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Results are displayed on a standard brain template (MNI152) and left is shown on the left. A table reporting main cluster peaks for these contrasts is shown in* Appendix 12 *(p231 of this thesis).*

### *5.4.1.3 Listening ROI analyses*

An analysis was also performed on the same cortical ROIs that were used in the studies presented in Chapters 3 and 4 (see Figure 3.1). For the listening tasks, correcting for multiple comparisons, there was a main effect of Language in the left anterior PT ($F = 17.679$ (1,20), $p < 0.005$) and a Language x Session interaction in the left posterior PT ($F = 18.116$ (1,20), $p < 0.005$) (Figure 5.10) and in the right anterior PT ($F = 14.168$, (1,20) $p < 0.01$). The overall picture was that activity for Non-native Listening did not decrease between Sessions, and actually increased in the right anterior PT, but activity in response to Native non-words tended to decrease. There was a strong trend for a Language x Session interaction in both

frontal opercula, $p < 0.01$ uncorrected ($p = 0.08$, corrected), with increasing activity for Non-native across Sessions.

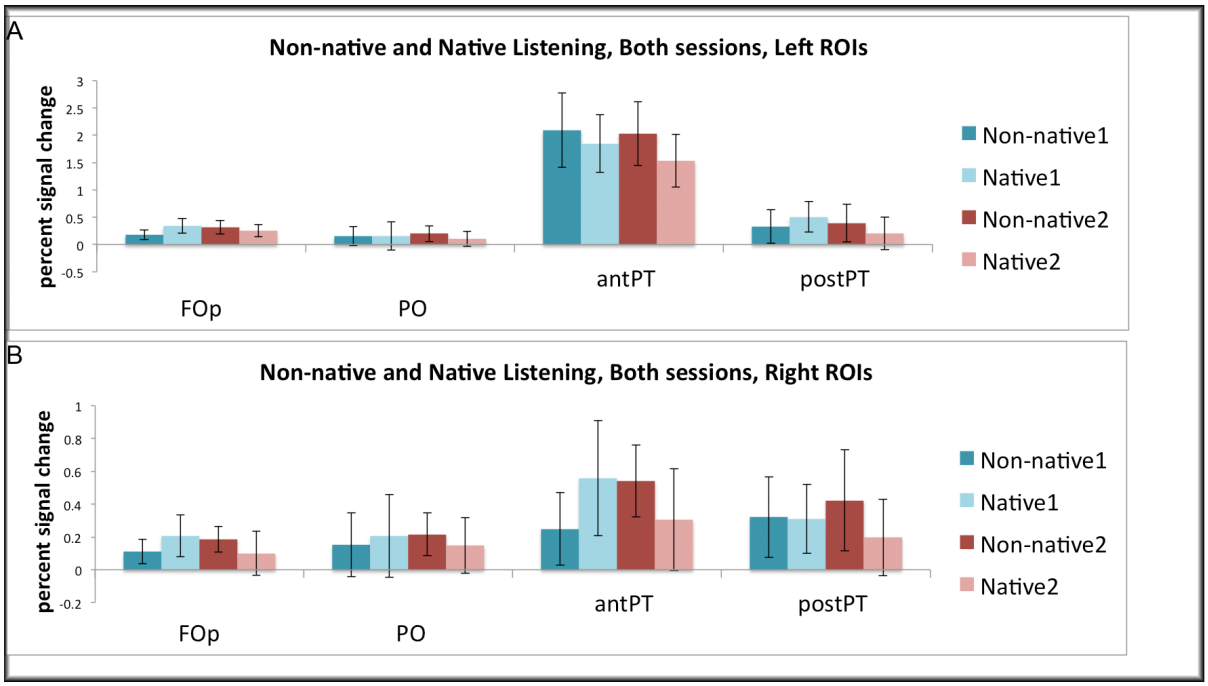



***Figure 5.10, ROI analyses of Native and Non-native Listening in both Sessions***

*A: Left ROIs, B: Right ROIs. Session 1 is shown in green and Session 2 in red, darker shades for Non-native Listening and lighter shades for Native Listening. Error bars display 95% confidence intervals.*

### 5.4.2 ANOVA: Language (Native, Non-native) x Session (Session 1 and Session 2) for the Repeating task

The main effect of Language was observed in extensive bilateral frontal, temporal and parietal regions (Figure 5.11A). The main effect of Session was observed in a number of cortical regions, but was most evident in bilateral basal ganglia (Figure 5.11B). The basal ganglia signal was also most evident in the Language x Session interaction (Figure 5.11C), and the cortical regions demonstrating a significant interaction were also very similar in distribution to those observed in the main effect of Session.

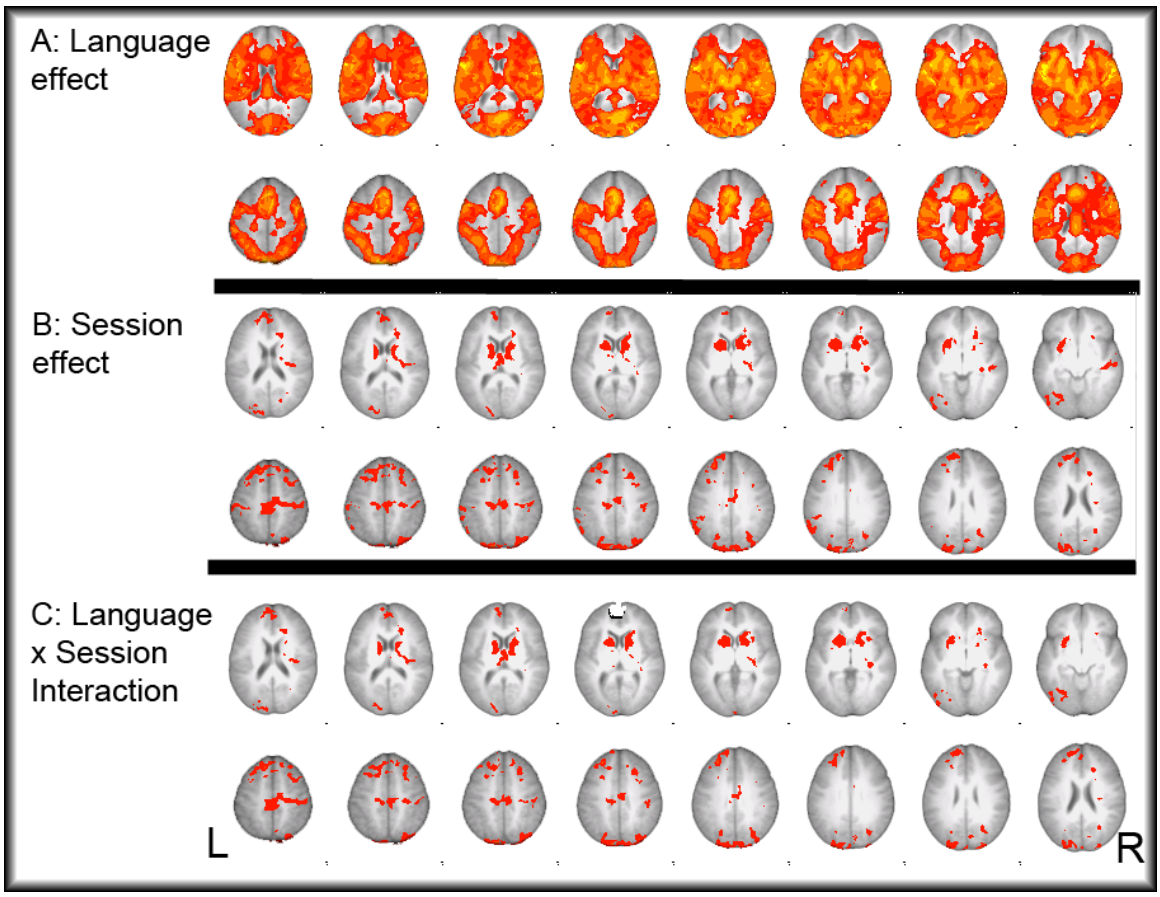



***Figure 5.11, Repeating ANOVA***

*A: Main effect of Language, B: Main effect of Session, C: Language x Session Interaction. Results are displayed on a standard brain template (MNI152) and axial slices are shown with left on the left. In each section A, B and C, slices are shown from in 4mm decrements from 18 to -10 for the top panels from 50 to 22 for the lower panels. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. A table reporting main cluster peaks for these contrasts is shown in Appendix 13 (p232 of this thesis).*

Subsequent to the 2 x 2 ANOVA, direct contrasts between conditions were analysed, commencing with contrasts of each Language versus Rest, and then direct contrasts between Native and Non-Native in Sessions 1 and 2. As with the contrasts from the Listening task, these thresholded direct contrasts were performed to provide more detailed information than the 2 x 2 ANOVA, again to reduce the possibility of Type 1 errors.

#### *5.4.2.1 Contrasts of Non-Native Repeating with Rest and Native Repeating with Rest – Both Sessions*

For Non-native Repeating in both Sessions there was extensive bilateral activity across motor and sensory cortices, the basal ganglia and the cerebellum (Figure 5.12A). This pattern of activity was also observed in both Sessions for Native Repeating (Figure 5.12B).

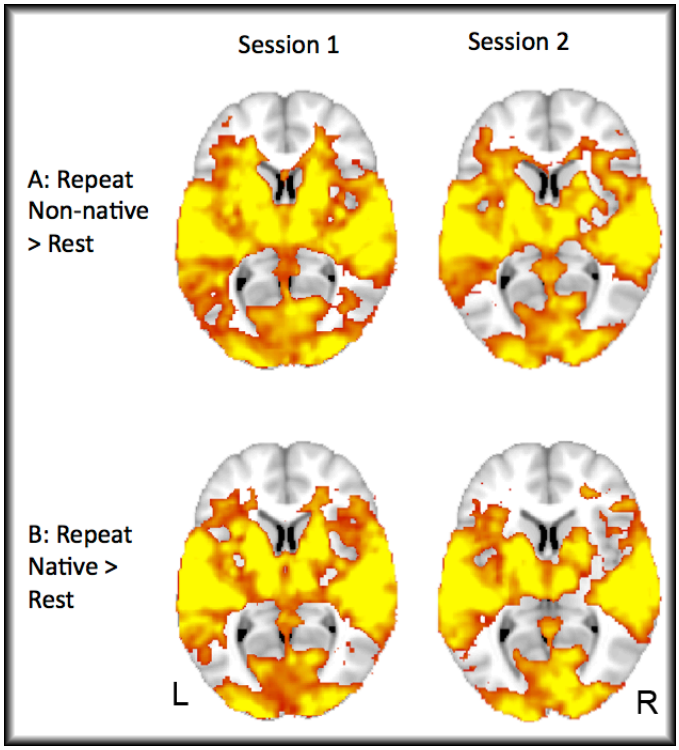



***Figure 5.12, Non-native Repeating > Rest and Native Repeating > Rest***

*A: Non-native Repeating > Rest, B: Native Repeating > Rest. Axial slices are shown, z = 5. Images are presented in neurological orientation, with left on the left, displayed on a standard brain template (MNI152). Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. A table reporting main cluster peaks for these contrasts is shown in* Appendix 14 *(p232 of this thesis).*

#### *5.4.2.2 Direct contrasts of Native Repeating with Non-native Repeating – Sessions 1 and 2*

In both Sessions, comparing across languages revealed greater activity for Non-native Repeating than Native Repeating in motor, auditory and somatosensory regions (Figure 5.13A). Native showed increased activity, relative to Non-native, in the default mode network (DMN) (Figure 5.13B). The DMN is distributed between ventral medial prefrontal cortex, posterior cingulate/retrosplenial cortex, inferior parietal lobule, lateral temporal cortex and dorsal medial prefrontal cortex (Buckner et al., 2008). It is more active in the absence of goal-directed tasks and less active during an effortful and focused task (Seghier and Price, 2012). Therefore, the greater activity of the DMN during Native Repeating is indirect evidence that this task required less task-dependent attention and 'effort' than Non-native Repeating. This is an incidental observation, and will not be discussed further.

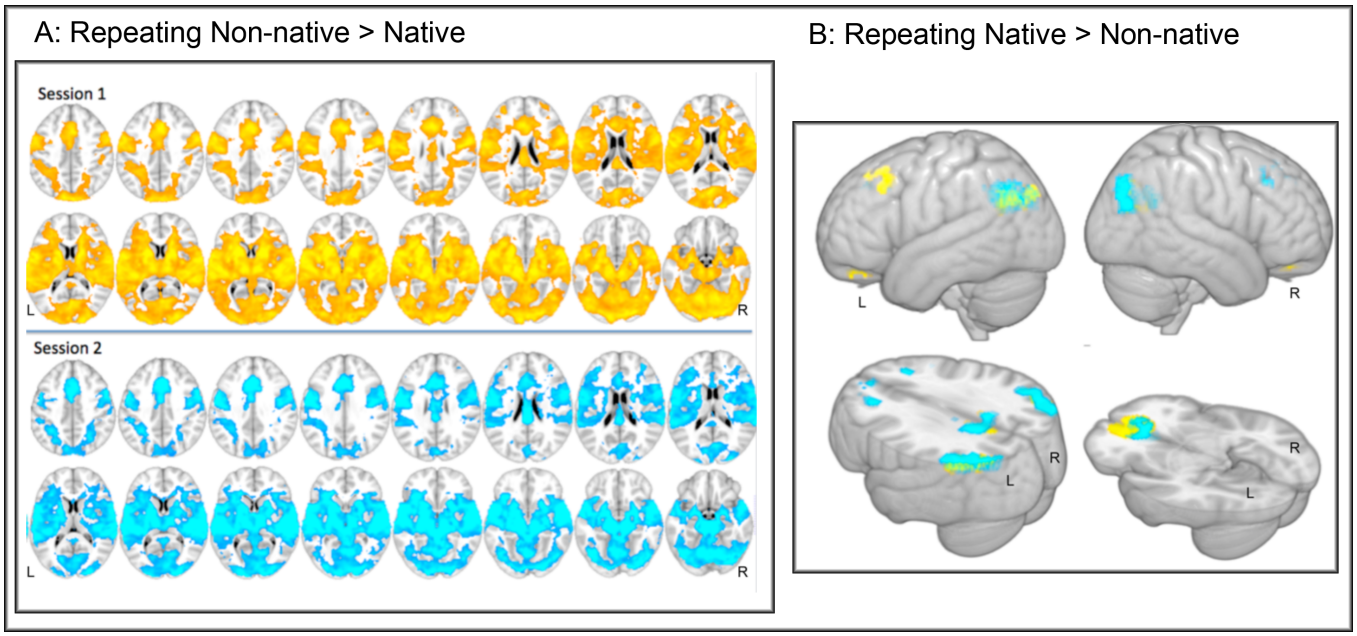



***Figure 5.13, Non-native Repeating > Native Repeating, and Native Repeating > Non-native Repeating in both Sessions***

*A: Repeating Non-native > Native. Session 1 (upper section) is shown in yellow and Session 2 (lower section) in blue. Images from each Session are displayed on a standard brain template with two rows of axial slices in 4mm decrements* (upper row, from z = 43 to 15; lower row, from z = 11 to -17). *Areas of activation in the second Session are broadly similar to those in the first Session, but less widespread. Images are presented in neurological orientation, with left on the left. B: Repeating Native > Non-native. Session 1 is shown in yellow and Session 2 in blue. Images are presented on a standard rendered brain template and with the render clipped from the top to reveal activity in more medial regions. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. A table reporting main cluster peaks for these contrasts is shown in* Appendix 15 *(p233 of this thesis).*

### *5.4.2.3 ROIs from a whole-brain Language x Session interaction for the Repeating task*

The main interest of Repeating relates to the Session and Language x Session effects in the basal ganglia, shown in Figure 5.14A. ROIs were used to determine the profiles of activity across conditions. The positioning of these 5mm spherical ROIs is described in the Methods section of this Chapter (Section 5.2.3.5, p131), and is summarised in Figure 5.14B. As the positioning of these ROIs was based on the whole-brain ANOVA, no statistics are reported for these analyses, although 95% confidence intervals for the effect sizes are included (Figure 5.15). What was evident was that although activity in these striatal regions was present during Native Repeating (see also Figure 5.12B), it was considerably greater during Non-Native Repeating; and activity during Non-native Repeating declined sharply between Sessions 1 and 2.

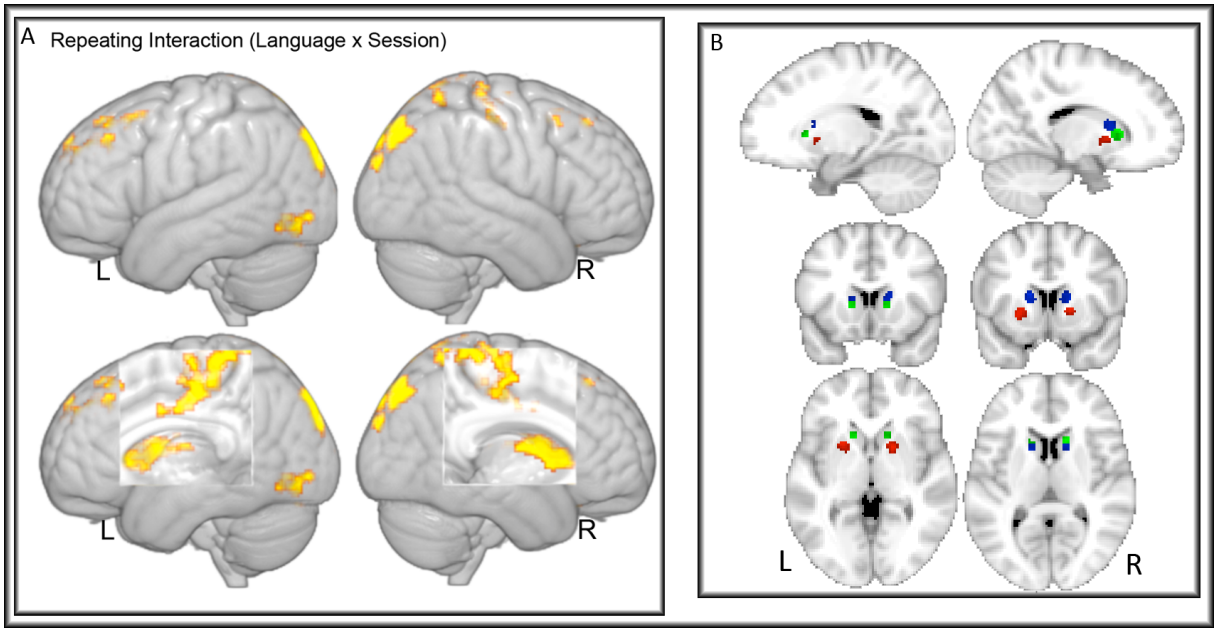



***Figure 5.14, Repeating Language x Session interaction and ROIs***

*A: The whole-brain interaction is shown on a rendered standard brain (top panel) and with a cut-out to reveal the basal ganglia activity (bottom panel. B: ROI masks are shown in both hemispheres for the anterior putamen (red), dorsal caudate (blue) and the ventral caudate (green). The top panel shows two sagittal slices, x = -18 and 16; the middle panel shows two coronal slices, y = 14 and 10 and the lower panel shows two axial slices, z = 0 and 8. ROI masks are displayed on a standard brain template (MNI152) and left is shown on the left.*

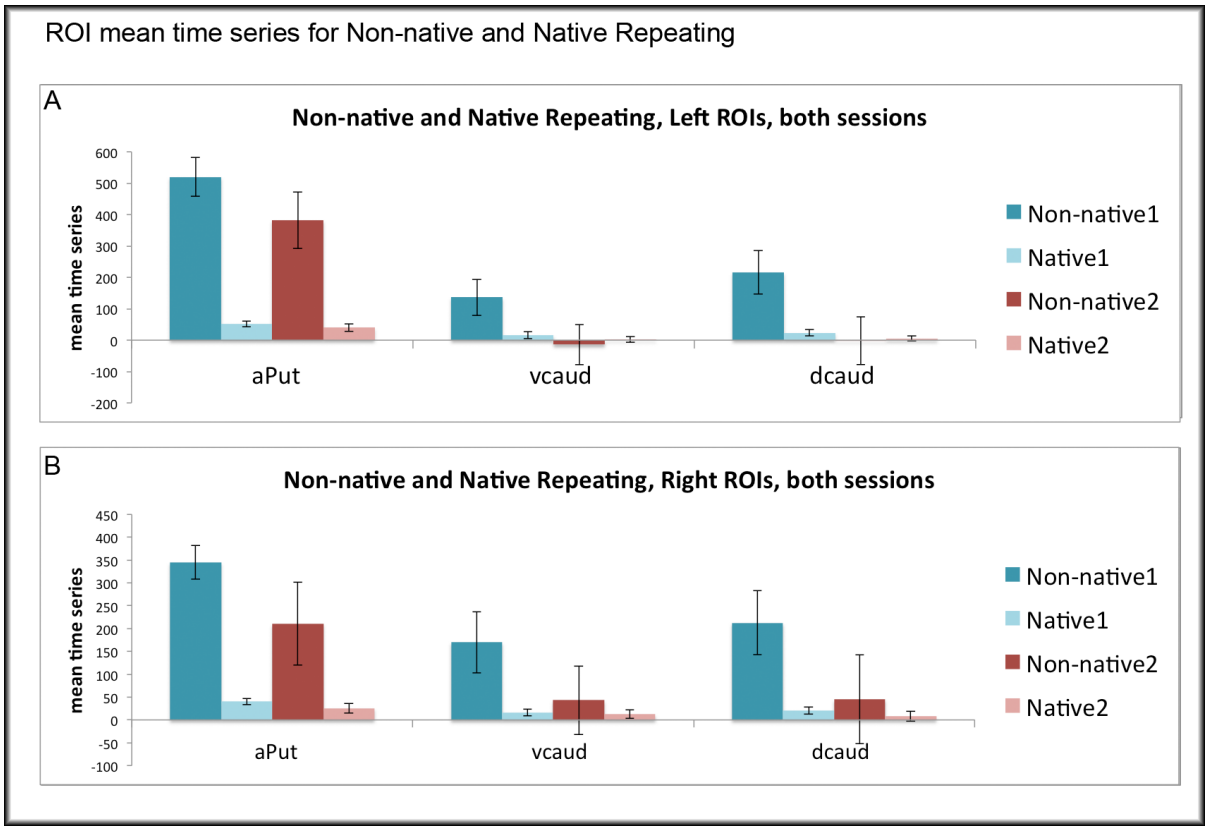



***Figure 5.15, ROI mean time series for Native and Non-native Repeating***

*A: Left ROIs, B: Right ROIs. Session 1 is shown in green and Session 2 in red, darker shades for Non-native and lighter shades for Native. Error bars represent 95% confidence intervals.*

Although there was a major effect between Sessions, a decline of activity in some of the basal ganglia regions was also evident across runs in Session 1, and to a lesser extent in Session 2 (Figure 5.16). As these ROIs were defined from the whole brain analyses, statistics to further investigate the changes between runs were not performed.

