The Bases of Difficulties in Spatial Hearing for Speech: Investigations using Psychoacoustic Techniques and Magneto-encephalography

Pádraig Thomas Kitterick

Submitted for the degree of Doctor of Philosophy

University of York
Department of Psychology

October 2008

Abstract

The experiments reported in this thesis investigated the bases of the difficulties that older adults report when trying to listen to what one person is saying when many other people are speaking at the same time. Experiments 1–4 examined the roles of voluntary and involuntary attention in a spatial listening task for speech among young normally-hearing listeners. When talkers started speaking one at a time, listeners could hear out a target phrase that was less intense than overlapping masker phrases. When talkers started speaking in pairs, listeners could attend to a less intense target phrase only when told in advance who to listen for, where they would speak from, or when they would speak. The distracting effect of the onset of a competing talker was effective over a broad time window. Experiment 5 investigated the relationships between performance on the spatial listening task and several predictors of performance among young and older normally-hearing adults. Poorer performance was related to self-reported difficulties with listening in everyday situations, poorer hearing sensitivity, and poorer performance on visual and auditory tasks of attention requiring fast speed of processing. Experiment 6 examined brain activity associated with successful performance on the spatial listening task using magneto-encephalography. Differences in cortical activity were identified at moments when attention had to be sustained on the target phrase, or when listeners had to resist distraction from the onset of a new masker phrase. Amplitudes, and/or latencies, of differences in brain activity arising in regions associated with attentional processes were related to performance. The results suggest that skills in attention contribute to the ability to listen successfully in multi-talker environments. Age-related difficulties with listening in those environments may arise due to a specific reduction in the ability to resist distraction or a general reduction in the speed at which information can be processed.

Contents

1 **Overview** **1**

1.1 Introduction 1

1.2 Overview of Following Chapters 2

2 **Hearing, Ageing, and Speech Perception** **5**

2.1 Introduction 5

2.2 Hearing and Ageing 6

2.2.1 Introduction 6

2.2.2 Hearing Difficulties 6

2.2.3 Speech Perception and Ageing 8

2.2.4 Impact of Hearing Difficulties on Quality of Life 19

2.2.5 Summary 21

2.3 Spatial Hearing and Speech Segregation 21

2.3.1 Introduction 21

2.3.2 Masking 22

2.3.3 Hearing with One Ear 23

2.3.4 Hearing with Two Ears 28

2.4 Summary 32

3 **Attention and Spatial Listening** **34**

3.1 Attentional Mechanisms 34

3.1.1 Introduction 34

3.1.2 Selective Attention 35

3.1.3 Shifting or Switching Attention 37

3.1.4 Voluntary and Involuntary Shifts of Attention 37

3.1.5 Abrupt Onsets and Attention Capture 38

3.2 Attention and Complex Listening Environments 40

3.2.1 Attentional Strategies in Complex Listening Environments 41

3.2.2 Uncertainty and Attention 43

3.3 Cortical Mechanisms of Attention 46

3.3.1 Fronto-Parietal Network 46

3.3.2 Auditory Selective Attention 51

3.3.3 Attention and Oscillatory Activity . . . 53
3.3.4 Attention to Speech in Noise . . . 57
3.4 Cognitive Effects of Ageing . . . 63
3.4.1 Changes in Central Processing . . . 63
3.4.2 Temporal Processing . . . 64
3.4.3 Age-related Changes in Attentional Processes . . . 65
3.4.4 Differences in Voluntary and Involuntary Processing . . . 65
3.4.5 Attention-related Changes in Neural Activation . . . 66
3.5 Summary . . . 67
3.6 Conclusions from the Literature Review . . . 69

4 Reconstructing Cortical Activity with High Temporal Resolution 70
4.1 Introduction . . . 70
4.1.1 Notation Style and Conventions . . . 72
4.2 The Inverse Problem . . . 72
4.3 The Forward Solution . . . 73
4.4 Minimum Norm . . . 77
4.4.1 Introduction . . . 77
4.4.2 Finding an Inverse . . . 77
4.4.3 Moore-Penrose Pseudoinverse . . . 78
4.4.4 Source Weighting . . . 80
4.4.5 Regularisation . . . 80
4.4.6 Noise Estimates . . . 84
4.4.7 Anatomical Constraints . . . 85
4.4.8 Depth Weighting . . . 86
4.4.9 Summary . . . 88
4.5 Linearly Constrained Spatial Filtering . . . 89
4.5.1 Introduction . . . 89
4.5.2 Designing a Spatial Filter . . . 89
4.5.3 Factors which Influence the Spatial Filter . . . 91
4.5.4 Depth Weighting . . . 91
4.5.5 Characterising Distributed Neural Activity . . . 92
4.5.6 Anatomical Constraints . . . 93
4.5.7 Correlated Sources . . . 94
4.5.8 Summary . . . 96
4.6 Comparison . . . 96
4.6.1 Summary . . . 99
4.7 Application to the Current Research . . . 100
4.8 Summary . . . 101

5 Voluntary and Involuntary Attention in a Multi-talker Environment 102
5.1 Introduction 103
  5.1.1 Benefits of Knowing 'Where' to Listen 103
  5.1.2 Benefits of Knowing 'Who' to Listen to 103
  5.1.3 Benefits of Knowing 'When' to Listen 104
  5.1.4 The Present Experiments 104
5.2 Experiment 1 105
  5.2.1 Introduction 105
  5.2.2 Methods 106
  5.2.3 Results 111
  5.2.4 Discussion 112
5.3 Experiment 2 112
  5.3.1 Introduction 112
  5.3.2 Methods 113
  5.3.3 Results 115
  5.3.4 Discussion 116
5.4 Experiment 3 118
  5.4.1 Introduction 118
  5.4.2 Methods 119
  5.4.3 Results 121
  5.4.4 Discussion 121
5.5 Experiment 4 123
  5.5.1 Introduction 123
  5.5.2 Methods 123
  5.5.3 Results 125
  5.5.4 Discussion 126
5.6 General Discussion 129
  5.6.1 Attentional Effects of Phrase Onsets 129
  5.6.2 Temporal Uncertainty 130
  5.6.3 Energetic Masking 131
5.7 Conclusions 131
5.8 Summary 132

6 Relationships among Age, Hearing Level, Attentional Abilities, and Performance on a Spatial Listening Task 134
6.1 Introduction 134
  6.1.1 Summary of Hypotheses 137
6.2 Methods 138
  6.2.1 Participants 138
  6.2.2 Spatial Listening Tasks 138

6.2.3 Attentional Test Battery 140
6.2.4 Self-reported Difficulties 142
6.2.5 Design 143
6.2.6 Procedure 143
6.2.7 Analyses 144
6.3 Results 149
6.3.1 Hearing levels 149
6.3.2 Spatial Listening Task 149
6.3.3 Attentional Measures 152
6.3.4 Self-reported Difficulties 154
6.3.5 Individual Differences 155
6.4 Discussion 158
6.4.1 The Roles of Hearing Loss and Attentional Abilities in Determining Spatial Listening Performance 159
6.4.2 Multi-talker Listening 162
6.4.3 Effects of Hearing Sensitivity 163
6.4.4 Attentional Factors 163
6.4.5 Self-reported Difficulties 164
6.5 Conclusions 165
6.6 Summary 166

7 Cortical Activation Patterns During a Spatial Listening Task 168
7.1 Introduction 169
7.1.1 Summary of Hypotheses 171
7.2 Methods 172
7.2.1 Participants 172
7.2.2 Spatial Listening Task 172
7.2.3 Stimuli & Recordings 173
7.2.4 MEG & EEG Recordings 175
7.2.5 MRI Acquisition 177
7.2.6 Co-registration & Headmodel 177
7.2.7 Design 178
7.2.8 Procedure 178
7.2.9 Preparation of MEG Data 178
7.2.10 Analysis 185
7.3 Results 197
7.3.1 Behavioural Performance 197
7.3.2 Response Latencies 198
7.3.3 Sensor-space 198
7.3.4 Source-space 209

7.3.5 Individual Differences 216
7.4 Discussion 222
7.4.1 “Onset processing” 223
7.4.2 “Call-sign processing” 224
7.4.3 “Onset attention” 225
7.4.4 Neural Bases of Focussing Attention and Resisting Distraction . . 225
7.4.5 Neural Correlates of Performance 233
7.5 Direction of Source-space Effects 235
7.6 Conclusions 235
7.7 Summary 235

8 Summary and General Discussion 237
8.1 Recap of Research Aims 237
8.2 Summary of Findings 238
8.2.1 Main Findings of Experiments 1–4 238
8.2.2 Main Findings of Experiment 5 238
8.2.3 Main Findings of Experiment 6 239
8.3 General Discussion 240
8.3.1 Effects of Age: Assessing the Impact of Cognitive Ability 240
8.3.2 Accounting for Individual Differences in Performance 241
8.4 Future Research 242
8.4.1 Understanding the Nature of Difficulties in Multi-talker Tasks . . 242
8.4.2 Disassociating Hearing Loss and Cognitive Difficulties 244
8.4.3 Improving Performance on Demanding Listening Tasks 245
8.5 Conclusion 247

Appendices

A Vectors and Matrices: Notation, Terminology and Properties 248
A.1 Vectors 248
A.1.1 Definition 248
A.1.2 Vector Operations 248
A.1.3 Norm of a Vector 248
A.2 Matrices 248
A.2.1 Definition 248
A.2.2 Norm of a Matrix 249
A.2.3 Eigenvalues and Eigenvectors 249
A.2.4 Singular Values and Vectors 249
A.2.5 Square Matrices 250
A.2.6 Covariance Matrices 250
A.2.7 The Matrix Inverse 251

A.2.8 Singularity 251
A.2.9 Condition Number 251
A.3 System of Linear Equations 252

B Calculations 254
B.1 Moore-Penrose Pseudoinverse 254
B.2 LCMV Optimised Spatial Filter 255
B.3 Equivalence 258

C Effects of Voice Characteristics and Location on Speech Perception 260
C.1 Talker Intelligibility 260
C.1.1 Introduction 260
C.1.2 Methods 260
C.1.3 Results 261
C.1.4 Discussion & Conclusions 261
C.2 Effect of Location on Speech Recognition 261
C.2.1 Introduction 261
C.2.2 Methods 262
C.2.3 Results 263
C.2.4 Discussion & Conclusions 263

D Hearing Health Questionnaire 264

E Speech, Spatial, and Qualities of Hearing Questionnaire 265

F Assessing the Discrimination Point of Call-sign Keywords 269
F.1 Introduction 269
F.2 Methods 269
F.2.1 Participants 269
F.2.2 Stimuli 270
F.2.3 Training 270
F.2.4 Procedure 270
F.2.5 Design 271
F.2.6 Analysis 271
F.3 Results 271
F.4 Discussion & Conclusions 272

G Cortical SPMs: Peak Activations 274

Glossary 284

References 286

List of Figures

2.1 Cross-section of the human cochlea. 7
2.2 Keyword identification performance in speech and noise. 25

3.1 The multiple-demand (MD) network. 49
3.2 Comparison of evoked and induced data. 54

4.1 Comparison of magnetic field strengths 71
4.2 MEG scanner and magnetically-shielded room. 72
4.3 The magnetic field generated by a current dipole 75
4.4 Spherical head models 76
4.5 A schematic representation of the regularisation problem in MEG 81
4.6 An example of an estimated L-curve 82
4.7 Effect of the choice of regularisation parameter 84
4.8 Sensitivity of MEG as a function of source depth. 87
4.9 Graphical representation of the LCMV spatial filter focused on a source of interest. 89
4.10 Data covariance matrix for two sources at varying degrees of correlation 95
4.11 A graphical representation of the source covariance matrices for Minimum Norm and LCMV Spatial Filtering 98

5.1 Experiment 1: The loudspeaker array. 107
5.2 Experiment 1: Sequence of phrases 108
5.3 Experiment 1: Mean and individual SRTs for each condition. 111
5.4 Experiment 2: Sequence of phrases 113
5.5 Experiment 2: Mean and individual SRTs for each condition. 115
5.6 Experiment 2: Estimated mean advantages in Experiments 1 and 2. . . . 116
5.7 Experiment 3: Schematic representation of target-masker asynchrony . 119
5.8 Experiment 3: Mean and individual SRTs for a range of target-masker asynchronies. 122
5.9 Experiment 4: Schematic representation of the target-masker asynchrony.124
5.10 Experiment 4: Mean and individual SRTs across a range of masker-target asynchronies. 126

5.11 Experiment 4: Results from comparable conditions across Experiments 1–4. 127

6.1 Experiment 5: Test of Everyday Attention: Map Search 140
6.2 Experiment 5: Example question from the SSQ. 143
6.3 Experiment 5: Pure-tone thresholds for young and older adults. 150
6.4 Experiment 5: SLT performance. 150
6.5 Experiment 5: Scatterplot of attentional and spatial listening performance measures. 157
6.6 Experiment 5: Models of spatial listening performance. 160

7.1 Experiment 6: Sequence of phrases 172
7.2 Experiment 6: Loudspeaker array used for the MEG recordings. 174
7.3 Experiment 6: The head and torso simulator used to record the stimuli. . 174
7.4 Experiment 6: Measuring the frequency response of ER30 tube-phones. 175
7.5 Experiment 6: Timing of keyword onsets. 176
7.6 Experiment 6: The magnetically-shielded scanning room. 179
7.7 Experiment 6: MEG data post-processing stream. 180
7.8 Experiment 6: MEG event processing stream. 181
7.9 Experiment 6: Definition of MEG data analysis windows. 182
7.10 Experiment 6: Selection of non-target data windows. 183
7.11 Experiment 6: Cortical model processing stream. 186
7.12 Experiment 6: MEG sensor groups for the sensor-space analyses. 190
7.13 Experiment 6: Example of the permutation analysis technique applied to evoked sensor data. 191
7.14 Experiment 6: Average FFT of evoked and induced data. 192
7.15 Experiment 6: Hemispheric sensor groups for spatial filtering analysis. . 195
7.16 Experiment 6: Latencies of the behavioural responses. 198
7.17 Experiment 6: Evoked power frequency spectra for the three epoch pairs. 199
7.18 Experiment 6: Evoked data within 0.5–30 Hz for the three epoch pairs. . 201
7.19 Experiment 6: Summary of evoked analysis on the sensor data. 202
7.20 Experiment 6: Induced data for the three epoch pairs. 205
7.21 Experiment 6: Decreases in β and γ power. 206
7.22 Experiment 6: Comparison of induced data for “Onset processing” epoch pair. 207
7.23 Experiment 6: Summary of induced analysis. 208
7.24 Experiment 6: Minimum-norm results - “Onset attention”. 210
7.25 Experiment 6: Minimum-norm results - “Call-sign processing” (0.5–4 Hz, Early window). 212
7.26 Experiment 6: Minimum-norm results - “Call-sign processing” (0.5–4 Hz, Late window). 213

7.27 Experiment 6: Minimum-norm results - “Call-sign processing” (4–8 Hz). 214
7.28 Experiment 6: Spatial filtering results for the “Onset attention” comparison. 215
7.29 Experiment 6: Spatial filtering results from 4–30 Hz. 216
7.30 Experiment 6: Spatial filtering results from 40–80 Hz. 217
7.31 Experiment 6: Locations of neural correlates of performance from the minimum-norm analysis. 221
7.32 Experiment 6: Scatter plots of performance during MEG imaging and MEG power differences in sensor- and source-space. 223
7.33 Experiment 6: Overlapping data windows. 229

8.1 Network of correlations between measures related to hearing, attention, and cortical activity. 242
8.2 Proposed attention-demanding visual task. 245

F.1 Appendix F: Results of the gating experiment. 272

List of Tables

3.1 The neural basis of three attentional functions. 47

4.1 Summary of comparison between Minimum norm and LCMV spatial filtering techniques. 99

5.1 Experiment 2: Differences in energetic masking between Experiments 1 and 2 114

6.1 Experiment 5: Descriptive information of participants. 138
6.2 Experiment 5: Adaptive procedure for the spatial listening task. 139
6.3 Experiment 5: Details of the 8 subtests of the TEA. 141
6.4 Experiment 5: Summary of the 10 measures of the TEA. 142
6.5 Experiment 5: 10 sub-scales of the SSQ. 143
6.6 Experiment 5: Correlation matrix for the different conditions of the SLT. 151
6.7 Experiment 5: Factor analysis of SLT data. 151
6.8 Experiment 5: Age-related differences in the TEA measures. 152
6.9 Experiment 5: Regression analysis of SLT performance. 154
6.10 Experiment 5: Self-reported scores and group differences for the SSQ sub-scales. 155
6.11 Experiment 5: Relationships between TEA subtests, SLT performance, and BEA levels. 156
6.12 Experiment 5: Relationships between pure-tone thresholds and SLT performance. 161

7.1 Experiment 6: Summary of epoch pairs for the MEG data analyses. . . . 189
7.2 Experiment 6: Summary of MEG data analyses. 196
7.3 Experiment 6: Individual differences in performance and hearing level. . 217
7.4 Experiment 6: Correlations between MEG performance and subtest of the TEA. 218
7.5 Experiment 6: Details of the windows containing neural correlates of performance. 219
7.6 Experiment 6: Correlations between sensor-space data and performance. 220
7.7 Experiment 6: Details of ROIs from the minimum-norm analysis. 220

7.8 Experiment 6: Neural correlates of performance extracted from the minimum-norm analysis. 222

C.1 Appendix C: SRTs for each of the 8 CRM voices. 261
C.2 Appendix C: SRTs for the three target locations. 263

G.1 Appendix G: Peak differences for “Onset attention” (minimum-norm). . 275
G.2 Appendix G: Regions of interest from 0.5–4 Hz in the early “Call-sign processing” window (minimum-norm). 276
G.3 Appendix G: Regions of interest from 0.5–4 Hz in the late “Call-sign processing” window (minimum-norm). 277
G.4 Appendix G: Peak differences for “Call-sign processing” (minimum-norm, 4–8 Hz). 278
G.5 Appendix G: Peak differences for “Onset attention” (spatial filtering, 8–13 Hz). 278
G.6 Appendix G: Peak differences for “Onset attention” (spatial filtering, 13–30 Hz and 40–80 Hz). 279
G.7 Appendix G: Peak differences for “Call-sign processing” (spatial filtering, 4–8 Hz). 280
G.8 Appendix G: Peak differences for “Call-sign processing” (spatial filtering, 8–13 Hz). 281
G.9 Appendix G: Peak differences for “Call-sign processing” (spatial filtering, 13–30 Hz and 40–80 Hz). 282
G.10 Appendix G: Co-ordinates of ROIs from the minimum-norm analysis. . . 283

Acknowledgements

I would like to sincerely thank my supervisor, Quentin Summerfield, for providing me with the opportunity to undertake this research and for his continuous support and encouragement.

I would like to acknowledge the other members of my research committee, Peter Bailey and Gary Green, for their help and support over the course of my PhD. Without exception, their criticisms have been highly constructive and their suggestions have been helpful.

I would also like to thank the staff of the York Neuroimaging Centre, including André, Becky, Will, Mark, and Michael. I am especially grateful to André Gouws for his technical support and assistance with MEG, and Will Woods for his comments on Chapter 4.

Many thanks to Ralph Holme and Catriona Crombie of the RNID for their support and for their assistance in recruiting older adult participants.

I would like to thank Paul, Rosie, Cindy, Vicky, Isa, Paula, Sam, and the many other friends that I have made in the Department of Psychology over the past three years for keeping me smiling and sane.

Finally, I would like to sincerely thank my parents for instilling in me the belief that what you can achieve is only limited by what you set out to achieve, and for providing me with the love and support without which I would certainly not have achieved this.

This thesis is dedicated to the memory of Dr. Anne Leahy.

Financial Support

This PhD was generously supported by a studentship from the Royal National Institute for Deaf People (RNID).

Additional funds were provided by the guarantors of *Brain* which enabled me to attend the Mid-Winter Meeting of the Association for research in Otolaryngology in Denver, Colorado (February 2007).

Declaration

This thesis comprises the candidate's own original work and has not, whether in the same or different form, been submitted to this or any other University for a degree. All experiments were designed by the candidate with assistance from the supervisor. All testing and analyses were conducted by the candidate.

Conference Presentations & Proceedings

Kitterick, P. T., & Summerfield, A. Q. (2006). The Role of Attention in the Spatial Perception of Speech. *Paper presented at the Annual Meeting of the British Society of Audiology*, Cambridge, UK.

Kitterick, P. T., & Summerfield, A. Q. (2007). Difficulties with Listening in Noisy Environments: Looking Beyond the Ear. *Paper presented at the RNID Hearing for the Future Conference*, London, UK.

Kitterick, P. T., & Summerfield, A. Q. (2007). The Role of Attention in the Spatial Perception of Speech. *Paper presented at the 30th Mid-Winter Meeting of the Association for Research in Otolaryngology*, Denver, USA.

Kitterick, P. T., Paul, I., Bailey, P. J., & Summerfield, A. Q. (2007). Preliminary Evidence of a Cortical Network for Spatial Listening in Multi-talker Environments as revealed by Magnetoencephalography (MEG). *Paper presented at the Annual Meeting of the British Society of Audiology*, London, UK.

Kitterick, P. T., Paul, I., Bailey, P. J., & Summerfield, A. Q. (2008). Preliminary Evidence of a Cortical Network for Spatial Listening in Multi-talker Environments as revealed by Magnetoencephalography (MEG). *Paper presented at the 31st Mid-Winter Meeting of the Association for Research in Otolaryngology*, Phoenix, USA.

Kitterick, P. T., Bailey, P. J., & Summerfield, A. Q. (2008). Individual Differences in Cortical Activation During Spatial Attention to Speech. *Paper presented at the Annual Meeting of the British Society of Audiology*, York, UK.

Chapter 1

Overview of Thesis

This chapter presents an introduction to the thesis, an outline of the main experiments, and a brief overview of each of the chapters.

1.1 Introduction

Older adults report difficulties with listening in noisy environments. Such difficulties have been associated with self-reported handicap, specifically in situations in which high attentional demands are placed on the listener (Gatehouse & Noble, 2004). These situations involve the focussing of attention on what one person is saying when many other people are speaking at the same time, or the switching of attention between talkers. Difficulties with coping in complex listening situations can result in avoidance, leading to isolation and a negative impact on quality of life (Noble, Ter-Horst, & Byrne, 1995).

A significant body of research concluded that the speech perception difficulties experienced by older adults are principally due to the decline in peripheral sensitivity that accompanies the natural ageing process (e.g. Plomp & Mimpen, 1979; van Rooij & Plomp, 1992). However, recent studies by Gatehouse, Akeroyd, Deas, Glover, Howell, & Lawson (2006) and Helfer & Freyman (2008) have suggested that cognitive decline should be considered as a significant additional contributor to age-related difficulties in speech perception.

The manner in which cognitive functions contribute to the ability to listen to what one person is saying when many other people are speaking at the same time is not fully understood. The overall goal of this thesis was to provide a better understanding of the role that cognitive processes play in perceiving speech in complex listening environments. The first aim was to develop an attention-demanding task of spatial listening for speech which more closely resembled the environments in which listeners report difficulties in everyday life, compared to tasks developed previously for multi-talker listening research. The second aim was to use the task to examine the relationships between speech-perception deficits, difficulties that older adults report in everyday life, and variables which may contribute to those difficulties, including hearing loss and cognitive ability.

The third aim was to translate the listening task so that it could be performed while brain activity was recorded using magneto-encephalography (MEG). The ability to measure the cortical activity associated with successful performance on an attention-demanding listening task would provide insights into the nature of cognitive processing associated with the task. Using MEG, it is possible to record cortical activity with high temporal precision. The technique also has the potential to resolve cortical sources with sub-centimetre resolution (Sekihara, Sahani, & Nagarajan, 2005). No previous study has applied this imaging technique to a multi-talker spatial listening task.

The first four experiments evaluated performance in several versions of a spatial listening task for speech among young normally-hearing listeners. The benefits from knowing *who* to listen for, *where* they would speak from, and *when* they would speak were studied individually and in combination with each other. The experiments also examined the distracting effect of a person starting to speak at the same time as, or shortly before or after, the person to whom the listener wishes to attend. The fifth experiment compared the performance of young and older normally-hearing adults on the spatial listening task. The experiment examined the relationships between performance, self-reported difficulties with listening in everyday situations, hearing sensitivity, and measures of auditory and visual attention. This experiment also identified the condition of the spatial listening task which exposed the largest individual differences in performance. The sixth experiment examined the location and time-course of cortical activity associated with successful performance of this condition of the task using MEG, and identified neural correlates of performance.

1.2 Overview of Following Chapters

Chapter 2: Hearing, Ageing, and Speech Perception

This chapter examines the prevalence of hearing loss and speech perception difficulties amongst elderly people, and the impact that such difficulties can have on their quality of life. The chapter presents evidence that speech perception difficulties are associated with ageing, and that age-related deficits in cognitive function contribute to those difficulties. The chapter considers evidence that attention-demanding tasks of speech perception which occur in everyday situations contribute to self-perceived disability amongst the elderly. The chapter provides an overview of the mechanisms which facilitate spatial listening and speech segregation.