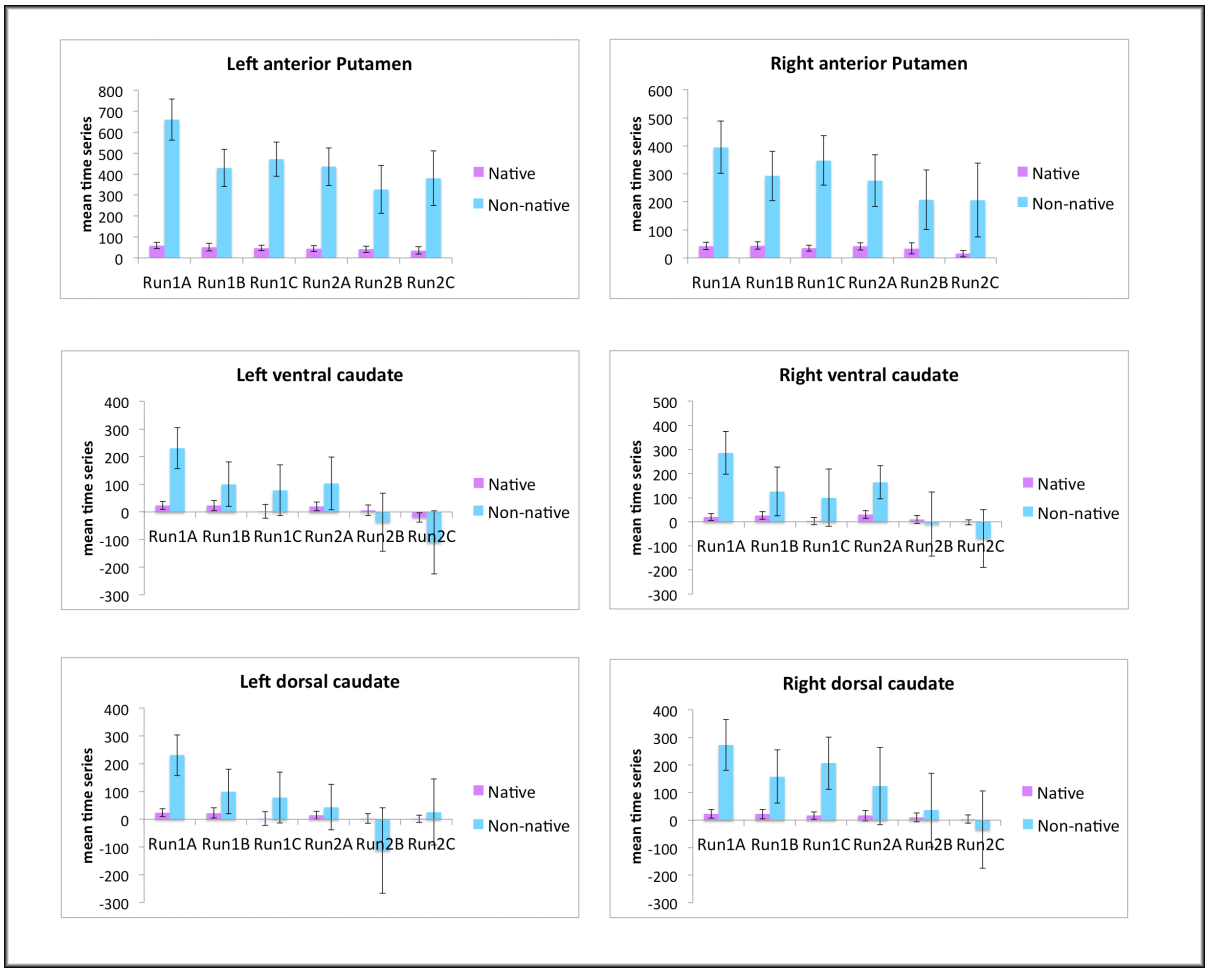



***Figure 5.16, ROI activity across all runs***

*ROI plots showing the mean time series for Native and Non-native Repeating in each of the three runs for both scanning Sessions. Native is shown in pink and Non-native in blue. Error bars display 95% confidence intervals.*

### 5.4.3 Interactions for each Non-native language separately with Native

Separate whole-brain 2 (Mandarin/German/Spanish and Native) x 2 (Session 1 and Session 2) ANOVAs were performed. The main effect of Language was widely distributed and very similar to that for all three languages analysed together, as illustrated in Figure 5.11A. There was a main effect of Session for all three speech sound groups that was most evident in the left basal ganglia for Spanish and in the left and right basal ganglia for Mandarin and German. There were significant interactions in the anterior striatum of both hemispheres for Mandarin and German but not for Spanish (Figure 5.17A). A table reporting main cluster peaks for these contrasts is shown in Appendix 16 (p234 of this thesis). The behavioural scores (Figure 5.4) had demonstrated that Spanish was rated the least accurately articulated of the three speech sound groups, both before and after training. This suggests that striatal activity in vocal learning was modulated by overall proficiency (Figure 5.17B).

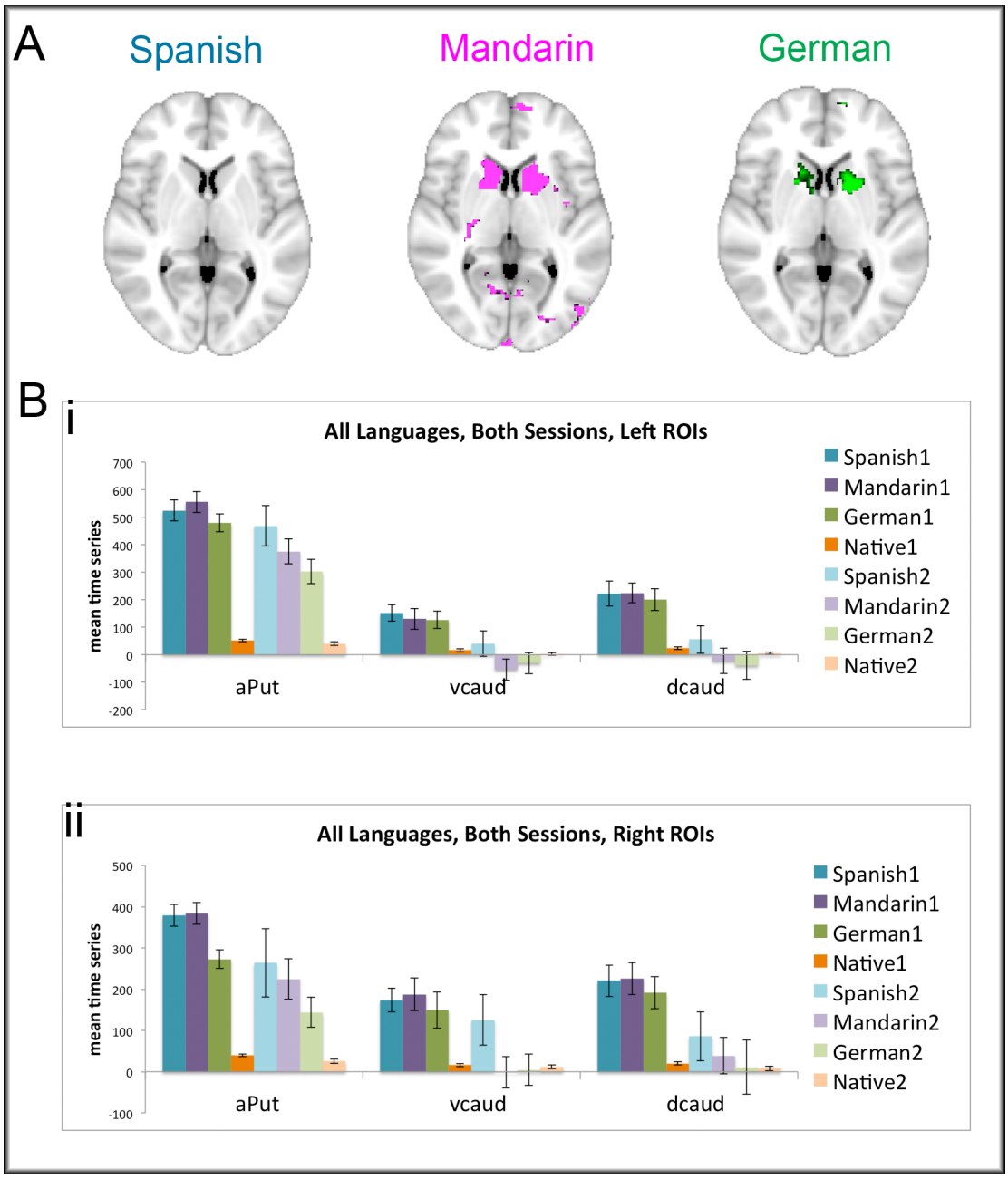



***Figure 5.17, Interactions for each Non-native language separately with Native***

*A: Axial slices of a standard brain template (MNI152), z = 4, showing the Interaction of Session x Language for each Non-native language separately with Native. Spanish revealed no activity from the whole-brain analysis; Mandarin is shown in pink and German in green. B: ROI activity shown for each language separately, with Left ROIs on the upper panel (Bi) and Right ROIs on the lower panel (Bii). Spanish is shown in blue, Mandarin in purple, German in green and Native in orange. Session 1 is shown in darker shades and Session 2 in lighter shades. Error bars display 95% confidence intervals.*

### *5.4.3.1 Cortical region of interest analyses: Repeating*

An analysis was also performed on the same cortical ROIs that were used in the studies presented in Chapters 3 and 4 and in the Listening component of this study. Within these ROIs, correcting for multiple comparisons, there was a main effect of Language (Non-native > Native) in left and right frontal operculum (left: F = 11.768 (1,20), p < 0.05; right: F = 11.408 (1,20), p < 0.05), left anterior PT (F = 34.936 (1,20) p < 0.0001), left parietal operculum (F= 11.842 (1,20), p < 0.005) and right posterior PT (F = 14.244 (1,20), p < 0.05). There were no main effect of Session in any ROIs, nor any Language by Session interactions (Figure 5.18).

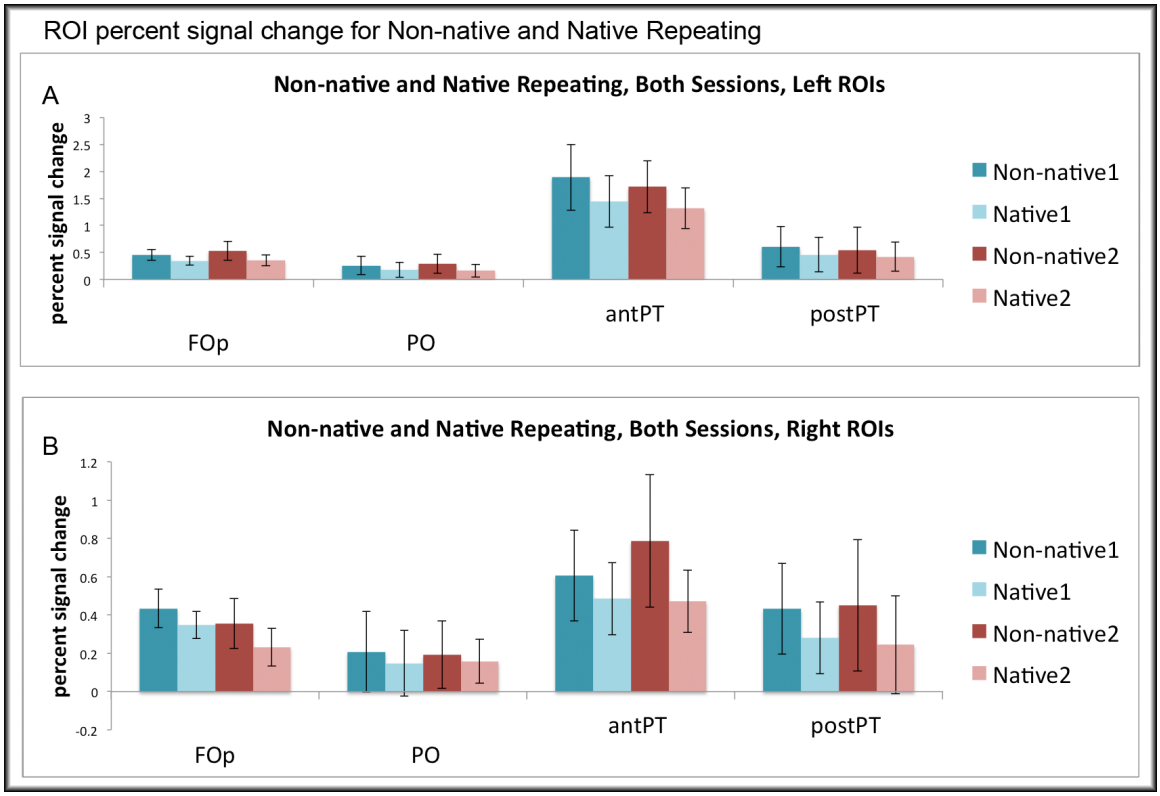



***Figure 5.18, ROI percent signal change for Repeating***

*Mean effect sizes for Native and Non-native Repeating in both Sessions. Session 1 is shown in green and Session 2 in red, darker shades for Non-native Repeating and lighter shades for Native Repeating. Error bars display 95% confidence intervals. A: Left ROIs. B: Right ROIs.*

### 5.4.4 Whole-brain results with speech score

#### *5.4.4.1 Non-native Repeating > Rest*

The contrast of Non-native Repeating > Rest difference across Sessions was run with each participant's individual speech score difference as an additional variable, to investigate how changes in brain activity were linked to improvements in behavioural score. Activity was observed in left primary sensorimotor cortex, the anterior and posterior left superior temporal gyrus, left and right frontal operculum and right lateral cerebellum (Figure 5.19).

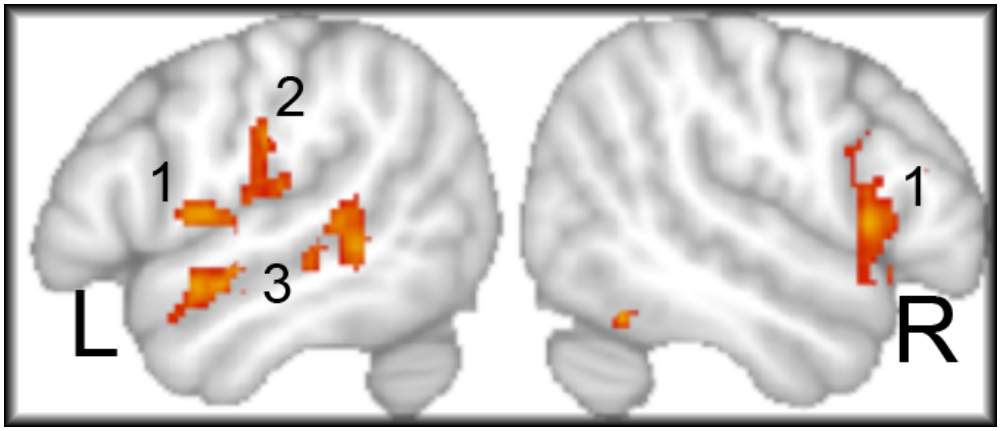



***Figure 5.19, Non-native Repeating, with speech score***

*Two sagittal slices are shown, x = -52 and 52, revealing activity in 1) bilateral frontal opercula, 2) left primary sensorimotor cortex and 3) anterior and posterior left superior temporal gyrus. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Results are displayed on a standard brain template (MNI152) and left is shown on the left. A table reporting main cluster peaks is shown in* Appendix 17 *(p235 of this thesis).*

#### *5.4.4.2 Whole-brain results with separate speech scores*

Within this network of regions with changes during Repeating in response to proficiency, each language contributed differently. There was some overlap between Spanish and German within left primary sensorimotor cortex and left inferior frontal operculum, whereas improvement in Mandarin sounds was responsible for reduction in activity in the left anterior superior temporal sulcus. Improvement in the Spanish

sounds correlated with a reduction in activity in left posterior superior temporal sulcus. The correlation plots in Figure 5.20 confirm that as the speech score improved, activity decreased.

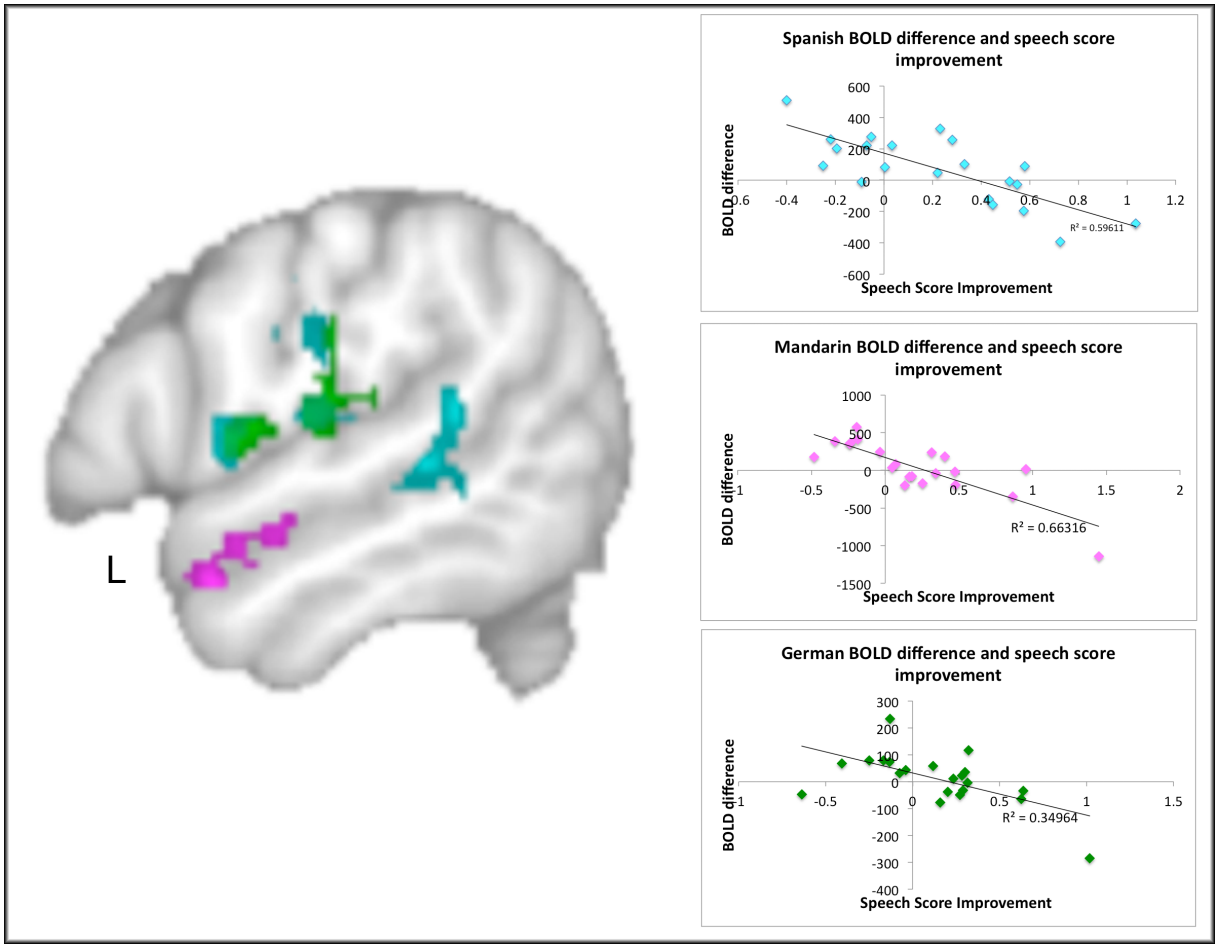



***Figure 5.20, Repeating individual languages, with speech score***

*Spanish is shown in cyan, Mandarin in violet and German in green. Activity is shown on a sagittal slice of a standard brain template (MNI152), x = -54. Z-statistic images were thresholded using clusters determined by z > 2.3 and a corrected cluster significance threshold of p < 0.05. Spanish showed activity in left posterior STS, left primary sensorimotor cortex and left frontal operculum. The peak voxel correlation between the BOLD difference across Sessions and the speech score improvement was r = -0.77 (voxel coordinates -58, -14, 14). Mandarin showed activity in anterior STS, correlation = -0.81 (voxel coordinates -64, -10, -6). German showed activity in left primary sensorimotor cortex and left frontal operculum, correlation = -0.59 (voxel coordinates -40, -26, 28). A table reporting main cluster peaks for these contrasts is shown in* Appendix 17 *(p235 of this thesis).*

## 5.5 Discussion

The pre-scanning phoneme discrimination task demonstrated that participants could detect differences between the non-native speech sounds included in this study, with performance near ceiling for Mandarin and German and well above chance for Spanish. Therefore, I have concluded that the participants were able to form accurate auditory templates of the non-native speech sounds they were required to repeat. Based on this assumption, the vocal training the participants received was directed at the formation of these auditory memories and then modifying their speech output pathways to match more closely accuracy of articulation to these novel templates.

The listening trials required the participants to hold themselves in readiness to repeat the first auditory stimulus (whether a non-native word or native non-word) in each trial, and then inhibit this preparation when they heard a second stimulus rather than seeing the visual prompt to repeat. The nature of this task resulted in activity within a bilateral fronto-parietal system for both non-native words and native non-words, and activity within this system was modulated across Sessions. Converging evidence from other studies that have investigated temporal ordering and 'chunking' of motor sequences (whether the movements involve the finger or the articulators) have observed activity in ventral prefrontal and intraparietal sulcal activity, usually bilateral (Bohland and Guenther, 2006; Jubault et al., 2007; Koechlin and Jubault, 2006; Majerus et al., 2006). This bilateral fronto-parietal system was engaged more fully for native non-words than non-native words during Session 1, but this difference in activity had reduced by Session 2. The obvious explanation for this finding is that motor planning of syllable sequencing was more readily achieved for the sequences of familiar syllables within bisyllabic native non-words at Session 1, but this proficiency improved for bisyllabic non-native words after training, and fronto-parietal activity equalised for native non-word and non-native word Listening trials. This demonstration of changing activity in domain-general fronto-parietal networks, presumably governing working memory and attentional processes, is discussed below.

The opposite interaction in the Listening trials, which was not apparent in the 2 x 2 ANOVA but only when employing direct contrasts between conditions (and therefore

needs to be treated with caution), revealed a change in activity at the left and right temporo-parietal junctions. However, the ROI analysis showed that this was a relatively trivial finding. The study described in Chapter 4 had already established greater activity in these bilateral regions for L2 relative to L1. This increased activity for non-native words was present across both Listening Sessions in this study, but declined for native non-words. The simplest explanation is that the novelty of native non-words produced activity at Session 1, but familiarisation with the presentation of these non-words at Session 1 resulted in reduced activity by Session 2; even though a different set of non-words was used between Sessions the range of phonemes within the non-words was the same.

Articulating the native non-words and non-native words activated extensive bilateral cortical and subcortical, cerebral and cerebellar systems, comprising motor, sensory and higher order systems – which included the bilateral fronto-parietal networks observed in the Listening trials. There were two broad findings. The first was that most of this system was more active for the non-native words, and this was evident at both Sessions. This replicates the finding from the study presented in Chapter 4. The second, and the one that will occupy much of the remaining discussion, was the effect of training on activity in the basal ganglia, and more specifically the anterior parts of the striatum: the head of the caudate nucleus and the anterior putamen in each hemisphere.

This study required participants to learn to pronounce single non-native words, with repeated practice on the articulatory movements required, making self corrections based on anticipated (encoded as efferent feedforward motor copies) and actual sensory feedback. This resulted in overall improvement on the trained words, but variable success across languages and across individual speech sounds within languages. Correlations between behavioural improvement and changes in activity across Sessions for all languages together were observed in the left and right frontal opercula, left ventral sensorimotor cortex, left superior temporal sulcus (STS) and right lateral cerebellum. The specific requirements of the different types of non-native sounds resulted in different patterns of activation for the correlations between the BOLD response and improvement in behavioural performance. Improvement in producing the novel pitch shifts associated with the tonal words of Mandarin

correlated with a reduction in activity in the anterior STS, known to be sensitive to the intelligibility of an auditory stimulus. In contrast, improvement in behavioural performance on the novel vowel sounds in German and novel consonant sounds in Spanish, probably requiring greater dependence on the subtle placement of the articulators, correlated with a decline in activity in ventral sensorimotor cortex. However, precisely which aspects of the speech production tasks result in specific reductions in activity in these regions must be speculative without further studies.

Repetition of non-words is impaired in children with language impairment (Weismer et al., 2000). This study induced similar repetition difficulties in normal participants, by requiring them to use unfamiliar articulatory movements. Gathercole (2006) argues that there are no accurate phonological representations for non-words and it is therefore difficult to repeat a non-word and match it to the phonological input, the heard non-word. One study using non-word repetition with children with Specific Language Impairment (SLI) interpreted difficulties in this task as impaired working memory and phonological storage, rather than impaired articulation (Gathercole and Baddeley, 1990). Gathercole and Baddeley (1990) claimed that articulation in these children was not impaired, as they were not affected by the complexity of the non-words, nor the rate of articulation. In contrast, another study by Bishop and colleagues (1996) found that children with SLI did demonstrate effects of non-word complexity, even when excluding children with poor or atypical articulation. Although these psychological studies did not have a neuroanatomical component, it was evident from this study, during both preparing to repeat in the Listening trials and overt repetition in the Repeating trials, that repeating bisyllabic native non-words and bisyllabic non-native words activate fronto-parietal networks associated with working memory and attention; and to a greater degree in repeating non-native words. It therefore seems possible that the results from this study might inform the possible role of impaired working memory in children with SLI.

Patients with developmental verbal dyspraxia (DVD), such as affected members of the 'KE' family, have articulatory difficulties, in both the planning and production of sequenced mouth movements and are also impaired on repetition of non-words (Watkins et al., 2002a). DVD has been associated with a single point mutation in *FOXP2* in family KE (Fisher and Scharff, 2009; Lai et al., 2001; Warker, 2009;

Watkins et al., 2002a; Watkins et al., 2002b), and in affected members there is reduced volume of the caudate nuclei (Vargha-Khadem et al., 2005; Watkins et al., 2002b). Knockdown of *FoxP2* in zebra finch Area X, a basal ganglia structure necessary for song learning, results in structural and functional striatal abnormalities and an inability to imitate the adult song accurately (Haesler et al., 2007). These findings further emphasise the role of the anterior striatum in vocal learning.

By engaging in sequential motor actions, successful performances are encoded as long-term procedural memories, which can automatically be retrieved even in the presence of interference by other concurrent task demands. During motor habit learning, striatal activity is highly variable, but as habits form, activity peaks at the beginning and end of the motor sequence (Howe et al., 2011). Trial-and-error motor learning engages cortico-basal ganglia circuits to develop the learning of action sequences, and repetition of successful behaviour results in reaching optimum performance (Graybiel, 2005). This means that the variance of the motor performance decreases; that is, the signal-to-noise ratio (SNR) within the system increases. Plasticity in the striatum appears to be necessary for the emergence of start/stop related activity and reduction in behavioural variability.