Chapter 3: Attention and Spatial Listening

This chapter outlines the attentional mechanisms that are involved in multi-talker listening, where the listener may need to focus on information from one talker, or follow several streams of information, in the presence of irrelevant speech or noise. The chapter presents evidence to support the suggestion that attention is important in spatial listening, that abrupt isolated stimulus onsets capture attention automatically, and that reducing uncertainty about properties of a target stimulus affects the attentional strategies that listeners adopt. The chapter provides an overview of the cortical mechanisms which have been associated with attention, and discusses studies which have examined cortical activity during speech-in-noise tasks. The chapter presents evidence that attentional strategies change with increasing age, and that those changes may reflect a decrease in automatic processing among elderly adults.

Chapter 4: Reconstructing Cortical Activity with High Temporal Resolution

This chapter presents an overview of methods used to localise the sources of neural activity which give rise to changes in the extra-cranial magnetic fields measured with MEG. The chapter discusses two such techniques, minimum-norm estimation and spatial filtering, their underlying fundamental concepts, and issues related to their application to MEG data. The chapter discusses the strengths and limitations of the techniques, and considers the application of the techniques to the analysis of brain activity associated with performance on a spatial listening task for speech.

Chapter 5: Voluntary and Involuntary Attention in a Multi-talker Environment

This chapter reports four experiments which examined the ability of listeners to perceive speech in a multi-talker environment using a spatial listening task for speech. In the task, several phrases are presented in an overlapping sequence in which a new phrase starts every 800 ms. Each phrase contains several keywords—a unique call-sign, a colour, and a number. The task required participants to attend to each new phrase onset and to determine whether the call-sign matched a call-sign that had been allocated to them. Participants reported the colour and number keywords from the phrase containing their allocated call-sign. Experiments 1 and 2 examined the benefits of providing prior information about the phrase containing the target call-sign—*who* would speak it, *when* it would be spoken, and *where* it would be spoken from. In Experiment 1, phrases started one at a time. In Experiment 2, phrases started in pairs. Using the same task, Experiments 3 and 4 examined the distracting effects of a phrase which started shortly before or after a target phrase. In Experiment 3, the onset of the target phrase was unpredictable in time. An asynchrony was introduced between the onset times of the target phrase and a paired masker phrase by varying the onset time of the target phrase. In Experiment 4, uncertainty about when the target phrase would be spoken was reduced by placing it within a temporally-regular sequence of phrases. A target-masker asynchrony was introduced by varying the onset time of the masker phrase.

Chapter 6: Relationships among Age, Hearing Level, Attentional Abilities, and Performance on a Spatial Listening Task

The experiment reported in this chapter (Experiment 5) investigated the relationship between performance on the tasks of spatial listening from Experiments 1 and 2, attentional ability as measured using an attentional test battery, hearing sensitivity, and self-reported difficulties with listening in everyday situations. The attentional battery included purely visual, purely auditory, and audio-visual tests. Hearing sensitivity was assessed by pure-tone audiometry. A questionnaire elicited self-reported measures of listening difficulties in everyday environments. Young and older normally-hearing adults were tested to examine age-related differences in the ability to cope with complex listening situations.

Chapter 7: Cortical Activation Patterns During a Spatial Listening Task

This chapter reports an experiment (Experiment 6) which examined the neural bases of focussing attention and resisting distraction in a multi-talker environment using MEG. The spatial listening task from Experiment 1 was adapted so that the task could be performed while lying supine in an MEG scanner. Data were collected for the groups of young and older adults who participated in Experiment 5. Neural activity was examined at key moments in the spatial listening task: when participants had to discriminate between target and non-target call-signs, when attention had to be sustained on a target phrase, and when the onset of a new phrase had to be ignored. The experiment examined the relationships between differences in MEG power at those key moments and performance in the spatial listening task.

Chapter 8: Summary and General Discussion

This chapter presents a summary of the main findings and conclusions from the six experiments. Several issues arising from the research are discussed, and directions for future research are proposed.

Chapter 2

Literature Review: Hearing, Ageing, and Speech Perception

This chapter examines the prevalence of hearing loss and speech perception difficulties amongst elderly people, and considers the impact that such difficulties can have on their quality of life. The chapter presents evidence that speech perception difficulties are associated with ageing, and that age-related deficits in cognitive function contribute to those difficulties. Evidence that attention-demanding tasks of speech perception contribute to self-perceived disability amongst the elderly is also considered. An overview of the mechanisms which facilitate spatial listening and speech segregation is provided.

2.1 Introduction

In this review, I will focus on the bases of difficulties that elderly people experience with the spatial perception of speech in noise, with particular emphasis on the role of attention. I will begin by looking at the prevalence of hearing loss and speech perception difficulties among elderly people, and the impact that such difficulties can have on their quality of life. I aim to show that speech perception difficulties are associated with ageing, that older adults report experiencing difficulties in attention-demanding tasks of speech perception which occur in everyday situations, and that those difficulties contribute to self-perceived handicap among the elderly. In light of this, I will then examine the mechanisms which facilitate spatial listening and speech segregation. I aim to show that there is a wide range of factors involved in our ability to segregate speech from background noise, or competing talkers.

2.2 Hearing and Ageing

2.2.1 Introduction

In this section I aim to show that hearing loss is prevalent in elderly adults and that impaired hearing has a direct effect on the ability to understand speech. I will discuss the evidence that cognitive factors are also likely to contribute to this deficiency in speech perception, particularly in demanding listening environments which contain high levels of background noise. I intend to highlight studies which have shown links between hearing disability and self-perceived handicap, particularly in listening situations which impose high attentional demands.

2.2.2 Hearing Difficulties

2.2.2.1 Natural Hearing Loss

Hearing sensitivity declines naturally with age. This natural loss, or *presbycusis*, occurs gradually and has the largest effect at mid to high frequency ranges (approximately above 2 kHz). The causes of presbycusis are complex—there are many possible contributors which can occur over an extended period of time and in a large variety of environments. Prolonged exposure to moderate-intensity noise, e.g. traffic, noisy equipment, loud music, is an obvious factor for inclusion. Alterations to the blood supply, linked with high blood pressure, heart disease, genetic predisposition, and circulatory problems may also play a role in peripheral auditory impairment. These factors can result in functional changes not only to the cochlea, but also to the middle and outer ears.

2.2.2.2 Peripheral Physiological Changes

Four types of presbycusis were identified by Schuknecht & Gacek (1993): sensori-neural, neural, strial (metabolic), and conductive. Sensori-neural presbycusis involves loss of sensory and supporting cells in the organ of Corti in the cochlea (Figure 2.1). Loss of outer hair cells is predominant, although inner hair cell loss is also observed among older people (Scholtz, Kammen-Jolly, Felder, Hussl, Rask-Andersen, & Schrott-Fischer, 2001). Neural presbycusis is an atrophy (cell death) of spiral ganglion cells in the auditory nerve. This affects output from the cochlea as a whole, leading to a uniform loss of sensitivity across frequency. This category of hearing loss is distinct from the high-frequency loss often associated with age-related hearing loss. However, neural presbycusis can lead to a decrease in speech discrimination that is disproportionate to the loss of sensitivity. Strial presbycusis involves atrophy of the stria vascularis which maintains the metabolic health of the cochlea. Again, the whole cochlea may be affected, but speech discrimination is usually preserved (Schuknecht

& Gacek, 1993). Finally, conductive (mechanical) presbycusis involves a thickening and/or stiffening of the basilar membrane. It is generally more severe in the basal turn which results in pronounced high-frequency loss. The four categories of presbycusis are not exclusive, and may co-occur (Scholtz et al., 2001).

2.2.2.3 Prevalence of Natural Loss

The most authoritative study of the prevalence of adult hearing impairment in the UK was conducted by Davis (1989). He reported that 37% of adults between the ages of 61–70 years and 60% between 71–80 years had a hearing loss in both ears of, or exceeding, 25 dB HL. A Nordic study of 1409 75-year olds in Finland, Sweden, and Denmark (Hietanen, Era, Henrichsen, Rosenhall, Sorri, & Heikkinen, 2005) found that 84–92% of men and 71–76% of women had at least a mild hearing impairment, defined as 21 dB HL or greater.

In addition to audiological measures, Davis (1989) collected self-reported measures of difficulties with hearing in quiet. The results showed that 15% of adults between the ages of 61-70 years reported at least a slight difficulty with hearing in quiet in their better ear, and 28% in their worse ear. The prevalence increased to 25% (better ear) and 38% (worse ear) between the ages of 71–80 years. A postal study in Scotland also examined self-reported difficulties through the use of a questionnaire (Hannaford, Simpson, Bisset, Davis, McKerrow, & Mills, 2005). The results showed that 43% of men and 22% of women between 60–74 years, and 56% of men and 41%

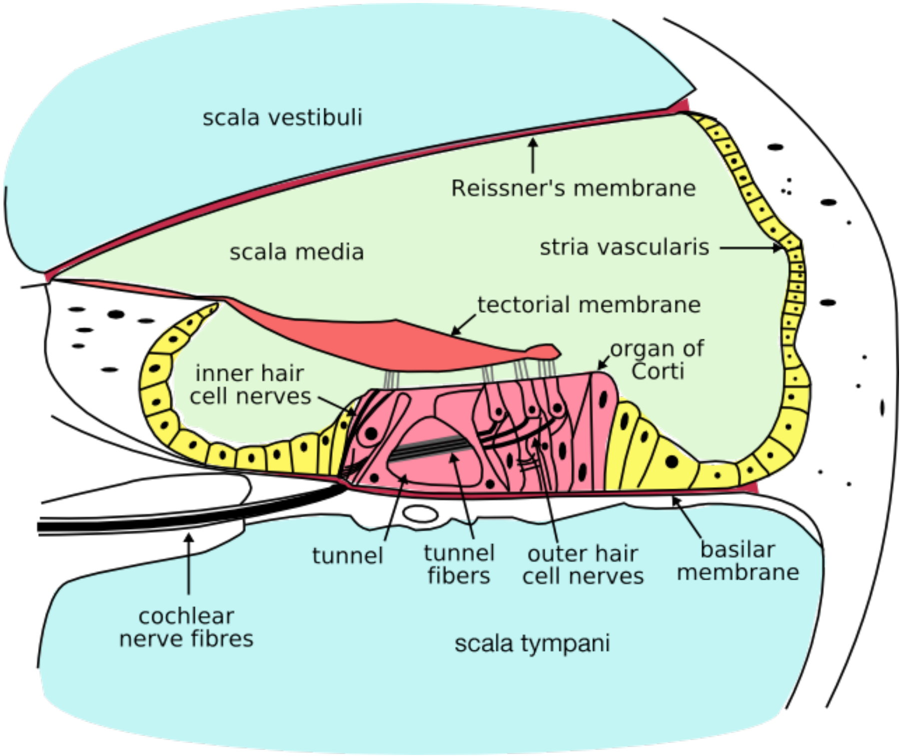

Figure 2.1. Cross-section of the human cochlea showing key features of the inner ear, including the inner and outer hair cells (adapted from Ropshkow, 2004).

of women over 75 years reported difficulty with hearing. Hietanen et al. (2005) found that between 41–57% of men and 28–37% of women aged 75 reported at least minor difficulties with hearing. From these results, it is clear that hearing loss and self-reported hearing difficulties are common amongst people over the age of 60, with up to a half of all men and a third of all women affected.

Hearing difficulties may be more prevalent than statistics would suggest. A longitudinal study was carried out over the course of four years on 2150 adults aged between 40–79 years in Japan (Uchida, Nakashima, Ando, Niino, & Shimokata, 2003). Pure-tone audiometric thresholds were obtained, along with a questionnaire on hearing problems. Uchida et al. (2003) found that elderly people tended to underestimate their level of hearing impairment more than middle-aged adults. Despite the fact that decline in auditory peripheral sensitivity in men can occur at a rate of twice that in women, and that the onset occurs earlier in life for men (Pearson, Morrell, Gordon-Salant, Brant, Metter, Klein, & Fozard, 1995), Uchida et al. (2003) reported that this under-estimation of hearing difficulties was more common among men than women.

2.2.2.4 Gender Differences in Sensitivity with Age

In general, natural hearing loss is more pronounced in men than in women. The Baltimore Longitudinal Study on Ageing (Pearson et al., 1995) measured the hearing levels of men and women between the ages of 17 and 90 years of age for a period of over 23 years. The findings revealed that the rate of decline in hearing sensitivity is almost twice as rapid for men as women at most frequencies and ages. The onset of this natural decline is generally later in women, around 50 years of age, with the decline appearing from age 30 in men. This imbalance between the degree of hearing loss in men and women has also been found in other studies of ageing (Davis, 1989; Hietanen et al., 2005; Uchida et al., 2003). The nature of the age-related decline in sensitivity also differs between men and women in terms of frequency characteristics. The decline occurs at all frequencies in men from onset at around 30 years of age, while for women it is detectable at 500 Hz and 8 kHz at 30 years and at other frequencies between 60-70 years of age (Pearson et al., 1995). The root cause of this difference in onset age of natural decline between genders has yet to be determined.

2.2.3 Speech Perception and Ageing

2.2.3.1 Introduction

Elderly people experience greater difficulties in understanding speech than younger listeners (Chaba, 1988; Martin & Jerger, 2005). While speech-perception performance in quiet is relatively well preserved (Davis, 1989; Wiley, Cruickshanks, Nondahl, Tweed, Klein, & Klein, 1998), as long as high-frequency information is still audible and talkers are familiar (Yonan & Sommers, 2000), difficulties arise when following a conversation in the presence of other talkers or background noise. Several studies have shown that speech comprehension is poorer in elderly adults compared to younger listeners in the presence of other talkers (Duquesnoy, 1983; Humes, Lee, & Coughlin, 2006), speech babble (Dubno, Dirks, & Morgan, 1984), and random noise (Bronkhorst & Plomp, 1989; Helfer & Wilber, 1990). The extent of such difficulties can depend on many different factors, including the audibility of the speech, whether words or sentences are used as stimuli, and the type of noise background; e.g. speech or speech-like noise (Gordon-Salant, 2005).

In this section, I will outline the main factors implicated in age-related declines in speech perception. I aim to show that older adults experience difficulties with perceiving speech when there is background noise, and that age-related difficulties with speech perception are common. I intend to highlight several studies which have examined the ability of older adults to hear what one person is saying in the presence of background noise which includes speech. I will present evidence that such difficulties have been linked with an age-related decline in cognitive performance, and those difficulties in turn are related to self-perceived handicap.

2.2.3.2 Causal Factors

A commission of the Committee on Hearing, Bioacoustics, and Biomechanics (CHABA) of the National Research Council (Chaba, 1988) reviewed the available studies on the difficulties in speech comprehension experienced by older adults. Their findings included three possible explanations for these difficulties: 1) changes in the auditory periphery, including sensori-neural hearing loss, 2) changes in the way sensory information is processed in the ascending auditory nervous system, between the brainstem and the primary auditory areas of the cortex, and 3) a decline in overall cognitive performance.

The first of these factors, peripheral hearing loss, has been consistently associated with deficits in speech perception (Festen & Plomp, 1983; Helfer & Wilber, 1990; Humes, 1996; Jerger, Stach, Pruitt, Harper, & Kirby, 1989; Plomp & Mimpen, 1979; van Rooij & Plomp, 1992). Reduced input to central auditory processes at high frequencies, which is a characteristic of age-related hearing loss, affects the comprehension of many speech sounds. For example, Humes (1991) reported that many fricatives and some stops, both voiced and voiceless, are mainly discriminable on the basis of frequency information over 2 kHz. Thus, despite the fact that individuals with high-frequency hearing loss may report 'hearing' speech, limitations in the ability to detect or resolve high-frequency information usually lead to difficulties in understanding speech (Humes, 1991).

Another facet of hearing loss which affects speech perception is the loss of outer hair cells, defined earlier as sensory-neural presbycusis (Schuknecht & Gacek, 1993).

This loss has the effect of reducing the frequency resolution of the organ of corti with a broadening of the bandwidths of auditory filters (Tyler, Hall, Glasberg, Moore, & Patterson, 1984), independent of age (Peters & Moore, 1992; Sommers & Humes, 1993). One consequence of this broadening is an increase in the amount of mutual masking between concurrent signals (Darwin, 2006; Moore, 1986).

Deficits in several cognitive functions have also been linked to speech perception difficulties (Gordon-Salant, 2005). Studies which have examined the relationship between cognition and speech perception in young and older adults are discussed in Section 2.2.3.5 of this chapter.

2.2.3.3 Prevalence of Difficulties in Speech Comprehension

Ageing studies have examined the prevalence of difficulties with speech perception in the presence of background noise. Davis (1989) reported that 35% of 61–70 year olds and 44% of those between the ages of 71–80 years reported difficulties understanding speech in noise. Hannaford et al. (2005) found that 44% of men and 26% of women between 60–74, and 61% of men and 43% of women over 75 reported difficulties when following a conversation in the presence of background noise. Thus, approximately 50% of adults over the age of 60 years of age in these studies have reported difficulties listening to speech in the presence of background noise.

2.2.3.4 Nature of Age-related Difficulties in Speech Perception

Studies which have examined the ability of older adults to perceive speech against a background of noise have highlighted aspects of those tasks which pose a particular difficulty for older adults. In this section, several of those aspects will be discussed, including the ability of older adults to ignore irrelevant speech, their susceptibility to distraction based on talker/masker gender differences and to informational masking, their reliance on voice familiarity, and the relationship between speech perception difficulties and hearing loss.

By varying the language in which speech maskers were spoken, Tun, O'Kane, & Wingfield (2002) reported that normally-hearing older adults have greater difficulty in ignoring background speech when it is understandable. Meaningful target and masker speech sentences were presented diotically over headphones, and the language of the masker phrases was varied from English to Dutch. For both maskers, the speech was syntactically- and semantically-correct. None of the participants spoke Dutch. Older adults were affected to a greater extent by the presence of English speech distractors compared to the Dutch maskers. This effect remained after controlling for the participants' speech-in-noise performance. No difference was found between the distracting effects of the two masker types for the young adults. Furthermore, when the participants were asked to indicate whether lists of words had been presented in either target or distractor sentences, the young adults were more accurate at recalling the content of the masker stimuli. The results suggest that there is an age-related increase in susceptibility to distraction from competing speech messages which are meaningful. The superior recognition performance of the young adults was interpreted as a reflection of more extensive processing of the non-attended stimuli compared to older adults. Despite this detailed processing, the young adults were able to resist distraction from the meaningful stimuli more successfully than the older adults. Thus, Tun et al. (2002) suggested that executive control is important in resisting distraction, and that decreased executive function in older adults makes them more susceptible to distraction, particularly when the background speech is intelligible.

In situations in which a high cognitive demand is not placed on the listener, there is evidence that older adults do not show a deficit in the ability to ignore irrelevant speech. Li, Daneman, Qi, & Schneider (2004) examined the effects of irrelevant speech and noise maskers in a word recognition task in young normally-hearing and older near normally-hearing adults. Pairs of target and masker sentences were presented using two loudspeakers, positioned 45° to the left and right of the participant. The target stimuli comprised a set of syntactically correct but semantically meaningless sentences (e.g. "The goose can kick a street"), each of which contained three key-words. The masker stimuli included speech-spectrum steady-state noise stimuli and speech stimuli. The masker speech stimuli comprised similar meaningless sentences which were mixed together—each stimulus was derived from two different sentences. Participants had to repeat the target phrase out loud, and trials were scored on the number of correct keywords. The precedence effect was used to vary the perceived location of the target and masker phrases. The target phrases were presented 3 ms earlier on the right than on the left—participants perceived the phrase as being located on the right. The perceived location of the masker phrases was varied by introducing a 3 ms lag for the phrase on the right, or on the left, or by omitting the lag. This created perceived masker locations on the left, right, and at the mid-point between the loudspeakers respectively. The signal-to-noise ratio was varied randomly between -12, -8, -4, and 0 dB.

Li et al. (2004) found similar levels of release from masking in both age groups when the masker was perceived at a different location than the target, and a similar difference in the size of the masking release between noise and speech maskers across the two groups. The only age-related difference was that the older adults required a higher signal-to-noise ratio (SNR) to achieve the same performance level as the young adults. As the study used the precedence effect to perceptually segregate the phrases, the difference between the signals arriving at the ears was similar in all conditions, which in turn meant that the amount of energetic masking of the target speech stimulus by the speech or noise masker stimuli was equivalent across conditions. Therefore, differences in performance arose purely as a consequence of the perception that the stimuli were segregated in space; i.e. the differences arose in the ascending auditory pathway due to binaural interactions rather than at the periphery. Thus, Li et al. (2004) interpreted the results as evidence that older adults do not have a greater difficulty at ignoring or inhibiting irrelevant speech while attending to other speech when cognitive demands are relatively low.

Thus study is limited in its relevance to understanding the difficulties that older adults report in everyday situations. First, the task did not engage cognitive functions related the analysis of semantic, linguistic, and contextual cues in the stimuli. This is principally due to the use of semantically-meaningless sentences both for the target and speech masker stimuli. Secondly, the location of the target phrase, on the right, was fixed and therefore predictable. Thus, although the participant had to focus their attention on the location of the target phrase and sustain their attention on it while ignoring the masker stimuli, participants did not have to specifically shift their attention to the target location on each trial. Thirdly, the onset of the masker and target phrases was predictable—each trial was initiated by the participant pressing a button which started the masker and 1 sec later the target phrase was presented. The participant was not required to maintain a state of vigilance to detect an unpredictable phrase onset which is often required in everyday situations. Thus, although the results of Li et al. (2004) suggest that not all cognitive functions are significantly affected by the ageing process, they cannot be generalised to the attention-demanding environments in which older adults report experiencing difficulties.

Everyday listening situations containing multiple talkers can be approximated by using syntactically-correct and semantically-meaningful speech stimuli for both target and masker phrases. In one such study, Helfer & Freyman (2008) examined the effect of multiple speech maskers on the perception of speech in young normally-hearing adults and older adults with a range of hearing sensitivities. The audiometric thresholds within the older adult group ranged from normal to moderate high-frequency loss. The target stimuli were sentences, each with a specific topic and spoken by a female talker. Each sentence contained three key-words of either one or two syllables. Speech maskers were constructed using sentences with different topics to those of the target stimuli and comprised two male or two female talkers. Noise maskers were constructed by modulating white noise with the envelope extracted from the speech maskers. The SNR and spatial location of the masker stimuli, either coincident with the target or separated by 60°, were varied. Unlike Li et al. (2004), the target phrase was was identified by its topic, which was presented visually prior to and during the auditory presentation of the target and masker sentences. The task was to verbally repeat the target phrase and accuracy was based on the number of key-words correctly reported. Helfer & Freyman (2008) observed that overall, older adults performed worse at the task than the young adult group. The smallest differences between the age groups arose when the target and maskers were both spoken by female talkers and presented from the same spatial location. Helfer & Freyman (2008) pointed out that this condition produces the maximum amount of informational masking, as there was a high probability of confusion between target and masker phrases. When the same target and masker stimuli were spatially-separated, both age groups received comparable amounts of release from informational masking. This result is compatible with the results of Li et al. (2004); i.e. older adults do not exhibit a higher level of susceptibility to confusion between highly-similar signals compared to young normally-hearing adults.

The largest differences between young and older listeners were identified when the masker talkers were of a different sex to the target talker. Helfer & Freyman (2008) interpreted this effect as being independent from informational masking, as little confusion should arise between the target and masker phrases when spoken by talkers of a different sex. Moreover, no effect of spatial separation was found for these stimuli. The distracting effects of opposite-sex maskers have been observed in previous studies of age-related differences in speech perception (Humes et al., 2006; Peters, Moore, & Baer, 1998). Helfer & Freyman (2008) provided two possible explanations for this age-related decrement in performance with opposite-sex maskers. The first is an attentional hypothesis in which older adults are unable to ignore the semantic content of the masker stimuli, or to allocate sufficient resources to the target stimuli. The second hypothesis is that the older adults cannot take advantage of “spectral fluctuations” in either of the speech maskers. In the same-sex masker condition, the performance of both groups is impaired by informational masking but in the opposite-sex condition, the younger adults were possibly able to take advantage of differences in the frequency content of the target and masker speech stimuli, even if transient, to segregate the target from the mask speech. This ability might be impeded for older adults due to a decline in frequency resolution which has been linked to cochlear hearing loss, independent of age (Peters et al., 1998). In summary, the results of Helfer & Freyman (2008) suggest that while older adults are not more susceptible to informational masking than young normally-hearing adults, i.e. confusion between two similar-sounding speech signals, they are poorer at ignoring an opposite-sex masker. This decrement in performance could be related to changes in the cochlea, leading to a decrease in frequency resolution when decoding auditory signals, or to a central deficit in the allocation of attentional resources to the task of inhibiting the distracting information.