The evidence for vocal learning comes almost exclusively from studies on songbirds (Kao et al., 2005; Ölveczky et al., 2005), with some evidence from vocal learning in non-human mammals (Barnes et al., 2005). It has been established that the anterior forebrain pathway (AFP) of songbirds is central to song learning, but that once song has become 'crystallised' continued accurate songbird production is much less dependent on the AFP (for discussion of this topic, with references, see Chapter 1). Further, the output from the AFP to the dorsal song production pathway introduces trial-to-trial variability. Motor learning is dependent on an optimal amount of noise, and studies on 'stochastic resonance' are appearing in the literature on human motor learning (Mendez-Balbuena et al., 2012). The AFP is the homologue of the cortical-striatal-pallidal-thalamic-cortical loops in the mammalian brain (see Chapter 1), and so it can be inferred that in the human activity in this circuit would decline with vocal learning. Erich Jarvis (Jarvis, 2004) has already extrapolated the findings from the avian literature, and in particular the electrophysiological and molecular studies on Area X (the avian striatum, also incorporating some pallidal-like neurons) and song

learning, to hypothesise that speech learning is dependent on the anterior striatum. Jarvis's specification of the anterior, as opposed to the posterior, putamen was not evident from his review. However, it must originate from studies on motor learning in non-human primates and rats. These studies have demonstrated that the anterior striatum is involved in motor learning, but once sequences of motor movements become habitual, the maintenance of these procedural memories is dependent on the posterior putamen (Graybiel, 2005; Miyachi et al., 1997; Yin et al., 2009).

In the present study, behaviourally, the SNR increased with training on vocal production of the non-native words, although least for those words the participants found most difficult, namely Spanish. This was associated with a decline in activity in the anterior striatum, seen both as a main effect of Session but also as a Session x Language, with significant training-induced reductions in basal ganglia activity on the non-native words. This study has demonstrated the rapid plasticity in the striatum that occurs following training in speech sounds. Modulation by an early learning effect was demonstrated in the scanning Session before training, in which activity within the left and right striatum declined across the first two runs.

Bilateral human lesions (e.g. strokes) that are confined to the anterior striatum are very rare. In the most widely cited case (Pickett et al., 1998), there was impairment of native language speech, along with higher order language and cognitive impairments which must relate to the presence of multiple cortical-basal ganglia loops in the human: limbic, associative and sensorimotor (Yin and Knowlton, 2006). The single case of Pickett and colleagues was not studied to determine whether new motor learning was particularly impaired.

No previous functional neuroimaging study in humans has demonstrated novel language learning in relation to striatal activity, and it is the first to confirm Jarvis's hypothesis (Jarvis, 2004). Rauschecker and colleagues did not find activation in the caudate nucleus for covert articulatory learning, but they suggested that perhaps overt speech is necessary to observe caudate activity in vocal learning (Rauschecker et al., 2008). Although the human striatum may have a role in maintaining previously acquired speech production, the inferences made from the songbird literature would indicate that the striatum is most active when the novel speech is first encountered. A

motor learning task involving sequences of finger movements in humans, Jueptner and Weiller (1998) demonstrated that the anterior striatum (putamen and the caudate nucleus) and premotor cortex were involved in learning a novel sequence of finger movements, whereas posterior putamen and sensorimotor cortex were activated as participants made overlearned automatic movements. The results from this study did not demonstrate this double dissociation between premotor and motor cortex and anterior and posterior striatum, but this is probably because the pronunciation of non-native words were not overlearned after five days of training. It may be impossible to recruit normal participants for long periods of training on vocal learning (foreign language students will want to learn vocabulary and syntax whilst learning pronunciation, and lexical semantic and linguistic learning will introduce confounds that prevent interpretation in terms of vocal learning alone); but if it were possible, such a study would be a natural extension of this study. Of particular interest would be to determine whether activity never declines to the level of L1 even though it may reduce with increasing use of late-acquired L2. This possibility is suggested by cortical, thalamic and cerebellar regions remaining more active when skilled late bilinguals speak in their L2 compared to L1, presented in Chapter 4.

In summary, this Chapter has demonstrated cortical and subcortical changes in response to non-native listening and repeating after a short period of training, indicating a distributed system involved in the motor-sensory learning of foreign speech. The following Chapter investigates distributed networks in more detail, by mapping the whole-brain connectivity of the temporo-parietal junction.

# 6 Parallel competing systems in the control of speech

In many imaging studies the default assumption is that the change in neural activation measured with functional MRI in a given region reflects a single underlying neural signal. By implication, discrete regions in the speech network can be given specific functional roles: e.g., perisylvian temporal and parietal regions may be involved in integrating sensory and motor information for speech production (Golfinopoulos et al., 2010; Guenther et al., 2006; Ventura et al., 2009). However, the spatial resolution of fMRI means that the signal within any voxel reflects, at a minimum, the net activity of many tens of thousands of synapses on many thousands of neurons. An alternative view questions whether the measured neural activation evoked by speech within a given region reflects the summation of multiple competing signals that carry different functional information (Leech et al., 2012; Smith et al., 2012). The most appropriate functional description may, therefore, not be at the level of brain regions but in terms of these signals, which may span multiple regions or may overlap with other signals. For the analyses in this Chapter I used data from three separate groups of participants and tasks, two involving speech production tasks (presented in Chapters 3 and 5 of this thesis) and one obtained while participants were 'at rest'. The resting state dataset was collected by the Traumatic Brain Injury group in the Computational, Cognitive and Clinical Neuroimaging Laboratory (C3NL) at Imperial College.

## 6.1 Aim and hypothesis

The aim of this study was to investigate the distinct signals from posterior left perisylvian 'language' cortex, in response to different types of speech production and at 'rest'. The hypothesis was that posterior left perisylvian 'language' cortex can be separated into distinct signals that may make functionally different contributions to speech. Specifically, I hypothesised that production of familiar and unfamiliar speech, or the presence or absence of sensory feedback during overt or covert speech, would differentially modulate the separate signals.

## 6.2 Methods

### 6.2.1 Participants

A different set of right-handed participants was used in each of the three studies, all native speakers of English:

- Study 1 (Chapter 5 of this thesis), *n* = 21, ages 19-40 years, ten females;
- Study 2 (Chapter 3 of this thesis) *n* = 17, ages 21-61 years, eight females;
- Study 3 *n* = 16, ages 26-58 years, eight females.

### 6.2.2 Functional MRI acquisition and analysis

Acquisition parameters were the same as outlined in Chapter 2. The three datasets varied in the number of TRs and in the use of sparse sampling. For Study 1 (Chapter 5), there were three runs, each of 105 volumes, using sparse sampling. For Study 2 (Chapter 3) there were 120 TRs in a single run, using sparse sampling. Study 3 contained 300 TRs, again in a single run, with continuous data acquisition.

### 6.2.3 Functional MRI experimental procedures

For Study 1, procedures were as outlined in Chapter 5. For Study 2, procedures were as outlined in Chapter 3. For Study 3 participants were asked to close their eyes and lie still. There was no other explicit task.

### 6.2.4 Data analysis

Preprocessing was carried out using the methods explained in Chapter 2, Section 2.2.6. To remove non-neural noise, variance associated with motion and the time courses of white matter and cerebro-spinal fluid (CSF) were removed from the whole brain functional data using the methods defined in section 5.2.3.3 of the previous Chapter (p131).

### 6.2.5 Defining the left perisylvian language area

The temporo-parietal region was defined using the Harvard-Oxford probabilistic atlas within FSL, for the structures planum temporale, parietal operculum, superior and middle temporal lobe regions and supramarginal gyri and angular gyri (see Figure 6.1A for the mask). A temporal concatenation group independent component analysis (ICA) (Beckmann et al., 2005) was then performed on the speech task data

within this mask. This approach produced spatial maps within the left temporo-parietal junction, identifying voxels that co-vary together. Each map corresponds to spatially and temporally distinct patterns of neural activity or sources of physiological or scanner noise. The probabilistic ICA approach allows for components to be spatially and temporally overlapping (Beckmann et al, 2005). The analysis was initially constrained to extract 10 independent components.

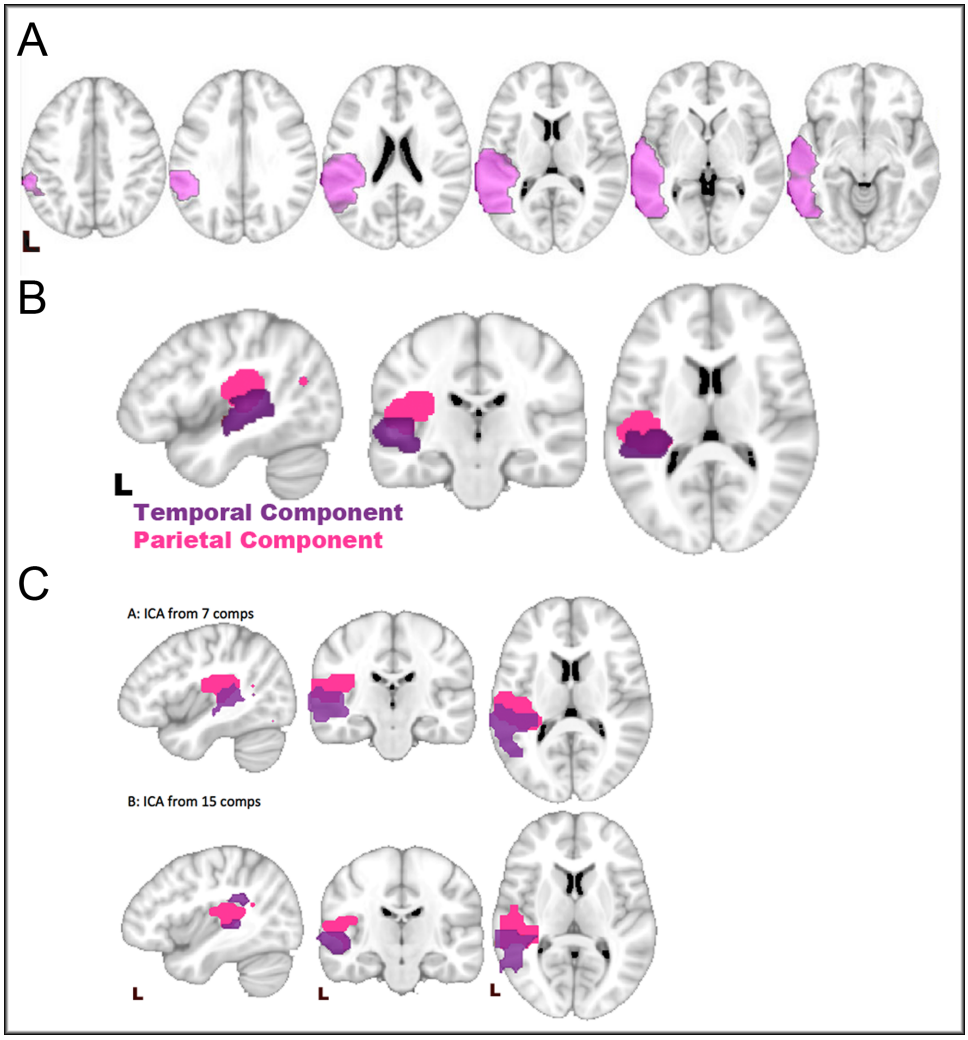



***Figure 6.1, ROI masks***

*A: The mask used for the original ICA. B: The superior temporal (pSTG) component is shown in purple and the ventral anterior parietal (vAPL) component in pink. The coloured overlays are displayed on sagittal (x = -45mm, left image), coronal (y = -22mm, centre image) and axial (z = 12mm, right image). C: The pSTG component is shown in purple and the vAPL component in pink. A shows the two components from an ICA of 7 components (MNI: -41, 22, 12) and B shows them from an ICA of 15 components (MNI: -39, -19, 8). Images are displayed on a standard brain template (MNI152), with left on the left.*

### 6.2.6 Decomposing the left perisylvian language area into discrete signals

The timecourse for each spatial map was calculated, controlling for the variance explained by the other spatial maps, by using a general linear model with the fMRI data simultaneously including the 10 spatial maps from the ICA as a design matrix. This technique identifies 10 independent signals from within the perisylvian region for each participant. A second GLM was used, this time using whole-brain fMRI data. This enables correlations between each of these components with activity across the rest of the brain. The timecourses of the sub-regions calculated previously were simultaneously included in the design matrix to generate a set of whole-brain statistical maps. This approach is a variant on the dual regression approach (Zuo et al., 2010), and has been used by others (Bonnelle et al., 2011; Leech et al., 2011; Sharp et al., 2011). The statistical maps calculated by this analysis provide a whole-brain voxelwise measure of functional connectivity with each of the signals while controlling for variance from the other signals. Components were classed as artefacts and excluded from further analysis if the majority of voxels were in white matter, ventricles or outside the brain (Leech et al., 2012; Smith et al., 2009).

Components that were considered to be neural signal and not noise were then entered into a higher-level general linear model (Beckmann et al., 2003) to compute parametric and non-parametric statistics (random permutation testing), based on the combined spatial maps from the previous step. Both the parametric and non-parametric statistics provided highly consistent results and the results section in this paper reports only the former. Statistics were corrected for multiple comparisons using family-wise error corrections. The individual timecourses generated for each component were also entered into a general linear model with the experimental design matrix for the two different speech datasets with rest as a baseline. This resulted in statistical measure of whether each component's timecourse was modulated by the different task conditions. For more traditional mass-univariate statistics, analysis was carried out using FLAME (FMRIB's Local Analysis of Mixed Effects) (Beckmann et al., 2003; Woolrich et al., 2004). Figure 6.2 displays the steps taken to find sub-regions and their whole-brain functional connectivity.

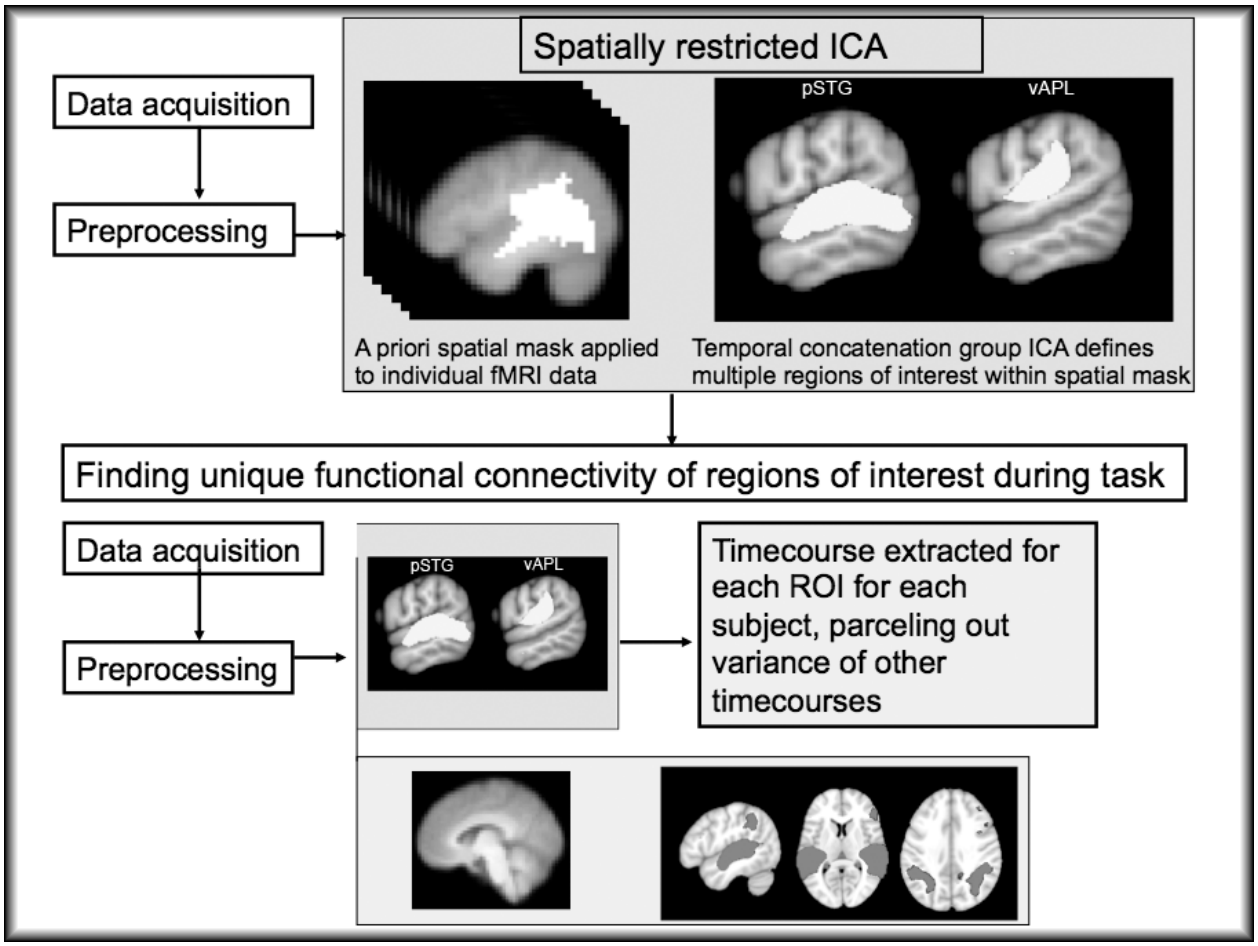



***Figure 6.2, Methods used for finding sub-regions within the left perisylvian language area and finding whole-brain functional connectivity with these sub-regions***

## 6.3 Results

### 6.3.1 Study 1

Twenty-one participants were scanned with fMRI while performing a speech production task, producing bisyllabic non-words or unknown non-native words that differed in place and manner of articulation (see Methods in Chapter 5, Section 5.2.2). I decomposed the FMRI data into ten sub-signals within the left temporo-parietal cortex using independent component analysis (Beckmann et al., 2005; Leech et al., 2012). From those components, two were defined as being of theoretical interest in speech production, replicating regions observed previously (Dhanjal et al., 2008; Simmonds et al., 2011a). One region was located primarily in superior temporal auditory cortex, including planum temporale (pSTG) and the other was predominantly within the ventral anterior parietal lobe (vAPL), including second-order

somatosensory cortex in the parietal operculum (Figure 6.1B). Although these two components were centred more in either the temporal or parietal lobe, they had considerable overlap. This was repeated with 7 and 15 components with qualitatively similar results (Figure 6.1C).

To characterise these signals, I investigated their functional connectivity with activity across the rest of the brain. I calculated participant-specific timecourses for each sub-signal and entered these simultaneously into a general linear model with the whole brain fMRI dataset (Beckmann et al., 2003; Leech et al., 2012). This resulted in a functional connectivity statistical map for each sub-signal. One considerable benefit of this approach is the use of multiple regression both in defining the participant-specific timecourses and subsequently in the resultant statistical maps, thereby isolating connectivity specific to that signal from competing signals and non-neural noise (Zuo et al., 2010).

Although the regions associated with the parietal operculum and superior temporal signals are spatially overlapping, the resulting whole-brain connectivity maps are spatially anti-correlated, suggesting that they compete to some degree (Figure 6.3). The pSTG signal was positively connected with activity in the superior temporal gyri, component of the inferior parietal lobes (predominantly dorsal, extending up to the intraparietal sulci), and along most of the length of the inferior frontal gyri. The bilateral distribution appeared largely symmetrical across the cerebral hemispheres (Figure 6.3). The parietal signal positively correlated with the postcentral gyri, the posterior half of insular cortex, posterior inferior parietal cortex (predominantly the angular gyri), and midline posterior cortex, where posterior cingulate cortex, anterior precuneus and retrosplenial cortex lie in close spatial relationship to one another. Again, this bilateral signal appeared largely symmetrical (Figure 6.3). The negative connectivity results showed an anti-correlated spatial pattern: the temporal signal with bilateral parietal opercular and postcentral cortex, anterior midline cortex, encompassing the supplementary motor area and anterior cingulate gyrus, and the cerebellum; and the parietal signal with the superior temporal sulci, extending up into the inferior frontal gyri, inferior parietal cortex (predominantly dorsal, extending up to the intraparietal sulci) and the pre-supplementary motor area. Once again, the bilateral anti-correlated signals appeared largely symmetrical.

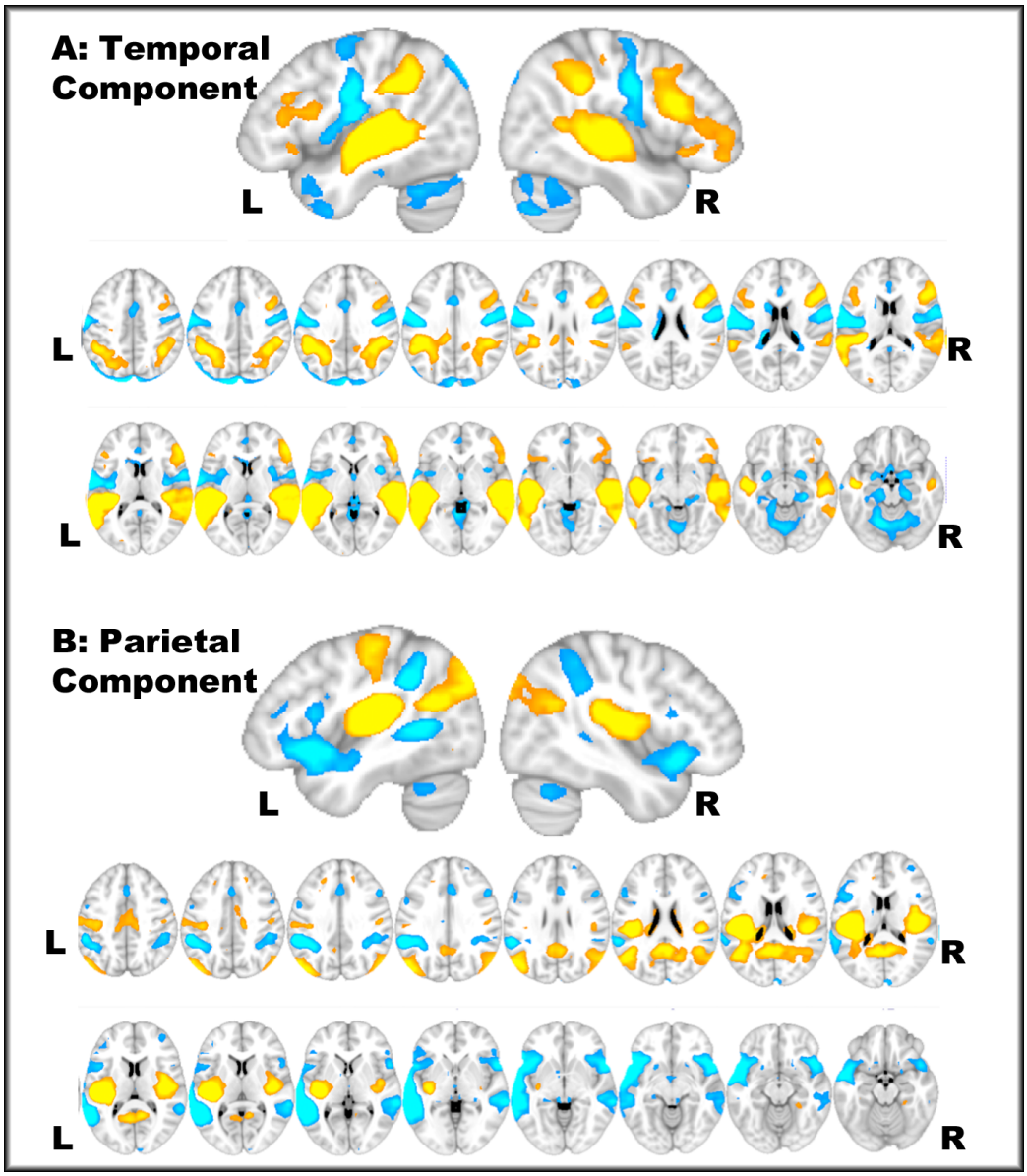



***Figure 6.3, Whole-brain connectivity maps***

*Whole-brain connectivity maps for the two components: the superior temporal (pSTG) component (A); and the ventral anterior parietal (vAPL) component (B). Positive connectivity maps are shown in yellow and negative connectivity (anti-correlation) maps in blue. For each component, the overlays are displayed on slices from a standard brain template (MNI152), with two sagittal slices through the left (x = -44mm) and right (x = 44mm) hemispheres, and two rows of axial slices (upper row, z = 43 to 15mm, in 4mm decrements, and lower row, z = 11mm to -17mm, again in 4 mm decrements). The statistical threshold for the overlays was set at $p < 0.01$, corrected for multiple comparisons using a correction for familywise error rate (FWE).*

The signals can also be characterised in terms of how they are modulated by different task demands. There was increased activity for all speech production tasks relative to rest for both signals, indicating their involvement in speech production. However, the signals were differentially modulated by the specific types of speech produced; in particular, the type and place of articulation (Figure 6.4). This was evident from a significant interaction between the language produced and the region, with the highest activity in pSTG for producing Mandarin tones (novel pitch shifts) and highest activity in the vAPL for producing novel German vowels (involving subtle lip control). A 4 (one native and three non-native languages) X 2 (region) analysis-of-variance (ANOVA) demonstrated a main effect of Language ($F = 4.94$, $df$ (3,60) $p < 0.005$, main effect of Region ($F = 23.123$, df (1,20), $p < 0.001$, and an interaction ($F = 8.423$, df (3,60), $p < 0.001$). Two-tailed, paired *t*-tests revealed that the temporal component showed significantly more activity when producing Spanish (mean = 15.97) and Mandarin (mean = 16.79), compared to German (mean = 15.33) or native (mean = 15.21) speech sounds. Spanish > Native ($t(20) = 2.393$, $p < 0.05$), Mandarin > Native ($t(20) = 6.247$, $p < 0.00001$), Mandarin > Spanish ($t(20) = 3.124$, $p < 0.005$), Spanish > German ($t(20) = 2.228$, $p < 0.05$), Mandarin > German ($t(20) = 5.148$, $p < 0.00005$). The parietal component showed significantly greater activity for producing German (mean = 4.87) compared to native (mean = 4.11) and Mandarin (mean = 4.35) speech sounds. German > Native ($t(20) = 2.484$, $p < 0.05$) and German > Mandarin ($t(20) = 2.217$, $p < 0.05$).

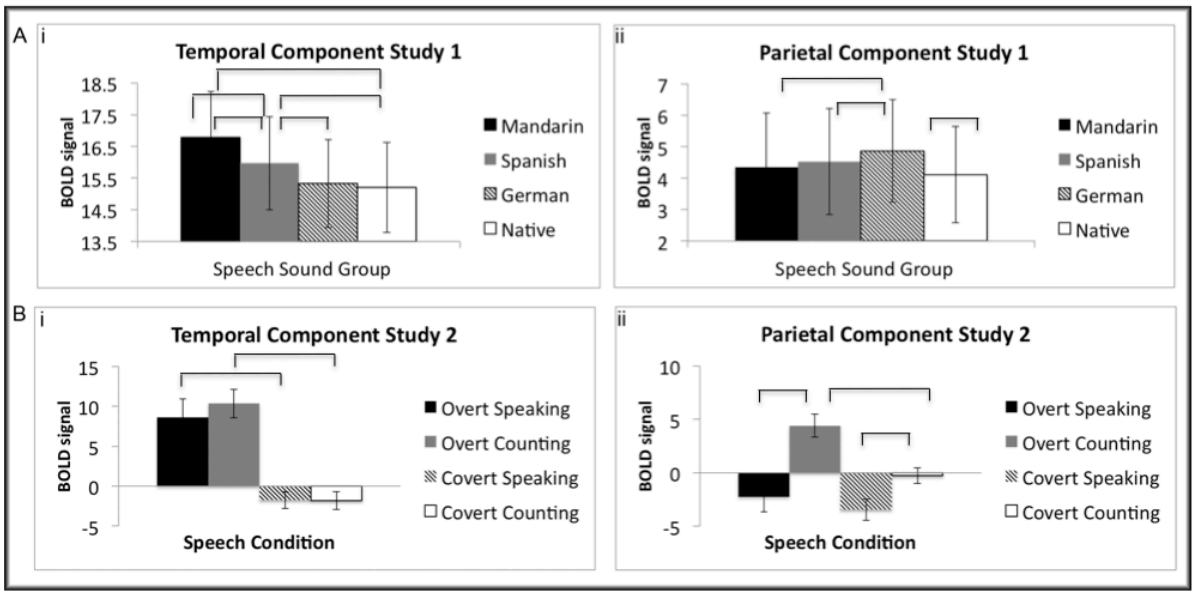



***Figure 6.4, Activity for the temporal and parietal signals in response to different speech production tasks***

*Activity for the temporal (pSTG) and parietal (vAPL) signals in response to different speech production tasks. A: activity during the four conditions in Study 1 (one native and three non-native speech production tasks). B: activity during the four conditions in Study 2 (two overt and two covert speech production tasks). Error bars represent standard error of the mean. Brackets indicate significant differences between conditions.*

There was also a third neural signal, which displayed a pattern of connectivity consistent with the well-characterised default model network (midline anterior and posterior cortices and lateral inferior parietal cortices, Figure 6.5A). This demonstrated a relative task-evoked deactivation for all speech sounds (Figure 6.5B).

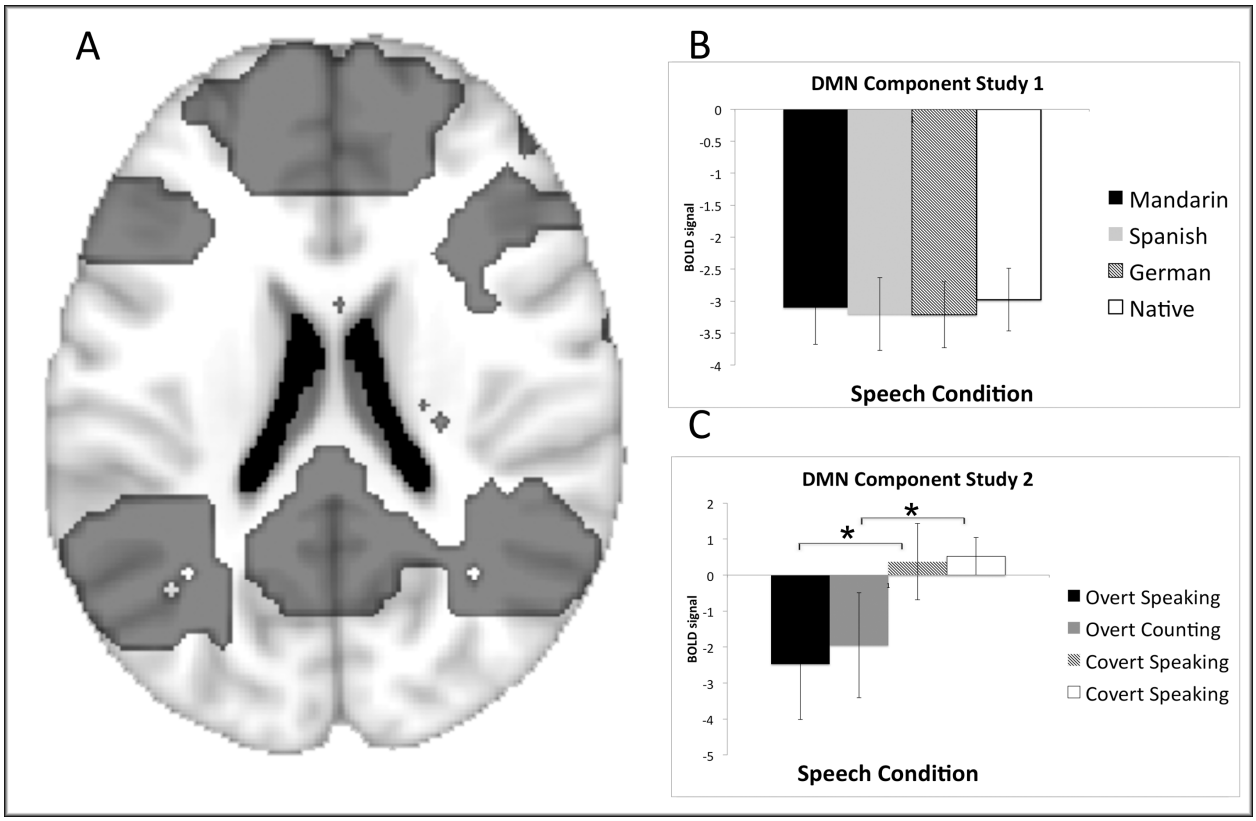



***Figure 6.5, Default mode network component***

*A: The default mode component identified by the ICA, displayed on an axial slice (z = 21) from a standard brain template (MNI152). B: activity during the four conditions in Study 1 (one native and three non-native speech production tasks). There were no significant differences between any of the languages. C: activity during the four conditions in Study 2 (two overt and two covert speech production tasks). Task (Speaking, Counting) by Type (overt, covert) interaction: main effect of Type F = 21.037, df (1,16), p < 0.001). Post-hoc two-tailed paired t-tests: (mean Covert Speaking = 0.375, Overt Speaking = -2.476, t(16) = -4.488, p < 0.001); (mean Covert Counting = 0.525, Overt Counting = -1.952, t(16) = -3.666, p < 0.005). Error bars represent standard error of the mean. *significant differences between conditions.*

### 6.3.2 Study 2

The pSTG and vAPL also showed highly distinct patterns of activation revealed by a second speech production dataset. In contrast to the first experiment, during which single words were produced in a constrained manner, 17 participants produced self-generated overt and covert propositional speech (Overt Speaking and Covert Speaking) cued by visually presented words, and overt and covert non-propositional speech (Overt Counting and Covert Counting). Applying the same regions as defined with the single-word production task above resulted in a similar pattern of functional connectivity for the two tasks (Figure 6.6ii): connectivity in bilateral temporal regions and inferior frontal regions were shared in the two tasks. A more detailed and quantitative comparison of spatial maps across datasets showing similarity across tasks is presented in the section, "*Comparison between studies"* below (Section 6.3.4, p177). A task by region interaction revealed a main effect of Task ($F = 15.421$, df (3,48), $p < 0.001$), main effect of Region ($F = 19.961$, df (1,16), $p < 0.001$) and an interaction ($F = 14.618$, df (3,48), $p < 0.001$). The activity within the pSTG signal was higher relative to rest for both overt speech production tasks, propositional and non-propositional (Figure 6.4B). Two-tailed paired *t*-tests revealed that the temporal component was significantly more active when producing overt speech than covert speech (mean Overt Speaking = 8.62, Covert Speaking = 1.76, $t(16) = 3.381$, $p < 0.005$), and overt counting than covert counting (mean Overt Counting = 10.34, Covert Counting = -1.80, $t(16) = 6.321$, $p < 0.00001$). In contrast, the parietal signal showed a relative deactivation for propositional speech, both overt and covert, and activated only for overt, but not covert, counting (Figure 6.4B). Two-tailed paired *t*-tests revealed that overt counting was significantly higher than covert counting (mean Overt Counting = 4.42, Covert Counting = -0.28, $t(16) = 3.209$, $p < 0.005$) and higher than overt speaking (mean Overt Speaking = 4.42, Overt Speaking = -2.25, $t(16) = 3.186$, $p < 0.01$) and covert counting was significantly higher than covert speaking (mean Covert Counting = -0.28, Covert Speaking = -3.46, $t(16) = 2.585$, $p < 0.05$).

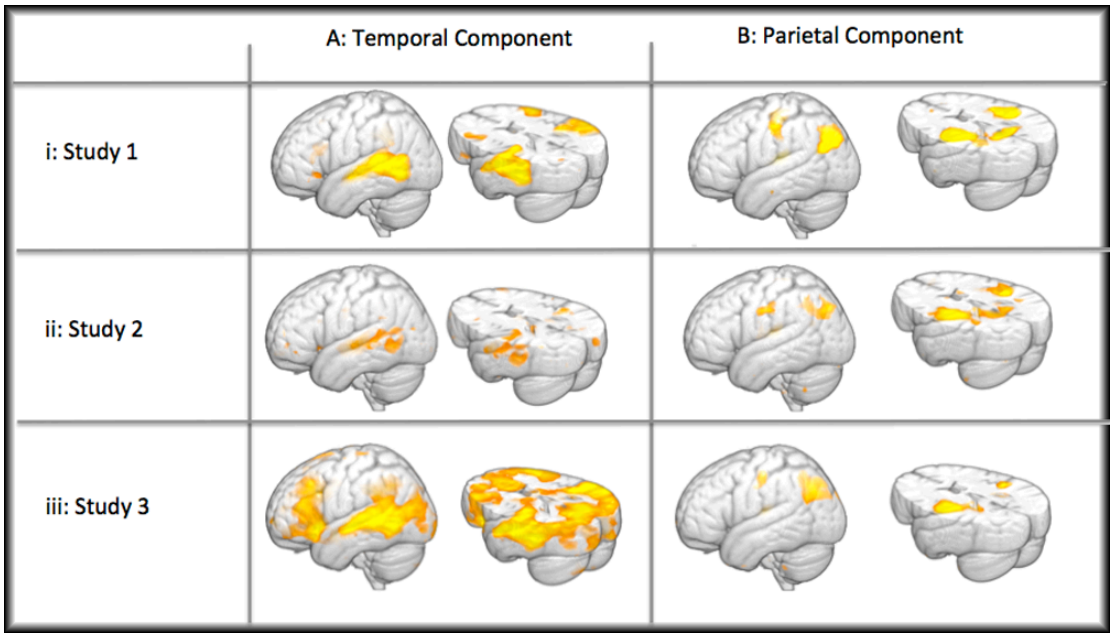



***Figure 6.6, Whole-brain connectivity maps for the two components for the three datasets***

*Whole-brain positive connectivity maps for the two components, temporal (A) and parietal (B) for the three datasets, i: Study 1, ii: Study 2, iii: Study 3. The top panel (Study 1) was thresholded for multiple comparisons and the lower panels (Studies 2 and 3) are shown to compare at a network level rather than at a voxel level. Images are presented on a rendered standard brain template (left hemisphere shown) and with the render clipped from the top to reveal activity in more medial regions.*

### 6.3.3 Study 3

Although the primary focus of the analysis was on regions involved in speech production, it is possible that the signals are present even in the absence of active speaking. To investigate this, the same functional connectivity analysis, with the same regions, was conducted on a third dataset of 16 participants scanned at 'rest'. The participants lay in the scanner with their eyes closed, without exposure to stimuli and with no explicit task. The resulting pattern of functional connectivity was, again, similar to that evoked by the speech production tasks (Figure 6.6iii): for the pSTG signal, bilateral temporal regions and inferior frontal regions; and for the vAPL signal, bilateral basal parietal and medial and lateral parietal regions. This suggests that the superior temporal and the parietal signals are present in some form, and have similar whole-brain functional connectivity, even in the absence of a speech task.

### 6.3.4 Comparison between studies

The results are reported at lenient statistical thresholds ($p < 0.01$ uncorrected) because of the focus on comparing the spatial pattern of functional connectivity, rather than assessing if a specific voxel is active or inactive. Furthermore, the choice of a statistical threshold is somewhat arbitrary given differences in imaging parameters and the differences in the sizes of the datasets. In this context, statistical thresholds correcting for multiple comparisons at the voxel or cluster level would be statistically invalid given the null hypothesis. To provide a more robust comparison of the similarity in the spatial pattern across the datasets, spatial correlation (Leech et al., 2012; Smith et al., 2009) showed that the functional connectivity maps acquired from all three datasets were highly similar (pSTG signal: all three $r$ values > 0.5, $p < 0.0001$; vAPL signal: all three $r$ signals > 0.6, $p < 0.0001$). Spatial correlation also showed that for all three datasets, the functional connectivity for pSTG and vAPL was significantly anti-correlated (all three $r$ values < -0.15, $p < 0.01$). As a further demonstration of the qualitative similarity of the whole-brain functional connectivity patterns, peaks of the original functional connectivity maps were located and the z-statistics across all three datasets plotted (see Figure 6.7 for the temporal (pSTG) component and Figure 6.8 for the parietal (vAPL) component). These figures again show both the qualitative similarity across the three datasets, and that in many regions the two components are anti-correlated.

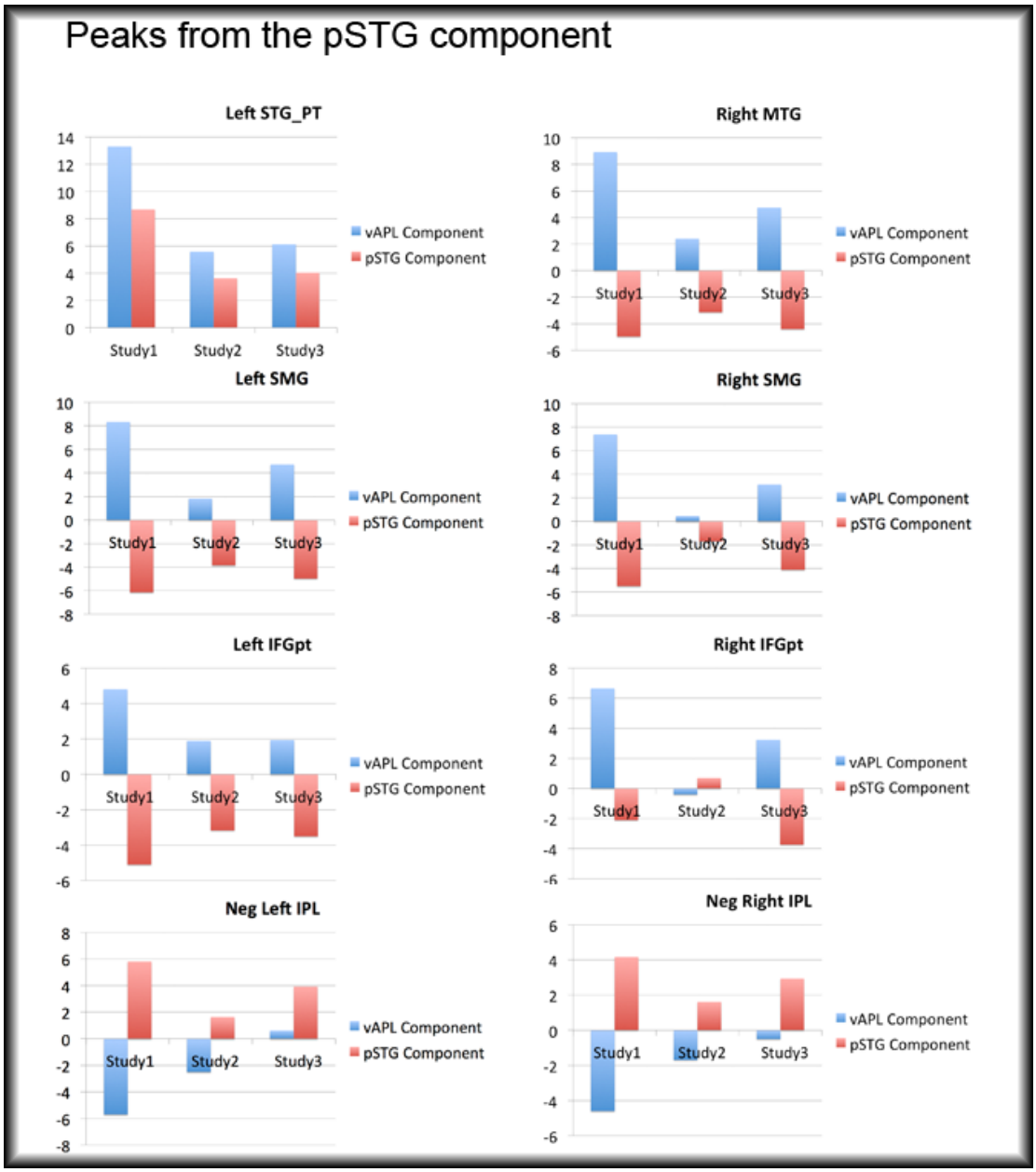



***Figure 6.7, Qualitative similarity of the whole-brain functional connectivity patterns for the temporal (pSTG) component***

*Peaks of the original functional connectivity maps were located and the z-statistics across all three datasets plotted. The plots demonstrate similarity across the three datasets and anti-correlation across the two components. Signal change coefficients are shown for the aVPL component in blue and for the pSTG component in red. From top to bottom and left to right, plots are shown for the left superior temporal gyrus / planum temporale (STG_PT), right middle temporal gyrus (MTG), left and right supramarginal gyri (SMG), left and right inferior frontal gyrus / pars triangularis (IFGpt) and left and right inferior parietal lobe (IPL).*

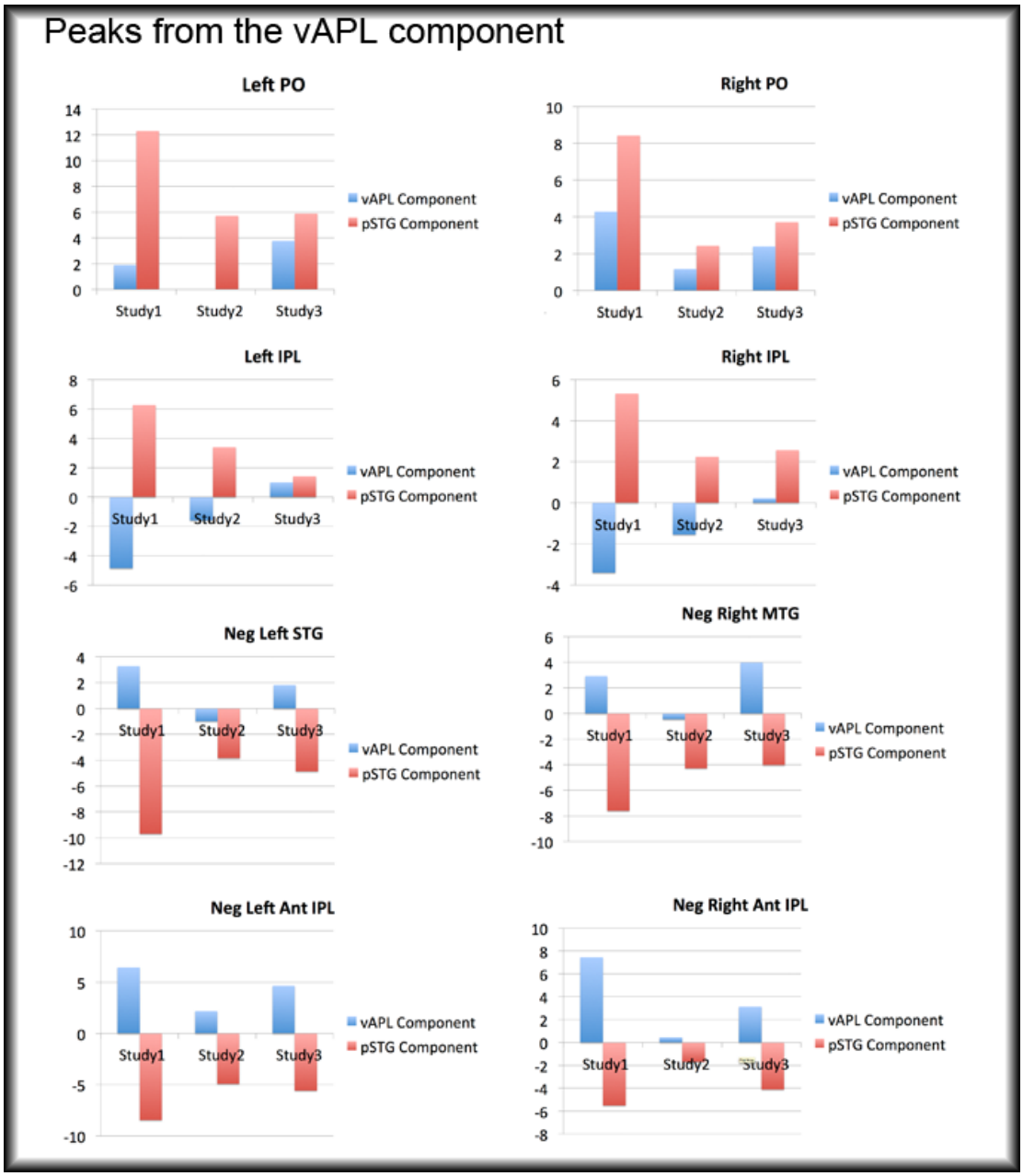



***Figure 6.8, Qualitative similarity of the whole-brain functional connectivity patterns for the parietal (vAPL) component***

*Peaks of the original functional connectivity maps were located and the z-statistics across all three datasets plotted. The plots demonstrate similarity across the three datasets and anti-correlation across the two components. Signal change coefficients are shown for the parietal component in blue and for the temporal component in red. From top to bottom and left to right, plots are shown for left and right parietal operculum (PO), left and right inferior parietal lobe (IPL), left superior temporal gyrus (STG), right middle temporal gyrus (MTG), and left and right anterior inferior parietal lobe (antIPL).*

### 6.3.5 Study 1: comparing across sessions

For Study 1, data from the second fMRI scanning Session (post-training) were also used for calculating whole-brain functional connectivity maps in response to the different speech tasks. A 4 x 2 (4 Languages x 2 Sessions) ANOVA was performed for each of the two components. There was a main effect of Language in the temporal component ($F = 6.633$, df (3,60) $p < 0.001$) but no Session effect and no interaction (Figure 6.9).

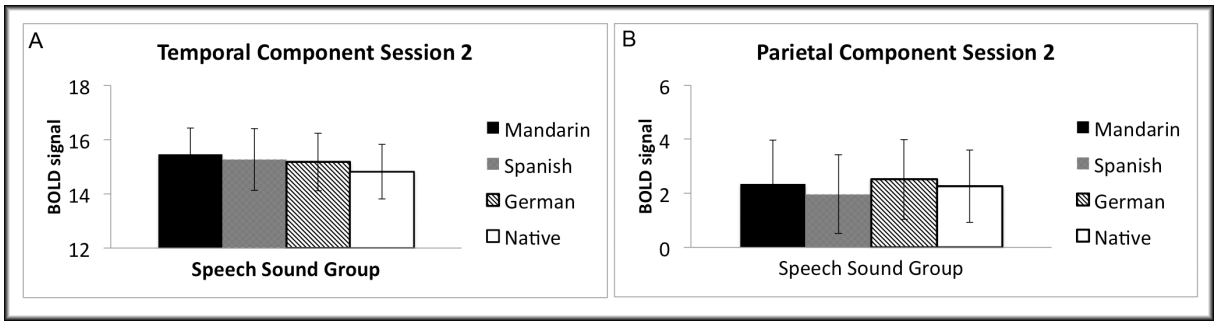



***Figure 6.9, ROI activity for Repeating all languages for both components in Session 2***

*A: Temporal Component and B: Parietal Component activity during Session 2 for the three non-native languages (Mandarin, Spanish and German) and for native. Error bars represent standard error of the mean.*

### 6.3.6 Study 1: comparing different levels of performance

When looking at trials in which participants were either above or below their median score for accuracy of speech production, in both Sessions of Study 1, there was a main effect of Language for the temporal component ($F = 8.380$ df (2,40), $p < 0.001$). The Non-native trials which were below the median for accuracy resulted in greater activity than Native in Session 1 (mean Non-native-below = 16.41, Native = 15.21, $t(20) = 4.807$, $p < 0.0001$, two-tailed, paired *t*-test) (Figure 6.10). There was no main effect of Session nor a Language by Session interaction in either component.