In everyday situations, it is common for the listener to have some familiarity with the voice of the person they are attending to. It has been suggested that the ability to take advantage of knowledge about voice characteristics changes with age, and may contribute to age-related speech perception difficulties. Yonan & Sommers (2000) familiarised a group of young normally-hearing and older mildly hearing-impaired listeners with four talkers, two male and two female. Participants were presented with sentences spoken by the four talkers and the name of the talker was visually presented. An identification test was performed to assess the ability to associate a name with each talker and also the ability to discriminate between the talkers. Older adults were poorer at recognising the four voices compared to the young adults. It was suggested that deficits in memory may decrease the ability of older adults to store and recall voice characteristics for later use as a cue in speech segregation and perception. To examine whether older adults were able to use talker familiarity to aid speech perception in noise, sentences were presented in white noise at different SNRs (-5, 0, and +5 dB). Half of the sentences were spoken by the four talkers familiar to the participants, and half were spoken by novel talkers. Yonan & Sommers (2000) found that the older adults benefited from talker familiarity as much as or more than the young adults, particularly at more adverse SNRs. The results suggest that older adults rely on a familiarity with voices to extract and segregate speech from a background of noise, despite requiring longer to become familiar with a voice and exhibiting deficits in storing and recalling voice characteristics.

Johnsrude, Mackey, Alexander, Macdonald, & Carlyon (2008) reported that not only can talker familiarity aid the process of listening to a familiar talker in the presence of other unfamiliar talkers, but it can also be used to aid the process of listening to an unfamiliar talker in the presence of familiar talkers, and that this ability changes with age. Pairs of phrases were presented to older adult listeners, and each phrase contained three unique key-words—a call-sign, a colour, and a number. One of the phrases contained the call-sign 'Baron', the target phrase, and participants were instructed to report the colour and number keywords in that phrase. The talkers who spoke the target and paired masker phrases were both unfamiliar, or either the target or masker talker was familiar to the listener. Johnsrude et al. (2008) found that for adults from 45–60 years of age, performance was higher when either the target or masker talker was familiar, compared to when both were unfamiliar. Thus, the listeners were able to benefit from familiarity with either the talker they were listening to, or the talker that they had to ignore. For the older listeners (60–79 years of age), performance only improved when the target talker was familiar relative to when both target and masker talkers were unfamiliar. The results suggest that the ability to take advantage of knowledge about the vocal characteristics of talkers changes with age, specifically the ability to ignore a familiar talker and attend to an unfamiliar talker. It is possible that deficits in this ability may contribute to difficulties that older adults experience in multi-talker environments.

The comorbidity of hearing loss (Davis, 1989; Hannaford et al., 2005) and speech perception difficulties (Wiley et al., 1998) among older adults suggests that the difficulties may largely be due to impaired input to the auditory system. Dubno et al.

(1984) examined the effects of age and hearing loss on the speech perception ability of two groups of listeners: adults less than 44 years of age and greater than 65 years of age. Both groups contained individuals with a range of hearing sensitivities, from thresholds within normal ranges to moderate sensori-neural hearing loss. SRTs were measured adaptively for words and for high- and low-predictable sentences, in quiet and against a background of multi-talker babble. Age-related deficits in performance in the speech-in-noise tasks were observed which were independent of hearing loss. No such effect was found for the speech-in-quiet tasks. Dubno et al. (1984) suggested that the ability of adults to perceive speech against a background of speech is affected by both age and hearing loss. The authors speculated that central factors may have contributed to the age-related differences in performance, including working memory and feature extraction, although changes within the auditory system which are not measured by the audiogram, such as a decrease in the frequency selectivity (Patterson, Nimmo-Smith, Weber, & Milroy, 1982), may also play a role. These findings provide evidence that hearing loss alone is insufficient to account for the deficits in speech perception amongst older adults.

In summary, there is evidence that the bases of difficulties with speech perception that older adults experience are numerous. In this section, evidence has been presented which suggests that the ability of older adults to ignore irrelevant speech which is comprehensible is impaired. Although there is evidence that they do not exhibit a greater susceptibility to informational masking, they display deficits in their ability to ignore background speech when it is spoken by a talker of a different gender to the talker they are attending to. Familiarity with a talker provides a benefit to older adults in situations in which the familiar talker is being attended to. Finally, evidence was presented that difficulties with speech perception among older adults cannot be fully accounted for by peripheral hearing loss alone.

2.2.3.5 Cognitive Factors which Affect Speech Perception

Many cognitive factors may affect the perception of speech, including working memory, attention, and speed of processing (Gordon-Salant, 2005). Through self-reported measures, Gatehouse & Noble (2004) identified that difficulties with tasks of speech perception which include a high attentional component could not be fully explained by degree of hearing loss. The implication is that there is a significant contribution from deficits in central processing to difficulties in listening to speech in noisy environments, and that age-related difficulties with speech perception cannot be accounted for solely by a decline in peripheral sensitivity.

Age-related deficits in cognitive performance are often accompanied by a decline in hearing sensitivity. It is therefore unclear whether sensory impairment contributes to the impairment of cognitive function or is an independent processes. Zekveld, Deijen, Goverts, & Kramer (2007) examined the links between cognitive deficits and cochlear hearing loss across adults ranging from 24 to 72 years of age. To avoid confounding cognitive measures with sensori-neural loss, the cognitive performance of the participants was assessed using purely visual tasks. Furthermore, the material for all of the tasks was exclusively non-verbal. The cognitive assessment comprised an IQ test, and several aspects of the Cambridge Neuropsychological Test Automated Battery (CANTAB) including pattern recognition, sustained attention, and spatial working memory tests. Higher IQ scores were found to be related to improved performance on the sustained attention task, and older age was related to lower working memory performance. Hearing loss was found to be related to the strategy that participants used during the working memory task—individuals with higher levels of hearing loss were found to adopt search strategies that were considered efficient for the working memory task. No significant correlations were found between hearing loss and performance on any of the cognitive tasks. Zekveld et al. (2007) concluded that hearing loss was not a predictor of deficits in sustained attention, pattern recognition, or working memory and that previous studies which have found links between sensori-neural hearing loss and cognitive factors may have failed to control for auditory and verbal confounds in their cognitive tasks. The correlation between hearing loss and working memory strategy was interpreted as evidence that individuals with hearing loss adopt strategies which make extensive use of cognitive processes, such as working memory, to compensate for the degraded sensory input. Thus, although a decline in hearing sensitivity leads to a greater reliance on higher-level cognitive functions, it does not predict a decline in cognitive performance. However, the reliance of older adults with hearing loss on strategies which place a high load on cognitive resources may expose any deficits in cognitive ability.

Age-related difficulties with speech perception have been linked to cognitive deficits arising in central auditory processes and higher-level processes which are independent of any single sensory modality (Chaba, 1988). Distinguishing between the effects of either group of processes requires that the ability of older adults to perceive speech in noisy environments is assessed together with performance on non-auditory cognitive tasks. In one such study, George, Zekveld, Kramer, Goverts, Festen, & Houtgast (2007) examined the relationships between speech reception thresholds (SRTs), auditory factors including hearing sensitivity, and non-auditory factors in the form of text reception thresholds (TRTs) and age. SRTs were measured by presenting a sentence and a noise masker monaurally to the participant's better ear and by adaptively altering the ratio between the level of the speech and masker. Stationary and modulated (square-wave) noise maskers were used. The text reception threshold (TRT) was measured by adaptively varying the amount of a visually-presented sentence that was masked by a vertical grating. Participants included normally-hearing and hearing-impaired older adults. George et al. (2007)

found that within the group of normally-hearing older adults, TRTs explained a significant proportion of the variance in the SRTs but hearing sensitivity did not. In contrast, performance in the hearing-impaired group was strongly related to hearing sensitivity, while TRTs accounted for only a small but significant proportion (9%) of SRT variance when modulated noise was used. The results support the hypothesis that individual differences in the ability to segregate speech from background noise arise, at least in part, from deficits in higher-level modality-independent cognitive functions.

Deficits in working memory capacity may also limit speech perception performance in complex environments. Working memory has been suggested to be a system of limited capacity which is responsible for facilitating the detailed processing of sensory information (Baddeley, 2003). Humes et al. (2006) examined the relationship between working memory capacity, as measured by a digit-span test, and the ability of hearing-impaired older adults to segregate two concurrent speech messages. Two speech messages were presented to the same ear in one condition, or one to each ear in the other condition. The task was to report information from one of the messages, identified by a location (left/right), a gender, or by a key-word in the phrase. The cue was presented either before (*selective attention*) or after (*divided attention*) the messages were presented. Individual differences in performance amongst the hearing-impaired elderly listeners were correlated with the measure of working memory capacity—this relationship was found in both the selective and divided attention conditions. Those differences were independent of hearing sensitivity. It has been suggested that working memory is an important part of the cortical network which enables listeners to selectively attend to one of multiple concurrent streams of information (Knudsen, 2007), a process which is central to coping with multi-talker listening environments. The findings of Humes et al. (2006) suggest that non-auditory cognitive functions are critical when extended processing must be carried out on auditory information, such as when multiple speech streams must be segregated.

Working memory has been shown to play an important role in speech perception in noise even in the absence of hearing impairment. Pichora-Fuller, Schneider, & Daneman (1995) assessed speech-in-noise performance of normally-hearing young and older participants with near normal hearing using speech stimuli from the Revised Speech Perception in Noise (SPIN-R) test (Kalikow, Stevens, & Elliott, 1977; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). The test comprises spoken sentences with predictable and unpredictable final words presented against 8-talker speech babble. The final word of the sentences is either predictable or unpredictable, based on the information provided in the sentence. Participants were required to report the final word of each sentence immediately after it was presented. The level of the background babble was varied to estimate the intelligibility function for each listener. In addition, participants had to remember the final words and recall them after a number of intervening sentences. This task was used to assess working memory capacity. The older adult listeners exhibited poorer performance than the young adults on both the word recognition and the working memory task. Pichora-Fuller et al. (1995) suggested that the lower memory scores for the older group were not evidence of a general cognitive decline, but resulted from the reallocation of cognitive resources from more central tasks such as working memory to support the processing of auditory information.

A reallocation of resources, as suggested by Pichora-Fuller et al. (1995), may affect the contributions of higher level functions to the performance of complex listening tasks. The processes linked to attention and working memory act on the information resulting from sensory processing. In turn, the results of those cognitive functions can be used to tune the information selection processes to improve the SNR of the attended stream of information (Knudsen, 2007). For hearing-impaired listeners, obtaining a sufficiently high SNR may require the allocation of more resources to sensory processes as compared to when the input is not impaired, as suggested by Pichora-Fuller et al. (1995). While such an increased allocation does not necessarily imply that the resources are reallocated from other processes, such as those related to cognitive functions, it does imply that additional processing stages are introduced to facilitate a more detailed analysis of the degraded sensory information. Introducing an additional processing stage has been found to be particularly detrimental to the speed of processing in older adults (Verhaeghen & Cerella, 2002). Therefore, impoverished sensory input may result in a decrease in the speed at which tasks can be performed based on that input, and the performance cost should be greater for older adults. The allocation of more resources to the processing of auditory information does not eliminate the possibility that the quality of the output from those sensory processes may still be degraded. Knudsen's model suggests that poor quality sensory information would affect how higher-level cognitive functions act on that information. In relation to speech perception, hearing loss may necessitate the extended use of top-down 'bias signals' to adjust the SNR of the information to which the participant wishes to attend to (speech), and to suppress or ignore the irrelevant aspects of the sensory input (other speech and/or noise) (Knudsen, 2007). Alternatively, the degraded input may necessitate the deliberate use of attentional strategies which reduce the load on the processes of working memory, as suggested by Zekveld et al. (2007).

This section has presented evidence that deficits in cognitive functions contribute to age-related speech perception difficulties. Those functions include attention, working memory, and the speed at which information is processed. Evidence for the absence of a link between hearing loss and cognitive deficits has also been discussed. In the context of these studies, it is important to recognise that even when auditory and cognitive deficits are taken into account, deficits in the ability to perceive speech in the presence of noise still cannot be fully accounted for. From a meta-analysis of studies which have simultaneously examined auditory and non-auditory contributions to difficulties with perceiving speech in noise, Houtgast & Festen (2008) suggested that such variables rarely explain more than 70% of the variance in performance on speech-in-noise tasks. Houtgast & Festen (2008) suggested several possible explanations for the 'missing 30%' including the choice of inappropriate statistical models and the effect of measurement error on the data. However, it is also possible that additional variance in SRTs could be accounted for by including a more diverse selection of predictor variables. Based on the evidence presented in this section, it is relevant to examine the relationship between speech perception and a wider range of cognitive measures than has previously been considered. Chapter 6 presents an experiment which addressed this issue.

2.2.4 Impact of Hearing Difficulties on Quality of Life

Difficulties related to hearing can impose stress on a person's life. To qualify this association, it is helpful to refer to the definition of the term disability as "an alteration of an individual's capacity to meet personal, social, or occupational demands because of an impairment or functional limitation" (Cocchiarella, Turk, & Andersson, 2000). Situations which involve listening to speech in the presence of background noise, comprising speech or non-speech sounds, are common in many everyday environments. Thus, difficulties with listening to speech in noise may affect a person's ability to cope with many situations in everyday life.

The ability to localise individual sound sources within complex environments can be important, not only to anticipate and avoid hazards, but also to follow conversations among a group of talkers. Noble et al. (1995) examined self-reported localisation difficulties amongst normally-hearing and hearing-impaired adults. They found a significant relationship between localisation difficulty and feelings of confusion and loss of concentration. Listeners who reported these difficulties experienced a need to remove themselves from environments in which they would often fail to identify and segregate competing sources of sound. However, Noble et al. (1995) did not find evidence that the subjective need to leave difficult listening environments was necessarily associated with the ability or opportunity to leave those situations, resulting in stress and handicap.

Gatehouse & Noble (2004) developed the Speech, Spatial, and Qualities of Hearing Scale (SSQ) to elicit self-reported difficulties with listening in everyday environments. Using the SSQ, the links between self-reported difficulties and an independent self-reported measure of handicap were examined. The SSQ was designed to examine self-reported disabilities in several aspects of hearing, including listening to speech in a variety of contexts, localising stationary and moving sound sources, segregating talkers, and dividing attention between multiple talkers. Gatehouse & Noble (2004) sought to identify those difficulties associated with commonly occurring listening situations which reflect the hearing demands of everyday life. After controlling for hearing loss, they found strong correlations between handicap and difficulties in situations that impose high demands on attention, such as multi-talker environments, in which listeners must divide their attention between talkers, switch their attention between talkers, or sustain the focus of their attention on a talker while ignoring competing talkers.

Asymmetry of hearing loss can have an impact on localisation and spatial hearing over and above the effects of each unilateral loss. Inter-aural differences, important for localising and segregating sounds distributed in space, may no longer be useful cues as information from sound sources might not be audible in both ears due to the asymmetry in sensitivity. Benefits that arise from binaural interactions could also be affected in the same way, as the information from a sound source is degraded more for one ear than for the other ear. Noble & Gatehouse (2004) compared the self-reported difficulties of two groups of elderly adults using the SSQ: one with a similar degree of hearing loss in both ears, and the other with an inter-aural difference in hearing level of 10 dB or greater, averaged over 0.5, 1, 2, and 4 kHz. They found that self-reported difficulties in the group with asymmetric hearing loss were consistently greater than the symmetric hearing loss group. This was most evident on measures relating to direction, distance, and movement in spatial listening. The perception of speech, particularly in group conversations, was also adversely affected by the asymmetric hearing loss. Self-reported scores on an independent measure of handicap in the group with asymmetric hearing loss were related to difficulties in spatial listening. Thus, it is not only the average magnitude of hearing loss that can affect a listener's ability to cope with complex listening environments, but also the difference in the degree of hearing loss between the two ears.

Despite evidence for a link between handicap and hearing disabilities related to the spatial perception of speech (Gatehouse & Noble, 2004), traditional hearing disability inventories fail to examine situations which involve selective attention and switching attention. Gatehouse & Noble (2004) suggested that the role of attention on auditory and cognitive abilities had not been explored through self-reported measures. Accordingly, Gatehouse et al. (2006) examined the relationship between attention, as measured with the Test of Everyday Attention (TEA) (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996), and self-reported auditory disability, as measured with the SSQ. The TEA comprises tests of selective attention, sustained attention, divided attention, and attention switching. The tests are divided into several sub-tests which are based on everyday activities, such as searching for instances of a symbol in a map or searching a page of a telephone directory for a particular number. Some tests are visual, some auditory, and some involve both modalities. Gatehouse et al. (2006) extracted two visual factors ("visual search" and "visual executive control") and one auditory factor ("auditory composite attention") from the TEA data, as measures of visual and auditory attention respectively. Together with a pure-tone audiogram, these three measures explained independent components of the variance in self-reported disability scores obtained using the SSQ. The audiogram had a large influence on all SSQ measures relative to the influence of the attentional measures. After variance associated with the audiogram had been accounted for, visual attention had an additional influence on all of the SSQ measures. Auditory attention accounted for a significant proportion of the residual variance on two sub-scales of the SSQ: *speech-in-speech contexts* and *multiple speech-streams with attention switching*. The results suggested that auditory-specific attention has an impact on self-reported difficulties in more demanding listening situations, particularly those that involve segregating concurrent speech streams and/or switching attention between streams. The relationships between self-reported listening difficulties, attentional ability as measured by the TEA, and performance on an attention-demanding task of listening for speech are examined in Chapter 6.

2.2.5 Summary

Difficulties related to hearing are common amongst elderly adults over 60 years of age. An important aspect of these difficulties relates to speech perception. Evidence that cognitive factors, in addition to age-related decline in auditory peripheral sensitivity, are involved in difficulties with speech perception has been presented. The implications of hearing difficulties for quality of life for elderly adults have been discussed, particularly those difficulties related to spatial listening for speech in situations which demand a high level of attentional control.

2.3 Spatial Hearing and Speech Segregation

2.3.1 Introduction

Effortless interpersonal communication is something that most young listeners take for granted. We converse face to face, over telephones, and through other media. The process of focusing on information from one source of sound and ignoring distracting sources is essential for communication to take place, and occurs without conscious effort. A common, and relevant, example is listening to what one person is saying while other people are speaking at the same time. Understanding how we achieve this was described as the "Cocktail Party Problem" by Cherry (1953). He asked two questions: "How do we recognize what one person is saying when others are speaking at the same time (the 'cocktail party' problem)? On what logical basis could one design a machine ('filter') for carrying out such an operation?" Although much has been learned about the processes which underpin these tasks, from both psychological and physiological point-of-views, resolution of the 'problem' remains incomplete.

A feature of most everyday environments is that sound sources tend to be separated in space, giving rise to differences in the spectro-temporal pattern of stimulation between the ears. Listeners are well equipped to take advantage of these differences. They are also capable of using monaural information, that is, information from one ear in isolation, to segregate concurrent speech streams, or to determine the elevation of a sound source. While spatial separation has been identified as an important factor in overcoming the difficulties of listening in noisy environments (Arbogast, Mason, & Kidd, 2002; Cherry, 1953; Ebata, 2003; Yost, 1997), there are many other factors which make important contributions to solving the problem, including spectral separation, spectral profile, harmonicity, temporal separation, temporal onsets and offsets, and temporal modulations (Yost, 1997).

In this section, I will discuss the mechanisms that facilitate spatial hearing and speech segregation, using either one or two ears. I aim to show that there is a wide range of factors which contribute to our ability to segregate speech from background noise, or other talkers, including low-level interactions at the auditory periphery, i.e. the cochlea, and more centrally in primary auditory processing mechanisms within the brain.

2.3.2 Masking

The term masking refers to the process by which thresholds for detecting or discriminating one sound are raised by the presence of another sound or sounds (ANSI, 1960). One type of masking occurs when the frequency spectra of sounds overlap. If frequency components of two sounds, a signal and masker, fall within the same critical band on the basilar membrane within the cochlea (Figure 2.1) then the presence of the masker reduces the change in the activity on the auditory nerve produced by the addition of the signal. When this happens, the ear is incapable of resolving both components and one is 'masked' by the other. This form of masking is commonly referred to as *energetic* masking, being principally the result of a physical interaction between the signals.

Energetic masking is distinct from informational masking (Kidd, Mason, Deliwala, Woods, & Colburn, 1994; Leek, Brown, & Dorman, 1991). Informational masking (IM) can be loosely defined as "masking that cannot be accounted for by energetic masking" (Arbogast, Mason, & Kidd, 2005) or alternatively as a difficulty in separating multiple sources of sound when the sources are mostly resolved at the auditory periphery (Watson, Kelly, & Wroton, 1976). It usually occurs when a target stimulus is similar to a masker (Arbogast et al., 2005), although large individual differences in IM due to target-masker similarity have been observed (Durlach, Mason, Kidd, Arbogast, Colburn, & Shinn-Cunningham, 2003). Informational masking can also arise when there is uncertainty about the target (Arbogast et al., 2002; Brungart & Simpson, 2004) or masker (Neff & Green, 1987) stimuli. Whether or not the effects of both similarity and uncertainty should be associated with IM remains unresolved (Durlach et al., 2003). Informational masking has been suggested as arising at a high level within the auditory system, due to "limitations imposed by more central processing" (Watson et al., 1976), rather than due to a loss of information at the auditory periphery such as might result from energetic masking. The precise nature of these "central processes" and their role in IM across tasks of varying complexity are as yet undetermined.

The effects of both types of masking depend on whether stimuli are presented *monaurally*, when information is presented to a single ear, *diotically*, when the same information is presented to both ears, or *dichotically*, when different information is presented to the two ears. They are discussed within the following sections. I will focus on work involving mostly speech or speech-like maskers. Speech-like maskers include random noise maskers which are spectrally-shaped to match the long-term spectrum of a stimulus or set of stimuli, and random noise that is modulated by the amplitude envelope of a speech stimulus. These modulated noise maskers are useful when examining speech intelligibility in multi-talker listening environments as they can account for the energetic masking caused by a speech masker while providing no informational masking.

2.3.3 Hearing with One Ear

2.3.3.1 Monaural Processing

Listeners can take advantage of differences in the patterns of stimulation from a sound source between the ears. I will discuss in a later section how these interaural differences are used in resolving the horizontal position (azimuth) of a source. In contrast, localisation in the vertical plane is aided by the effects of the pinnae on sounds arriving at our ears. The pinna is the part of the ear which lies outside the ear canal. The complex folds of the pinnae primarily affect the high-frequency components of sounds arriving at the ears (above about 5 kHz) due to interference between the incoming and reflected sounds (Akeroyd, 2006). As a result of this filtering, the spectral profile of the sound is modified with energy amplified at some frequencies and attenuated at others, creating a pattern of spectral peaks and dips that varies with sound source elevation and thereby provides spectral cues that can be used for sound source localisation (Moore, 2003). Traditionally, the processing of spectral cues from the pinnae has been thought of as being performed monaurally (Wightman & Kistler, 1992). However, evidence exists which suggests that spectral information from both ears is combined to form an overall percept of the location of a sound, perhaps in the form of a binaural 'weighted' average (Hofman & Van Opstal, 2003).

2.3.3.2 Speech Segregation with One Ear

When sound is presented monaurally or diotically, listeners are capable of segregating concurrent sources to a limited degree compared to when different information is presented to each ear. When we listen with one ear or when the same information is presented to both ears, there is no disparity between the ears that can be used to localise sources to different points in space. The cues that are used to segregate sound sources in monaural or diotic listening situations include differences in fundamental frequency and voice characteristics (Brungart, 2001).

Effect of Talker Similarity

Without spatial information to segregate sources, the similarity of voices plays a crucial role in our ability to differentiate between information from individual talkers. By examining the effects of same-talker, same-sex, and different-sex speech maskers on speech intelligibility, Brungart (2001) concluded that informational masking plays an important role in diotic listening conditions with two talkers. At negative SNRs, performance dropped 15–20% from different-sex masker to same-sex masker, and by the same amount again from same-sex masker to same-talker masker (Figure 2.2). The results showed that performance decreased as the similarity between target and masker was increased. It would be expected that the increase in energetic masking would be larger between different-sex and same-sex maskers compared to between the same-sex and same-talker maskers. The difference in performance between the same-sex and same-talker maskers suggested that informational masking had a large effect on performance than energetic masking, when spatial information was not available.

The more similar the target and masker, specifically when masker and target phrases are spoken by the same talker, the more energetic masking is increased due to the presence of energy in similar frequency bands in both target and masker. By using speech-like noise maskers, the effects of energetic masking can be examined separately from any informational masking that might be caused by the speech stimuli from which the noise maskers were generated. Brungart (2001) also compared the intelligibility of speech stimuli in the presence of two speech-like noise maskers: the first was random noise spectrally shaped to match the average long-term spectrum of the speech stimuli (noise masker), and the other was random noise modulated by the amplitude envelope of a randomly chosen speech stimulus (noise-modulated masker). He found that both noise and noise-modulated maskers showed the typical pattern in which accuracy of identifying target sentences decreased monotonically as the SNR decreased. The speech maskers gave a different pattern of results—performance was independent of SNR at negative SNRs (Figure 2.2). This finding suggests that the listeners were able to use the target-masker level difference as a cue to distinguish the target from the masker, and this advantage outweighed the increase in energetic masking of the target by the masker as the SNR was decreased. The large difference in performance curves for speech and non-speech maskers suggested informational masking had a larger effect compared to the energetic masking created by the noise maskers. Thus, in speech-on-speech masking conditions, the results suggest that informational masking, rather than energetic masking, influences performance, particularly at negative SNRs.