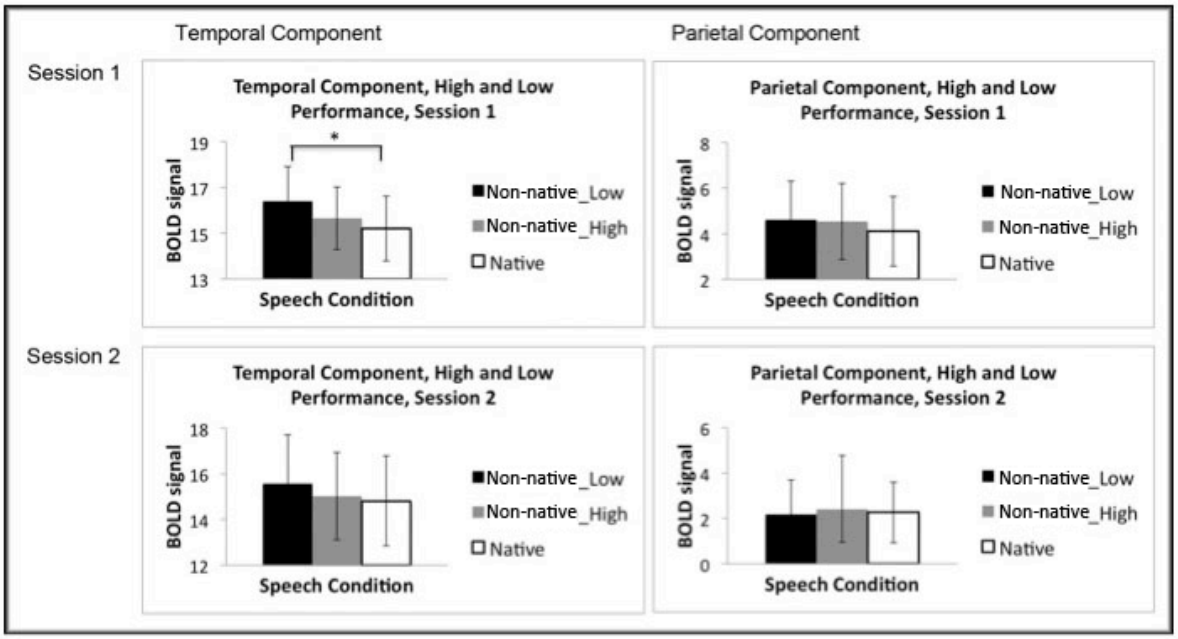



***Figure 6.10, Activity for different levels of performance for both components in both Sessions***

*Activity is shown for the Non-native sounds, above (Non-native_High) and below (Non-native_Low) the median of each participant's performance, as well as Native. Top panel shows results for Session 1 (A: temporal component, B: parietal component). Bottom panel shows results for Session 2 (C: temporal component, D: parietal component). Error bars represent the standard error of the mean. *Non-native_Low showed significantly greater activity than Native for the temporal component in Session 1 ($p < 0.0001$).*

### 6.3.7 Study 1: Listening

Activity was also measured for the Listening conditions of Study 1. For the temporal component, there was a main effect of Language ($F = 4.774$, df (1, 20), $p < 0.05$ and a Language by Session interaction ($F = 6.855$, df (1,20), $p < 0.05$) (Figure 6.11). A two-tailed paired t-test revealed that Native Listening resulted in significantly greater activity than Non-native in Session 1 (Native = 27.22, Non-native = 24.48, t (20) = -3.819 ($p < 0.001$) in Session 1). There were no main effects or interaction in the parietal component.

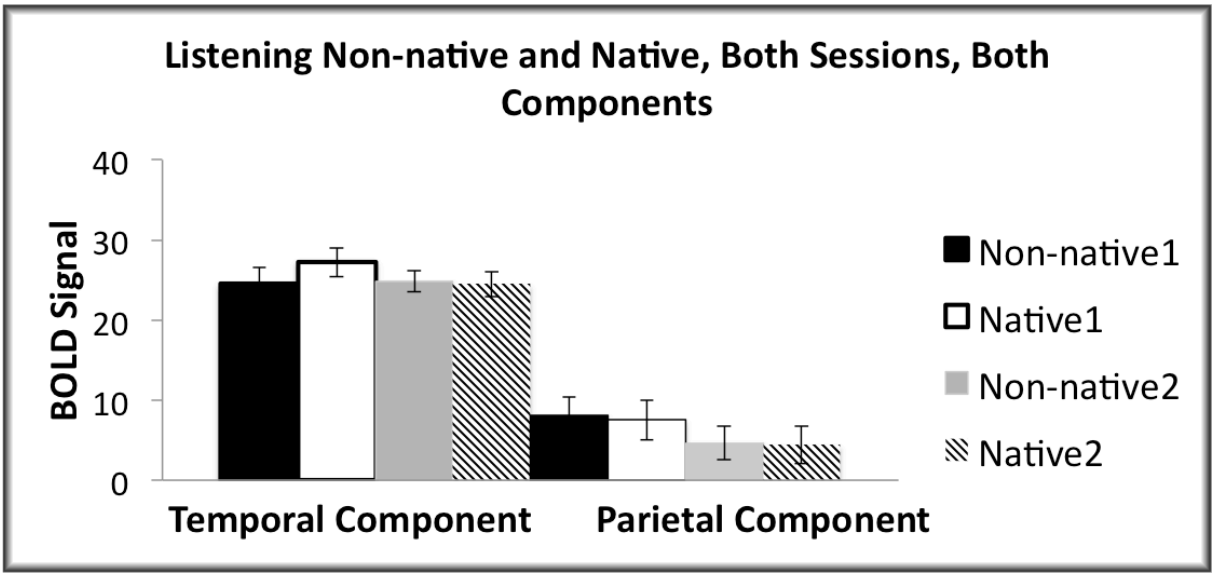



***Figure 6.11, Listening for both components in both Sessions***

*Activity for Listening in Session 1 is shown in black (Non-native) and white (Native) and in Session 2 in grey (Non-native) and with diagonal lines (Native). The temporal component is shown on the left and the parietal component on the right. Error bars represent standard error of the mean. *Native showed significantly greater activity than Non-native for the temporal component in Session 1 ($p < 0.001$).*

## 6.4 Discussion

These findings have a number of implications for understanding the neural systems underlying speech production and, more generally, for how the brain accomplishes other overlearned tasks. First, the results showed that, in contrast to most univariate analyses of perisylvian regions, there are two discrete but overlapping signals centred around the posterior end of the Sylvian fissure: one centred on the planum temporale and the other on the parietal operculum. Although the former region is usually considered to be auditory association cortex and the latter to be somatosensory, these regions are probably not strictly unimodal (Smiley et al., 2007). Second, these signals appear to be functionally different, evident from both the anti-correlated pattern of functional connectivity and the manner in which they were differentially modulated by speech tasks. Third, these signals were present in 'resting scan' data, with similar patterns of functional connectivity in the absence of a speech task. This suggests that speech involves modulating signals that are ongoing, rather than initiating activity in a set of brain regions necessary for speech. This last finding has broader implications for how the brain accomplishes complex tasks in general. I consider each of these findings in turn.

In this Chapter I demonstrated that the pSTG and vAPL signals overlap within left perisylvian cortex in a region encompassing area Spt, localised to the posterior half of the left planum temporale (Hickok et al., 2011). As discussed in previous Chapters, this region is involved in speech production and non-verbal vocal tract sounds (Hickok et al., 2011; Pa and Hickok, 2008), integrating the neural code of the auditory forms of words with the neural code for their motor (articulatory) forms. Most functional imaging studies have investigated a one-to-one mapping between this specific cortical region and auditory-motor transformation during speech production. In contrast, the studies presented here demonstrate that within area Spt there are at least two systems with quite different remote functional connections, both intra- and inter-hemispheric. This work is consistent with other studies indicating the role of multiple regions in the planum temporale, with dissociations of activity in response to auditory input and speech output (Tremblay et al., 2011). Similarly, these findings are consistent with other studies that have demonstrated the context-dependent response of perisylvian second-order somatosensory association cortex to overt

speech (Dhanjal et al., 2008; Simmonds et al., 2011a). The univariate analyses presented in Chapter 3 did not find activity in the planum temporale or the parietal operculum during Covert Speaking, a result also revealed by the multivariate analyses presented here. This again suggests that overt speech is required for revealing the local systems involved in the integration of speech motor plans with the sensory consequences of articulation.

The demonstration of the extensive perisylvian pSTG and vAPL networks is also consistent with the clinical literature on conduction aphasia (Buchsbaum et al., 2011); the syndrome in which patients with infarction of the left temporo-parietal junction make sound-based errors during speech production, repetition and reading aloud whilst retaining normal speech comprehension. The broad lesions resulting in conduction aphasia well match the extensive functional systems observed here.

As well as differences in distribution of connectivity, the two left posterior perisylvian signals displayed an anti-correlated spatial pattern, with the negative connectivity of each signal substantially overlapping the other signal's positive connectivity in a region that included area Spt. The existence of anti-correlated networks is emerging as an important concept in understanding the brain's broad functional architecture; for example, the anti-correlation between the default mode network and task-positive attentional networks (Fox et al., 2005; Leech et al., 2012; Smith et al., 2012). The datasets presented here demonstrate that the motor-sensory control of speech involves parallel competing signals converging on a large region of left temporo-parietal cortex.

One interpretation of this specific example of network anti-correlation is that the two sensory modalities alternate in importance during the generation of speech, depending on the type of utterance. For the single word production task (Study 1), all conditions, whether producing a non-native word or native non-word, led to increased activation for both signals, implying that both are involved in speech production. However, the specific requirements of the different types of non-native sounds resulted in different patterns of activation for the two signals, hinting at their different functional roles. Thus, producing the novel pitch shifts (sensed in the auditory domain) associated with the tonal words of Mandarin resulted in the greatest activity

in the network centred on the pSTG. The pSTG component also demonstrated a significant main effect of Language for Listening (Native > Non-native) and a Language by Session interaction (Native decreased after training, whereas Non-native did not change). In contrast, the generation of novel vowel sounds in German, with dependence on the subtle placement of the lips (sensed in the somatosensory domain), resulted in the greatest activity in the network centred on the vAPL. Precisely which aspects of the speech production tasks result in the different signal modulations must be speculative without further studies, but the message from the pattern of results is that the functional modulations of the two signals is likely to depend on the specific speech production demands at any one moment. The temporal component was the only one that showed any effect of behavioural proficiency, with significantly greater activity for Non-native below average repeating than Native; a difference that disappeared after training. Further, this result is consistent with recent behavioural work whereby in any one individual, preference for one or other sensory feedback while speaking may predominate (Lametti et al., 2012).

The functional dissociation evident from the anti-correlation of the signals was accompanied by the further dissociation on the basis of the manner in which they communicate with the rest of the brain. That the signal centred on the left pSTG had connectivity with bilateral superior temporal (auditory) cortex, and the one centred on the vAPL had connectivity with bilateral somatosensory cortex in the postcentral gyri, relates to the previous discussion of the sensory modality that may dominate depending on the specific utterance. However, there were additional components of both connectivity patterns that were located in high-order inferior frontal and parietal cortex. Their contribution to speech control, by modulating activity in temporo-parietal cortex, cannot be determined from this study, although the contribution of inferior frontal and dorsal inferior parietal cortex may plausibly relate to attention. The signal correlating with the vAPL region included posterior inferior parietal cortex and medial posterior cortex, including posterior cingulate cortex. These regions are typically thought to be part of the default mode network (Greicius and Menon, 2004). However, the parietal opercular signal reported here was significantly activated by speech production, unlike the deactivation expected if it formed part of the default mode network; and there was a separate signal in an expected spatial pattern for the

default mode network with relative deactivation during the speech production tasks (Figure 6.5). It would seem, therefore, that speech production involves communication with a local system within medial posterior cortex. This is consistent with the growing body of work suggesting the extensive functional heterogeneity of posterior cingulate cortex (Dastjerdi et al., 2011; Hagmann et al., 2008; Leech et al., 2012; Leech et al., 2011; Margulies et al., 2009; Seghier and Price, 2012).

The second study demonstrated that the vAPL signal was deactivated during sentential volitional speech, both overt and covert, but not during counting. This result is consistent with previous work reporting deactivation in secondary somatosensory cortex during propositional speech (Dhanjal et al., 2008). The finding of increased activity in vAPL for non-word and non-native single word speech production replicates previous research on non-native propositional speech production (Simmonds et al., 2011a). The superior temporal signal had a completely different pattern of modulation by task, whereby only overt speech resulted in increased activation. This suggests that activity in posterior perisylvian cortex is not independent of the speech task. Future studies incorporating speech tasks that variably manipulate dependency on feedback from sensory, linguistic and semantic systems will further refine our knowledge about the function of posterior perisylvian cortex.

By considering what happens at “rest”, it is evident that the two neural signals appear to be ongoing; they (or signals with a highly similar spatial pattern of functional connectivity) occur even in the absence of speech production. This result contrasts with the normally implicit, and sometimes explicit, conclusion that speech is a well-described neural system that changes from being in a rest state to becoming active when producing or comprehending speech. The present finding fits better with a view of speech production and comprehension as a special case of the normal ongoing functioning of the brain. Under this view, neural regions, such as auditory association cortex, have specific functional roles in, for example, decoding speech representations. However, understanding the highly complex process of speech involves integrating these specific speech functions with the other neural functions the brain performs. The networks, although relatively stable and observable across

tasks, are modulated differently by different task situations. It is this modulation that becomes apparent as increased activation evoked by a task.

The focus of these analyses has been on one region of the language network, albeit one that has received considerable emphasis in speech production over the last decade. Future studies may concentrate on other regions involved in speech, such as anterior perisylvian cortex, including Broca's area, which is discussed in more detail in the final Chapter of this thesis. In addition, these studies have only considered speech production and rest and not speech comprehension or non-speech tasks. It will be of interest to discover how the signals reported here are affected by other tasks. It is also possible that there are more signals within posterior perisylvian cortex than those found at the dimensionality analysed here. Although the same division of the region into pSTG and vAPL was also found at 7 or 15 components (Figure 6.1C), it is possible that a higher resolution could further fractionate this region revealing more discrete functional signals.

# 7 Thesis summary and future directions

## 7.1 Summary of results and main findings

As laid out in Chapter 1, there were four main aims to this thesis, and these were to investigate:

- the use of overt and covert speech production to examine motor-sensory control;
- the response of motor-sensory control systems to speaking a late-acquired proficient non-native language;
- the role of language proficiency on speech motor-sensory control;
- and, the distinct signals from posterior left perisylvian 'language' cortex, in response to different types of native and non-native speech production.

With regard to the first aim, Chapter 3 investigated common systems active for both overt and covert speech, as well as identifying functional components that are only involved in overt speech. Propositional and non-propositional speech tasks were included. Overt propositional speech revealed a pattern of activity that comprised the sum of activity from overt non-propositional speech (motor-sensory but not linguistic aspects) and covert propositional speech (linguistic but not motor-sensory aspects). Activity during both forms of propositional speech was observed in the pre-SMA, anterior cingulate cortex, left dorsolateral prefrontal cortex (including classic Broca's area), inferolateral temporal cortex and lateral cerebellar regions. Covert propositional speech did not result in activity in motor-sensory regions, but the same higher-order regions recruited for overt propositional speech were observed. This finding does not support claims from previous studies about the visualisation of the 'sensorimotor interface' integrating speech-related motor and sensory activity during covert speech. Instead, the results from this study suggest that overt speech is required for revealing the local systems integrating speech motor plans with the auditory and somatosensory consequences of articulation. This study identified the broad systems involved in overt speaking, and the role of these systems in non-native speech production was investigated in the subsequent Chapters of this thesis.

To address the second aim, Chapter 4 presented functional imaging data from highly proficient non-native speakers of English who had all acquired L2 after L1 had been established. Although there was a range of first languages, this did not affect results. Different native languages had no demonstrable influence on differences in articulating in L1 relative to the common L2 of English. Therefore, I did not find evidence that the motor-sensory consequences of speaking in L2 are dependent on L1, or are related to, for example, differences in stress patterns between languages. Speech (L1 + L2) contrasted with the rest baseline replicated the findings from Chapter 3 for overt propositional speech by revealing activity in bilateral primary somatosensory and auditory regions, the left frontal operculum and the cerebellum. Directly contrasting the production of L2 with that of L1 revealed premotor and sensory regions, including the left temporo-parietal junction (both the parietal operculum and the planum temporale), the basal ganglia and thalami, and midline vermal cerebellum. An increase in activity in anterior medial prefrontal cortex reflected the increased cognitive demands of producing L2. Region of interest analyses at the temporo-parietal junction confirmed and extended the impression from the whole-brain analyses. Producing L2 resulted in significantly greater activation than L1 in the posterior planum temporale, the parietal operculum and the frontal operculum.

To meet the third aim of this thesis, the training study presented in Chapter 5 was designed to extend the findings from the previous Chapter by investigating the role of proficiency in motor-sensory control of non-native speech. In both Sessions (pre- and post-training), non-native, relative to native, word repetition resulted in increased activity in widespread bilateral cortical and subcortical regions. The listening conditions revealed an unexpected result of fronto-parietal activity that was greater for native than non-native words, but only prior to training on non-native words. I attributed this activity to working memory and attentional processes. The results for Repeating showed a main effect of Language, Non-native > Native, throughout cortical and subcortical regions in both cerebral hemispheres. The main effect of Session was observed in the left and right striatum, and this was also the predominant anatomical location of the Language by Session interaction. Region of interest analysis confirmed significant Language by Session interactions for German and Mandarin but not for Spanish. The behavioural data demonstrated learning for all

languages but Spanish was rated the least accurate, both before and after training, indicating that striatal activity in vocal learning is modulated by proficiency. Further, in the scanning Session before training the activity within the left and right striatum declined across runs, indicating modulation by an early learning effect.

The final aim of this thesis was addressed by Chapter 6, which used three separate fMRI datasets to decompose the neural activity within the left posterior perisylvian speech region into discrete signals. Two overlapping signals were identified, one centred on the posterior superior temporal gyrus and the other on the adjacent ventral anterior parietal lobe. The connectivity of the temporal component was with bilateral superior temporal and inferior frontal regions, whereas that of the parietal component was with other parietal regions, lateral and medial. Surprisingly, the signals displayed robust spatial anti-correlation, in which the negative connectivity of each signal substantially overlapped the other signal's positive connectivity, suggesting that these two systems operate in close competition. Therefore, both components are involved in speech production, but their patterns of evoked activity dissociate in response to different demands, and are also evident when participants are not engaged in speech. These findings indicate that the neural architecture underlying speech production relies upon modulating ongoing, parallel, competing signals in distributed networks across bilateral temporal, parietal and inferior frontal cortices.

Previous studies of bilingualism have concentrated on differential demands on linguistic or cognitive processes between languages. The results presented here demonstrate that the motor-sensory demands on producing a language learned after the period of 'innate' language acquisition of infancy have an impact; that is, the consequences of speaking with an accent. Learning to speak a second language after the normal period of 'innate' language acquisition has functional consequences on cortex involved in auditory and somatosensory feedback control of articulation, as well as on subcortical structures within the basal ganglia. Figure 7.1 presents an outline of the cortical motor-sensory systems involved in speech production. In L1 these systems are highly tuned and efficient, but in a foreign language, the feedforward and feedback pathways are likely to be less efficient. This can be because of less efficient processing in feedforward motor pathways, from less

efficient sensory predictions, or from resulting inefficient sensory feedback, or a combination of all three.

This thesis has revealed networks involved in the motor-sensory learning of foreign speech, which include motor-sensory regions not previously implicated as important for bilingualism as well as inferior frontal, superior frontal, superior temporal and sub-cortical regions previously related to bilingual cognitive control and linguistic processing. It seems likely that the elevated activity in these motor-sensory areas relates to relative inefficient integration of feedforward and feedback signals that underlies a persisting accent when speaking in L2.

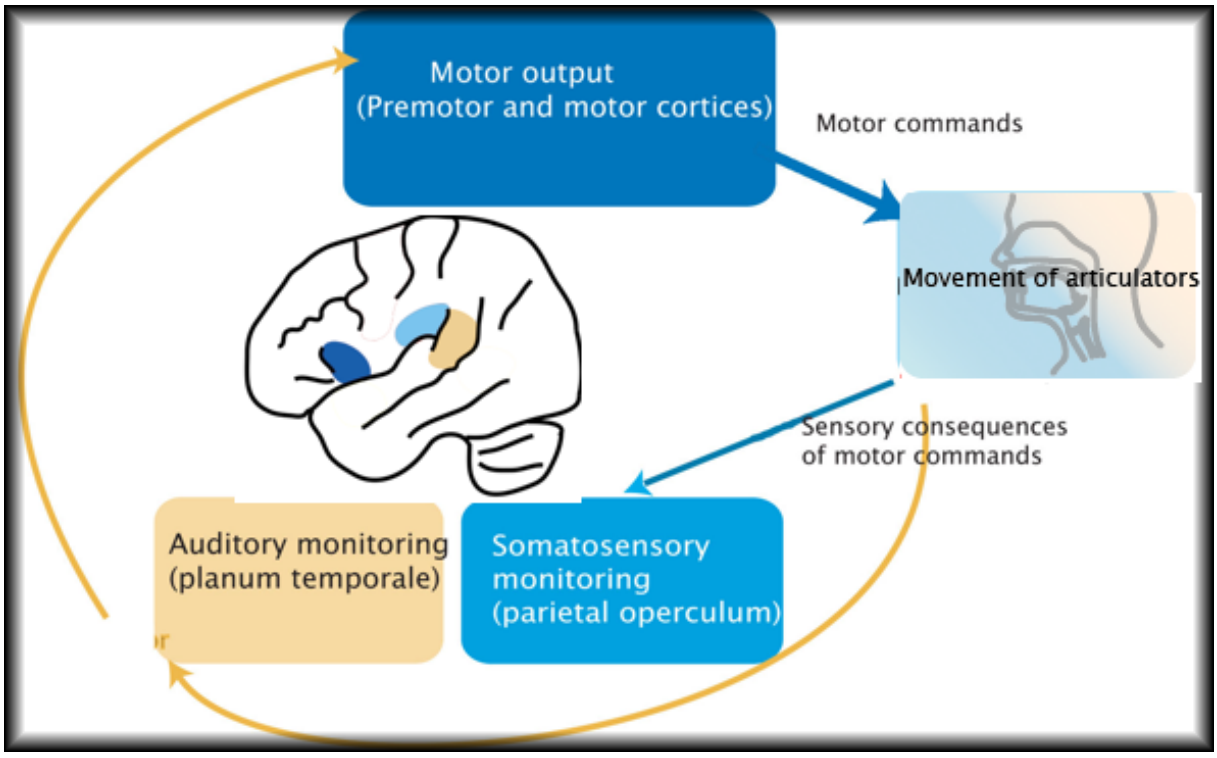



***Figure 7.1, A schematic diagram of cortical motor-sensory systems involved in speech production***

*The motor-sensory systems involved in speech production use feedforward motor commands (shown in dark blue) and feedback sensory monitoring, both somatosensory (shown in pale blue) and auditory (shown in beige), as well as motor-sensory control from subcortical structures within the basal ganglia (not shown here). In bilingualism, L1 and L2 are hypothesised to use the same motor-sensory control systems, although the studies presented in this thesis reveal different responses within these systems for L2, both novel and proficient. Figure adapted from Simmonds et al., (2011b).*

## 7.2 Future directions

The work presented here has raised a number of predictions for investigation in future studies. Firstly, the findings from the training study demonstrated the role of the striatum in human vocal learning. Previous studies investigating motor learning in rodents and vocal learning in songbirds, as well as human studies of motor learning, have proposed different involvement of specific striatal regions at various stages of learning. This could be examined in more detail by acquiring data at a higher resolution with a smaller voxel size. Higher resolution scanning would be advantageous for investigating striatal effects in more detail as larger voxels likely overlap the adjacent and relatively small structures, thereby increasing partial volume effects (Limbrick-Oldfield et al., 2012). Imaging in the coronal plane would also allow visualisation of midbrain dopaminergic regions, which project strongly to the striatum and influence striatal plasticity. An additional method to be included when investigating the striatum in more detail relates to physiological monitoring. For the data presented in Chapter 5, I filtered the data with time series from CSF and white matter to try and reduce the amount of noise in the data, but this could be greatly improved through the use of physiological monitoring, with electrocardiagram pads and a respiratory belt. Physiological noise can be removed retrospectively, using a method such as RETROICOR (retrospective image correction) (Glover et al., 2000; Limbrick-Oldfield et al., 2012) and this avoids the need for cardiac gating. This method models the likely effect of cardiac and respiratory cycles by specifying the time of data acquisition relative to each physiological phase (Brooks et al., 2008; Glover et al., 2000). By modelling these physiological regressors in the GLM, genuine effects are more likely to be detected, with less chance of accepting false positives (Harvey et al., 2008). Improving the signal using these methods may lead to identification of specific striatal structures involved in different stages of learning, by looking at rapid changes across runs as well as longer changes across Sessions.