From these results, Brungart drew three conclusions about diotic listening tasks: (1) the effect of any kind of masker, speech or non-speech, on the accuracy of identifying words in target sentences is negligible when the target is at least 10 dB more intense than the maskers, (2) the SNR has a large effect on speech intelligibility in the range of SNRs from +10 to 0 dB when the masker is speech (virtually no effect with noise maskers), and (3) speech intelligibility is independent of SNR in the range from 0dB to -10dB for speech maskers (noise maskers have a large effect). This complex relationship between SNR and intelligibility, differing with masker type, highlights the important role that informational masking plays in speech segregation.

Segregating Multiple Speech Streams

Successfully segregating more than two talkers using only monaural cues is challenging, and is largely determined by the amount of energetic masking of a target talker by the competing talkers. Drullman & Bronkhorst (2000) presented a target phase and multiple speech maskers monaurally. The number of competing talkers, including

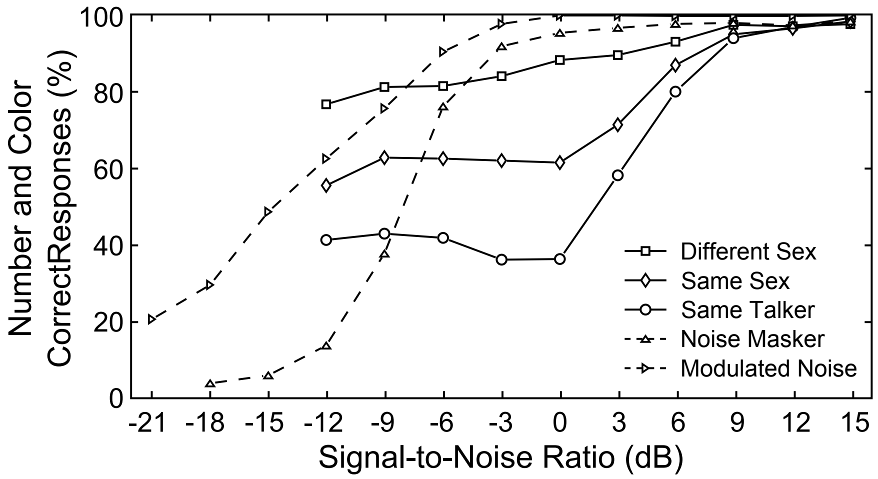

Figure 2.2. Percentage correct identification of two keywords (colour and number) within a target phrase as a function of SNR. Dotted-lines refer to speech-like noise maskers (spectrally shaped ‘noise’ masker and a modulated noise masker) (adapted from Brungart, 2001)

male and female talkers chosen at random, was varied. Overall performance for isolated word identification with one masker was approximately 48%, similar to studies which have measured keyword identification in sentences in the presence of a single masker (Brungart, 2001; Stubbs & Summerfield, 1990). Scores for sentences were poorer than isolated words, at 36% (SNR = 0 dB). Performance decreased by approximately 30% to 6% for entire sentences, and 30% to 18% for words, when a second competing talker was introduced (SNR = -3 dB). The introduction of a third competing talker (SNR = -4.8 dB) reduced performance by an additional 5%, and the addition of a fourth masker (SNR = -6 dB) had no further effect on performance. By increasing the number of maskers, the amount of spectro-temporal overlap with the target increases, increasing the amount of energetic masking. The results of Drullman & Bronkhorst (2000) suggest that this masking is the factor which limits intelligibility when multiple speech signals are presented monaurally or diotically.

The relationship between target-masker similarity and the number of concurrent talkers was examined together in a diotic listening task by Brungart, Simpson, Ericson, & Scott (2001). They found that the effect of altering the similarity between target and masker(s) was somewhat consistent across different numbers of maskers. Performance was best when the target was different from the masking voices (different sex) and worst when it was very similar (same talker). A systematic effect of varying the similarity of the target and masker(s) (same-talker to same-sex to different-sex) arose when the target level was higher than the levels of the masker(s), i.e. at positive SNRs. At SNRs of 0 dB or below, performance depended much less on the similarity between target and masker(s). An unexpected finding was that performance decreased when only one of the maskers was of a different sex to the target than when the maskers were of the same sex as the target, so-called “odd-sex distraction”. Brungart et al. (2001) suggested that this effect arises from the listener’s propensity to attend to the most salient talker in a group.

In summary, when listening to multiple talkers with one ear, the ability to hear what one person is saying is limited by the energetic masking of that talker by the other voices. Similarity between talkers affects performance when the target is more intense than the maskers, and the presence of a single masker of a different gender to the other talkers can be particularly distracting.

Use of Level Differences

Brungart et al. (2001) also examined the effects of presenting more than one speech masker diotically. At a given positive SNR, performance was found to *improve* as the number of maskers was increased from 2 up to 4. This effect was larger when the maskers were of a different gender to the target compared to when both target and masker were spoken by talkers of the same gender. The additional maskers may have made the task of segregating the target speech easier by reducing the intelligibility of the masker speech, and therefore reducing the amount of informational masking it created. However, at negative SNRs, this ability to use level differences to hear out the target was found only in the one target–one masker condition, where performance was largely independent of SNR. The same result was also reported by Drullman & Bronkhorst (2000). When a second masker was added, performance decreased more steeply as the SNR was decreased, suggesting that listeners were unable to use the difference in level to segregate the target phrase. The addition of further maskers made little difference to this trend. Therefore it would seem that while differences between target and maskers, such as their fundamental frequency and other voice characteristics, create large effects at positive SNRs, it is the number of masking talkers that determines performance at negative SNRs.

Use of Voice Familiarity

Brungart et al. (2001) examined the effect of providing listeners with prior information about the voice characteristics of a target talker. A speech target was presented simultaneously with multiple speech maskers. Listeners were either provided with information about the target talker by listening to several stimuli spoken by that talker, or did not know in advance which talker would be the target. Performance improved with target-familiarity when the maskers comprised both male and female talkers, and when all maskers were of a different gender to the target phrase. There was almost no improvement in performance when the target and maskers were spoken by talkers of the same gender. The absence of a benefit of familiarity to segregating multiple talkers of the same gender lead Brungart et al. (2001) to suggest that the prior information was used mainly to determine the gender of the target talker, rather than to extract more subtle voice characteristics to segregate the target talker.

2.3.3.3 Summary

In summary, in diotic listening conditions, the vocal characteristics of competing speech signals can have large effects on the intelligibility of a target speech signal, especially at positive SNRs. This is the case even when the number of talkers is increased from two to four. The effect of increasing the number of competing talkers at negative SNRs is to decrease performance, with the differences in talker characteristics between target and maskers continuing to have a substantial but smaller effect than at positive SNRs. The evidence reviewed in this section suggests that both voice similarity and the number of talkers are important factors that influence speech segregation with one ear, and therefore do not rely on binaural processing.

2.3.4 Hearing with Two Ears

2.3.4.1 Binaural Processing

The binaural system is important for spatial hearing. It enables listeners to localise sources of sound, and compared with monaural listening, it enables listeners to improve speech perception in noise. Listeners can take advantage of the fact that the pattern of stimulation from a sound source differs between the ears. As sound-waves travel in air, properties of the waves are altered depending on the distance they must travel. A sound originating on the left side of a listener has further to travel to reach the right ear, so an inter-aural time difference (ITD) occurs. In a similar way, the intensity of the sound decreases with distance, giving rise to an inter-aural intensity difference (IID, or inter-aural level difference, ILD, when measured in dB). However, a greater contribution to ILDs originates from the effect of the head on sound waves. Diffraction of the waves (i.e. the bending of waves around the head) is either partial or does not occur at all, creating an acoustic “shadow” on the far side of the head from a source. This lack of diffraction gives rise to an ILD. The size of the ILD is affected not only by distance from the source (when the source is very close, within about a meter, to the listener) but also by its frequency and angle (Shaw, 1974; Wightman & Kistler, 1989). ILDs are generally more pronounced at high-frequencies, due to the fact that the wavelengths of high-frequency sounds are small in relation to the size of the human head. Frequencies whose wavelengths are materially smaller than the distance between the ears are not diffracted by the head, and an ILD arises.

The usefulness of ITDs depends on frequency. Given that sounds positioned to the extreme left or right of a listener generate ITDs of around 690 μs (Moore, 2003), the localisation of sounds with periods shorter than half that duration is ambiguous. For example, if a sinusoidal sound has a period of 150 μs and an ITD of 150, 300, or 450 μs, it is difficult for the auditory system to tell which period arriving at one ear corresponds to which period at the other ear without the aid of additional information such as onset timing or harmonicity. Therefore, ITDs are usually more informative for lower frequency sounds, which have longer periods. The effect of creating an ITD between sounds presented over headphones is that the sound is lateralised within the head, and the position along the mid-line, the line connecting the two ears, is determined by varying the ITD. The same effect is obtained by varying ILDs through varying levels of a stimulus presented over headphones.

This simple division wherein ILDs are the primary cues for localising energy above about 1500 Hz, while ITDs are used below 1500 Hz, is known as the “Duplex Theory” (Rayleigh, 1907). The theory is an over-simplification of how the auditory system localises sound sources, especially when applied to complex sounds such as natural and speech sounds, but largely holds for pure-tone stimuli. In the real world, we use both ILDs and ITDs to localise sound sources. However, our use of one cue over the other depends on many factors, including the stimulus and the amount, and kind, of background noise (Akeroyd, 2006).

To assess the relative importance of ITD and ILD cues, Wightman & Kistler (1992) manipulated the ITD cues of Gaussian noise bursts so that they conflicted with the associated ILD information. They showed that ITD is the dominant cue for localising broadband stimuli in the horizontal plane. The frequency-specific nature of this dominance was indicated by a further experiment which used band-limited stimuli. The greater the amount of energy below about 1.5 kHz, the stronger the effect of ITD on perceived spatial location. Thus, the results suggest that ITDs are the dominant cue for localisation, as long as sufficient acoustical information is available at low frequencies.

For high-frequency sounds and sounds whose location cannot be determined by ITD, the integration of ILD and spectral cues is necessary for accurate localisation. Lorenzi, Gatehouse, & Lever (1999) examined the ability of listeners to localise a click-train in the frontal-horizontal plane in quiet and in the presence of a broadband white-noise masker. They varied the horizontal location (−90°, 0°, and +90°), frequency content (low-pass, high-pass, and broadband), and SNR of the click-train. They found that listeners take advantage of whichever cues, either high-frequency ILDs, low-frequency ITDs, or spectral cues, are available but favour high-frequency ILD and spectral cues. When noise was present on the listeners' left or right sides, listeners were found to rely on high-frequency ILD and spectral cues. This finding contrasts with the suggestion of Wightman & Kistler (1992) that ITDs are the dominant cue in quiet. Therefore, Lorenzi et al. (1999) suggested that low-frequency ITD cues are less resistant to noise than high-frequency ILD cues when noise is lateralised. While models like the "Duplex Theory" (Rayleigh, 1907) suggest a basis for the use of inter-aural cues, listeners take advantage of both level and time differences and spectral cues that arise from source location and the effects of our own head to localise sounds accurately when noise is present.

2.3.4.2 Speech Perception with Two Ears

Bilateral Asymmetry in Speech Processing

There is evidence for asymmetries in the auditory system—a greater accuracy for speech perception is observed when speech stimuli are presented to right ear (Right-ear Advantage, or REA) over the left. Conversely, there is a greater accuracy at reporting non-linguistic stimuli such as complex tone bursts with different fundamental frequencies, when presented to the left ear (Left-ear Advantage, or LEA) over the right (Jerger & Martin, 2004). This pattern is commonly attributed to hemispheric asymmetry. Two models have been proposed to account for this asymmetry: 1) the classic 'structural' model of Kimura (1967) (referenced in Jerger & Martin, 2004) whereby the asymmetries exist in the connections between the cochlea and the central auditory system (CAS) such that information at each ear is “better represented” in the contra-lateral cortical hemisphere and therefore the right ear input has a more direct, and hence swifter, connection to the areas of linguistic processing in the left hemisphere, and 2) the ‘attentional’ model proposed by Kinsbourne (1978) (referenced in Jerger & Martin, 2004) which argues for an attentional bias towards the ear or ‘space’ contra-lateral to that of the activated hemisphere. There is strong evidence to support the structural model, both anatomical and functional, but there is evidence that attention can have large effects on the behavioural manifestations of the asymmetry (Jerger & Martin, 2004). Advantages have been found for speech stimuli presented in the right hemi-field in the presence of simultaneous speech maskers, both for free-field presentation (Kidd, Arbogast, Mason, & Gallun, 2005) and for virtual headphone presentation using HRTFs (Bolia, Nelson, & Morley, 2001).

Independence of Input to the Two Ears

Some of the earliest experiments in dichotic listening were those of Cherry (1953). He presented two different recordings of continuous speech by the same talker, first to both ears simultaneously and then one recording to each ear. He observed the ability of listeners to use differences between signals arriving at the ears to segregate multiple concurrent sources. When both recordings were presented diotically (the same input to both ears), listeners found it difficult to focus on just one of the recordings. However, overall performance was quite good, perhaps because listeners could replay the recordings as many times as they wished, or were able to use differences in the fundamental frequency or linguistic content of the two recordings (Section 2.3.3, p. 23). When the two speech streams were presented separately, one to each ear, the task became trivial—rejecting the input at one ear, as well as switching attention between the ears, were both reported to be simple tasks. These early experiments showed that the auditory system is capable of using information arriving at the ears either relatively independently, allowing attention to be directed to one ear only, or in combination as in most natural situations where binaural cues are used to resolve source location and distance.

Drullman & Bronkhorst (2000) compared dichotic presentation, where different signals were presented to the two ears, to diotic presentation, where the same signal was presented to both ears, for a target speech stimulus with speech maskers. They found that performance in identifying both individual words and entire sentences varied with presentation mode. For dichotic presentation, increasing the number of maskers to the ear contra-lateral to the target phrase had no effect on performance. This result is in line with Cherry’s experiments (1953), in showing that listeners can suppress information from one ear without affecting the ability to attend to information presented to the other ear. When the number of maskers was increased in the same ear as the target talker, performance dropped in an identical way to the monaural condition. In other words, the presence of maskers in the ear contralateral to the target in binaural presentation had no impact on performance in segregating the target—only maskers in the target ear had an effect. Thus, listeners are capable of suppressing competing speech information from one ear independent of how many maskers are presented to the contra-lateral ear.

Effects of Spatial Separation

In multi-talker environments, the simultaneous presence of many speech streams means that not only is there a high probability of peripheral masking at the cochlea due to the fact that most speech contains energy across the same frequency range, but when multiple talkers are speaking simultaneously, informational masking can arise. The spatial separation of sources has been shown to provide a 'release' from both informational and energetic masking.

Release from masking can be attributed to two effects. The first is the "better-ear advantage"; i.e. the fact that the target-masker ratio can be more favourable at one ear than the other as a result of the spatial separation of the sources. The second is other inter-aural differences which are processed by the binaural parts of the auditory system (Kidd et al., 2005). The advantage from spatial separation varies for different kinds of masker, e.g. speech-shaped noise or speech, and the attentional 'state' of the listener. Maximum releases from masking for a speech target between 8-18 dB have been reported for unmodulated speech-shaped noise (Arbogast et al., 2002; Bronkhorst & Plomp, 1988) and speech-modulated noise (Bronkhorst & Plomp, 1992) maskers. With regard to speech-on-speech masking, Hawley, Litovsky, & Colburn (1999) examined the effect of proximity of a speech masker to a speech target in the free-field and in a virtual auditory space. They found that the intelligibility of the speech target could be reduced by as much as 55% by increasing the proximity of the speech masker. This effect of spatial separation was more influential on overall intelligibility than the number of competing talkers present.

In reviewing the evidence for the effect of spatial separation from noise and speech maskers on speech intelligibility, Ebata (2003) reported that the benefit from spatial separation of sources has been measured at everything from a few decibels to over 10 dB. The review encompassed a variety of presentation methods, including in the free-field and over headphones, using ITDs, ILDs, and HRTFs. Ebata (2003) concluded that in general spatial release from masking is 3–6 dB greater when unmodulated noise maskers are used, compared to speech maskers. This difference arises largely due to the high level of energetic masking introduced by noise maskers. Like earlier reviews (Bronkhorst, 2000; Yost, 1997), Ebata (2003) identifies the spatial separation of speech sources as an important contributing factor for the perception of speech in a multi-talker environment.

Spectro-Temporal Grouping

The problem of speech segregation may not be about just *detecting* but also *allocating* each "local spectro-temporal" feature to the appropriate voice (Darwin, 2006). Features, or 'glimpses' (Cooke, 2006), of the target voice need to be integrated and related to each other to ensure that the speech of the target voice is perceived as a coherent stream. Spatial cues play an important role in tying these various features together into a unified percept. However, in the frequency domain, the use of binaural cues such as ITDs to group concurrent frequency regions of a sound is difficult at best. This result has been shown with competing formants, each created with two noise bands, separated first by presentation ear, where identification was successful, and then by ITDs, whereby listeners failed to use the inter-aural cue to group the different bands with similar ITDs into separate perceived objects (Culling & Summerfield, 1995). Harmonicity and onset timing are more salient in integrating features across frequency (Darwin, 2006). In sequential grouping, binaural cues play a more obvious role. For example, Sach & Bailey (2004) observed that both ITDs and ILDs can be used to group spatially separated competing rhythms, even when both comprise tones of the same pitch. The results suggested that it was the difference in perceived spatial location, rather than the value of either binaural cue, that was central to the segregation.

2.4 Summary

- Hearing difficulties are common among adults over 60 years of age, and include difficulties with the perception of speech in noise.
- Deficits in cognitive function and sensori-neural hearing loss have been associated with age-related difficulties with speech perception.
- Older adults report experiencing difficulties with listening in everyday environments in which they must divide, switch, and sustain their attention, and difficulties in those environments are related to the perception of handicap.
- The ability to segregate information from multiple talkers who are speaking simultaneously involves the use of monaural and binaural processes, both at the auditory periphery and within the brain.
- When listening with one ear, similarity between talkers and the number of talkers that are speaking affect the ability of a listener to segregate information from multiple talkers.
- When listening with two ears, listeners can take advantage of the spatial separation of talkers to focus attention on an individual talker.

- When multiple talkers are speaking at the same time, information from one talker can be obscured by two types of masking: informational and energetic.
- Difficulties with segregating information from multiple talkers who are sufficiently separated in space and speaking at the same time are principally due to confusion between the talkers resulting from informational masking, rather than difficulties in resolving the information from each talker due to energetic masking.

Chapter 3

Literature Review: Attention and Spatial Listening

This chapter outlines the various attentional mechanisms that are involved in multi-source listening, where the listener may need to focus on information from one source, or follow several sources of information, in the presence of irrelevant sources or noise. The chapter discusses studies which show that attention is important in spatial listening and that uncertainty about the location of a target source and the occurrence of abrupt isolated onsets can affect the attentional strategies that listeners adopt. The chapter provides an overview of the cortical mechanisms which have been associated with attention, and discusses studies which have examined cortical activity while listeners are attending to speech in noise. The chapter presents evidence that attentional strategies change with increasing age, and that those changes may reflect a decrease in automatic processing in elderly adults.

3.1 Attentional Mechanisms

In this section, I will examine the evidence that attention plays an important role in spatial listening for speech. I will start by describing the attentional mechanisms that are involved in focusing on information from one source in the presence of other competing sources, or dividing attention between several competing sources. I will discuss the effects that uncertainty about sound sources, both targets and maskers, and abrupt onsets have on attentional control. Finally, I will present evidence that attentional mechanisms and strategies change with advancing age.

3.1.1 Introduction

Much early work on attention was conducted in the auditory domain. The experiments of Cherry (1953), for example, used a dichotic listening task, where listeners were required to attend to speech presented to one ear while ignoring speech presented to the other. Cherry's work raised important questions about how listeners deal with incoming sensory information, including (Driver, 2001): 1) How much information do listeners obtain from unattended signals? 2) Is it possible to divide attention between two or more signals and, if so, for what kinds of signals can listeners do this? In this section, I will discuss the theories and studies which have examined how listeners attend to information selectively. I will show that the extent to which unattended stimuli are processed is still controversial. I will also discuss studies which have explored listeners' ability to switch attention. I aim to show that abrupt onsets, such as those which occur in multi-talker listening situations as new talkers start to speak, are important in the allocation of attentional resources.

3.1.2 Selective Attention

Solving the cocktail party problem involves segregating the target source from multiple concurrent sources that are usually spatially distributed. Therefore, selective attention plays a role. If concurrent sources can be separated using the physical cues available to the listener, information from the target is focused on and information from other sources is suppressed. Theoretical models will be discussed which have attempted to explain how selective attention operates on incoming sensory information, particularly with regard to the amount of processing that unattended information receives.

3.1.2.1 Filter Theory of Attention

Broadbent's filter theory of attention (1958, referenced in (Driver, 2001)) proposed that processing of sensory information is organised in two stages. The first, 'pre-attentive' stage, which is applied to all stimuli, encompasses the extraction of physical properties, such as pitch, for audition, and colour, for vision. This process is performed in a parallel fashion, on all stimuli at the same time. In the second, 'attentive' stage, non-physical semantic features are processed, such as the meaning of words. This second stage has a more limited capacity than the first. Therefore, processing is carried out on a limited number of stimuli. The process of determining which stimuli receive more detailed processing is determined by a selective filter, and there must be sufficient differences between stimuli so that they can be separated and 'examined' by the selective filter (Driver, 2001).

Relating this conclusion back to the classic dichotic listening task, we can see how if there are sufficient differences between the two speech stimuli in terms of physical properties, determined by aspects such as location, spectro-temporal properties, and fundamental frequency, then it is possible to selectively filter out the desired speech stream. Cherry (1953) found that when listeners attended to a speech stimulus in one ear while ignoring a competing speech stimulus presented to the other ear, the listeners could recall very little, limited to general properties such as gender, about the non-attended message. This result is compatible with Broadbent's filter theory, as the non-attended message would be subject only to the first stage of processing, which provides the listener with only basic physical properties of the stimulus.

3.1.2.2 Processing of Unattended Information

One of the main areas of contention between different models and theories of attention is the extent to which unattended stimuli are processed for both physical properties and semantic content. Most theories of attentional processing can be crudely broken down into two general groups: early and late selection. Broadbent's filter theory (1958) is an example of early selection, where the bridge between attended and non-attended stages occurs early on in the processing of the stimuli. An alternative theory, or 'late-selectionist' view, suggests that stimuli actually receive a great deal more processing in the first stage, but that the processes of selective attention make the more detailed information about non-attended stimuli unavailable to the listener. Only stimuli which enter the second stage interact with awareness, response, and memory (Driver, 2001). To give a concrete example, the mention of one's own name in an unattended speech stream usually results in a reorienting of attention, suggesting that, for certain stimuli at least, there is semantic processing of stimuli even when they are not the focus of attention, a process which cannot be accounted for by Broadbent's model (Deutsch & Deutsch, 1963)

A different approach to the question of this more extensive processing of unattended stimuli was presented by Treisman (1960; 1969). In her account, inputs from unattended stimuli are "attenuated", not simply removed. The likelihood of information from these inputs receiving attention depends on the current context and on the intrinsic importance of stimuli, with some words, such as our own name, being more likely to receive attention (Driver, 2001). Finally, in a theory which brings together evidence for both early and late selection theories, Lavie & Tsal (1994) found that the perceptual load of a task influenced the effectiveness of distractor stimuli. In their account, the level to which unattended stimuli are processed depends on the amount of resources that the attended stimuli consume, which is dictated by the perceptual load of the task being performed.

In summary, Broadbent's filter theory proposed that a selective filter is applied early in the processing chain, so that unattended stimuli receive only low-level processing, mainly for the extraction of physical properties. The premise of many alternative theories to Broadbent's (Deutsch & Deutsch, 1963; Lavie & Tsal, 1994; Treisman, 1960) has been that if such an 'early' selective attentional filter does indeed exist, then certain information from unattended stimuli leaks through, facilitated by many factors including context, priming, and perceptual load. Lachter, Forster, & Ruthruff (2004) examined the difficulties involved in controlling attention in an experimental setting, and in determining whether or not information from unattended sources does indeed slip through such an early selective filter. They suggested that shifts in attention are difficult to control for or detect, and can possibly explain some of the observations which disagree with the notion of an early selective filter. Thus, the question of how much processing unattended stimuli receive is still the subject of debate.

3.1.3 Shifting or Switching Attention

In multi-talker listening environments, not only do listeners focus their attention on individual sources, but they must also frequently switch between sound sources when following a conversation with several talkers. Therefore, knowledge about how listeners shift their attentional focus is relevant in relation to the cocktail party problem.

Once listeners can segregate two speech streams, the process of switching between them is easily accomplished (Cherry, 1953). In terms of Broadbent's selective attention filter (Broadbent, 1958), if stimuli can be segregated based on their physical properties, then it is possible for listeners to alter the parameters of the selective filter, to attend to one stimulus or another. If this switching between stimuli is performed rapidly enough, it could give rise to the apparent ability to attend to multiple stimuli at the same time. However, it might be possible to split attentional resources between multiple concurrent sources, providing the amount of resources needed to process each stimulus is small (Broadbent, 1958).

3.1.4 Voluntary and Involuntary Shifts of Attention

Early experiments on attention included attempts to estimate the time required to switch attention between inputs. Cherry & Taylor (1954) measured the ability to repeat a speech message while continuously switching the ear to which the message was presented. By determining the fastest rate at which the stimulus was switched between the ears that permitted accurate tracking, they suggested that the time to shift attention from one ear to another was approximately a sixth of a second.

This early theory did not distinguish between voluntary and involuntary shifts of attention, and therefore did not consider that the two different categories of attentional shifts might have varying time costs associated with them. Most of the work in the area of voluntary versus involuntary shifts of attention has been in vision, and involves varying the time between the onsets of a cue and target, i.e. stimulus onset asynchrony (SOA). Involuntary shifts have been found to take less than 100 ms, whereas voluntary shifts can range between 150–500 ms (Lachter et al., 2004). Wolfe, Alvarez, & Horowitz (2000) showed that voluntary shifts in visual attention are up to an order of magnitude slower than involuntary shifts dictated by the salience of stimuli. If this is consistently true, then it might be to the advantage of a listener to adopt an attentional strategy of allowing the focus of their attention to be modulated involuntarily by salient stimuli, rather than deliberately attending to certain stimuli.

The effect of maintaining such an alert attentional state on detecting a target item does not necessarily mean that information related to the target item is accessed more rapidly. The adoption of a vigilant state of attention may lead to faster response times to a target item (Posner, Snyder, & Davidson, 1980) but comes at a cost of more errors. Being in a state of alertness may not affect access to information about the target but the speed of processing of the target within the attentional system. This increased speed of processing may be indicative of the use of information which is lower in quality and therefore results in more frequent errors (Posner & Petersen, 1990).

Mondor & Zatorre (1995) examined the amount of time necessary to focus auditory attention towards a cued spatial location and the effect of distance between spatial locations. They found that as the time between cue and target was increased, the time required to make a discrimination judgement about the target decreased. This finding suggests that time is required to shift attentional focus to a cued location. Mondor & Zatorre (1995) also found that the distance of the attention shift did not affect the time necessary to perform the shift.

3.1.5 Abrupt Onsets and Attention Capture

Given that involuntary shifts of attention are more rapid than voluntary ones, allowing attention to be captured by a person who starts to speak might be a more efficient strategy in a multi-talker environment than a strategy which involves actively monitoring for new talkers and shifting attention to a new talker voluntarily. However, if it is more advantageous to adopt a stimulus-driven approach to coping with multiple sources, any of which may contain relevant information, then what factors in a cocktail-party environment might cause involuntary shifts in attention? As suggested above, one such factor might be the onsets of new voices, which are plausible candidates for stimuli that might cause involuntary shifts of attention because an onset usually signals a source of new information within an environment, e.g. a person starting to speak. In the visual domain, Yantis & Jonides (1984) examined the effect of abrupt onsets on attention by presented participants with an array of letters, which either appeared abruptly ('onset' stimuli) or were gradually unmasked over a period of 80 ms ('no-onset' stimuli). The task was to indicate whether a target letter ('E' or 'H') had been presented. Reaction times were greater when no-onset targets were displayed, in comparison to onset targets, regardless of how many distractor stimuli were presented at the same time as the target. Yantis & Jonides (1984) suggested that isolated abrupt onsets automatically capture attention and are allocated attentional resources. They describe this effect in terms of an 'abrupt-capture' model, in which an item with an abrupt onset is processed first, and then serial searching commences over the remaining stimuli.

In the experiments by Yantis & Jonides (1984), only one abrupt onset occurred on each trial. If it were the case that attention was drawn to all abrupt onsets, i.e. that an abrupt onset alone was sufficient to 'grab' attentional resources, then it might be expected that two simultaneous abrupt onsets would cause attention to be divided between them, or possibly to be switched between them rapidly, or for one of the stimuli to be missed. The involuntary 'capture' of attention by abrupt onsets would also be expected to be independent of the current attentional state of the perceiver. Kahneman, Treisman, & Burkell (1983) presented participants with words graphically at spatially uncertain locations and measured the time taken to name the words when presented alone, or with a patch of random dots; i.e. a distractor which was irrelevant to the target. They found that when the distractor onset was simultaneous with the target word onset, the time required to name the target word increased significantly. They also found that when the offset of the distractor was abrupt and simultaneous with the onset of the target, performance was also impaired. In contrast, if the distractor was presented before the target but its offset was either before or after the onset of the target, performance improved compared to when the target was presented alone. Kahneman et al. (1983) attribute such changes in performance associated with simultaneous onsets or onset-offset pairings to a 'filtering cost'. This cost represents the additional time needed to discriminate between the simultaneous salient events. Thus, it appears that isolated abrupt onsets can automatically capture attentional resources, but when such onsets occur simultaneously with other onsets or abrupt offsets, then an overhead, in the form of extra time required to select an event to attend to, is incurred.

Effects similar to those found by Kahneman et al. (1983) from varying the temporal order of target and masker onsets on the ability to detect the target have been observed in the auditory domain. Bacon & Moore (1986) presented a 20 ms pure-tone signal at 1 kHz and a 400 ms pure-tone masker whose frequency was varied around that of the signal. The onset of the signal was either simultaneous with the masker onset, in the middle of the masker, or the signal was positioned so that the offsets of signal and masker were simultaneous. Psychophysical tuning curves (PTCs) were measured by varying the level of the masker using an adaptive procedure to estimate the level at which the signal was masked for several masker frequencies. The PTCs were found to be broadest when the onsets were simultaneous, and sharpest when the signal was presented in the middle of the masker; the curves were slightly broader when the offsets were simultaneous. This broadening was most pronounced for those masker frequencies above the frequency of the signal. Thus, when the onsets were simultaneous, high-frequency maskers were more successful in masking the signal at lower levels compared to when the onsets were separated in time. Bacon

& Moore (1986) concluded that frequency selectivity improved when the signal and masker onsets or offsets were temporally separated compared to when the onsets were simultaneous, and that central auditory processes may have contributed to the observed effects.

Yantis & Jonides (1990) examined whether or not abrupt onsets always capture attention automatically. The stimuli and task were identical to those used by Yantis & Jonides (1984). Participants were cued to a possible target location, after which the target and distractor letters appeared. The validity of the cue, the position and number of distractors, and the onset type (abrupt or gradual) of the target were varied. Yantis & Jonides (1990) found that the effectiveness of an abrupt onset at capturing attention was modulated by the participant's attentional state. By presenting a perfectly valid cue before the onset of the stimuli, they also found that effects of abrupt onsets can be negated by highly focused attention to a location other than that of the onset. In other words, highly focused attention is resistant to abrupt onsets.

In summary, both stimulus onsets and offsets can modulate attention. When abrupt onsets are sufficiently isolated from other abrupt onsets, they can 'grab', or 'capture', attention. This effect can be diminished significantly when attention is focused on a spatial location at the time of the onset. Simultaneous onsets can delay the time necessary to attend to a particular stimulus, possibly because of an overhead imposed by the need to deliberately select which event to attend to. The attentional effects of phrase onsets in a multi-talker environment will be reported in Chapter 5.

3.2 Attention and Complex Listening Environments

There are several important questions related to how attention influences performance in cocktail-party situations, which include: 1) how important is the ability to divide attention or rapidly switch attention between different sources in hearing out information within a cocktail-party environment? 2) How much advantage is gained from directing attention to a location within such an environment? 3) What kinds of information about a source are necessary for a listener to focus their attention on it? The original experiments by Cherry (1953) on understanding binaural speech separation were focused on understanding the processes behind the cocktail-party problem. However, his paradigm is sufficiently different from a real-world listening environment that it might not tell us much about the roles that selective attention and shifting of attention actually play when multiple sources are spatially separated in addition to background noise. In this section, I will discuss studies which have examined the role of attention in complex listening situations, particularly those which involve multiple spatially-distributed sources, and the effect that various types of uncertainty can have on attention.

3.2.1 Attentional Strategies in Complex Listening Environments

Arbogast & Kidd (2000) examined the role of attention in a complex multi-source listening environment. Stimuli were presented through seven loudspeakers arranged at 30° intervals on an arc in the free-field. On each trial, the listener was instructed to attend to one of the seven locations, and one target and six maskers were presented. On a proportion of the trials the target was at that attended location, otherwise it was presented at an unattended location. The target consisted of eight contiguous tone bursts, each of which lasted 60 ms, whose frequencies were arranged to form either a rising or falling sequence. Participants made a rising/falling discrimination based on the target stimulus. Arbogast & Kidd (2000) found a decrease in response accuracy and an increase in response time when the target was presented at unattended locations.

They attributed both effects to the use of an attentional strategy which would involve listeners dividing their attention and allowing it to be grabbed by salient stimulus onsets. This conclusion is in agreement with evidence from vision, which suggests that using dispersed attention in visual search tasks increases the speed in finding a target over serial searching (Yantis & Jonides, 1990). Arbogast & Kidd (2000) suggest that if the first stimulus to grab attention is not the target, then listeners must disengage attention from the location of that stimulus, once it has been identified as a non-target, and spread their attention to try to identify the target at other possible locations. This strategy would be more efficient than searching through each of the possible locations one at a time. For visual search, Yantis & Jonides (1984) suggested that if the first stimulus to capture attention is not the target, then a serial search begins through each of the remaining stimuli, resulting in an increase in the time needed to find the target.

Arbogast & Kidd (2000) also attributed the effects found to the presence of an auditory spatial filter, which increases the SNR at and around (to a more limited degree) the focus. The use of such a filter would be advantageous in a multi-source high-uncertainty environment. Mondor & Zatorre (1995) found that the distribution of auditory attention, when focused on a particular location, declined with distance from the focal point (greater response times with increasing distance). Arbogast & Kidd (2000) failed to find this gradual decline, with performance being equally poor, and response time increasing by a comparable amount, at the sampled locations on the left and right sides of the location on which attention was focused. Arbogast & Kidd (2000) suggested that the absence of a gradual effect on performance and responses times with increasing distance from the attended location indicated that the locations closest to the attended location (±30°) marked the edges of the attentional filter; i.e. the locations beyond which the filter had its maximum effect, regardless of distance from the attended location.

Attending to multiple simultaneous speech streams, as is required when talkers start to speak simultaneously, places a high load on the attentional resources of the listener, particularly that of divided attention. Shafiro & Gygi (2007) examined the effects of increasing the load on divided attention in a multi-talker listening task. The load was manipulated by increasing the number of target phrases that had to be identified within a group of phrases, all of which started simultaneously. On each trial, 2-4 phrases were selected from the CRM corpus (Bolia, Nelson, Ericson, & Simpson, 2000). The phrases take the form: “Ready CALL-SIGN go to COLOUR NUMBER now”. There are eight possible call-signs, four colours, and eight number key-words. One phrase contained the call-sign ‘Baron’. The mixture of phrases was presented diotically and participants were instructed to report the colour and number key-words from phrase containing the ‘Baron’ call-sign only when they had identified a specified number of target call-signs with the phrases of that trial. The design of the experiment allowed for the divided attention aspect of the task, in the form of target call-sign detection sensitivity, and the selective attention component, as the accuracy of reported colour and number key-words, to be examined independently. Shafiro & Gygi (2007) found that detection sensitivity for the call-sign key-words and performance at reporting key-word information from the phrase containing ‘Baron’ decreased as the number of target call-signs, or attentional load, was increased. The effect of attentional load was found to have a significant effect on performance even when the number of irrelevant talkers was varied and the configuration of gender across the target and irrelevant talkers was manipulated. Possibly, the load on attentional resources would be reduced if multiple talkers did not start to speak at the same time. The extent to which the onsets of a target phrase and a masker phrase, presented against a background of additional speech maskers, must be separated in time to sufficiently reduce the attentional load and therefore lead to an increase in target identification performance is unknown. This issue was examined in Chapter 5.

This spatial distribution of sources is a key feature that listeners exploit to separate the signal arriving at the ears into distinct components, or streams (Cherry, 1953; Ebata, 2003; Yost, 1997). The process involves the use of inter-aural cues to segregate information from multiple sources, and also the use of attention to focus on the location of the source which is of interest to the listener. This focusing of attention has been compared to a spatial filter, whose effectiveness degrades as distance from the focal point increases (Mondor & Zatorre, 1995). In an environment where many people are speaking at the same time, the location of the talker to whom a listener wishes to attend can often change. This is also the case for the talkers that the listener wishes to ignore. Allen, Carlile, & Alais (2008) reported that listeners can take advantage of spatial separation to segregate multiple talkers even when those location cues were not persistent. Three CRM phrases were presented on each trial, one of which contained the call-sign “Baron”. The participants had to identify the phrase containing “Baron”, the target phrase, and report its colour-number co-ordinate. Three loudspeakers were used to present the stimuli, positioned at −30°, 0°, and +30°, where 0° was directly in front of the participant. The phrases were either all presented 1) all at 0°, 2) each from a different loudspeaker with the target at 0°, 3) all at 0° for 700 ms, after which the phrases separated as in (2), or 4) all separated as in (2) for 700 ms, after which the masker phrases were presented with the target phrase as 0°. The delay of 700 ms was sufficient to encompass the "Ready CALL-SIGN" portion of the phrases, but not the colour-number co-ordinate. Allen et al. (2008) found that, compared to the condition when all phrases were collocated at 0° (1), the largest release from masking was found when the target and maskers were continually separated (2), and when they were separated after 700 ms (3). Both results are compatible with the use of the distinct spatial location of the target phrase (0° in both conditions) to hear out the colour-number co-ordinate and ignore the irrelevant information contained in the masker phrases. However, when the phrases were initially separated but then all moved to 0° (4), the location of the target phrase was not unique while the colour-number co-ordinate was spoken, diminishing the effectiveness of the location cue. Despite this, significant release from masking was found for that condition. In other words, initial exposure to the target phrase at a unique location provided a significant benefit in hearing out the information in the phrase, even when the target phrase was masked by two other phrases at the same location while that information was being spoken. This result suggests that listeners can take advantage of brief moments in which talkers are spatially separated, possibly to extract cues other than location based on vocal characteristics. Those cues can then be used to segregate that talker from other talkers in the event that cues based on location are no longer valid.

3.2.2 Uncertainty and Attention

The highly-uncertain task used by Arbogast & Kidd (2000) highlights the relationship between benefits from focused or 'selective' attention and the degree of uncertainty about the target. It has been shown that uncertainty about a target talker, in terms of fundamental frequency (Arbogast et al., 2002), identity (Brungart et al., 2001), or location (Kidd et al., 2005), can degrade speech recognition performance with multiple talkers, especially when target and masker voices are similar in terms of fundamental frequency and informational content, leading to informational masking. Kidd et al. (2005) examined the ability of participants to report two key-words from a target phrase in the presence of two competing speech messages. They systematically varied uncertainty about the location from which the target phrase would be presented. The target phrase was identified by a 'call-sign', a single key-word located near the start of the phrase, which was cued before target presentation ('selective attention') or after ('divided attention').

When the target location was known, the listeners made very few errors, indicating that it was possible to segregate the sound sources. The effect of uncertainty about target location was to increase the rate of misdirection of attention. From the analysis of errors, Kidd et al. (2005) found that there was surprisingly little switching or shifting of attention between phrases—even if listeners knew that the attended phrase was not the target, they still tended to report information from the non-target phrase as it had been the focus of their attention. Thus, there was a penalty, in terms of a decrement in performance, associated with the misdirection of attention. This result highlights the important role that attention can play in uncertain listening situations, both through facilitating and inhibiting source identification.

Kidd et al. (2005) suggested that the link between attention and uncertainty about target location could be explained in terms of Lavie's 'Perceptual load' theory of attention (Lavie, Hirst, de Fockert, & Viding, 2004). The theory relates the cognitive and perceptual load of a task to the interference effects of non-attended stimuli, or distractors. If the perceptual load of the task is high, there are fewer resources left to processes unattended stimuli. However, in cocktail party tasks such as those used by Kidd et al. (2005), it is the cognitive load that is high, due to the heavy use of working memory and the requirement of dividing attention, elicited by the high levels of uncertainty, rather than the perceptual load, i.e. there is only one target and it can be segregated from concurrent maskers. According to perceptual load theory, such task conditions would increase the chance of interference from distractors (Kidd et al., 2005).

The complexity of the environment also modulates the effect of attention in multi-talker listening tasks. Increasing the number of talkers within an environment increases the uncertainty about a target talker, unless the listener has prior knowledge about the target talker. Therefore the number of talkers and known information about a target can potentially alter the benefit that attention plays in such tasks. As has been shown previously (Brungart & Simpson, 2002; Kidd et al., 2005; Yost, Dye, & Sheft, 1996), increasing the number of concurrent talkers increases the processing load associated with identifying and segregating a target stimulus. The study by Kidd et al. (2005) highlighted the beneficial role that the directed focusing of attention can play in such situations.

Brungart & Simpson (2004) examined the effects of different kinds of uncertainty about a speech masker on a key-word identification task, similar to that used by Kidd et al. (2005). The target phrase was presented to the right ear, and either one masker was presented to the same ear as the target (monaural condition) or two maskers were presented, one to each ear (dichotic condition). Uncertainty about the talker who would speak the masker phrases and the content (keywords) of the masker phrases were varied systematically. Reducing uncertainty about the voice of the masker(s) or the content of the masker in the contra-lateral ear to the target did not affect performance. Fixing the talker of the masking phrase in the target ear had a small effect on performance. However, reducing uncertainty about the content of the masking phrase in the target ear increased performance at 0 dB SNR by approximately 20%. This result, along with the fact that almost all incorrect responses corresponded to the content of the masker phase presented to the target ear, was interpreted by Brungart & Simpson (2004) as evidence that the listeners were able to understand both phrases, but had difficulty determining which information belonged to which phrase. Thus, even when the listeners knew who to ignore, it was the similarity between the target and masker phrases, resulting in informational masking, rather than difficulty in segregating the talkers, possibly due to energetic masking, which dominated performance.

In contrast to uncertainty about *where* a stimulus will be presented from or *who* will speak it, uncertainty about *when* a stimulus will appear has been shown to have limited effects on performance. Leibold, Neff, & Jesteadt (2005) examined the ability of participants to detect a 1 kHz tone within a temporally-regular sequence of five non-overlapping maskers. The maskers were either broadband noise or complex tones with 1, 2, or 10 randomly chosen components which did not overlap the target tone. The target tone was always presented simultaneously with one of the maskers and all stimuli were 100-ms long. The position of the target tone was either varied randomly from trial to trial or was constrained to appear only in the first, third, or fifth slots. With both kinds of maskers, there was little or no benefit from reducing uncertainty about the temporal position of the target tone, apart from when the target tone appeared in the fifth and final position in the sequence. The absence of an effect of uncertainty about when the target stimulus would appear may have been due, in part, to the presentation of stimuli in a temporally-regular sequence, or due to the brief period over which the participants had to sustain their attention on the sequence.

In summary, uncertainty about *where* a target phrase will appear can be detrimental to speech perception performance in multi-source listening environments. Focused attention to, or the use of a spatial filter at, a known target location can assist listeners in ignoring irrelevant information from sources at other locations. By focusing on the wrong source location, which is more likely as uncertainty about target location increases, listeners either take longer to find a target, or confuse information presented at the attended location with that of the target. This confusion arises when there is sufficient similarity between target and masker, in terms of location, fundamental frequency, voice characteristics, and informational content. Small benefits may arise from knowing *who* will speak a target phrase. There is limited evidence that knowing *when* to listen provides an advantage in detecting a target stimulus. The relative benefits from knowing *who*, *where*, and *when* were examined within a single listening task for speech in Chapter 5.

3.3 Cortical Mechanisms of Attention

It is evident from the diversity and complexity of the theoretical models seeking to explain attention that there is no single nor simple representation of attentional processes within the brain. Just as the assertion that "everyone knows what attention is" (James, 1890) is incompatible with the continuing lack of a formal definition of what attention is and encompasses, the notion that there is a well-defined region of the brain for attention is not supported by several decades of research. Perhaps due to the complexity of attention, or due to the important role that attention plays in cognition, a large amount of research has been carried out in an effort to understand the cortical manifestations of selective attention, the switching and shifting of attentional focus, and the effects of distraction. In this section, I will outline the results of neuroimaging studies which have used electro-encephalography (EEG), magneto-encephalography (MEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI) to study the effects of attention at a cortical level. First, I will discuss the key regions of the brain which have been consistently implicated in attentional control, both from a top-down, or executive control, and a bottom-up, or stimulus-driven, perspective. Next, I will examine studies which have specifically examined the neural correlates of attention to auditory stimuli using a range of neuroimaging techniques. I will then outline studies which have used MEG and EEG techniques to examine the role of oscillatory activity, particularly at high frequencies (> 40 Hz), in the communication between the many cortical regions involved in attention. Finally, I will present the results of research which has used complex tasks of speech perception to examine the neural correlates of attention during "cocktail-party" listening.

3.3.1 Fronto-Parietal Network

In a seminal paper, Posner & Petersen (1990) brought together the results of then recent advances in neuroimaging through the use of EEG and PET in an attempt to understand the neural basis of attention. Posner & Petersen (1990) outlined three "fundamental findings" central to an understanding of attention. First, attention is a system of cortical processes which is distinct from but connected to systems which process sensory information. It is therefore an independent system which interacts and communicates with an array of cortical processes. Secondly, attention comprises a network of anatomical areas rather than being localised to a single cortical region and is not a result of the entire brain "operating as a whole". Thirdly, different areas in the attentional network carry out different kinds of processing; i.e. the individual components of the network are related to specific cognitive functions.

Posner & Petersen (1990) identified three cortical structures central to the processes of disengaging, shifting, and focussing of attention: the parietal cortex, mid-brain structures such as the superior colliculus, and the thalamus. Although the number of cortical regions implicated in attention has expanded since that review (Table 3.1) these three areas are still considered central to attentional processing (Posner, 2008). In visual research, attention to a location can decrease response times to items/events at that location (Posner et al., 1980). Attention also increases the amplitude of ERPs evoked by items/events compared to the same items/events presented at an unattended location (Hillyard, Hink, Schwent, & Picton, 1973). Using PET, such enhanced responses have been localised to the posterior parietal cortex (Corbetta, Miezin, Shulman, & Petersen, 1993). The parietal cortex has been found to be activated in many tasks which recruit attention processes (Duncan, 2006). Lesion studies have shown that damage to the parietal cortex can lead to deficits in the ability to disconnect attention from a source in the contralateral hemifield, therefore implicating the region in attentional shifts (Posner et al., 1980). Damage to the parietal lobes bilaterally has been found to affect the ability to shift attention between locations in both auditory and visual space (Phan, Schendel, Recanzone, & Robertson, 2000).

The parietal system can also be shown to play different roles in each hemisphere. While both parietal cortices are implicated in the process of shifting attention, lesion studies have suggested that there are hemisphere differences in their contribution to selective attention. The studies suggested that in the visual domain the right hemisphere is important for attention to low spatial frequencies, or at the 'global' level, and in the left to high spatial frequencies, or the 'local' level (Posner & Petersen, 1990). In the mid-brain, damage to the superior colliculus can introduce deficits in shifting attention (Posner, Petersen, Fox, & Raichle, 1988) and thalamic lesions can lead to difficulties focussing attention on the contralateral side to the lesion (Posner & Petersen, 1990).

Network	**Structures**
Orienting	Superior parietal
	Temporal parietal junction
	Frontal eye fields
	Superior colliculus
Alerting	Locus coruleus
	Right frontal
	Parietal
Executive control	Anterior cingulate
	Lateral ventral
	Prefrontal
	Basal ganglia

Table 3.1. Cortical and sub-cortical regions implicated in three attentional functions, or networks (adapted from Posner, 2008).

In relation to maintaining a general state of vigilance, Posner & Petersen (1990) identified the right hemisphere as linked with complex and attention-demanding tasks and, from a range of lesion and split-brain studies, as being important for adopting an alert attentional state. More specifically, the mid-frontal regions of the right cerebrum have been identified in tasks requiring continuous alertness (Deutsch, Papanicolaou, Bourbon, & Eisenberg, 1987; Posner, 2008). These processes involved in maintaining alertness can affect other attentional systems, specifically attentional systems located in the posterior parietal lobe of the right hemisphere (Posner & Petersen, 1990). In summary, from a large body of neuroimaging research, attention-related processes have been linked to several cortical and sub-cortical structures including, but not limited to, the parietal lobe (disengaging/shifting), the mid-brain (shifting), the thalamus (selecting/filtering), and the frontal lobe (vigilance).

As a result of the broad consensus across neuroimaging studies of the involvement of parietal and frontal areas in a diverse set of cognitive tasks, Duncan (2006) referred to the activation of the fronto-parietal network as the multiple-demand (MD) pattern. The "pattern" of cognitive demands which activated the network was found to include response conflict, task novelty, working memory, and perceptual difficulty (Figure 3.1). This meta-analysis of 20 studies lead Duncan (2006) to conclude that these regions are generally involved in demanding cognitive activity. More specifically, they may be involved in the selection of information through a process of competitive selection which can be biased by top-down demands related to task or behavioural goals. Duncan & Owen (2000) suggested that the network is highly flexible. Although certain regions have been found to be more frequently observed under certain conditions, the different regions within the network can adopt different roles based on task demands. This generality of function was identified particularly in pre-frontal regions of the cortex. It is congruent with the notion that while attention can influence and modulate activity in regions specialised in the processing of incoming information, there are regions of the cortex which are required for general attentional processes, including but not limited to the selection of, orientation to, and continuous detection of information (Posner & Petersen, 1990).