Secondly, the work presented in this thesis investigating the motor-sensory control of speech production in normal participants could be extended into clinical studies of movement disorders due to degeneration in the basal ganglia. There are two opposing types of motor symptoms of basal ganglia disorders: hypokinetic, with difficulty in initiating movements, such as in Parkinson's disease (PD), and

hyperkinetic, making unwanted movements, such as in Huntington's disease (Graybiel, 1996). Motor learning has been investigated in both these patient groups (Heindel et al., 1988; Pisani et al., 2005) but vocal learning in basal ganglia disorders has not yet been studied. For example, in Parkinson's disease the learning of sequential finger movements is impaired, and vocal learning in PD patients may mirror this impairment, but this has not yet been investigated. The investigation of vocal learning in PD patients would provide insight into the specific striatal structures involved in different stages of learning, aided by the more detailed scanning techniques discussed above. Parkinson's disease has a posterior-to-anterior and between-hemisphere gradient of severity (Nandhagopal et al., 2009), and so by including patients with varying levels of severity, it would be possible to identify the specific contributions to learning of particular striatal regions. As there is decline of dopaminergic function in all striatal regions, in addition to the topography of severity, longitudinal studies of vocal learning in PD patients would offer further insight into the role of the striatum in vocal learning. The effect of administration of the Parkinson's treatment levodopa could also be studied in terms of both vocal learning proficiency and learning-related striatal plasticity. Greater understanding of the different contributions of specific striatal regions at separate stages of vocal learning will provide further insights into both the pedagogic and clinical aspects of motor learning discussed in the Introduction to this thesis.

Finally, the connectivity analysis presented in Chapter 6 warrants further investigation. Chapter 6 identified parallel competing systems in distributed networks that support speech production and showed dissociated patterns of evoked activity in response to different speech demands, being differentially modulated by the type and place of articulation. These systems originated from the temporo-parietal junction and it would be an interesting extension to investigate different systems originating from a more anterior region within perisylvian cortex and from the striatum. I have already begun this work by running similar analyses with an original mask from Broca's area (Brodmann Areas 44 and 45). From the ten components identified in the first step, five related to signal rather than noise (Figure 7.2). Figure 7.3 displays the whole-brain connectivity maps of these five signals, along with plots of responses to Native and Non-native Listening and Repeating, with significant differences summarised in the figure legend. Investigating the task responses of these components in more

detail, as well as using data from other Chapters presented in this thesis, will provide greater understanding of the function of these networks.

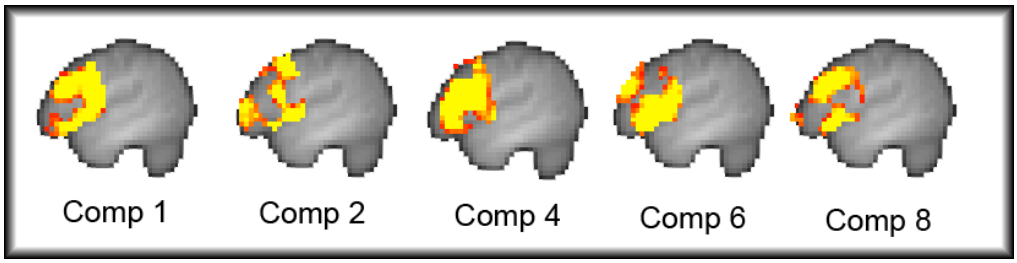



***Figure 7.2, Masks of the five components from the ICA, shown on sagittal slices of the left hemisphere***

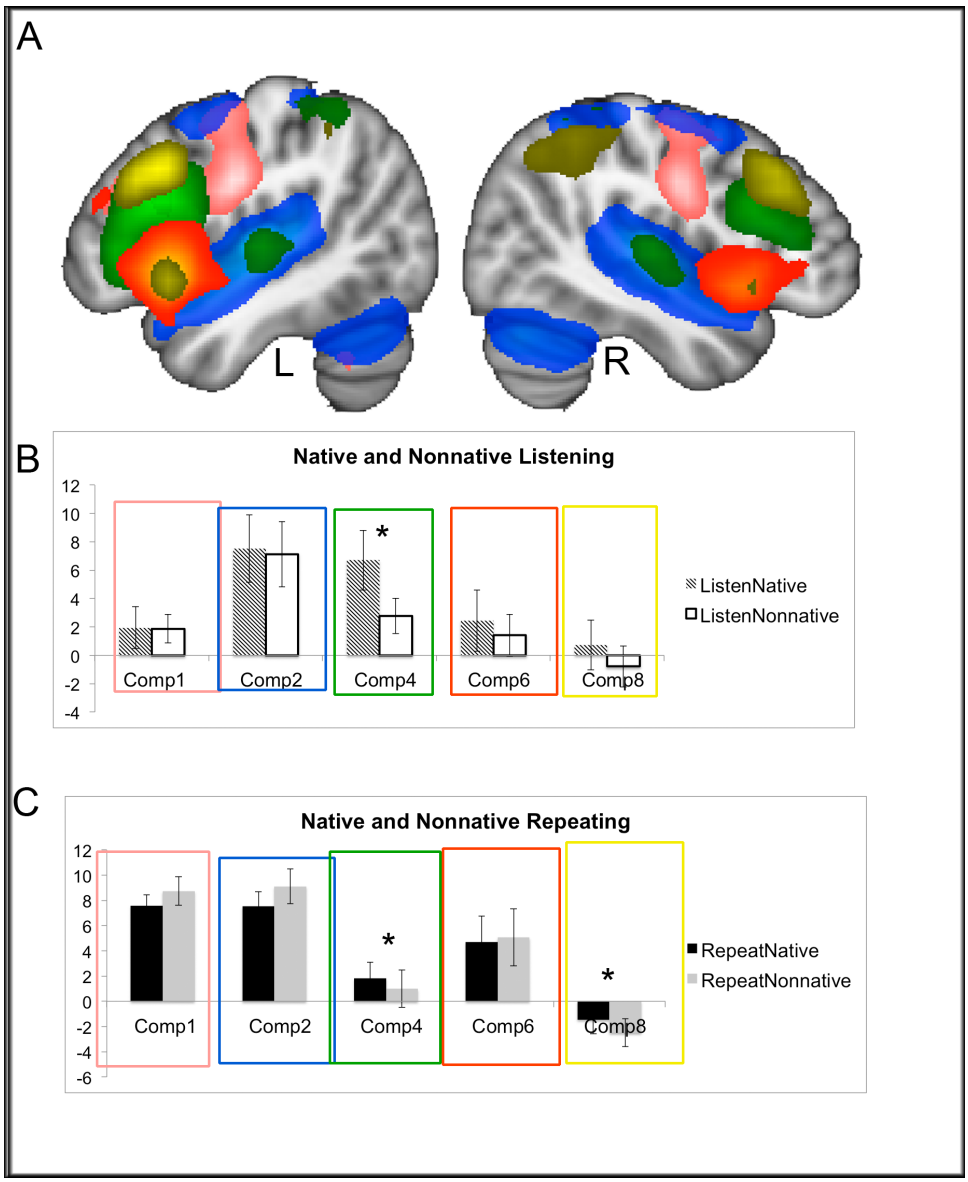



***Figure 7.3, Connectivity patterns and task responses for components from a Broca's area mask (legend overleaf)***

**Figure 7.3, Connectivity patterns and task responses for components from a Broca's area mask (Figure on previous page)**

*A: The connectivity patterns from the five components of interest are displayed on left and right sagittal slices (x = -46 and 46) of a standard brain template (MNI152). Component 1 is shown in pink, 2 in blue, 4 in green, 6 in red and 8 in yellow. B: Activity in each component for Listening (Native shown with diagonal lines and Non-native shown in white). Correcting for multiple comparisons, two-tailed paired t-tests revealed a significant difference for component 4 (Native > Non-native, $p < 0.0005$). C: Activity in each component for Repeating (Native shown in black, Non-native shown in grey). Correcting for multiple comparisons, two-tailed paired t-tests revealed significant differences for component 2 (Non-native > Native, $p < 0.00005$) and for component 4 (Native > Non-native, $p < 0.005$). Error bars represent 95% confidence intervals. * significant language differences.*

## 7.3 Conclusion

The main goal of my research was to investigate the motor-sensory learning of foreign speech, with the general hypothesis that speaking in a non-native language, relative to native, requires greater engagement of motor-sensory control systems. In addition it was predicted that proficiency in non-native speech production leads to reduction in activity in these cortical and subcortical regions. The work presented in this thesis confirms these hypotheses, and brings a new understanding to the persistence of a foreign accent in a second language learnt in adulthood. It is likely that the persistent accent is the result of a failure to achieve the same proficiency in integrating motor feedforward and sensory feedback control of articulation in a second language as in the first. The work presented here also demonstrates the involvement of the striatum in human vocal learning as an effect of early learning, occurring within minutes, and modulated by proficiency. Following the observation of the striatal effects of learning foreign speech, I therefore propose a future clinical study investigating vocal learning in patients with impaired striatal activity. Furthermore, the study presented in Chapter 6 indicates that speech production relies upon modulating ongoing, parallel, competing signals in distributed networks and suggests that further networks could be identified from connectivity patterns originating in premotor or striatal regions. The preliminary data presented in Chapter 7 seem to support this prediction and build on the findings from the studies presented here, providing a more extensive understanding of the motor-sensory learning of foreign speech.

# 8 References

Abutalebi, J., Annoni, J.-M., Zimine, I., Pegna, A.J., Seghier, M.L., Lee-Jahnke, H., Lazeyras, F., Cappa, S.F., and Khateb, A. (2008). Language control and lexical competition in bilinguals: an event-related fMRI study. Cereb Cortex *18*, 1496-1505.

Abutalebi, J., Cappa, S.F., and Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. Biling Lang Cogn *4*, 179-190.

Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. J Neurolinguist *20*, 242-275.

Abutalebi, J., and Green, D. (2008). Control mechanisms in bilingual language production: neural evidence from language switching studies. Lang Cognitive Proc *23*, 557-582.

Abutalebi, J., Rosa, P.A.D., Tettamanti, M., Green, D.W., and Cappa, S.F. (2009). Bilingual aphasia and language control: a follow-up fMRI and intrinsic connectivity study. Brain Lang *109*, 141-156.

Ackermann, H., and Hertrich, I. (2000). The contribution of the cerebellum to speech processing. J Neurolinguist *13*, 95-116.

Ackermann, H., and Riecker, A. (2010). The contribution(s) of the insula to speech communication: a review of the clinical and functional imaging literature. Brain Struct Funct *214*, 5-6.

Aglioti, S., Beltramello, A., Giradi, F., and Fabbro, F. (1996). Neurolinguistic and follow-up study of an unusual pattern of recovery from bilingual subcortical aphasia. Brain *119*, 1551-1564.

Ahlsen, E. (2006). Introduction to neurolinguistics (Amsterdam/Philadelphia, John Benjamins Publishing Company).

Alario, F.X., Goslin, J., Michel, V., and Laganaro, M. (2010). The functional origin of the foreign accent. Psychol Sci *21*, 15-20.

Aliu, S.O., Houde, J.F., and Nagarajan, S.S. (2008). Motor-induced suppression of the auditory cortex. J Cogn Neurosci *21*, 791-802.

Archibald, J. (1988). Second language phonology (Amsterdam/Philadelphia, John Benjamins Publishing Co.).

Archibald, J. (1998). Second language phonology, phonetics and typology. Stud Second Lang Acq *20*, 189-211.

Arslan, L.M., and Hansen, J.H.L. (1997). A study of temporal features and frequency characteristics in American English foreign accent. J Acoust Soc Am *102*, 28-40.

Avery, P., and Ehrlich, S. (1992). Teaching American English pronunciation (Oxford, England, Oxford University Press).

Awad, M., Warren, J.E., Scott, S.K., Turkheimer, F.E., and Wise, R.J.S. (2007). A common system for the comprehension and production of narrative speech. J Neurosci *27*, 11455-11464.

Baddeley, A.D. (1986). Working memory (Oxford, Oxford University Press).

Bamiou, D.-E., Musiek, F.E., and Luxon, L.M. (2003). The insula (Island of Reil) and its role in auditory processing. Brain Res Rev *42*, 143-154.

Barch, D.M., Sabb, F.W., Carter, C.S., Braver, T.S., Noll, D.C., and Cohen, J.D. (1999). Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. Neuroimage *10*, 642-657.

Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z., and Graybiel, A.M. (2005). Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. Nature *437*, 1158-1161.

Basho, S., Palmer, E.D., Rubio, M.A., Wulfeck, B., and Muller, R.A. (2007). Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech. Neuropsychologia *45*, 1697-1706.

Beckmann, C.F., DeLuca, M., Devlin, J.T., and Smith, S.M. (2005). Investigations into resting-state connectivity using independent component analysis. Philos Trans R Soc Lond B Biol Sci *360*, 1001-1013.

Beckmann, C.F., Jenkinson, M., and Smith, S.M. (2003). General multilevel linear modeling for group analysis in FMRI. Neuroimage *20*, 1052-1063.

Binder, J.R., Desai, R.H., Graves, W.W., and Conant, L.L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex *19*, 2767-2796.

Birn, R.M., Bandettini, P.A., Cox, R.W., Jesmanowicz, A., and Shaker, R. (1998). Magnetic field changes in the human brain due to swallowing or speaking. Magnet Reson Med *40*, 55-60.

Birn, R.M., Bandettini, P.A., Cox, R.W., and Shaker, R. (1999). Event-related fMRI of tasks involving brief motion. Hum Brain Mapp *7*, 106-114.

Bishop, D.V.M., North, T., and Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. J Child Psychol Psyc *37*, 391-403.

Blakemore, S.-J., Frith, C.D., and Wolpert, D.M. (2001). The cerebellum is involved in predicting the sensory consequences of action. Neuroreport *12*, 1879-1884.

Blank, S.C., Scott, S.K., Murphy, K., Warburton, E.A., and Wise, R.J.S. (2002). Speech production: Wernicke, Broca and beyond. Brain *125*, 1829-1838.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. Linguistic Analysis *20*, 3-49.

Boersma, P., and Weenink, D. (2007). Praat: doing phonetics by computer (version 4 D.5.25) [Software] (Available from http://www.praat.org [2007]).

Bohland, J.W., and Guenther, F.H. (2006). An fMRI investigation of syllable sequence production. Neuroimage *32*, 821-841.

Bolhuis, J.J., Okanoya, K., and Scharff, C. (2010). Twitter evolution: converging mechanisms in birdsong and human speech. Nat Rev Neurosci *11*, 747-759.

Bonnelle, V., Leech, R., Kinnunen, K.M., Ham, T.E., Beckmann, C.F., De Boissezon, X., Greenwood, R.J., and Sharp, D.J. (2011). Default mode network connectivity predicts sustained attention deficits after traumatic brain injury. J Neurosci *31*, 13442-13451.

Booth, J.R., Wood, L., Lu, D., Houk, J.C., and Bitan, T. (2007). The role of the basal ganglia and the cerebellum in language processing. Brain Res *1133*, 136-144.

Briellman, R.S., Saling, M.M., Connell, A.B., Waites, A.B., Abbott, D.F., and Jackson, G.D. (2004). A high-field functional MRI study of quadri-lingual subjects. Brain Lang *89*, 531-542.

Brooks, J.C.W., Beckmann, C.F., Miller, K.L., Wise, R.G., Porro, C.A., Tracey, I., and Jenkinson, M. (2008). Physiological noise modelling for spinal functional magnetic resonance imaging studies. Neuroimage *39*, 680-692.

Brown, S., Ingham, R.J., Ingham, J.C., Laird, A.R., and Fox, P.T. (2005). Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. Hum Brain Mapp *25*, 105-117.

Buchsbaum, B.R., Baldo, J., Okada, K., Berman, K.F., Dronkers, N., D'Esposito, M., and Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - An aggregate analysis of lesion and fMRI data. Brain Lang *119*, 119-128.

Buchsbaum, B.R., and D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. J Cogn Neurosci *20*, 762-778.

Buchsbaum, B.R., Olsen, R.K., Kock, P., and Berman, K.F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron *48*, 687-697.

Buckner, R.L., Andrews-Hanna, J.R., and Schacter, D.L. (2008). The brain's default network. Ann N Y Acad Sci *1124*, 1-38.

Bygate, M. (1987). Speaking (Oxford, Oxford University Press).

Callan, D.E., Jones, J.A., Callan, A.M., and Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory- and auditory/orosensory internal models. Neuroimage *22*, 1182-1194.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science *280*, 747-749.

Chee, M.W., Tan, E.W.L., and Thiel, T. (1999). Mandarin and English single word processing studied with functional magentic resonance imaging. J Neurosci *19*, 3050-3056.

Chee, M.W.L., Hon, N., Ling Lee, H., and Siong Soon, C. (2001). Relative language proficiency modulates BOLD signal change when bilinguals perform semantic judgments. Neuroimage *13*, 1155-1163.

Chee, M.W.L., Soon, C.S., and Lee, H.L. (2003). Common and segregated neuronal networks for different languages revealed using functional magnetic resonance adaptation. J Cogn Neurosci *15*, 85-97.

Chen, J.L., Penhune, V.B., and Zatorre, R.J. (2009). The role of auditory and premotor cortex in sensorimotor transformations. Ann N Y Acad Sci *1169*, 15-34.

Clarke, C.M., and Garrett, M.F. (2004). Rapid adaptation to foreign-accented English. J Acoust Soc Am *116*, 3647-3658.

Costa, A., and Santesteban, M. (2004). Lexical access in bilingual speech production: evidence from language switching in highly proficient bilinguals and L2 learners. J Mem Lang *50*, 491-511.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J.T., Aso, T., Urayama, S., Fukuyama, H., Stockton, K.*, et al.* (2006). Language control in the bilingual brain. Science *312*, 1537-1540.

Crosson, B., Sadek, J.R., Bobholz, J.A., Gökçay, D., Mohr, C.M., Leonard, C.M., Maron, L., Auerbach, E.J., Browd, S.R., Freeman, A.J.*, et al.* (1999). Activity in the paracingulate and cingulate sulci during word generation: an fMRI study of functional anatomy. Cereb Cortex *9*, 307-316.

Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. Neuroimage *9*, 179-194.

Danesi, M. (1994). The neuroscientific perspective in second language acquisition research: a critical synopsis. IRAL *32*, 201-229.

Dastjerdi, M., Foster, B.L., Nasrullah, S., Rauschecker, A.M., Dougherty, R.F., Townsend, J.D., Chang, C., Greicius, M.D., Menon, V., Kennedy, D.P.*, et al.* (2011). Differential electrophysiological response during rest, self-referential, and non-self-referential tasks in human posteromedial cortex. Proc Natl Acad Sci U S A *108*, 3023-3028.

Dauer, R.M. (2005). The lingua franca core: a new model for pronunciation instruction? TESOL Quart *39*, 543-550.

De Jong, N.H., and Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. Behav Res Methods *41*, 385-390.

Derwing, T.M. (2003). What do ESL students say about their accents? Can Mod Lang Rev *59*, 547-566.

Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E., Matthews, P.M., and Tyler, L.K. (2000). Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. Neuroimage *11*, 589-600.

Dhanjal, N.S., Handunnetthi, L., Patel, M.C., and Wise, R.J.S. (2008). Perceptual systems controlling speech production. J Neurosci *28*, 9969-9975.

Doupe, A.J., and Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. Annu Rev Neurosci *22*, 567-631.

Doya, K., and Sejnowski, T.J. (1995). A novel reinforcement model of birdsong vocalization learning. In Advances in neural information processing systems 7, G. Tesauro, D.S. Touretzky, and T.K. Leen, eds. (Massachusetts Institute of Technology).

Dulay, H., Burt, M., and Krashen, S.D. (1982). Language two (New York, Oxford University Press).

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? J Exp Psychol Hum Percept Perform *25*, 1568-1578.

Eda-Fujiwara, H., Imagawa, T., Matsushita, M., Matsuda, Y., Takeuchi, H.-A., Satoh, R., Watanabe, A., Zandbergen, M.A., Manabe, K., Kawashima, T.*, et al.* (2012). Localized brain activation related to the strength of auditory learning in a parrot. PLoS ONE *7*, e38803.

Eickhoff, S.B., Schleicher, A., Zilles, K., and Amunts, K. (2006). The human parietal operculum. I. Cytoarchitectonic mapping of subdivisions. Cereb Cortex *16*, 254-267.

Eilers, P.H.C., and Goeman, J.J. (2004). Enhancing scatterplots with smoothed densities. Bioinformatics *20*, 623-628.

Eliades, S.J., and Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature *453*, 1102-1107.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterick, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S.*, et al.* (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron *33*, 341-355.

Fisher, S.E., and Scharff, C. (2009). *FOXP2* as a molecular window into speech and language. Trends Genet *25*, 166-177.

Flege, J.E. (1981). The phonological basis of foreign accent. TESOL Quart *15*, 443-455.

Flege, J.E. (1995). Second language speech learning: theory, findings, and problems. In Speech perception and linguistic experience: issues in cross-language research, W. Strange, ed. (Timonium, MD, York Press), pp. 233-277.

Flege, J.E., and Hillenbrand, J. (1984). Differential use of temporal cues to the [s-z] contrast by native and non-native speakers of English. J Acoust Soc Am *79*, 508-517.

Flynn, S., and Manuel, S. (1991). Age-dependent effects in language acquisition: An evaluation of critical period hypotheses. In Point counterpoint: Universal Grammar in the second language, L. Eubank, ed. (Amsterdam, Benjamins), pp. 117-145.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Essen, D.C.V., and Raichle, M.E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A *102*, 9673-9678.

Fox, P.T., and Raichle, M.E. (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci U S A *83*, 1140-1144.

Francis, W.S. (1999). Cognitive integration of language and memory in bilinguals: semantic representation. Psychol Bull *125*, 193-222.

Frenck-Mestre, C., Anton, J.L., Roth, M., Vaid, J., and Viallet, F. (2005). Articulation in early and late bilinguals' two languages: evidence from functional magnetic resonance imaging. Neuroreport *16*, 761-765.

Frith, C., Friston, K.J., Liddle, P.F., and Frackowiak, R.S.J. (1991). A PET study of word finding. Neuropsychologia *29*, 1137-1148.

Galaburda, A., and Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. J Comp Neurol *190*, 597-610.

Gathercole, S.E. (2006). Nonword repetition and word learning: the nature of the relationship. Appl Psycholinguist *27*, 513-543.

Gathercole, S.E., and Baddeley, A.D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? J Mem Lang *29*, 336-360.

Ghosh, S.S., Tourville, J.A., and Guenther, F.H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. J Speech Lang Hear R *51*, 1183-1202.

Glover, G.H., Li, T.-Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med *44*, 162-167.

Golestani, N., Alario, F.-X., Meriaux, S., Le Bihan, D., Dehaene, S., and Pallier, C. (2006). Syntax production in bilinguals. Neuropsychologia *44*, 1029-1040.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., and Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. Cereb Cortex *17*, 575-582.

Golfinopoulos, E., Tourville, J.A., and Guenther, F.H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage *52*, 862-874.

Gracco, V.L., and Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw and laryngeal movements. J Neurosci *14*, 6585-6597.

Gracco, V.L., Tremblay, P., and Pike, B. (2005). Imaging speech production using fMRI. Neuroimage *26*, 294-301.

Graybiel, A. (1996). Basal ganglia: new therapeutic approaches to Parkinson's disease. Curr Biol *6*, 368-371.

Graybiel, A. (2000). The basal ganglia. Curr Biol *10*, R509-R511.

Graybiel, A. (2005). The basal ganglia: learning new tricks and loving it. Curr Opin Neurobiol *15*, 638-644.

Green, D.W. (2003). Neural basis of lexicon and grammar in L2 acquistion. In The lexicon-syntax interface in second language acquisition, R. van Hout, A. Hulk, F. Kuiken, and R. Towell, eds. (Amsterdam, John Benjamins), pp. 197-218.

Greicius, M.D., and Menon, V. (2004). Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation. J Cogn Neurosci *16*, 1484-1492.

Guenther, F.H. (2006). Cortical interactions underlying the production of speech sounds. J Commun Disord *39*, 350-365.

Guenther, F.H., Ghosh, S.S., and Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang *96*, 280-301.

Haesler, S., Rochefort, C., Georgi, B., Licznerski, P., Osten, P., and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of *FoxP2* in songbird basal ganglia nucleus Area X. PLoS Biol *5*, e321.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. PLoS Biol *6*, e159.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summerfield, A.Q., Elliott, M.R., Gurney, E.M., and Bowtell, R.W. (1999). "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp *7*, 213-223.

Harvey, A.K., Pattinson, K.T.S., Brooks, J.C.W., Mayhew, S.D., Jenkinson, M., and Wise, R.G. (2008). Brainstem functional magnetic resonance imaging: disentangling signal from physiological noise. J Magn Reson Imaging *28*, 1337-1344.

Heindel, W.C., Butters, N., and Salmon, D.P. (1988). Impaired learning of a motor skill in patients with Huntington's disease. Behav Neurosci *102*, 141-147.

Hernandez, A.E., Martinez, A., and Kohnert, K. (2000). In search of the language switch: an fMRI study of picture naming in Spanish-English bilinguals. Brain Lang *73*, 421-431.

Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci *15*, 673-682.

Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. Neuron *69*, 407-422.

Hickok, G., Okada, K., and Serences, J.T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol *101*, 2725-2732.

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. Trends Cogn Sci *4*, 1463-1467.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. Nat Rev Neurosci *8*, 393-402.

Hill, J. (1970). Foreign accents, language acquisition, and cerebral dominance revisited. Lang Learn *20*, 237-248.

Houde, J.F., and Jordan, M.I. (1998). Sensorimotor adaptation in speech production. Science *279*, 1213-1216.

Howe, M.W., Atallah, H.E., McCool, A., Gibson, D.J., and Graybiel, A.M. (2011). Habit learning is associated with major shifts in frequencies of oscillatory activity and synchronized spike firing in striatum. Proc Natl Acad Sci U S A *108*, 16801-16806.

Huang, J., Carr, T.H., and Cao, Y. (2001). Comparing cortical activations for silent and overt speech using event-related fMRI. Hum Brain Mapp *15*, 39-53.

Illes, J., Francis, W.S., Desmond, J.E., Gabrieli, J.D.E., Glover, G.H., Poldrack, R., Lee, C.J., and Wagner, A.D. (1999). Convergent cortical representation of semantic processing in bilinguals. Brain Lang *70*, 347-363.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: which suggested differences can we trust and what do they mean? Lang Learn *56*, 279-304.

Indefrey, P., and Levelt, W.J.M. (2004). The spatial and temporal signatures of word production components. Cognition *92*, 101-144.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition *87*, B47-B57.

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. J Neurosci *23*, 9541-9546.

Jarvis, E.D. (2004). Learned birdsong and the neurobiology of human language. Ann N Y Acad Sci *1016*, 749-777.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage *17*, 825-841.

Jernigan, T.L., Gamst, A.C., Fennema-Notestine, C., and Ostergaard, A.L. (2003). More "mapping" in brain mapping: statistical comparison of effects. Hum Brain Mapp *19*, 90-95.

Jones, J.A., and Munhall, K.G. (2005). Remapping auditory-motor representations in voice production. Curr Biol *15*, 1768-1772.

Jubault, T., Ody, C., and Koechlin, E. (2007). Serial organization of human behavior in the inferior parietal cortex. J Neurosci *27*, 11028-11036.

Jueptner, M., and Weiller, C. (1998). A review of differences between basal ganglia and cerebellar control of movements as revealed by functional imaging studies. Brain *121*, 1437-1449.

Kao, M.H., Doupe, A.J., and Brainard, M.S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature *433*, 638-643.

Kawashima, R., Okuda, J., Umetsu, A., Sugiura, M., Inoue, K., Suzuki, K., Tabuchi, M., Tsukiura, T., Narayan, S.L., Nagasaka, T.*, et al.* (2000). Human cerebellum plays an important role in memory-timed finger movement: an fMRI study. J Neurophysiol *83*, 1079-1087.

Kemeny, S., Ye, F.Q., Birn, R., and Braun, A.R. (2005). Comparison of continuous overt speech fMRI using BOLD and arterial spin labeling. Hum Brain Mapp *24*, 173-183.

Kim, K.H.S., Relkin, N.R., Lee, K.-M., and Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. Nature *388*, 171-174.

Klein, D., Milner, B., Zatorre, R.J., Meyer, E., and Evans, A.C. (1995). The neural substrates underlying word generation: a bilingual functional-imaging study. Proc Natl Acad Sci USA *92*, 2899-2903.

Klein, D., Milner, B., Zatorre, R.J., Zhao, V., and Nikelski, J. (1999). Cerebral organization in bilinguals: a PET study of Chinese-English verb generation. Neuroreport *10*, 2841-2846.

Klein, D., Watkins, K.E., Zatorre, R.J., and Milner, B. (2006). Word and nonword repetition in bilingual subjects: a PET study. Hum Brain Mapp *27*, 153-161.

Klein, D., Zatorre, R.J., Milner, B., Meyer, E., and Evans, A.C. (1994). Left putaminal activation when speaking a second language: evidence from PET. Neuroreport *5*, 2295-2297.

Koechlin, E., and Jubault, T. (2006). Broca's area and the hierarchical organization of human behaviour. Neuron *50*, 963-974.

Kovacs, A.M., and Mehler, J. (2009). Flexible learning of multiple speech structures in bilingual infants. Science *325*, 611-612.

Kovelman, I., Baker, S.A., and Petitto, L.-A. (2008). Bilingual and monolingual brains compared: a functional magnetic resonance imaging investigation of syntactic processing and a possible "neural signature of bilingualism". J Cogn Neurosci *20*, 153-169.

Krashen, S.D. (1982). Principles and practice in second language acquisition (Oxford, Pergamon Press).

Krashen, S.D., and Terrell, T.D. (1983). The natural approach: language acquisition in the classroom (London, Prentice Hall Europe).

Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. Nat Rev Neurosci *5*, 831-843.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos Trans R Soc Lond B Biol Sci *363*, 979-1000.

Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519-523.

Lamendella, J.T. (1977). General principles of neurofunctional organization and their manifestations in primary and non-primary langauge acquisition. Lang Learn *27*, 155-196.

Lametti, D.R., Nasir, S.M., and Ostry, D.J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. J Neurosci *32*, 9351-9358.

Leech, R., Braga, R., and Sharp, D.J. (2012). Echoes of the brain within the posterior cingulate cortex. J Neurosci *32*, 215-222.

Leech, R., Kamourieh, S., Beckmann, C.F., and Sharp, D.J. (2011). Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. J Neurosci *31*, 3217-3224.

Lehéricy, S., Ducros, M., Van de Moortele, P.-F., François, C., Thivard, L., Poupon, C., Swindale, N., Ugurbil, K., and Kim, D.-S. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. Ann Neurol *55*, 522-529.

Lenneberg, E. (1967). The biological foundations of language (New York, Wiley).

Leonardo, A. (2004). Experimental test of the birdsong error-correction model. Proc Natl Acad Sci U S A *101*, 16935-16940.

Levelt, W.J.M. (1989). Speaking: from intention to articulation (Cambridge, MA, MIT Press).

Levelt, W.J.M., Roelofs, A., and Meyer, A.S. (1999). A theory of lexical access in speech production. Behav Brain Sci *22*, 1-75.

Limbrick-Oldfield, E.H., Brooks, J.C.W., Wise, R.J.S., Padormo, F., Hajnal, J.V., Beckmann, C.F., and Ungless, M.A. (2012). Identification and characterisation of midbrain nuclei using optimised functional magnetic resonance imaging. Neuroimage *59*, 1230-1238.

Long, M.H. (1990). Maturational constraints on language development. Stud Second Lang Acq *12*, 251-285.

Loucks, T.M.J., Poletto, C.J., Simonyan, K., Reynolds, C.L., and Ludlow, C.L. (2007). Human brain activation during phonation and exhalation: common volitional control for two upper airway functions. Neuroimage *36*, 131-143.

Ludlow, C.L., and Cikaoja, D.B. (1998). Is there a self-monitoring speech perception system? J Commun Disord *31*, 505-510.

Lurito, J.T., Karekan, D.A., Lowe, M., Chen, S.H.A., and Mathews, V.P. (2000). Comparison of rhyming and word generation with fMRI. Hum Brain Mapp *10*, 99-106.

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., Vandewalle, G., Collette, F., and Maquet, P. (2006). The left intraparietal sulcus and verbal short-term memory: Focus of attention or serial order? Neuroimage *32*, 880-891.

Margulies, D.S., Vincent, J.L., Kelly, C., Lohmann, G., Uddin, L.Q., Biswal, B.B., Villringer, A., Castellanos, F.X., Milham, M.P., and Petrides, M. (2009). Precuneus

shares intrinsic functional architecture in humans and monkeys. Proc Natl Acad Sci U S A.

McArthur, T. (2002). The Oxford guide to world English (Oxford, England, Oxford University Press).

Mendez-Balbuena, I., Manjarrez, E., Schulte-Mönting, Huethe, F., Tapia, J.A., Hepp-Reymond, M.-C., and Kristeva, R. (2012). Improved sensorimotor performance via stochastic resonance. J Neurosci *32*, 12612-12618.

Miyachi, S., Hikosaka, O., Miyashita, K., Kárádi, Z., and Rand, M.K. (1997). Differential roles of monkey striatum in learning of sequential hand movement. Exp Brain Res *115*, 1-5.

Moser, D., Fridriksson, J., Bonhilha, L., Healy, E.W., Baylis, G., Baker, J.M., and Rorden, C. (2009). Neural recruitment for the production of native and novel speech sounds. Neuroimage *46*, 549-557.

Möttönen, R., and Watkins, K.E. (2011). Using TMS to study the role of the articulatory motor system in speech perception. Aphasiology *26*, 1103-1118.

Muñoz, C., and Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. Lang Teach *44*, 1-35.

Nandhagopal, R., Kuramoto, L., Schulzer, M., Mak, E., Cragg, J., Lee, C.S., McKenzie, J., McCormick, S., Samii, A., Troiano, A.*, et al.* (2009). Longitudinal progression of sporadic Parkinson's disease: a multi-tracer positron emission tomography study. Brain *132*, 2970-2979.

Nasir, S.M., and Ostry, D.J. (2006). Somatosensory precision in speech production. Curr Biol *16*, 1918-1923.

Nasir, S.M., and Ostry, D.J. (2009). Auditory plasticity and speech motor learning. Proc Natl Acad Sci U S A *106*, 20470-20475.

Neufeld, G. (1979). Towards a theory of language learning ability. Lang Learn *29*, 227-241.

Neville, H.J., and Bavelier, D. (2001). Specificity of developmental neuroplasticity in humans: evidence from sensory deprivation and altered language experience. In Toward a theory of neuroplasticity, C.A. Shaw, and J.C. McEachern, eds. (Hove, Taylor & Francis), pp. 261-274.

Nichols, T., and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. Stat Methods Med Res *12*, 419-446.

Nieto-Castanon, A., Guenther, F.H., Perkell, J.S., and Curtin, H.D. (2005). A modeling investigation of articulatory variability and acoustic stability during American English /r/ production. J Acoust Soc Am *117*, 3196-3212.

Nottebohm, F. (2002). Why are some neurons replaced in adult brain? J Neurosci *22*, 624-628.

Nottebohm, F. (2005). The neural basis of birdsong. PLoS Biol *3*, e164.

Ölveczky, B.P., Andalman, A.S., and Fee, M.S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol *3*, e153.

Ono, M., Kubik, S., and Abernathey, C.D. (1990). Atlas of the cerebral sulci (Stuttgart; New York, Georg Thieme Verlag; Thieme Medical Publishers, Inc.).

Pa, J., and Hickok, G. (2008). A parietal-temporal sensory-motor integration area for the human vocal tract: evidence from an fMRI study of skilled musicians. Neuropsychologia *46*, 362-368.

Palmer, E.D., Rosen, H.J., Ojemann, J.G., Buckner, R.L., Kelley, W.M., and Petersen, S.E. (2001). An event-related fMRI study of overt and covert word stem completion. Neuroimage *14*, 182-193.

Pandya, D.N., and Sanides, F. (1973). Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern. Anat Embryol (Berl) *139*, 127-161.

Paradis, M. (1983). Readings on aphasia in bilinguals and polyglots (Montreal, Marcel Didier).

Parker Jones, Ō., Green, D.W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., Lee, H.L., Ramsden, S., Gazarian, K., Prejawa, S.*, et al.* (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. Cereb Cortex *22*, 892-902.

Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J., and Evans, A.C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. Eur J Neurosci *8*, 2236-2246.

Penfield, W., and Roberts, L. (1959). Speech and brain mechanisms (Princeton, New Jersey, Princeton University Press).

Penhune, V.B., Zatorre, R.J., MacDonald, J.D., and Evans, A.C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. Cereb Cortex *6*, 661-672.

Petkov, C.I., Kayser, C., Augath, M., and Logothetis, N.K. (2009). Optimizing the imaging of the monkey auditory cortex: sparse vs. continuous fMRI. Magn Reson Imaging *27*, 1065-1073.

Pickett, E.R., Kuniholm, E., Protopapas, A., Friedman, J., and Lieberman, P. (1998). Selective speech motor, syntax and cognitive deficits associated with bilateral damage to the putamen and the head of the caudate nucleus: a case study. Neuropsychologia *36*, 173-188.

Pisani, A., Centonze, D., Bernardi, G., and Calabresi, P. (2005). Striatal synaptic plasticity: implications for motor learning and Parkinson's disease. Mov Disord *20*, 395-402.

Price, C.J., Green, D.W., and von Studnitz, R. (1999). A functional imaging study of translation and language switching. Brain *122*, 2221-2235.

Raichle, M.E., Fiez, J.A., Videen, T.O., MacLeod, A.-M., K., Pardo, J.V., Fox, P.T., and Petersen, S.E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. Cereb Cortex *4*, 8-26.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., and Shulman, G.L. (2001). A default mode of brain function. Proc Natl Acad Sci U S A *98*, 676-682.

Raizada, R.D.S., Tsao, F.-M., Liu, H.-M., and Kuhl, P.K. (2010). Quantifying the adequacy of neural representations for a cross-language phonetic discrimination task: prediction of individual differences. Cereb Cortex *20*, 1-12.

Ramnani, N., and Miall, C. (2001). Expanding cerebellar horizons. Trends Cogn Sci *5*, 135-136.

Rauschecker, A.M., Pringle, A., and Watkins, K.E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. Hum Brain Mapp *29*, 1231-1242.

Rodriguez-Fornells, A., Rotte, M., Heinze, H.-J., Nosselt, T., and Munte, T.F. (2002). Brain potential and functional MRI evidence for how to handle two languages with one brain. Nature *415*, 1026-1029.

Rosen, H.J., Ojemann, J.G., Ollinger, J.M., and Petersen, S.E. (2000). Comparison of brain activation during word retrieval done silently and aloud using fMRI. Brain Cogn *42*, 201-217.

Sakai, K.L. (2005). Language acquisition and brain development. Science *310*, 815-819.

Salisbury, R.F. (1962). Institute of social research in New Guinea. Anthropology News *3*, 1-2.

Scherag, A., Demuth, L., Rosler, F., Neville, H.J., and Roder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. Cognition *93*, B97-B108.

Schleppegrell, M. (1987). The older language learner. In ERIC digest (Washington, DC, Center for Applied Linguistics, ERIC identifier ED287313).

Scott, S.K., and Wise, R.J.S. (2003). Functional imaging and language: a critical guide to methodology and analysis. Speech Commun *41*, 7-21.

Scovel, T. (1969). Foreign accents, language acquisition and cerebral dominance. Lang Learn *19*, 245-253.

Scovel, T. (1988). A time to speak: a psycholinguistic inquiry into the critical period for human speech (New York, Newbury House Publishers).

Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., and Greicius, M.D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci *27*, 2349-2356.

Seghier, M.L., and Price, C.J. (2012). Functional heterogeneity within the default network during semantic processing and speech production. Front Psychol *3:10.3389/fpsyg.2012.00281*.

Sellars, C., Hughes, T., and Langhorne, P. (2005). Speech and language therapy for dysarthria due to non-progressive brain damage. Cochrane Db Syst Rev.

Sereno, M.I., and Huang, R.-S. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. Nat Neurosci *9*, 1337-1343.

Shallice, T. (1988). From neuropsychology to mental structure (Cambridge, UK, Cambridge University Press).

Sharp, D.J., Beckmann, C.F., Greenwood, R., Kinnunen, K.M., Bonnelle, V., De Boissezon, X., Powell, J.H., Counsell, S.J., Patel, M.C., and Leech, R. (2011). Default mode network functional and structural connectivity after traumatic brain injury. Brain *134*, 2233-2247.

Shiller, D.M., Sato, M., Gracco, V.L., and Baum, S.R. (2009). Perceptual recalibration of speech sounds following speech motor learning. J Acoust Soc Am *125*, 1103-1113.

Shuster, L.I., and Lemieux, S.K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. Brain Lang *93*, 20-31.

Simmonds, A.J., Wise, R.J.S., Dhanjal, N.S., and Leech, R. (2011a). A comparison of sensory-motor activity during speech in first and second languages. J Neurophysiol *106*, 470-478.

Simmonds, A.J., Wise, R.J.S., and Leech, R. (2011b). Two tongues, one brain: imaging bilingual speech production. Front Psychol *2: 10.3389/fpsyg.2011.00166*.

Simonyan, K., Ostuni, J., Ludlow, C.L., and Horwitz, B. (2009). Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. J Neurosci *29*, 14912-14923.

Smiley, J.F., Hackett, T.A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D.C., and Schroeder, C.E. (2007). Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. J Comp Neurol *502*, 894-923.

Smith, S.M. (2002). Fast robust automated brain extraction. Hum Brain Mapp *17*, 142-155.

Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R.*, et al.* (2009). Correspondence of the brain's functional architecture during activation and rest. Proc Natl Acad Sci U S A *106*, 13040-13045.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E.*, et al.* (2004). Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage *23*, S208-S219.

Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann, C.F., Jenkinson, M., Andersson, J., Glasser, M.F.*, et al.* (2012). Temporally-independent functional modes of spontaneous brain activity. Proc Natl Acad Sci U S A *109*, 3131-3136.

Snow, C.E. (2002). Second language learners and understanding the brain. In The langauges of the brain, A.M. Galaburda, S.M. Kosslyn, and Y. Christen, eds. (Cambridge, Massachussetts/London, England, Harvard University Press), pp. 151-165.

Sorensen, A.P. (1967). Multilingualism in the Northwest Amazon. Am Anthropol *69*, 670-684.

Sörös, P., MacIntosh, B.J., Tam, F., and Graham, S.J. (2010). FMRI-compatible registration of jaw movements using a fiber-optic bend sensory. Front Hum Neurosci *4*, 1-8.

Southwood, M.H., and Flege, J.E. (1999). Scaling foreign accent: direct magnitude estimation versus interval scaling. Clin Linguist Phonet *13*, 335-349.

Speedie, L.J., Wertman, E., Tair, J., and Heilman, K.M. (1993). Disruption of automatic speech following a right basal ganglia lesion. Neurology *43*, 1768-1774.

Stein, M., Federspiel, A., Koenig, T., Wirth, M., Lehmann, C., Wiest, R., Strik, W., Brandeis, D., and Dierks, T. (2009). Reduced frontal activation with increasing 2nd language proficiency. Neuropsychologia *47*, 2712-2720.

Taylor, B.P. (1974). Toward a theory of language acquisition. Lang Learn *24*, 23-35.

Thierry, G., and Wu, Y.J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. Proc Natl Acad Sci U S A *104*, 12530-12535.

Tourville, J.A., Reilly, K.J., and Guenther, F.H. (2008). Neural mechanisms underlying auditory feedback control of speech. Neuroimage *39*, 1429-1443.

Tremblay, P., Deschamps, I., and Gracco, V.L. (2011). Regional heterogeneity in the processing and the production of speech in the human planum temporale. Cortex *doi:10.1016/j.cortex.2011.09.004*.

Tremblay, S., Shiller, D.M., and Ostry, D.J. (2003). Somatosensory basis of speech production. Nature *423*, 866-869.

Vaid, J., and Hull, R. (2002). Re-envisioning the bilingual brain using function neuroimaging. In Advances in the neurolinguistics of bilingualism, F. Fabbro, ed. (Udine, Forum), pp. 315-355.

van den Noort, M., Bosch, P., and Hugdahl, K. (2006). Looking at second language acquisition from a functional- and structural MRI background. Proceedings of the 28th Annual Meeting of the Cognitive Science Society
Mahwah:NJ: Lawrence Erlbaum Associates, Inc
http://wwwcogscirpiedu/CSJarchive/Proceedings/2006/docs/p2293pdf, 2293-2298.

Vargha-Khadem, F., Gadian, D.G., Copp, A., and Mishkin, M. (2005). *FOXP2* and the neuroanatomy of speech and language. Nat Rev Neurosci *6*, 131-138.

Ventura, M.I., Nagarajan, S.S., and Houde, J.F. (2009). Speech target modulates speaking induced suppression in auditory cortex. BMC Neuroscience *10*, 58-69.

Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., and Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. Neuroimage *30*, 1414-1432.

Walsh, T.M., and Diller, K.C. (1978). Neurolinguistic foundations to methods of teaching a second language. IRAL *16*, 1-14.

Warker, J.A., Xu, Y., Dell, G. S. and Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. Cognition *112*, 81-96.

Warren, J.D., and Griffiths, T.D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. J Neurosci *23*, 5799-5804.

Watkins, K.E., Dronkers, N.F., and Vargha-Khadem, F. (2002a). Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. Brain *125*, 452-464.

Watkins, K.E., Smith, S.M., Davis, S., and Howel, P. (2008). Structural and functional abnormalities of the motor system in developmental stuttering. Brain *131*, 50-59.

Watkins, K.E., Varha-Khadem, F., Ashburner, J., Passingham, R.E., Connelly, A., Friston, K.J., Frackowiak, R.S.J., Mishkin, M., and Gadian, D.G. (2002b). MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain *125*, 465-478.

Weismer, S.E., Tomblin, J.B., Zhang, X., Buckwalter, P., Chynoweth, J.G., and Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. J Speech Lang Hear Res *43*, 865-878.

Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. Behav Res Meth Instrum Comput *20, 6-11.*

Wilson, S.M., and Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech production. Neuroimage *33*, 316-325.

Wise, R.J.S., Greene, J., Buchel, C., and Scott, S.K. (1999). Brain regions involved in articulation. Lancet *353*, 1057-1061.

Wise, R.J.S., Scott, S.K., Blank, S.C., Mummery, C.J., Murphy, K., and Warburton, E.A. (2001). Separate neural subsystems within 'Wernicke's area'. Brain *124*, 83-95.

Woolrich, M.W., Behrens, T.E.J., Beckmann, C.F., Jenkinson, M., and Smith, S.M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage *21*, 1732-1747.

Wu, Y.J., and Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. J Neurosci *30*, 7646-7651.

Wu, Y.J., and Thierry, G. (2011). Event-related brain potential investigation of preparation for speech production in late bilinguals. Front Psychol *2, doi 10.3389/fpsyg.2011.00114*.

Yetkin, F.Z., Hammeke, T.a., Swanson, S.J., Morris, G.L., Mueller, W.M., McAuliffe, T.L., and Haughton, V.M. (1995). A comparison of functional MR activation patterns during silent and audible language tasks. Am J Neuroradiol *16*, 1087-1092.

Yin, H., Mulcare, S., Hilàrio, M., Clouse, E., Holloway, T., Davis, M., Hansson, A., Lovinger, D., and Costa, R. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. Nat Neurosci *12*, 333-341.

Yin, H.H., and Knowlton, B.J. (2006). The role of the basal ganglia in habit formation. Nat Rev Neurosci *7*, 464-476.

Zuo, X.N., Kelly, C., Adelstein, J.S., Klein, D.F., Castellanos, F.X., and Milham, M.P. (2010). Reliable instrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach. Neuroimage *29*, 2163-2177.