The concept of competitive selection is a key part of attention. We are often bombarded by a large amount of information from multiple sources and our ability to select information is important to function in many everyday environments. In a recent review, Knudsen (2007) combined processes of competitive selection with three other key cortical functions to propose a cortical model of attention. The other processes comprised bottom-up saliency filters, working memory, and top-down sensitivity control. Working memory (WM) has been proposed as a core function necessary for the extended processing of information, decision making, perception, and the planning and performance of actions (Baddeley, 2003). The concepts of attention and WM are closely related, and many of the cortical regions identified in studies of attention have also been found in tasks specifically designed to load working memory. These regions include ventro- and dorso-lateral pre-frontal cortex and inferior parietal cortex (Baddeley, 2003; Knudsen, 2007).

Our ability to bias selection processes towards task- or behaviourally-relevant information is evident from everyday situations in which attention is focussed on a single source of information while ignoring other inputs which may be continuously changing. The source of this *bias* on competitive selection may result directly from WM. From a review of studies which have examined the neural correlates of top-down executive control, Knudsen (2007) proposed that information is produced by WM which is fed back into selection processes to improve the SNR of incoming information. This might be achieved by adjusting the location of attentional focus in one or more dimensions; e.g. space, frequency, etc. Functions which exert such a bias have also been localised to the pre-frontal cortex but the other regions involved are specific to the task being performed (Corbetta et al., 1993; Desimone & Duncan, 1995; Pugh, Offywitz, Shaywitz, Fulbright, Byrd, Skudlarski, Shankweiler, Katz, Constable, Fletcher, Lacadie, Marchione, & Gore, 1996). Another source of bias on the competitive selection of information for further processing and access to working memory is bottom-up saliency filtering. Stimuli which are particularly salient, either due to the current environment, their infrequency, or certain features such as their intensity (e.g. loudness or brightness), can 'grab' attention involuntarily. To reach the level of our consciousness they must pass through several stages of information selection unhindered. This is accompanied by strong patterns of neural activation at short latencies relative to the stimulus onset which implies that the bottom-up saliency filters are active at early pre-cortical stages of stimulus processing (Knudsen, 2007).

There are many areas outside the fronto-parietal network which have also been suggested as fundamental to attention. In considering target detection and the

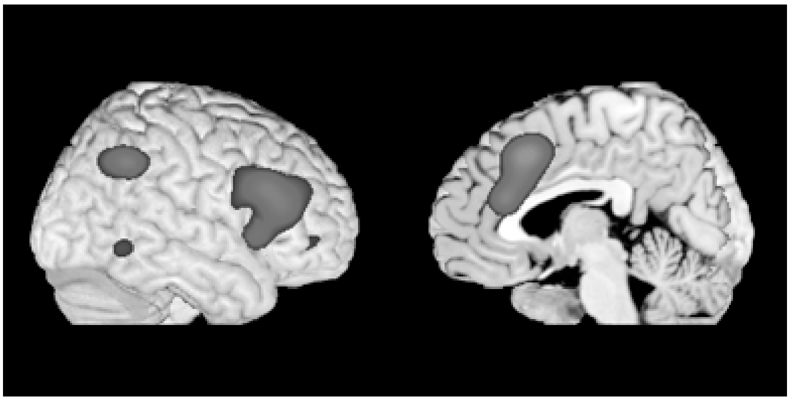

Figure 3.1. Cortical regions in the multiple-demand (MD) network activated by a range of cognitive tasks. The data were extracted as regions of interest from 20 studies and have been smoothed. The structures include the inferior frontal sulcus & gyrus, anterior insula, anterior cingulate, and the intraparietal sulcus (after Duncan, 2006).

focussing of attention, Posner & Petersen (1990) distinguished between a general alert state, in which monitoring multiple locations across multiple modalities does not introduce a noticeable penalty compared to monitoring a single modality, and a state in which attention is focussed on a target and is processing sensory information in relation to that target. The suggestion is that there is an independent global system for monitoring or maintaining a state of vigilance. In tasks of target detection, the anterior cingulate cortex (ACC) has been found to be modulated by the number of targets that must be detected. Posner et al. (1988) presented participants with single words visually which had to be categorised as dangerous (target items) or not dangerous. By varying the number of possible target items, the regional cerebral blood flow (rCBF) in ACC measured using PET was found to increase when the number of target items was also increased. Similar changes in other areas implicated in monitoring visually presented word forms such as the lateral aspect of the frontal cortex were not observed. The ACC has therefore been linked with both target detection, specifically when attention is focussed on generating actions; i.e. responding to target items. Posner & Petersen (1990) suggested that the known connections between the ACC and posterior parietal cortex and also from the ACC to pre-frontal cortex (Goldman-Rakic, 1988) makes it a prime candidate for involvement in tasks which involve attention and language processes, and that the role of the ACC is as an anterior attentional system which mediates the relationship between processes in frontal areas and the posterior system in the parietal lobe based on the overall processing load or the activity within the anterior systems.

In summary, several regions of the human cerebral cortex have been found to be activated across a range of tasks which require different aspects of attentional control. These regions have been identified using PET and fMRI and include a fronto-parietal network comprising the pre-frontal cortex and the parietal lobe, and also other regions such as the anterior cingulate cortex. Despite a large body of research, the components of this “multiple-demand” network are still not completely understood, both in terms of their individual functions and interactions. The description of the cortical representation of attention is far from complete, in part due to the complexity of the concepts underpinning attention but also because of variations in cortical activation across different tasks. Neuroimaging studies have provided a basis for the development and improvement of models of attention by examining how changes in attentional load and/or state affect neural activity, by identifying common regions recruited by a diverse range of attention-demanding tasks, and by studying how attention can modulate the activity of other cortical processes including sensory information processing.

3.3.2 Auditory Selective Attention

A large proportion of the attention literature is based on experiments in the visual domain. The suggestion that processes of attention recruited within the brain vary based on the task being performed (Knudsen, 2007) implies that studies which have used auditory stimulation are central to an understanding of the role of attention in complex listening tasks. Many early studies of auditory attention used EEG to record changes in the auditory ERP due to attentional manipulations. Hansen & Hillyard (1980) used sequences of tone pips at two frequencies, one low and one high, to examine the effects of attention on scalp-recorded ERPs. The task of the participants was to attend to one of the tone streams (low or high) and identify longer tones embedded in the sequence of shorter tones. A slow negative difference ERP, computed by subtracting the responses to the attended and non-attended shorter tones, was found to occur between 100–400 ms after the onset of the tones. This difference wave was referred to as the attention-related negativity, processing negativity, or the negative displacement (Nd). Hansen & Hillyard (1980) suggested that the Nd and the modulation of the auditory N_1 component of the auditory ERP reflected processes of selective attention. In addition to being evoked by attention to stimuli identified through pitch differences, more recent research has found that the Nd is also evoked by attention to stimuli that are distinguished by their spatial location (Teder-Sälejärvi & Hillyard, 1998). This finding supports the assertion that the Nd reflects cortical activity which encompasses more general processes of attention, rather than processing which is specific to a particular stimulus characteristic.

While such ERP results give an indication as to changes in the magnitude and extent of cortical activity, and also the synchrony between neurons which would lead to larger measured electric or magnetic fields, they do not provide direct information about the location of the generators. Measurement of ERPs cannot clearly distinguish between the number of neural sources or different cortical regions which give rise to ERPs, and those regions which may modulate their strength and latency. In a study on auditory attention, Frith & Friston (1996) used PET to examine the cortical regions which show enhanced rCBF as a result of selective attention. The results revealed no enhancement in auditory cortex when attention was directed towards a stream of tones compared to when attention was focused on a stream of visually-presented letters and the stream of tones was ignored by the listener. In contrast, increased rCBF was seen in visual areas when attention was direct towards the visual stream compared to the auditory stream. Frith & Friston (1996) suggested that the discrepency between the ERP and PET results might be due to an increased synchrony in auditory areas with selective attention leading to enhanced ERPs but not necessarily greater rCBF which is modulated by the number of active neurons. The study did identify the mid-thalamus in response to attention to tones only.

The authors suggested that this structure could be involved in organising the neural synchrony in auditory areas, leading to the enhanced ERPs measured in other studies (Hansen & Hillyard, 1980; Hillyard et al., 1973; Näätänen & Michie, 1979).

Stronger parallels can be drawn with the ERP literature and the localisation of ERP sources through the measurement of event-related fields (ERFs) with MEG. Magnetic fields resulting from neural activity are not distorted by the surrounding tissues and skull (Okada, Lauritzen, & Nicholson, 1987). One of the first studies to localise the generators of the Nd attention-related ERP component was conducted by Woldorff, Gallen, Hampson, Hillyard, Pantev, Sobel, & Bloom (1993). Participants were presented with short tone bursts at 1 kHz in the left ear and 3150 Hz in the right ear while MEG and EEG recordings were made over the left hemisphere. The task was to attend to a specified ear and identify rare tones which were less intense than the majority of the tones. Attention effects were calculated as the difference between ERPs/ERFs in the attend left and attend right conditions. Woldorff et al. (1993) found a difference wave between the attended and non-attended conditions as identified by previous ERP studies (e.g. Hillyard et al., 1973). Similar attention-related effects were observed in the electric and magnetic recordings. The effect was identified as having a dipolar pattern centred over the sylvian fissure and was localised lateral to Heschl's gyrus. Woldorff et al. (1993) concluded that the attention-related modulation of the N_1, and its magnetic equivalent the N_1m, are due to modulation of the generators of those components, which are located in primary auditory cortex. In addition, an earlier component starting as early as 20 ms after stimulus onset was identified as being modulated by attention. This component was localised close to the later component and was taken as evidence that attention can impose effects on sensory processing at very early latencies.

Using a similar paradigm to Frith & Friston (1996), Woodruff, Benson, Bandettini, Kwong, Howard, Talavage, Belliveau, & Rosen (1996) examined the blood oxygen level-dependent (BOLD) signal using fMRI when attention was focussed on either a stream of auditory stimuli or visual stimuli presented simultaneously. The stimuli comprised digits which were either spoken or shown on screen. Comparing the conditions in which participants attended to the auditory stream to conditions in which they focussed on the visual stream revealed increased BOLD in the sensory cortex of the modality being attended to. This finding that activity in early and primary auditory areas could be modulated by selective attention was also observed using PET by Alho, Medvedev, Pakhomov, Roudas, Tervaniemi, Reinikainen, Zeffiro, & Näätänen (1999) using an odd-ball task. Sequences of 400 Hz standard tones within infrequent 'deviant' 500 Hz tones were presented to the left and right ears independently together with a visual stream of letters also comprising standard and deviant stimuli. Participants were instructed to attend to the left steam, the right steam, or the visual stream. A comparison of rCBF in the attend-visual and each of the attend-auditory conditions revealed activation in prefrontal cortex and enhanced temporal activation in the hemisphere contralateral to the attended ear. The focus of this lateralised activity was on the superior temporal plane close to Heschl's gyrus. Thus, in contrast to Frith & Friston (1996), the results of Woodruff et al. (1996) and Alho et al. (1999) supported the hypothesis that had arisen from ERP research (Hillyard et al., 1973) that selective attention modulates the firing rate or number of neurons in primary auditory cortex. This increased firing rate may reflect the stimulus selection process or an enhancement of the SNR as a result of an earlier stimulus selection process. The contrast between the similar PET studies may have arisen due to the use of a more rapid stimulus presentation rate (10 Hz for Alho et al. vs a maximum rate of 0.6 Hz for Frith, et al.) which Frith & Friston (1996) identified as affecting rCBF. An alternative explanation is that the presentation of different tone sequences to each ear meant that participants had to actively ignore the contralateral auditory input, requiring a more complex and stronger attentional response. The prefrontal activation was linked to volitional control of spatial selective attention and Alho et al. (1999) concluded that it was responsible for controlling the selective tuning in auditory cortex.

The presentation of auditory stimuli while imaging the brain is complicated by the noise produced by imaging methods with high spatial resolution, such as fMRI. While techniques like EEG are silent and provide excellent temporal resolution, the localisation of activity is relatively difficult and inaccurate. These challenges are likely to have contributed to fewer studies of auditory attention compared to the visual literature. Nonetheless, fMRI has been successfully applied to the study of auditory attention, and several of those studies are discussed in Section 3.3.4.2.

3.3.3 Attention and Oscillatory Activity

The use of EEG and MEG allows for the examination of neural activity at a wide range of frequencies. While traditional analysis of event-related potentials (ERPs) typically involves activity below 30 Hz, recent studies have examined the role of high-frequency (or gamma-band, $> 30\ Hz$) activity in working memory and processes related to attention. These high-frequency signals are not necessarily tightly phase-locked to the onset of a stimulus. The signals are therefore referred to as being *induced* by the stimulus, rather than being *evoked* by it. Evoked signals are phase-locked to the stimuli or events within the stimuli and are preserved with averaging of the MEG data over trials; i.e. ERP analysis. In contrast, induced signals are not strictly phase-locked to the stimuli or events within the stimuli (Tallon-Baudry & Bertrand, 1999). Signal averaging, which is commonly used in the analysis of EEG and MEG data, has the effect of attenuating activity which is not tightly phase-located to the stimulus (Figure 3.2). Induced signals in the raw sensor data, or in 'sensor-space', can be preserved by transforming the data from individual trials into the frequency or time-frequency domains prior to averaging. Induced activation at the cortical level, or in 'source-space', can be estimated by calculating the data covariance matrix from the individual trials. An example of such a technique is *Spatial filtering* and is discussed in detail in Chapter 4.

Large populations of neurons are necessary for the creation of signals that can be measured with EEG and MEG (Jensen, Kaiser, & Lachaux, 2007). When such populations of neurons fire in synchrony, oscillatory signals can be measured at the surface of the scalp (EEG) or around the head (MEG). The synchronisation of neurons has been suggested as a mechanism for communication both within and between regions of the brain. As the core concept of attention involves the organisation of information through competitive selection, suppression, enhancement, and other processes, the relationship between oscillatory activity and information-centred cognitive processes such as attention is of great interest. Salinas & Sejnowski (2001) proposed that the synchronisation of firing patterns amongst a group of neurons is related to the effects on other neural circuits 'downstream' from the synchronous

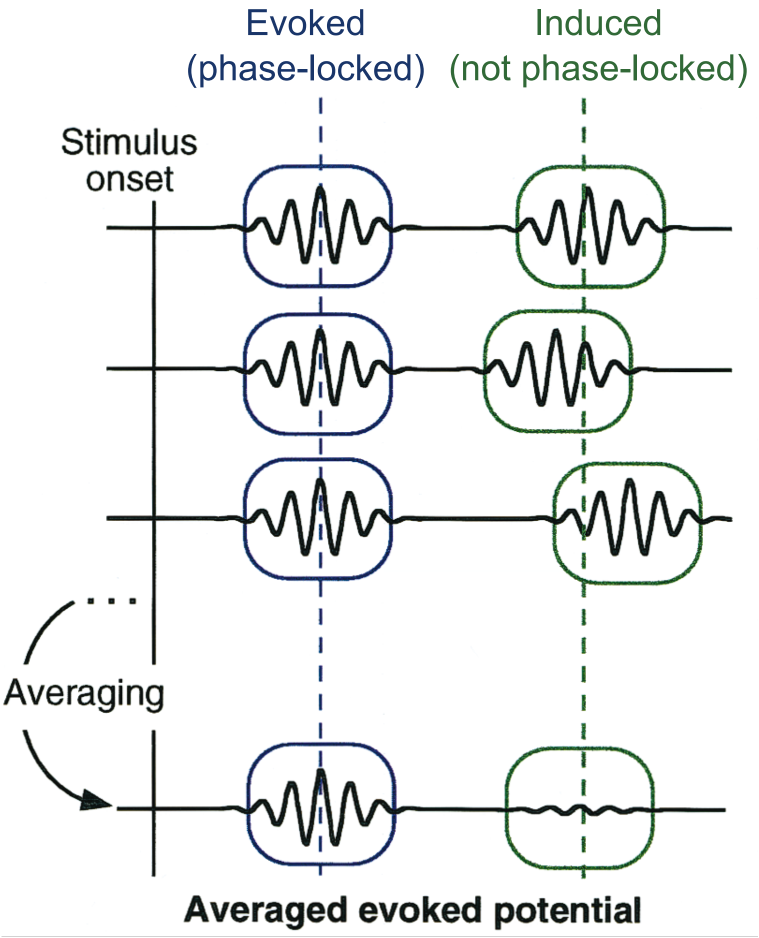

Figure 3.2. The effect of averaging on activity which is evoked by (phase-locked to) the stimulus and activity which is induced (non-phaselocked) by the stimulus. Time-domain averaging only preserves the evoked activity (adapted from Tallon-Baudry & Bertrand, 1999).

group. Key to this proposal is the fact that the pattern of output from a single neuron is dependent on both the firing rate of neurons from which it receives signals and also the correlation, or synchrony, between the input neurons. Neural synchrony may therefore be important for communication between distributed regions of the cortex by providing sufficient signal-to-noise for long-range transmissions.

Oscillatory activity at frequencies above 30 Hz has been linked with attentional processes. In particular, oscillations in the γ-band (*gamma* frequencies, usually between 40–80 Hz) have been implicated in selective attentional processes and has been found to be due in part to cortico-thalamic communication. By measuring magnetic fields over the right hemisphere using MEG, Ribary, Ioannides, Singh, Hasson, Bolton, Lado, Mogilner, & Llinás (1991) observed γ-band oscillations at 40 Hz in response to the presentation of a range of auditory stimuli. The oscillations were found to start in the thalamus and project to large regions of the cortex, and were suggested as being related to the synchronisation of a wide network of cortical regions. The focus of the oscillatory activity was found in auditory regions, suggesting that the oscillations were also related to the processing of incoming sensory information. Furthermore, Ribary et al. (1991) found that the coherent cortico-thalamic oscillatory pattern was insensitive to the rate of stimulus presentation, was tightly phase-locked to the onset of stimuli for a period up to 200 ms, and was observed for both steady-state 40-Hz tones and stimuli whose frequency content was chosen at random. These results suggest that 40-Hz oscillatory activity does not arise solely due to the content of stimuli, i.e. those with energy at 40 Hz, but rather a general function of the cortex. The widespread nature of the oscillatory activity and the cortico-thalamic connections may indicate that this function reflects the functional coupling of different regions of the cortex.

Tiitinen, Sinkkonen, Reinikainen, Alho, Lavikainen, & Näätänen (1993) examined the effect of attention on γ oscillations by presenting participants with sequences of standard tones (700 ms presentation rate) containing infrequent deviant tones of a different frequency using EEG. Independent streams were presented to both ears. The task was to indicate the presence of the deviant tones with a response. Participants either attended to the stimulus stream in one ear and ignored the stimuli in the other ear (attend condition) or read a book and ignored all auditory stimuli (reading condition). An analysis of the transient evoked response between 30–60 Hz revealed a peak of oscillatory activity centred on 40 Hz which was modulated by the attentional condition. In the attend condition, the largest response was observed for the attended stimuli by electrodes over the contralateral hemisphere, and a smaller response was observed for unattended stimuli by electrodes ipsilateral to the attended ear. The smallest responses were found for stimuli presented to either ear while the participant read a book. No affect of attention was found at lower frequency bands. Thus, Tiitinen et al. (1993) suggested that the 40 Hz transient response reflected an aspect of selective attention, which differed in each of the three attentional states: attending, ignoring/suppressing, and not attending.

Similar to the use of high-frequency oscillations to bind related information, the mechanisms by which information is passed between the distinct regions of the cortex associated with attentional processes are relevant to our understanding of attention. Gross, Schmitz, Schnitzler, Kessler, Shapiro, Hommel, & Schnitzler (2004) used the "attentional blink" effect, in which the detection threshold of a target stimulus is modulated by the occurrence of another target which precedes the second target by less than 500 ms. This effect is suitable for studying attention as the perceptual processing of the second target item is not inhibited by the preceding target, but rather the participant does not become consciously aware of the information arising from that processing (Luck, Vogel, & Shapiro, 1996). Participants were presented with a rapid sequence comprising 13 letters on a screen while neural activity was recorded using MEG. Each sequence contained up to two targets. When two targets were present, they were either separated by a single non-target item ($\sim$ 300 ms between target stimuli) or by 5 non-target items ($\sim$ 900 ms).

When comparing the differences in the response to target items which were correctly identified and the response to non-target items, a peak representing greater activation in response to the target items was found between around 15 Hz, within the β-band (*beta* frequencies, usually between 13–30 Hz), 400 ms after the onset of the items. Functional maps of the activity underlying the difference peak revealed activity in frontal, occipital, posterior parietal, and temporal regions. The parietal and temporal activation was right lateralised and the frontal activity was left lateralised. By examining the phase synchrony of the 15-Hz oscillations between these cortical regions in response to target and non-target items, a network of regions whose synchrony was modulated specifically by target items was identified. Gross et al. (2004) compared the level of phase synchrony in this target-related network on the trials containing two target stimuli separated by one item; i.e. those trials which were likely to give rise to an "attentional blink". They found that trials in which both targets were identified (no observed "attentional blink" effect) exhibited higher levels of phase synchrony between the target network compared to trials in which the second target was not identified; i.e. when the participants showed the "attentional blink" effect. This increase in synchrony was apparent before the onset of the first target stimulus. In summary, Gross et al. (2004) found enhanced β synchronisation in response to target items, and stronger β synchronisation throughout trials where participants sustained attention and hence did not display the "attentional blink".

In light of these two results, Gross et al. (2004) concluded that synchronisation of neural activity within the β frequency band was associated with a more alert attentional state which lead to the successful identification of both target stimuli. The localisation of the synchronised activity to areas previously identified in studies of attention provided further support for this hypothesis. The ability of enhanced communication within a network through phase synchrony in oscillatory activity to affect the detection sensitivity of task-related stimuli may indicate that the synchronisation affects processes of competitive selection. Such processes control the selection of information which receives further processing based on the capacity of the processing network and top-down goal-directed feedback from higher-level cognitive processes (Desimone & Duncan, 1995; Knudsen, 2007). Thus, executive function may exert a bias on the selection process in favour of information which is task or behaviourally relevant. Gross et al. (2004) suggested that this top-down directive may be directed to a diverse network of processes through the use of synchronisation. The frequency of oscillation may relate to the distance between the cortical structures being synchronised. Models of neural activity have suggested that frequencies in the γ-band are suitable for short range or local communication and that frequencies in the β-band are suitable for long-range communication (Bibbig, Traub, & Whittington, 2002).

The findings of Gross et al. (2004) suggest that oscillatory activity may be important in understanding the attentional state of the participants, and an insight into the manner in which information is selected or attended to. The relationship between γ- and β-band activity and attention was examined using MEG in Chapter 7.

3.3.4 Attention to Speech in Noise

3.3.4.1 ERP Research

Although the number of behavioural studies which have attempted to study performance in realistic complex listening situations has increased over the 50 years since Cherry described the ‘cocktail party’ listening situation (Cherry, 1953), and despite the recent technological advances in the resolution of MRI scanners and density of electro- and magneto-encephalography (EEG/MEG) arrays, there are very few studies which have examined the neural processes which underpin our ability to listen to speech in such complex situations. Further to this, even fewer imaging studies have attempted to recreate the attentional demands of these environments, which often require the dividing, switching, focussing, and sustaining of attention.

Perhaps due to the absence of scanning noise as compared to fMRI, the ERP literature contains a number of studies which have attempted to study the cortical response to continuous speech. One of the earliest ERP studies to examine cortical activation for attended and unattended continuous speech messages was conducted by Hink & Hillyard (1976). Participants were presented with two continuous speech streams over headphones, a female talker in the left ear and a male talker in the right ear, into which task-irrelevant probe stimuli had been inserted. The probe stimuli comprised synthesised phonemes and their fundamental frequency was matched to the voice of the talker with whom they were paired. Participants were instructed to attend to the left or right speech stream and were questioned on the contents to confirm that they had been attending. Hink & Hillyard (1976) found significantly larger peaks corresponding to the N_1 and P_2 components of the auditory ERP in response to the attended probe stimuli compared to the unattended probes. This finding was similar to previous studies which examined attentional effects on ERPs in response to simple tone bursts (Hillyard et al., 1973). The auditory N_1 was therefore suggested as a reflection of early selective attentional processes which also indexes the bias towards stimuli in the attended 'channel', which includes spatial and spectral domains. This 'enhancement' of the N_1 was later reinterpreted as being due to the coincidence of the Nd and the N_1, and was termed the 'Hillyard effect' (Näätänen, Gaillard, & Mäntysalo, 1978; Näätänen & Michie, 1979).

Woods, Hillyard, & Hansen (1984) also presented participants with dichotic speech material containing probes. The speech stimuli were constructed from a male and a female talker reading prose. The probes comprised speech material in the form of the CVC syllable /bʌt/ and the vowel segment /a/ recorded using the same talkers, and non-speech material comprising tone-bursts at frequencies corresponding to the fundamental frequencies of the two talkers and also average second formant frequencies for male and female talkers. The comparison of ERPs for the attended and unattended speech probes revealed a broad negative-going wave with two components: an early phase starting at 50–100 ms after the onset of the probe and a late phase with an onset between 250 and 300 ms. The attentional modulation was found to be similar to the Nd component. In contrast, attention to the non-speech second formant probes did not elicit a comparable effect, instead producing an attention-related transient positive peak. No attentional effect was observed for the tone probes at the fundamental frequency. Earlier studies had found that the Nd is elicited by stimuli which are similar to those receiving attention (Hansen & Hillyard, 1980; Näätänen & Michie, 1979). The attentional modulations observed by Woods et al. (1984) which were specific to speech probes were taken as evidence for the tuning of selective attention processes to speech or speech-like stimuli during speech listening.

So far, only ERP studies which have presented stimuli dichotically over headphones have been considered. In a recent study, Nager, Dethlefsen, & Münte (2008) extended the work of Woods et al. (1984) by presenting three concurrent speech messages, a female talker at 0° azimuth and two male talkers at ±70°, over headphones by recording the messages in the free field using a artificial head with microphones in each ear canal. Overlaid onto each speech stream was a sequence of probe stimuli, in the form of the syllable /da/. At each spatial location, the same talker was used for the continuous speech and probe syllables. Participants were instructed to attend to either the left- or right- lateralised speech message spoken by one of the male talkers and to ignore the other messages. To ensure that participants were attending, they had to answer questions about the attended speech stream between presentations. A comparison of the ERPs of attended and unattended probe stimuli revealed a negative deflection from 300–500 ms with a peak at $\sim$ 375 ms for the attended probes. The size of this deflection was found to be significantly modulated by attention. The topographical distribution of the negative ERP was similar to the Nd, although the latency was later than identified in previous studies which was attributed to the complexity of the task. The study demonstrated that classical attention-related modulations of the auditory ERP could be elicited by stimuli which were presented in a virtual auditory space. The use of these techniques provides a methodology to study more complex listening situations when only headphone presentation is available, such as is common in MEG, PET, and fMRI.

3.3.4.2 Selective Attention to Speech with fMRI and PET

Activation patterns related to the perception of individual phonemes, words, and sentences have received extensive research (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Hickok & Poeppel, 2000; Scott & Wise, 2003; Vigneau, Beaucousin, Hervé, Duffau, Crivello, Houdé, Mazoyer, & Tzourio-Mazoyer, 2006) and have identified several important structures in speech and language processing, including Wernicke's area on the superior temporal gyrus (Brodmann area 22) and Broca's area in the inferior frontal gyrus (Brodmann areas 44/45). In contrast, few fMRI studies have examined the pattern of metabolic changes within the cortex in response to speech-in-speech or speech-in-noise listening. These tasks have a strong attentional component that is not evoked by listening to speech in quiet and are therefore important in understanding the cortical processes involved in the perception of speech in everyday situations.

Of those studies which have examined speech in speech or speech in noise, many have used the classical dichotic listening paradigm (Cherry, 1953; Kimura, 1967; Treisman, 1969) in which different speech streams are presented to each ear and the participant is instructed to attend to one stream only. Hashimoto, Homae, Nakajima, Miyashita, & Sakai (2000) used fMRI to compare cortical activity when participants performed a dichotic listening task and a diotic listening task in which both ears received the same stimulus. Two types of stimuli were used: speech and non-speech. For the speech stimuli, sentences from a story were synthesised and divided into smaller phrases between 400–700 ms in length. These were the target speech stimuli. Non-target speech stimuli were created by reordering the syllables of the phrases. For the non-speech stimuli, 600 ms bursts of white noise were used as non-targets and the target stimuli contained an additional tone added to the noise. Scanning blocks comprised non-speech sounds presented diotically (control) followed by diotic and dichotic speech blocks in alteration. The diotic speech blocks either contained target or non-target speech stimuli, and the dichotic blocks either contained a target/non-target speech pair or two non-target speech stimuli. Participants indicated the presence of a target stimulus, either speech or non-speech. The contrast between diotic and dichotic blocks which contained target speech stimuli would highlight those areas of the cortex which were involved in selective attention to either ear and the processes of suppressing or ignoring input from the contralateral ear. For that contrast, enhanced activation in the dichotic condition was found for auditory areas bilaterally: secondary auditory cortex on the medial portion of Heschl's sulcus, the anterior superior temporal gyrus, and planum temporale. Increased activation for dichotic listening was also found in the inferior frontal gyrus bilaterally and in the anterior insula of the right hemisphere. The differences in auditory areas could have arisen due to the introduction of inter-aural differences only in the dichotic condition and the required use of selective attention and segregation processes. Furthermore, Hashimoto et al. (2000) found evidence that sub-regions within IFG, STG, and PT exhibited different responses to the experimental manipulations; i.e. selective attention and segregation in the dichotic condition and speech processing in both diotic and dichotic conditions. Therefore, while these regions may have been identified in studies of speech and language comprehension, it is not necessarily the same regions of cortex which are activated when additional attentional and segregation demands are introduced.

In a recent study, Nakai, Kato, & Matsuo (2005) examined such responses when participants had to follow a story in the presence of another talker. A spoken narrative was presented diotically in alternated blocks, which comprised a single target talker or the same talker and an additional distracting talker speaking simultaneously. The participant had to follow the story being spoken by the target talker across all blocks, even those containing the second talker. In one condition, the second talker was identical to the target talker but spoke a different narrative. In the other condition, the second talker was of the opposite sex as the target talker. Self-reported difficulty scores revealed that the participants found the same-voice distractor condition very difficult and the different-voice distractor condition moderately difficult. When only a single talker was present, activation was observed bilaterally along the transverse temporal gyrus and more laterally on the superior temporal gyrus (STG). These regions had also been found in a study of speech comprehension (Nakai, Matsuo, Kato, Matsuzawa, Okada, Glover, Moriya, & Inui, 1999). When the same-voice distractor talker was present and participants had to segregate and focus on only one of the speech streams using contextual information, a wide range of additional cortical areas were activated. The network included bilateral pre-central gyrus, bilateral middle frontal gyrus, bilateral frontal operculum, right supramarginal gyrus, and cingulate cortex. When compared with the single talker condition, significant additional activation was found in several areas including left posterior STG, left transverse temporal gyrus, bilateral frontal operculum, right posterior inferior frontal gyrus (IFG), and cingulate gyrus, many of which have been associated with the "Multiple demand" network activated in a diverse range of complex tasks (Duncan, 2006). In contrast, the condition in which the different-voice distractor was present activated the same regions as the single voice condition; i.e. bilateral STG. When this different-voice distractor condition was compared to the single talker condition, significantly more activation was found only in left and right posterior STG. Thus, the simple addition of a second talker had an extensive effect on the network of cortical areas involved in processing and attending to the target talker when the distractor talker was the same as the target but not when the two voices were different.

The experiment was very likely to have loaded working memory, as the context of the story was essential in segregating the two speech streams when spoken by the same talker. The activation of the right middle-frontal gyrus, IFG, and the pre-central gyrus supports this assertion (Baddeley, 2003). Although a situation with two concurrent identical talkers is not likely to occur in everyday life, situations in which multiple people are speaking at the same time are more likely to require the listener to rely on context to identify information relevant to the current conversation than if there was a single person talking. Comparing the two conditions with two voices, Nakai et al. (2005) found that anterior cingulate cortex, the pre-central and middle-frontal gyrii in the left hemisphere, and the middle-frontal gyrus, frontal operculum, and supplementary motor area in the right hemisphere were associated with the more difficult task of segregating the two identical voices. The activation of anterior cingulate cortex was taken as reflecting the use of selective attention in a complex task by aiding in the selection of relevant information based on feedback from executive control processes. The IFG and frontal operculum activation in the same-voice distractor condition suggests that detailed analysis of the prosody of the speech streams was necessary to support the segregation process, perhaps based on rhythm (Platel, Price, Baron, Wise, Lambert, Frackowiak, Lechevalier, & Eustache, 1997). In contrast with studies of selective attention to speech presented alone in words, sentences, or phonemes, or two speech streams presented independently to the two ears, the process of segregating speech which is overlapping in the auditory input involves a widespread and complex network of neural processes. The distributed nature of the regions identified by Nakai et al. (2005) emphasise the possibility that decrements in any one of multiple cognitive processes, including rhythm processing, syntactic and prosodic analysis of speech, competition management, and working memory could be involved in difficulties with "cocktail-party" listening situations.

Salvi, Lockwood, Frisina, Coad, Wack, & Frisina (2002) examined the differences in rCBF that arise when participants listen to speech in quiet and in noise. Stimuli were from the Speech in noise (SPIN) test (Bilger et al., 1984; Frisina & Frisina, 1997) and comprised sentences which contained a target word at the end which could be predicted from the context of the sentence. The background noise was multi-talker babble from the revised SPIN test (Bilger et al., 1984). Participants were instructed to speak the final word of each sentence, and to say ‘nope’ if they could not. This was also the case when the noise was presented on its own to match motor and speech production activity across conditions. As in previous studies (Hashimoto et al., 2000; Nakai et al., 1999), activation was found in STG bilaterally for the speech in noise condition, and the extent of activation in these regions was greater than in the speech in quiet condition. Activation to speech in noise was also seen in the thalamus and medial frontal areas close to the anterior cingulate. Both of these areas were found to be significantly more active in the speech in noise condition than in the speech in quiet. These areas have been associated with complex tasks and high attentional demands, particularly for shifts of attention and when there is competition between multiple sources of information (Posner & Petersen, 1990). The increased processing demands of segregating speech from the background noise and a possible increase in the alertness or arousal state of the participant in the more challenging task lead to activation of a wider attentional network through the thalamus which has projections to frontal and parietal cortices (Posner & Petersen, 1990).

Studies have also examined the cortical networks associated with selective attention to speech when information from another sensory input is ignored. The findings of Woodruff et al. (1996), discussed earlier, suggest that attention to a modality enhances the cortical response to stimuli within the areas of the cortex associated with sensory processing for that modality. Sabri, Binder, Desai, Medler, Leitl, & Liebenthal (2008) extended these findings to speech by presenting participants with spoken nouns and visually-presented Japanese characters simultaneously. Attention was either focussed on the auditory or on the visual stimuli, creating ‘attend auditory’ and ‘ignore auditory’ conditions. A comparison of the cortical responses, as measured with fMRI, in the auditory attend and ignore conditions revealed significantly greater BOLD signals bilaterally in superior and middle temporal gyrii, and superior temporal sulcus; i.e. areas involved in the processing of auditory stimuli. This finding was similar to that of Woodruff et al. (1996) who presented digits in the auditory and visual domains, and is a general finding in studies of auditory selective attention (Hashimoto et al., 2000; Nakai et al., 1999, 2005; Salvi et al., 2002; Woodruff et al., 1996). Significantly elevated attention-related BOLD signals were also found bilaterally in the inferior parietal lobe, in left supramarginal gyrus and right angular gyrus. Other regions which showed an attention-related enhanced response were left prefrontal cortex including the IFG, left orbitofrontal gyrus in the frontal lobe, bilateral MFG, right post-central gyrus, and cingulate cortex. Several of these areas were also identified when segregating two speech streams spoken by the same talker (Nakai et al., 2005) or when attending to a speech message in one ear while ignoring speech input to the other ear (Hashimoto et al., 2000). Thus, the distributed set of regions across the temporal, frontal, parietal, and cingulate regions of the cortex are implicated in the performance of complex tasks which involve selective auditory attention. The activation of the cingulate cortex is congruent with the selection and shifting between multiple informational sources, in line with previous findings (Salvi et al., 2002; Posner & Petersen, 1990).

In summary, attention to speech in noise recruits processes of attention, audition, language, and working memory. Therefore, the neural representation of performance on such tasks comprises a widespread network of cortical regions. Listening to speech in noise recruits 'classical' attentional areas, such as pre-frontal cortex, the parietal lobe, and the anterior cingulate cortex. Attention-related activation was also identified in auditory areas along the superior temporal gyrus. Of particular note is that the network of cortical regions recruited was determined by the task demands. In more difficult listening situations, STG activation was enhanced and additional areas were found to be recruited when the demands of the task were increased; i.e. when noise was introduced or when participants had to segregate two similar speech streams. Thus, to fully examine the cortical regions involved in "cocktail-party" listening, it is necessary to use tasks of spatial listening for speech which more closely recreate the high level of demand that arises in everyday situations.

3.4 Cognitive Effects of Ageing

3.4.1 Changes in Central Processing

There are many general changes that our central nervous systems undergo as we age. These include physiological changes, such as neural loss, loss of synaptic connections to other neurons due to depletion of dendritic masses, and dysfunction of both excitatory and inhibitory neuro-transmitter systems (Kok, 2000; Willott, 1996). There are also changes in the cochlea and spiral ganglion cells with advancing age (Schuknecht & Gacek, 1993). Therefore, the input to the central auditory system (CAS) changes with age, most often with a reduction of high-frequency information from the cochlea. As the behaviour of neurons can be substantially affected by changes to their normal synaptic inputs, it is likely that the CAS adapts to the altered input from the periphery (Willott, 1996).

One way in which it is possible to observe how the auditory system adapts to peripheral changes is through the re-organisation of tonotopic mappings. It has been well established that there is a tonotopic organisation in many parts of the auditory system, i.e. neurons only respond to a specific set of frequencies and such neurons are organised spatially within the nervous system in a low-to-high-frequency arrangement. These frequency mappings can be affected by a change in the input from the auditory periphery. For example, when presbycusis causes a diminished or completely absent input to those neurons which respond to high-frequency sounds, the neurons in question become responsive to lower frequencies (Willott, Aitkin, & McFadden, 1993).

Other general observations have been made about central processing in elderly adults, such as an overall slowing of all cognitive processes. This slowing tends to be present across a wide range of tasks, including those related to memory, reasoning, and spatial abilities (Kok, 2000). Processing speed, while not the exclusive factor, has been considered a major contributor to age-related effects across a wide range of cognitive tasks (Salthouse, 1996). Another common finding amongst elderly adults, referred to as the 'complexity effect', is that as task difficulty increases, response time also increases proportionally, at a greater rate than is found for young adults. Again, a decline in processing speed has been suggested as a contributing factor to this effect, as has a reduction in processing resources, and information loss (Kok, 2000).

3.4.2 Temporal Processing

The difficulties that elderly people experience in complex listening environments may relate to a decline in the ability to take advantage of gaps in background noise, i.e. it is the temporal complexity which poses the greatest problem. Such temporal fluctuations are a characteristic of multi-talker environments, and a failure to utilise them as temporal releases from masking could affect performance in such circumstances. Dubno, Horwitz, & Ahlstrom (2002) found that temporal release from masking, using modulated noise maskers, is greater for young listeners than for elderly adults. However, Souza & Turner (1994) have argued that the reduction in ability to take advantage of the temporal masking release that occurs with fluctuating noise maskers is primarily due to hearing loss, rather than age.

Changes in gap-detection are indicative of temporal processing ability (Akeroyd & Summerfield, 1999) and may therefore be related to or contribute towards speech perception directly. While the link between changes in temporal resolution and speech perception is unclear, it has been shown that there are age-related changes in temporal processing, and that those changes are not directly related to changes in auditory sensitivity (Snell & Frisina, 2000). Impairment in temporal analysis often accompanies hearing loss, but reduced peripheral sensitivity is not always indicative of poor temporal analysis performance in elderly people (Tyler, Summerfield, Wood, & Fernandes, 1982). Tyler et al. (1982) suggested that gap detection and stimulus duration discrimination could be important for speech perception, and could contribute to poor performance in elderly listeners. While Tyler et al. (1982) found relationships between their psychoacoustic measures and speech perception tasks, other studies have not found such strong relationships (Snell & Frisina, 2000). A fundamental issue related to this is that temporal processing is usually assessed with non-linguistic stimuli, which are thought to be processed in fundamentally different ways from speech, involving different functional areas of the brain (Martin & Jerger, 2005).

3.4.3 Age-related Changes in Attentional Processes

An important point about age-related changes in cognitive functioning is that many of the proposed causal factors imply that such changes are widespread and not specific to any subset of cortical or anatomical areas. If one considers the view that processes involved in attention are widespread, distributed across several functional areas of the brain, then the impact of the effects of ageing is likely to affect not only audition, but also other modalities and functional areas connected to auditory processing. Causal factors such as a global reduction in neural connectivity due to loss of white matter fibers (O'Sullivan, Jones, Summers, Morris, Williams, & Markus, 2001) and demyelination (Albert, 1993; Bartzokis, 2004) support this idea of a distributed change, which would be more prone to affect complex cognitive tasks, such as attention, which employ a larger amount of 'neural space' than simple ones (Kok, 2000).

3.4.4 Differences in Voluntary and Involuntary Processing

While it is difficult to decompose selective attention into facilitation and inhibition components, it has been suggested that elderly people are less successful at suppressing irrelevant or non-attended information (Kok, 2000). Selective attention filters are thought to become less acute with age; i.e. the inhibition of non-relevant information is affected.

Sommers (1997) examined the effects of varying phonetically relevant and non-relevant properties of speech stimuli, i.e. properties which would affect word recognition or properties which would be irrelevant to word recognition, for both young and elderly listeners. The task was to identify monosyllabic words presented monaurally over headphones. The stimulus properties relevant to word recognition were talker characteristics and speaking rate, which have been shown to affect word recognition when varied. The irrelevant property was overall amplitude. Sommers (1997) found that elderly adults performed consistently worse than young adults when the overall amplitude of the words was varied from trial to trial. The results suggest that the elderly listeners found it more difficult to ignore changes in the irrelevant stimulus property, overall amplitude of the speech, independent of age-related hearing loss. Sommers (1997) suggested that this could indicate a deficit in selective attention, or more precisely, an inability to identify or ignore stimulus properties which are irrelevant to the current task.

With regard to shifting attention, some studies have shown that both voluntary and involuntary shifts in attention are well preserved with age (Folk & Hoyer, 1992; Madden, 1990). Divided attention tasks are generally more complex and reliant on memory resources, so decreases in performance associated with divided attention tasks could therefore be explained by an increase in the number of processing operations required, resulting in an increase in task complexity, which can thus be attributed to the 'complexity effect' (Kok, 2000).

3.4.5 Attention-related Changes in Neural Activation

There is evidence of changes in attentional processes with age from recordings of neural activity. There are many ERP components which are thought to be associated with auditory attentional processes, both voluntary and involuntary. These include the mismatch negativity (MMN) (Näätänen, Paavilainen, Tiitinen, Jiang, & Alho, 1993) and the P300 (Picton, 1992; Soltani & Knight, 2000).

The MMN is thought to be a reflection of automatic, attention-independent, processing (Näätänen et al., 1993). It is elicited by infrequent, deviant, stimuli presented within a sequence of identical, standard, stimuli, referred to as an odd-ball paradigm. Age-related changes in the MMN have been found using a dichotic listening task (Woods, 1992). Woods (1992) used tone-bursts presented dichotically; one ear received stimuli at 1.3 kHz, and the other at 700 Hz. Participants, who included middle-aged and elderly adults, attended to a specified ear only, and responded to the presence of deviant stimuli, which were identical to the standard stimuli in the attended ear, except longer in length. The MMN was calculated by subtracting responses to standards from those to deviant stimuli. Woods (1992) found that the amplitude of the MMN was significantly reduced in the elderly adults, which he attributed to a lack of automatic processing, as suggested by previous studies (Ford & Pfefferbaum, 1991; Näätänen et al., 1993). Similar decreases in MMN amplitude have been found in elderly adults compared to young adults, also using a deviant detection task (Czigler, Csibra, & Csontos, 1992).

The P300 is also commonly elicited by odd-ball paradigms, with its amplitude varying with the relative novelty of stimuli (Picton, 1992). Two separate components of the P300 have been identified: the P3a which is thought to be related to the detection of novel events and is automatic in nature, and the P3b which has been associated with voluntary, top-down, target detection (Soltani & Knight, 2000). Age-related changes in the P300 have been observed in both visual and auditory tasks. Dujardin, Derambure, Bourriez, Jacquesson, & Guieu (1993) employed an odd-ball task to elicit a P300 using two consonant-vowel syllables (frequent 'DA', infrequent 'LI') in both young and elderly adults. In the auditory task, stimuli were presented diotically, and in the visual task the same syllables were presented on a screen. Participants responded to the infrequent stimuli. Dujardin et al. (1993) found that elderly adults showed a component with reduced amplitude and longer latency compared to the young adults. This difference was associated with a lack of automatic processing of information in the elderly, who reported having to work hard to perform the task. In contrast, the young adults reported having little difficulty with the task.

In vision, Madden, Spaniol, Whiting, Bucur, Provenzale, Cabeza, White, & Huettel (2007) conducted an fMRI study in which participants were presented with four letters, three of which were grey in colour, while one was red. The task was to identify which of two target letters (‘E’ or ‘R’) had been present in the array of letters. In the ‘guided’ condition, there was a 0.75 probability of the red letter being the target, providing a cue for top-down attentional control. In the ‘neutral’ condition, there was only a 0.25 probability of the red letter being the target. Madden et al. (2007) found positive correlations between performance and neural activation in the frontal eye field and superior parietal lobule in elderly adults, and neural activation in the fusiform gyrus in the young adults. The difference in functional areas which correlated with performance between the two groups of participants was interpreted as evidence for a significant age-related difference in top-down attentional control. Madden et al. (2007) suggested that top-down processes of attention remain intact with age, at least with regard to visual attention, and that elderly adults place more emphasis on top-down control processes compared with young controls. This again speaks for the change in processing ‘styles’ of elderly adults, relying on more top-down executive attentional control, possibly due to a deficiency in the more automatic attentional processes, as suggested by Dujardin et al. (1993).

In summary, there is evidence for an apparent increase in voluntary ‘effort’ with age for cognitive tasks, particularly those involving attention. This may, in turn, imply that a strategy of allowing one’s attention to be grabbed, i.e. a stimulus-centred approach, is no longer successful, supported by evidence for a lack of automatic, involuntary, processing in elderly people (Dujardin et al., 1993). The increased involvement of top-down control observed in elderly adults also suggests that a more voluntary, or directed, approach to shifting attention is required to maintain adequate levels of performance (Madden et al., 2007). The higher cost of voluntary attention shifts over involuntary ones in terms of processing speed has been shown (Wolfe et al., 2000), and in the light of the evidence present here, would suggest that a reliance on executive control of attention would contribute to slower, less accurate, performance amongst elderly adults.

3.5 Summary

- Models of attention largely differ in the degree to which unattended information is processed. ‘Early selection’ models theorise that unattended information is only processed at a low level to extract the basic features of stimuli. ‘Late selection' models suggest that unattended information receives more extensive processing, or that unattended information can 'leak through' a filter which usually prevents unattended information from receiving further processing.

- Attention plays an important role in the ability to listen in multi-talker environments. Focussing attention on the spatial location of a talker of interest, by knowing where they will speak, improves the ability of a listener to ignore irrelevant information arriving from other locations. Generally, smaller benefits arise from knowing who to listener for, or when to listen.

- Uncertainty about who, where, and when, often common multi-talker environments, can be detrimental to the ability of a listener to focus their attention on a talker of interest.

- Stimulus onsets can capture the focus of attention involuntarily. Highly-focussed attention improves the ability to resist distraction from the onsets of irrelevant stimuli.

- A common group of brain regions are associated with performance on a range of attentional-demanding tasks. This "multiple-demand" network of regions includes the pre-frontal cortex, the parietal lobe, and the anterior cingulate cortex.

- Attention to speech in the presence of noise or other speech is associated with the activation of a wide range of cortical regions including primary and association auditory cortices and those regions associated more generally with attention-demanding tasks.

- The oscillatory activity of neurons within the cortex has been linked to communication between brain regions, the current attentional state, and may provide an insight into the manner in which information is selected or attended to.

- Age-related changes in cognitive function include a reduction in the involuntary processing of information and an increase in amount of voluntary effort required to perform complex tasks. These changes may influence the attentional strategies that older adults use to cope with attention-demanding tasks, independent of the sensory modality of the stimuli.

- The reliance of older adults on the volitional control of attention has been associated generally with slower, less accurate, performance compared to younger adults.

3.6 Conclusions from the Literature Review

The premise of this review is that elderly adults experience difficulties with the perception of speech in environments which involve multiple talkers speaking concurrently. These difficulties are less apparent amongst young adults, to whom such listening environments pose few problems. This review has discussed evidence from ageing, spatial listening, speech perception, and attention research which has implications for understanding the difficulties that elderly people experience in multi-talker listening environments.

I have examined the importance of central auditory processes in segregating speech streams when presented to one ear, both ears, or when multiple speech streams are perceived to be distributed in space. In light of this, I have examined the evidence that attention plays an important role in multi-talker listening tasks, enabling listeners to focus attention on a single sound source location while suppressing information from other locations, and to switch attention between different sound source locations. The cognitive nature of these attentional mechanisms has been discussed, along with evidence that processes of attention are subject to age-related changes. Finally, various factors that can influence attention, both in terms of allocation of resources and the selection of successful attentional strategies, have been presented. These include abrupt onsets and offsets, and a decline in automatic processing in elderly adults, leading to an increased reliance on the slower processes of volitional control.