# 9 Appendix

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **ANOVA: Main effect of Overt/Covert** | *R primary auditory cortex* | *17886* | *7.99* | *48* | *-18* | *2* |
| | | *L primary auditory cortex* | *6641* | *8.09* | *-44* | *-18* | *0* |
| | | *premotor cortex* | *624* | *4.24* | *0* | *14* | *58* |
| B | **ANOVA: Main effect of Speaking/Counting** | *L inferior frontal gyrus* | *22860* | *7.19* | *-48* | *28* | *12* |
| | | *R caudate* | *785* | *4.03* | *16* | *16* | *2* |

***Appendix 1, Cluster table for the ANOVA in Chapter 3***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Overt Speaking > Rest** | *L superior temporal gyrus* | *25503* | *5.83* | *-62* | *-28* | *2* |
| | | *R primary auditory cortex* | *3554* | *5.16* | *44* | *-26* | *4* |
| B | **Covert Speaking > Rest** | *L inferior frontal gyrus* | *2045* | *4.47* | *-48* | *30* | *10* |
| | | *R inferior parietal lobule* | *1839* | *4.92* | *42* | *-58* | *-34* |
| | | *paracingulate cortex* | *1630* | *4.78* | *-4* | *20* | *38* |
| | | *white matter* | *549* | *4.01* | *-40* | *-46* | *-8* |
| C | **Overt Counting > Rest** | *L Heschl's gyrus* | *4196* | *5.68* | *-38* | *-24* | *6* |
| | | *R primary auditory cortex* | *3744* | *5.19* | *44* | *-24* | *4* |
| | | *cerebellum* | *1113* | *4.68* | *-12* | *-58* | *-22* |

***Appendix 2, Cluster table for the speech conditions > Rest in Chapter 3***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Overt Speaking > Covert Speaking** | *L secondary somatosensory cortex* | *5577* | *5.26* | *-48* | *-10* | *10* |
| | | *R primary auditory cortex* | *4132* | *4.85* | *40* | *-26* | *6* |
| | | *brainstem* | *4098* | *4.74* | *8* | *-40* | *-44* |
| | | *premotor cortex* | *414* | *3.62* | *-6* | *18* | *56* |
| | | *thalamus* | *385* | *3.45* | *4* | *-2* | *4* |
| B | **Overt Counting > Covert Counting** | *white matter* | *11591* | *5.89* | *-34* | *-30* | *8* |
| | | *cerebellum* | *3012* | *4.33* | *-16* | *-56* | *-22* |
| C | **Overt Speaking > Overt Counting** | *L inferior frontal gyrus* | *23397* | *5.58* | *-46* | *26* | *10* |
| | | *L superior parietal lobule* | *670* | *4.29* | *-24* | *-74* | *48* |
| D | **Covert Speaking > Covert Counting** | *cerebellum* | *6540* | *4.97* | *36* | *-60* | *-34* |
| | | *L inferior frontal gyrus* | *6284* | *5.06* | *-48* | *30* | *12* |
| | | *white matter* | *1060* | *3.53* | *-16* | *-4* | *12* |
| | | *R frontal orbital cortex* | *719* | *3.75* | *46* | *24* | *-12* |
| | | *R caudate* | *684* | *3.8* | *16* | *16* | *4* |
| | | *R middle temporal gyrus* | *457* | *3.83* | *48* | *-36* | *-4* |
| | | *white matter* | *423* | *3.26* | *-28* | *-68* | *6* |
| E | **Overt Counting > Overt Speaking** | *precuneus* | *4210* | *4.42* | *10* | *-64* | *32* |
| | | *R inferior parietal lobule* | *1587* | *4.18* | *56* | *-28* | *34* |
| F | **Covert Counting > Covert Speaking** | *precuneus* | *2216* | *3.58* | *-4* | *-54* | *34* |
| | | *R inferior parietal lobule* | *1552* | *3.69* | *52* | *-56* | *38* |

***Appendix 3, Cluster table for the direct contrasts in Chapter3***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Auditory Functional Localiser** | *R superior temporal gyrus* | *4459* | *8.69* | *68* | *-14* | *-2* |
| | | *L superior temporal gyrus* | *4387* | *8.78* | *-64* | *-24* | *4* |
| | | *visual cortex* | *326* | *4.05* | *14* | *-90* | *-4* |
| B | **Tongue Functional Localiser** | *L primary sensorimotor cortex* | *30471* | *16.6* | *-58* | *-6* | *28* |
| | | *R cerebellum* | *5355* | *16.8* | *18* | *-62* | *-24* |
| | | *L frontal pole* | *750* | *4.88* | *-38* | *42* | *20* |
| | | *R frontal pole* | *522* | *3.58* | *40* | *44* | *-2* |

***Appendix 4, Cluster table for the functional localisers in Chapter 4***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Speech (L1 + L2) > Rest** | *L primary sensorimotor cortex* | *6023* | *4.89* | *-50* | *-12* | *26* |
| | | *R secondary somatosensory cortex* | *4272* | *4.46* | *64* | *0* | *14* |
| | | *L cerebellum* | *1608* | *4.6* | *-12* | *-62* | *-22* |
| B | **L2 > L1** | *cerebellum* | *9484* | *4.36* | *2* | *-52* | *-18* |
| | | *L inferior frontal gyrus* | *5212* | *4.54* | *-52* | *8* | *10* |
| | | *paracingulate gyrus* | *2678* | *3.79* | *0* | *48* | *16* |

***Appendix 5, Cluster table for the speech conditions in Chapter 4***

## Training Session 1a: Articulatory Anatomy

Watch **presentation 1a** again.
Use the information in the presentation to complete the following sentences:

1) The soft palate is the back part of the …… of the mouth.

2) The teeth are used in sounds including […..]

3) The epiglottis separates the ……. from the ……

4) The tongue tip is the end of the tongue, nearest to the …….

5) The oral cavity is the inside of the …….

6) The alveolar ridge is the ……. part of the roof of the mouth, just behind the ……..

7) The tongue blade is the central part of the tongue, below the ……..

8) The nasal cavity is the area above the roof of the mouth, from the ……. to the ……

9) The pharynx is the cavity above the larynx, at the …….

10) The lips are used in sounds including [……], [……] and […….]

11) The mandible is the …….., and holds the ……..

12) The tongue back is the back part of the tongue, below the ……..

13) The vocal folds are stretched horizontally across the larynx and …….. to create sound. An example of a "voiced" sound is […….] and an example of a "voiceless" sound is [……]

14) The hard palate is central part of the ……..

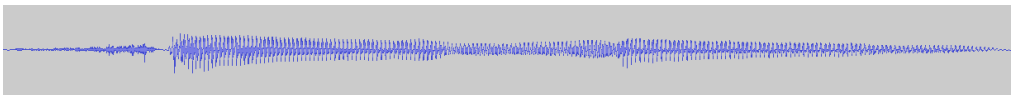
6

***Appendix 6, Example page of articulatory anatomy from the training materials used in Chapter 5***

## Training Session 4j  Mandarin Tones 1 and 4

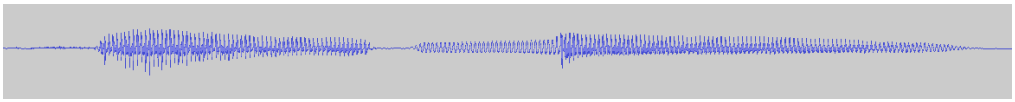
Below are the spectrograms for the sounds included in the presentation. Compare your spectrograms to these examples. Repeat your attempts of the speech sounds and try to get your spectrograms close to these examples.

Sound 1

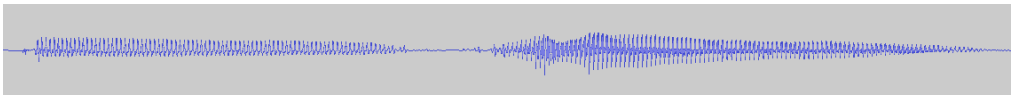


Sound 2

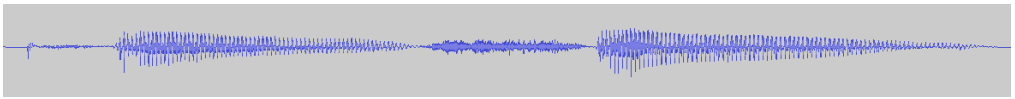
Sound 3



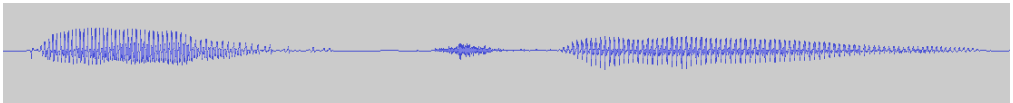
Sound 4



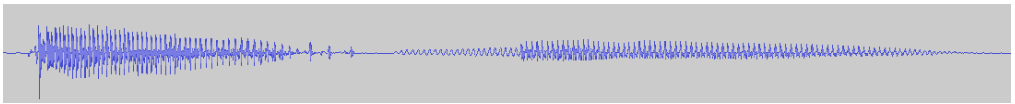
Sound 5



Sound 6

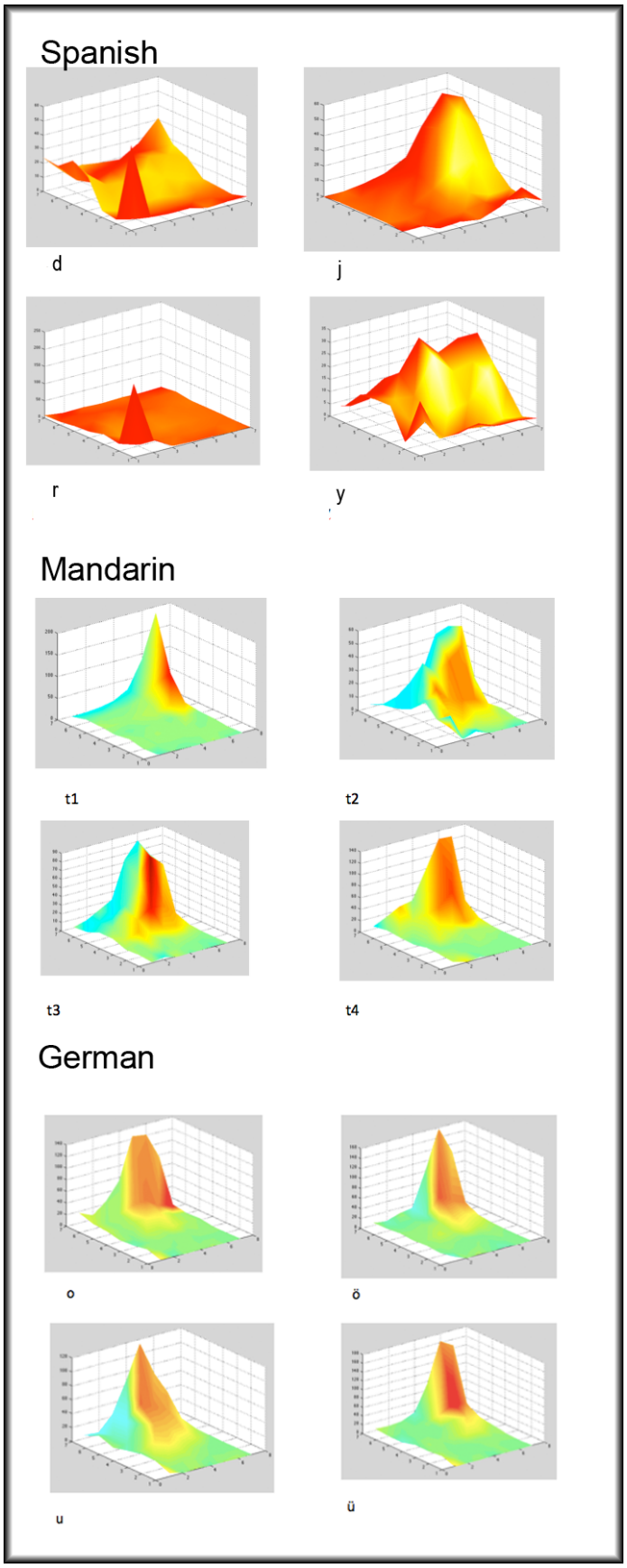


Sound 7

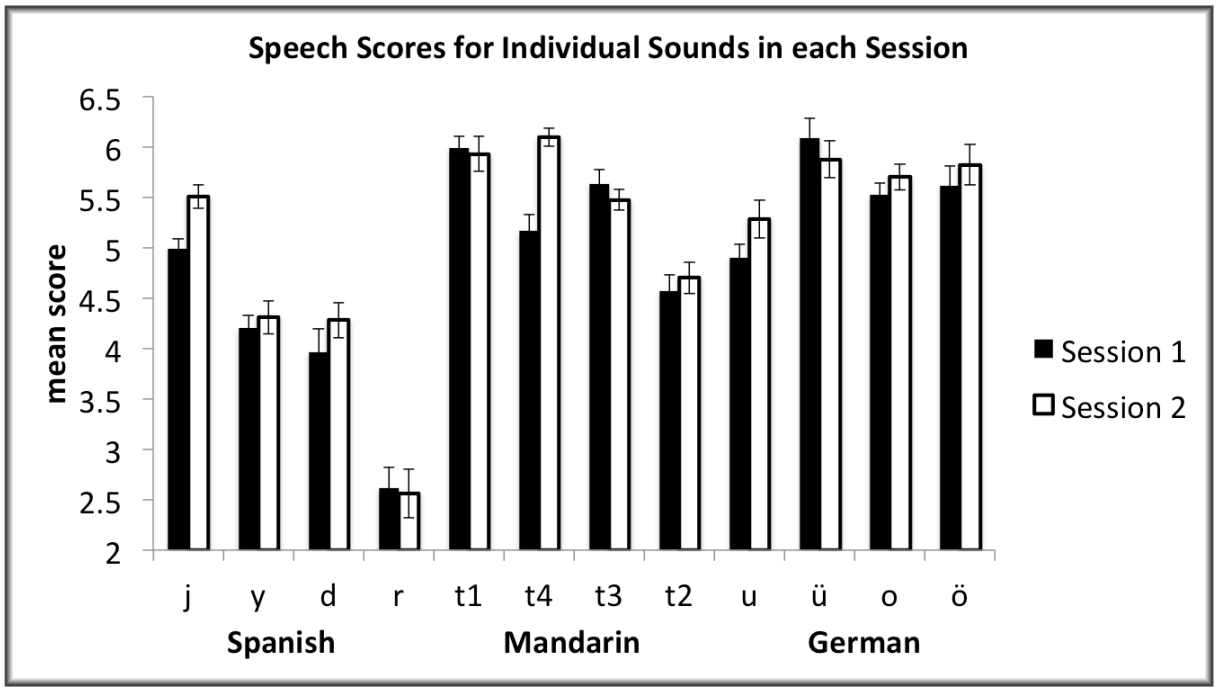




***Appendix 7, Example page of spectrograms from the training materials used in Chapter 5***




***Appendix 8, Speech rating correlations for the individual sounds in Chapter 5***




***Appendix 9, Speech scores for the individual sounds in Chapter 5***

*Session 1 is shown in black and Session 2 in white. From left to right are the mean scores are shown for the Spanish sounds (Trained: j and y; Untrained d and r), Mandarin sounds (Trained: t1 and t4; Untrained t3 and t2), and German sounds (Trained: u and ü; Untrained: o and ö). Error bars display 95% confidence intervals.*

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| **A** | **Listen ANOVA Main effect of Language** | *L inferior frontal gyrus* | *959* | *3.59* | *-54* | *32* | *4* |
| | | *R inferior frontal gyrus* | *606* | *4.87* | *50* | *30* | *4* |
| **B** | **Listen ANOVA Main effect of Session** | *R middle temporal gyrus* | *1199* | *5.39* | *70* | *-32* | *-2* |
| | | *R intraparietal sulcus* | *854* | *3.9* | *34* | *-52* | *36* |
| | | *L parietal operculum* | *752* | *3.42* | *-60* | *-12* | *8* |
| | | *R primary somatosensory cortex* | *504* | *2.16* | *48* | *-28* | *52* |
| | | *R middle frontal gyrus* | *416* | *1.66* | *36* | *10* | *42* |
| | | *L intraparietal sulcus* | *401* | *1.58* | *-44* | *-44* | *38* |
| | | *premotor cortex* | *357* | *1.31* | *-10* | *-20* | *54* |
| **C** | **Listen ANOVA Language x Session Interaction** | *R intraparietal sulcus* | *1463* | *4.26* | *36* | *-54* | *34* |
| | | *R frontal pole* | *474* | *3.64* | *42* | *32* | *28* |

***Appendix 10, Cluster table for the Listening ANOVA in Chapter 5***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Session 1 Listening Non-native > Rest** | *L Heschl's gyrus* | *60020* | *6.43* | *-42* | *-26* | *6* |
| B | **Session 1 Listening Native > Rest** | *L Heschl's gyrus* | *61420* | *7.12* | *-42* | *-24* | *6* |
| | | *R frontal pole* | *468* | *3.41* | *12* | *50* | *42* |
| C | **Session 2 Listening Non-native > Rest** | *R Heschl's gyrus* | *84102* | *7.28* | *50* | *-10* | *0* |
| D | **Session 2 Listening Native > Rest** | *L Heschl's gyrus* | *11230* | *7.27* | *-40* | *-24* | *4* |
| | | *L occipital pole* | *8665* | *6.0* | *-10* | *-100* | *-2* |
| | | *R Heschl's gyrus* | *6786* | *7.44* | *42* | *-22* | *8* |
| | | *R inferior frontal gyrus* | *1947* | *4.59* | *54* | *34* | *8* |
| | | *L inferior parietal lobule* | *872* | *4.06* | *-28* | *-76* | *48* |
| | | *R intraparietal sulcus* | *497* | *3.62* | *34* | *-48* | *48* |

***Appendix 11, Cluster table for the Listening > Rest conditions in Chapter 5***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Listening Session 1 Native > Non-native** | *L inferior frontal gyrus* | *2661* | *4.38* | *-46* | *22* | *14* |
| | | *R inferior frontal gyrus* | *2390* | *4.64* | *46* | *30* | *12* |
| | | *L middle temporal gyrus* | *1471* | *3.99* | *-52* | *-8* | *-16* |
| | | *R middle temporal gyrus* | *1447* | *4* | *52* | *-44* | *8* |
| | | *R intraparietal sulcus* | *1323* | *4.02* | *36* | *-54* | *36* |
| | | *L intraparietal sulcus* | *916* | *3.84* | *-34* | *-60* | *44* |
| | | *superior frontal gyrus* | *635* | *3.6* | *0* | *54* | *32* |
| | | *precuneus* | *483* | *3.15* | *2* | *-66* | *26* |
| | | *R superior temporal gyrus* | *332* | *3.91* | *50* | *-12* | *14* |
| B | **Session 2 Listening Non-native > Native** | *L parietal operculum* | *980* | *4* | *-60* | *-6* | *10* |
| | | *R primary auditory cortex* | *753* | *3.89* | *56* | *-12* | *8* |
| | | *cerebellum* | *284* | *3.62* | *24* | *-58* | *-16* |

***Appendix 12, Cluster table for the Listening contrasts in Chapter 5***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Repeat ANOVA Main effect of Language** | *cerebellum* | *139549* | *14.6* | *26* | *-62* | *-26* |
| B | **Repeat ANOVA Main effect of Session** | *primary motor cortex* | *3615* | *5.55* | *4* | *-32* | *64* |
| | | *L caudate* | *3213* | *5.13* | *-14* | *8* | *14* |
| | | *L superior parietal lobule* | *2167* | *6* | *-10* | *-86* | *44* |
| | | *L superior frontal gyrus* | *1895* | *5.85* | *-12* | *32* | *54* |
| | | *L occipital fusiform gyrus* | *565* | *4.72* | *-34* | *-80* | *-12* |
| | | *R middle temporal gyrus* | *432* | *4.83* | *28* | *32* | *50* |
| | | *L supramarginal gyrus* | *389* | *4.71* | *-58* | *-30* | *52* |
| | | *L inferior parietal lobule* | *384* | *4.29* | *-60* | *-56* | *32* |
| C | **Repeat ANOVA Language x Session Interaction** | *primary motor cortex* | *3084* | *5.3* | *4* | *-32* | *64* |
| | | *L caudate* | *2186* | *4.9* | *-14* | *8* | *14* |
| | | *L superior parietal lobule* | *1942* | *5.89* | *-10* | *-88* | *42* |
| | | *L superior frontal gyrus* | *1626* | *5.6* | *-12* | *32* | *54* |
| | | *L occipital fusiform gyrus* | *435* | *4.56* | *-34* | *-80* | *-12* |
| | | *R middle frontal gyrus* | *358* | *4.69* | *28* | *32* | *50* |

***Appendix 13, Cluster table for the Repeating ANOVA in Chapter 5***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Session 1 Repeat Non-native > Rest** | *precuneus* | *115918* | *7.32* | *-14* | *-64* | *22* |
| B | **Session 1 Repeat Native > Rest** | *R superior temporal gyrus* | *103743* | *6.77* | *56* | *-10* | *-10* |
| C | **Session 2 Repeat Non-native > Rest** | *R cerebellum* | *102888* | *7.57* | *34* | *-54* | *-34* |
| D | **Session 2 Repeat Native > Rest** | *R cerebellum* | *94331* | *7.2* | *36* | *-52* | *-32* |

***Appendix 14, Cluster table for the Repeating > Rest conditions in Chapter 5***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| **A** | **Session 1 Repeat Non-native > Native** | *cerebellum* | *115348* | *7.35* | *-14* | *-64* | *-22* |
| **B** | **Session 1 Repeat Native > Non-native** | *frontal medial cortex* | *2070* | *5.75* | *-8* | *46* | *-14* |
| | | *lateral occipital cortex* | *989* | *4.08* | *-58* | *-62* | *28* |
| | | *angular gyrus* | *680* | *4.52* | *50* | *-54* | *18* |
| | | *precuneus* | *642* | *4.11* | *-10* | *-60* | *14* |
| | | *L middle frontal gyrus* | *387* | *4.07* | *-24* | *28* | *38* |
| **C** | **Session 2 Repeat Non-native > Native** | *cerebellum* | *88878* | *7.56* | *34* | *-54* | *-34* |
| **D** | **Session 2 Repeat Native > Non-native** | *angular gyrus* | *1067* | *4.82* | *52* | *-54* | *24* |
| | | *cingulate gyrus* | *1009* | *4.31* | *4* | *38* | *-4* |
| | | *precuneus* | *971* | *4.82* | *-8* | *-54* | *42* |
| | | *lateral occipital cortex* | *914* | *4.3* | *-50* | *-55* | *24* |
| | | *frontal pole* | *830* | *5.09* | *-28* | *38* | *36* |
| | | *R middle frontal gyrus* | *743* | *4.52* | *28* | *22* | *44* |
| | | *premotor cortex* | *452* | *3.92* | *-6* | *-22* | *56* |
| | | *paracingulate* | *328* | *4.13* | *8* | *50* | *14* |

***Appendix 15, Cluster table for the Repeating contrasts in Chapter 5***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| **A** | **ANOVA Repeat Spanish/Native x Session Interaction** | *lateral occipital cortex* | *417* | *4.21* | *-10* | *-86* | *44* |
| **B** | **ANOVA Repeat Mandarin/Native x Session Interaction** | *lateral occipital cortex* | *5901* | *4.8* | *-12* | *-86* | *42* |
| | | *superior parietal lobule* | *3043* | *4.7* | *6* | *-48* | *76* |
| | | *L putamen* | *2851* | *4.71* | *-22* | *12* | *0* |
| | | *L frontal pole* | *1359* | *5.16* | *-26* | *40* | *36* |
| | | *R middle frontal gyrus* | *369* | *4.26* | *28* | *32* | *50* |
| **C** | **ANOVA Repeat German/Native x Session Interaction** | *L superior frontal gyrus* | *1784* | *4.76* | *-12* | *32* | *54* |
| | | *R primary motor cortex* | *1776* | *4.47* | *20* | *-32* | *70* |
| | | *lateral occipital cortex* | *1308* | *5.02* | *32* | *-82* | *44* |
| | | *L superior parietal lobule* | *1095* | *4.18* | *-46* | *-48* | *56* |
| | | *R middle frontal gyrus* | *1015* | *4.44* | *46* | *26* | *40* |
| | | *white matter* | *702* | *4.48* | *-14* | *-2* | *12* |
| | | *white matter* | *467* | *4.03* | *12* | *4* | *10* |

***Appendix 16, Cluster table for the separate Repeating ANOVAs in Chapter 5***

| | Contrast | Region | Voxels | Z-Max | Z-Max X (mm) | Z-Max Y (mm) | Z-Max Z (mm) |
|---|---|---|---|---|---|---|---|
| A | **Repeat Non-native > Rest difference across sessions, with Speech Score improvement** | *L secondary somatosensory cortex* | *1890* | *4.71* | *-54* | *2* | *6* |
| | | *R cerebellum* | *838* | *4.13* | *40* | *-48* | *-32* |
| | | *R temporal pole* | *639* | *4.16* | *56* | *16* | *-4* |
| | | *visual cortex* | *611* | *4.44* | *-2* | *-86* | *-10* |
| | | *white matter* | *293* | *3.34* | *16* | *-76* | *20* |
| | | *lateral occipital cortex* | *280* | *3.73* | *-28* | *-74* | *28* |
| B | **Repeat Spanish difference across sessions with Speech Score improvement** | *L occipital fusiform gyrus* | *989* | *4.8* | *-40* | *-72* | *-12* |
| | | *L secondary somatosensory cortex* | *621* | *4.25* | *-58* | *-14* | *14* |
| | | *R temporal-occipital fusiform cortex* | *617* | *3.89* | *42* | *-58* | *-10* |
| | | *white matter* | *604* | *3.49* | *-16* | *-6* | *12* |
| | | *L supramarginal gyrus* | *463* | *4.08* | *-52* | *-42* | *12* |
| | | *L premotor cortex* | *454* | *4.26* | *-10* | *-2* | *52* |
| C | **Repeat Mandarin difference across sessions with Speech Score improvement** | *R frontal pole* | *315* | *4.25* | *24* | *40* | *-18* |
| | | *L superior temporal gyrus* | *289* | *4.31* | *-64* | *-10* | *-6* |
| | | *R cerebellum* | *279* | *3.51* | *32* | *-48* | *-28* |
| D | **Repeat German difference across sessions with Speech Score improvement** | *white matter* | *423* | *3.63* | *-40* | *-26* | *28* |

***Appendix 17, Cluster table for the whole-brain results with speech score improvements for Non-native Repeating and each Language separately, in Chapter 5***