The difficulties that elderly listeners experience when understanding speech in complex listening environments have an appreciable impact on those listeners' own perceived disability in coping with situations which are common in everyday life. It is therefore relevant to conduct research which decomposes the nature of those difficulties, by relating performance in listening situations which reflect the complexity of real-world environments to possible changes in the neural mechanisms underlying the cognitive processes of speech perception and attention.

Chapter 4

Reconstructing Cortical Activity with High Temporal Resolution

This chapter presents an overview of methods used to localise the cortical sources of activity which give rise to the magnetic fields measured with MEG. The chapter discusses the minimum-norm estimation and spatial filtering techniques. The fundamental concepts of each technique are outlined, along with issues related to the analysis of MEG data specific to each technique. The chapter compares the strengths and limitations of the techniques, and discusses the application of the two techniques to the analysis of neural activity associated with performance on a spatial listening task for speech.

4.1 Introduction

The ultimate goal of any neuroimaging technique is to measure and subsequently reconstruct the activity of neural populations within either the cerebral cortex, the mid-brain, or the brainstem. Magneto-encephalography (MEG) is one such approach. It offers excellent temporal resolution, down to the millisecond level, and the potential for resolving sources separated by only a few millimetres. MEG measures the extra-cranial magnetic fields generated by intracellular current flow in the dendritic trunks of pyramidal neurons (Okada, 2003). These fields are extremely weak compared to the earth's own magnetic field (Figure 4.1), and are detectable only by using highly-sensitive sensors. In MEG, the sensors are super-conducting quantum interface devices (SQUIDs) which operate at temperatures close to absolute zero (4.2° K, −268.95° C) through immersion in liquid helium inside a cryogenic dewar. The sensitivity of the SQUIDs to environmental noise means that the MEG dewar is typically placed within a shielded room (Figure 4.2). The creation of a detectable signal requires a large number of neurons with similar orientations to be active simultaneously. It is estimated that at least a million active synapses, which could correspond to as little as 1 mm^3 of the cortex, are required to produce a magnetic field measurable with MEG (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). However, due to the fact that there may be neighbouring areas of current flow which are oriented in the opposite direction, as much as 40 mm^3 could be necessary (Chapman, Ilmoniemi, Barbanera, & Romani, 1984, referenced in Hämäläinen et al. (1993)).

There are a variety of approaches to estimating the activity of sources within the head which give rise to the magnetic fields observed with MEG. These approaches are referred to as *inverse solutions*—a term which refers to the inversion of the process used to measure the magnetic fields by estimating the sources underlying the field patterns. A widely used approach is *minimum-norm estimation* in which source activity is modelled in terms of a finite number of current dipoles whose locations are known, and belongs to a family of techniques called distributed dipole models. Dipolar sources are commonly used in MEG to represent localised current flow in the grey matter of the cortex.

Other methods assume that the activity of neural sources can be estimated using a finite number of modelled sources whose amplitudes and orientations are not known a priori. *Spatial filtering* is such a method, previously used in radar and auditory applications, which attempts to focus the entire MEG sensor array to estimate the extent of source activity at individual locations in the brain. A map of distributed activity can be determined by repeating the procedure over a grid of points, and is referred to as a *source scanning* technique. With this method, no assumptions about the number or location of sources are required.

Although these two *source analysis* methods attempt to address the same question, i.e. what was the nature of the distributed source activity in the cortex which

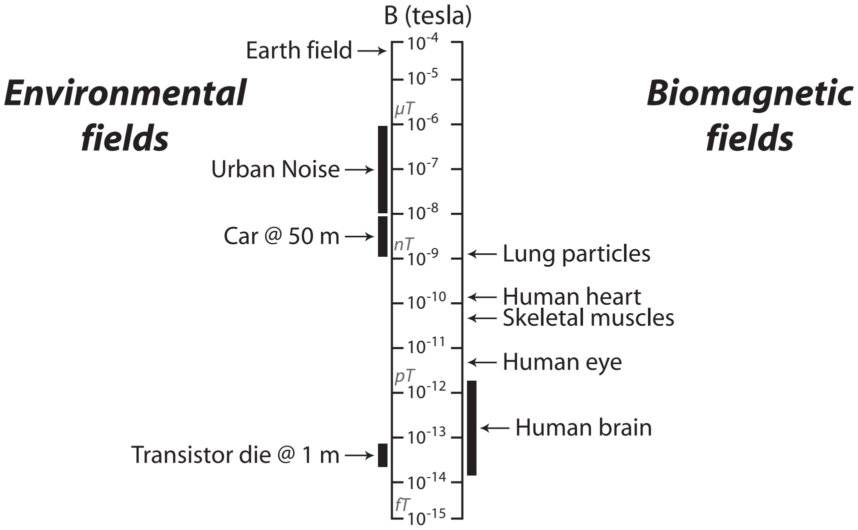

Figure 4.1. Comparison of environmental and biological magnetic field strengths on a log scale. The fields created by neural activity ("Human brain") are 7 orders of magnitude weaker than the earth's field (adapted from Vrba, 2002).

produced the measured magnetic fields, they vary considerably in their a priori assumptions, spatial resolution, localisation accuracy, and their performance under different conditions of use. A consequence is that the choice of analysis approach can greatly affect the source estimates obtained from MEG data and therefore must be considered carefully. This chapter will begin with an explanation of the *inverse problem* associated with MEG source analysis. Following that, the techniques of minimum-norm estimation and linearly constrained minimum variance (LCMV) spatial filtering will be explained in terms of their fundamental assumptions and mathematical formulations. The advantages and disadvantages of both approaches will be outlined and compared in the context of the current research.

4.1.1 Notation Style and Conventions

In this chapter, vectors and matrices will be employed to describe the mathematical bases of different approaches to solving the MEG inverse problem. Both are defined in Appendix A, along with their respective notations and terminology.

4.2 The Inverse Problem

The inverse problem may be described as follows: how can the location and activity of sources, which are assumed to be neurons within the cerebral cortex and the principle generators of the measured magnetic fields, be estimated from the data recorded by an array of sensors located around the head? Helmholtz (1853) showed that knowledge of the electromagnetic fields outside a spherical conductor is insufficient

Figure 4.2. An MEG dewar (left) positioned within a magnetically-shielded room (right).

to determine the distribution of sources within the conductor which gave rise to the fields. To relate that to the case of MEG, knowledge of the time-varying magnetic fields outside the human head is not sufficient to establish the exact pattern of neural activity which created the field patterns. The problem is said to be *ill-posed* because it does not have a single unique solution.

The non-uniqueness of the inverse problem can be illustrated in terms of source activity which we cannot measure. In the case of MEG, regions of neural activity are commonly modelled as current dipole sources within a spherically-symmetric conductor (Baillet, Mosher, & Leahy, 2001). If such a source is oriented radially with respect to the surface of the conductor, it lies along a line which passes through the centre of the conductor. This *radially-oriented* source does not produce a magnetic field outside the conductor (Sarvas, 1987). Only sources with a tangential component produce a measurable magnetic field outside spherically-symmetric conductors (Hämäläinen et al., 1993; Mosher, Leahy, & Lewis, 1999). This issue is relevant because spherical models of the head are often used in the analysis and interpretation of MEG data (Baillet et al., 2001). One can then see that to any estimate of the source activity based on the measured magnetic fields, an arbitrary number of radial sources could be added without affecting the validity of the estimate. In this way it is apparent that no unique solution to the distribution of current sources within a conductor can be deduced from the measured magnetic fields alone.

Solutions to the inverse problem can be found by imposing restrictions and/or a priori assumptions on the system. It is these underlying constraints which distinguish different source analysis techniques in terms of their spatial and temporal resolution, suitability to the analysis of focal or distributed activation patterns, and other factors which are discussed in this chapter. However, regardless of the approach that is adopted, the inverse problem in MEG remains highly *underdetermined*; i.e. the number of measurement sensors (typically less than 300) is much less than the number of sources (there are approximately 10^{10} cortical neurons with as many as 10^{14} synapses in the cerebral cortex (Hämäläinen et al., 1993)). The consequence is that the activity of all sources cannot be estimated independently of each other. Thus, the limiting factor in the reconstruction of cortical activity using MEG is ultimately the sensor array which is used to record the magnetic fields, in terms of its sensitivity, noise level, and density.

4.3 The Forward Solution

To solve the inverse problem we must first define the relationship between the activity of neurons within the cortex and the magnetic fields measured using a sensor array outside the head, also referred to as the *forward problem*. Consider a single neuron modelled as a single dipolar source within a spherically-symmetric conductor. Activation of the source results in a primary ionic current flow within the neuron itself and a secondary, or *volume*, current flow within the surrounding fluid which serves to avoid a build up of charge within the cell (Figure 4.3). Both currents produce a magnetic field *inside* the conductor. In a spherically-symmetric conductor, the net magnetic field due to the volume currents is zero when measured by a sensor oriented radially with respect to the conductor at a distance (Mosher et al., 1999; Wang & Kaufman, 2003). Therefore, the magnetic fields observed by a radially oriented sensor outside the conductor are a result of primary currents rather than volume currents.

If we know the location of the dipolar source, $\vec{r}_Q$, then its primary current flow, $J^P(\vec{r}_Q)$, within a spherically-symmetric conductor, V, can be related to the magnetic field, $\vec{B}(\vec{r})$, that would be measured by a sensor at location $\vec{r}$ outside the conductor according to the Biot-Savart law (Sarvas, 1987; Mosher et al., 1999):

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi}\int_{\mathbf{V}} J^P(\vec{r}_Q) \times \frac{\vec{r}-\vec{r}_Q}{\left|\vec{r}-\vec{r}_Q\right|^3}\mathrm{d}\vec{r}_Q \tag{4.3.1}$$

where μ_0 is the permeability of free space and **V** is the region defined by the conductor, V. Sarvas (1987) showed that even if the sensor is not oriented radially with respect to the surface of the conductor, the total magnetic field due to a source can still be calculated without specific reference to the volume currents. If the source is a current dipole with moment $\vec{Q}$, a vector which describes the orientation and strength of its activity, and is at the location $\vec{r}_Q$, then the magnetic field can be computed as (Sarvas, 1987):

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi}\frac{F\vec{Q}\times\vec{r}_Q - (\vec{Q}\times\vec{r}_Q\cdot\vec{r})\nabla F(\vec{r},\vec{r}_Q)}{F(\vec{r},\vec{r}_Q)^2} \tag{4.3.2}$$

where

$$F(\vec{r},\vec{r}_Q) = |\vec{a}|\,(|\vec{r}|\,|\vec{a}| + |\vec{r}|^2 - \vec{r}_Q\cdot\vec{r})$$

and

$$\begin{aligned}\nabla F(\vec{r},\vec{r}_Q) = {} & (|\vec{r}|^{-1}\,|\vec{a}|^2 + |\vec{a}|^{-1}\,a\cdot\vec{r} + 2\,|\vec{a}| + 2\,|\vec{r}|)\vec{r} \\ & - (|\vec{a}| + 2\,|\vec{r}| + |\vec{a}|^{-1}\,\vec{a}\cdot\vec{r})\vec{r}_Q\end{aligned}$$

with

$$\vec{a} = (\vec{r}-\vec{r}_Q)$$

Using Eq. 4.3.2, the relationship between current flow at the i^{th} source location and the magnetic field at the j^{th} sensor can be expressed as a vector of three values which describe the magnetic field due to three orthogonal unit dipoles at that source location; i.e. oriented in the x, y, and z directions:

$$\vec{\ell}_{i,j} = \begin{bmatrix} \ell_x & \ell_y & \ell_z \end{bmatrix} \tag{4.3.3}$$

The magnetic field, $\vec{B}_{i,j}$, due to source activity at the i^{th} source location and measured by the j^{th} sensor can then be calculated by multiplying each component of the vector in Eq. 4.3.3 by the strengths of the three orthogonal sources, expressed as a vector $\vec{s}$, and then combining them:

$$\vec{B}_{i,j} = \sum_{k}^{x,y,z} \vec{\ell}_{i,j}(k)\vec{s}_i(k) \tag{4.3.4}$$

For a given source location, the magnetic field vector (Eq. 4.3.3) can be calculated for each of the sensors in the array and, taken together, is referred to as the *forward field* for that source location. For multiple source locations and sensors, the forward fields can be arranged in matrix form so that the values relating to each sensor are contained in a single row, and the three values for each source location are arranged in adjacent columns. Therefore, with m sensors and n sources, we have an $m \times (n \times 3)$ matrix:

$$L = \left\{ \begin{array}{ccccccc} \ell_{1,1,x} & \ell_{1,1,y} & \ell_{1,1,z} & \cdots & \ell_{1,n,x} & \ell_{1,n,y} & \ell_{1,n,z} \\ \ell_{2,1,x} & \ell_{2,1,y} & \ell_{2,1,z} & \cdots & \ell_{2,n,x} & \ell_{2,n,y} & \ell_{2,n,z} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \ell_{m,1,x} & \ell_{m,1,y} & \ell_{m,1,z} & \cdots & \ell_{m,n,x} & \ell_{m,n,y} & \ell_{m,n,z} \end{array} \right\} \tag{4.3.5}$$

This matrix is commonly referred to as the *forward solution*, the *lead fields*, or

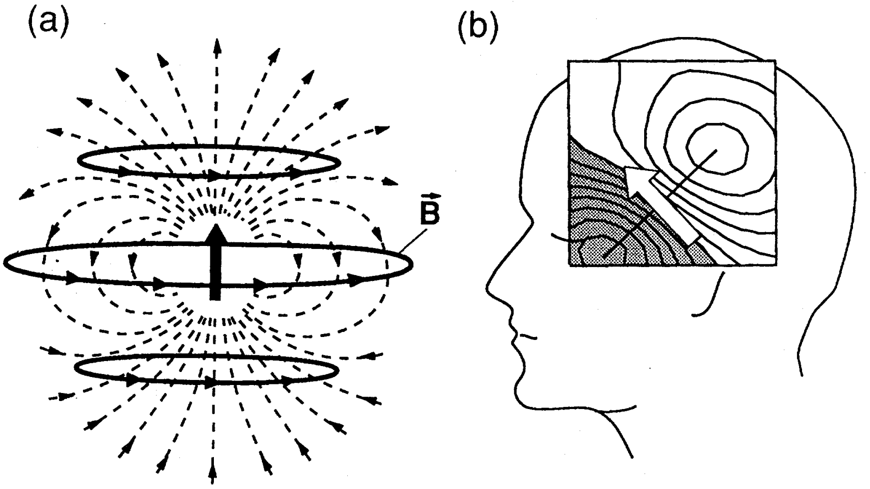

Figure 4.3. The magnetic field generated by a current dipole. (a) depicts the current dipole (arrow), the volume currents (dashed lines), and the resultant magnetic field $\vec{B}$. (b) shows an example of a magnetic field pattern measured outside the head due to a dipolar source (arrow) located beneath the mid-point of the line joining the extrema of the field (after Hämäläinen et al., 1993).

the *gain matrix*. Following common terminology, each row is referred to as the *lead field* for a particular sensor, which relates the resultant field at the sensor to the current flow across all the source locations, and each column is the *forward field* for a particular source location, which relates the current flow of a single source location to the magnetic field pattern across all the sensors (Ermer, Mosher, Baillet, & Leahy, 2001; Mattout, Phillips, Penny, Rugg, & Friston, 2006).

In practice, the calculation of the *lead fields* incorporates information about the model which approximates the conductor in which the source is located; in this instance, the human head. Two methods of approximating the human head include a single sphere which approximates the skull (*single-sphere model*) (Cuffin & Cohen, 1977) or multiple spheres, each of which best approximates the curvature of the skull under a given sensor (*multiple overlapping spheres model*) (Huang, Mosher, & Leahy, 1999) (Figure 4.4).

Using the matrix of lead fields, L, the relationship between the strength of activity across a set of source locations and the magnetic field at the j^{th} sensor can be expressed by expanding Eq. 4.3.4 to include n source locations:

$$\vec{B}_j = \sum_{i=1}^{n} \sum_{k}^{x,y,z} \vec{\ell}_{i,j}(k)\vec{s}_i(k) \tag{4.3.6}$$

Therefore, Eq. 4.3.6 is a linear model of how the strengths of activation of a discrete number of sources, $\vec{s}$, give rise to the measured magnetic fields, $\vec{B}$, at a given sensor. It is referred to as the *forward problem* of MEG. In matrix notation the relationship across multiple sensors can be written as:

$$d = Ls \tag{4.3.7}$$

where d is an $m \times 1$ vector containing the measured magnetic fields across the m sensors at a single time point, L is the $m \times (n \times 3)$ forward solution matrix which relates the 3 orthogonal components of n sources to the m measurements, and s is the

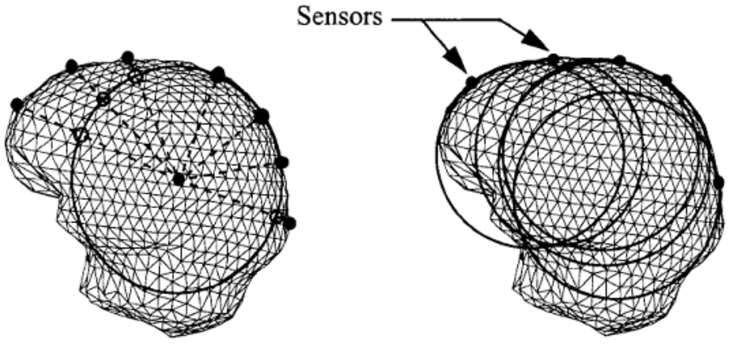

Figure 4.4. Spherical head models using a single best-fit sphere (left) and a 'sensor-weighted' or multiple local sphere model (right) (after Ermer et al., 2001).

$(n \times 3) \times 1$ vector of source strengths or dipole moments. Therefore, L is a *transposition* or *translation* operator between the sensor and source spaces. The formulation of the forward problem in Eq. 4.3.7 is an example of a system of linear equations. These systems are discussed in more detail in Appendix A.

If we assume that there is some additive noise across the sensors, then we have

$$d = Ls + \varepsilon \tag{4.3.8}$$

where ε is an $m \times 1$ vector of additive noise across the m sensors. In practice, ε can be estimated by measuring the magnetic field at the MEG sensors in the absence of a human participant.

As d is known (the measured MEG data), L can be calculated, and ε can be estimated, Eqs. 4.3.7 and 4.3.8 have one unknown: the source strengths, s. The aim of source analysis techniques in MEG is therefore to invert the *forward problem* to create an expression for the source strengths, s, based on the forward solution, L, and the measured data, d. Matrix notation will be used for the remainder of this chapter to describe methods for estimating s, the unknown column vector of source strengths.

4.4 Minimum Norm

4.4.1 Introduction

If we presume that the magnetic signals recorded using MEG can be described completely by a finite number of dipolar sources whose locations and orientations are known a priori, then the remaining parameters to be estimated are the source strengths, s. As the relationship between the source strengths and the measured fields is linear (Sarvas, 1987), the problem to be solved can be stated in terms of the system of linear equations in Eq. 4.3.7. Therefore, linear estimation techniques can be used to solve for the values of the source strengths. One such technique is based on least-squares estimation, or minimising the error between modelled and observed data, and is referred to as *minimum-norm estimation.* The assumptions and constraints under which minimum-norm estimates are computed will be described, and the advantages and also the limitations of such a linear approach will be discussed.

4.4.2 Finding an Inverse

We start by restating the forward problem from Eq. 4.3.7 which determines the relationship between the source strengths which we want to estimate, s, and our measured data at a single time point, d, using the lead fields, L:

$$d = Ls \tag{4.4.1}$$

The simplest approach to solving the inverse problem is to find a solution, s, which explains the measured data completely. That is, we will not distinguish between the model of the source activity and the noise within the sensor measurements. Our goal is then to invert Eq. 4.4.1 by finding an inverse operator, represented by a matrix W, which relates the source strengths to the measured data:

$$s = Wd \tag{4.4.2}$$

As our forward solution is of $m \times (n \times 3)$ dimensions, the inverse operator W would be an $(n \times 3) \times m$ matrix. The simplest solution for the source strengths, s, that can be derived from the forward problem in Eq. 4.4.1 is:

$$s = L^{-1}d \tag{4.4.3}$$

which is the obtained by simply multiplying both sides of Eq. 4.4.1 by the inverse of L; i.e. L^{-1}. However, L^{-1} only exists if L is a square matrix (the number of rows and columns are equal) and is of full rank; i.e. either all the columns or all the rows are linearly independent of one another (Matrix inverses are discussed in detail in Appendix A, Section A.2.7, p. 251). The number of observations in MEG is less than the number of sources; i.e. $m < n$, and L is not square. Therefore, the true matrix inverse, L^{-1}, does not exist, and Eq. 4.4.3 has no solution for MEG.

One might be tempted to alter the number of sources in our model to make L a square matrix. This would artificially constrain the problem, as we know that for MEG our estimate of the number of sources, n, is greater than the number of measurements, m. Neither would such an approach necessarily lead to a unique solution, as our measurements might not be completely independent, which would make the matrix L singular or near singular, and therefore non-invertible. In other words, L would be a square matrix but L^{-1} would not exist (Wang & Kaufman, 2003).

4.4.3 Moore-Penrose Pseudoinverse: The Solution of Minimum Norm

Instead of artificially constraining the number of sources in our model to make our forward solution, L, square and invertible, we can use the *generalised inverse*, L^{+}. Unlike the true matrix inverse, L^{-1}, the generalised inverse can be calculated for non-square matrices such as our forward solution. The expression for the solution, s, would then become:

$$s = L^{+}d \tag{4.4.4}$$

This inverse operator, L^{+}, is also termed the *pseudoinverse* or *Moore-Penrose Pseudoinverse* (Moore, 1920; Penrose, 1955). As discussed previously, the *inverse* *problem* is ill-posed, and a unique solution does not exist. A unique solution can only be found by imposing constraints on the linear system in Eq. 4.4.1. The Moore-Penrose inverse, L^+, has been shown to provide a least squares solution to a system of linear equations (Penrose, 1956). Thus, the solution, s, given by Eq. 4.4.4 is that solution which minimises the difference between the modelled data and the observed data, or the average squared error in prediction (Hays, 1974):

$$\min_s \left\{\|Ls - d\|^2\right\} \tag{4.4.5}$$

where $\|\cdot\|$ indicates an ℓ_2 norm and the notation $\min_y\{f(y)\}$ refers to the minimisation of the expression $f(y)$ with respect to y. In the case of MEG, it is not guaranteed that the least squares constraint will lead to a unique solution, s; i.e. there may be an infinite number of solutions which satisfy Eq. 4.4.5. If our measurements, and therefore the rows of the forward solution, L, are not independent, an infinite number of solutions which satisfy Eq. 4.4.5 will exist. However, the solution derived using the Moore-Penrose Pseudoinverse is unique and is the solution which has the smallest Euclidean or ℓ_2 norm, i.e. the square root of the sum of the squared components or the power of the solution, of all possible solutions which satisfy Eq. 4.4.5 (Mosher, Baillet, & Leahy, 2003; Sarvas, 1987). For that reason, it is referred to as the solution of minimum norm (Hämäläinen et al., 1993) or as the *minimum norm least squares* (MNLS) solution to an underdetermined system of linear equations (Wang & Kaufman, 2003).

In summary, there are two constraints which we impose on our system of linear equations (Eq. 4.4.1) to arrive at the minimum norm solution: the fit of the modelled data to the measured data and the minimisation of solution norm; i.e.:

$$\min_s \left\{\|Ls - d\|^2 + \|s\|^2\right\} \tag{4.4.6}$$

If we assume that the measurements are independent of each other, then the *lead fields* or the rows of the lead field matrix, L, are also independent. Under that assumption and when $m < n$, the pseudoinverse can be calculated as (Golub & Pereyra, 1973; Barnett, 1990) (see Appendix B.1, p. 254):

$$L^+ = L^T(LL^T)^{-1} \tag{4.4.7}$$

where L^T is the transpose of L; when its rows are replaced by its columns and *vice versa*. Substituting Eq. 4.4.7 into Eq. 4.4.4 yields an expression for the *classical* minimum norm solution for MEG:

$$s = L^T(LL^T)^{-1}d \tag{4.4.8}$$

4.4.4 Source Weighting

It is often desirable to incorporate a priori information about the source model into the solution. This can be achieved by modifying the constraints on the solution in Eq. 4.4.6 as follows:

$$\min_s \left\{ \|Ls - d\|^2 + (s^T C_s^{-1} s) \right\} \tag{4.4.9}$$

where C_S is a source weighting matrix. The matrix can be used to incorporate information about which sources are expected to be active from fMRI evidence (Dale & Sereno, 1993), to incorporate depth-weighting (Fuchs, Wagner, Kohler, & Wischmann, 1999), as part of iterative focusing methods (Gorodnitsky, George, & Rao, 1995; Liu, Schimpf, Dong, Gao, Yang, & Gao, 2005), or to specify estimates of covariance between sources (Mattout et al., 2006). Some of these techniques are discussed later in this chapter. Minimising this expression leads to the *weighted* minimum norm solution (Barnett, 1990; Iwaki & Ueno, 1998; Tarantola, 2004):

$$s = C_S L^T (L C_S L^T)^{-1} d \tag{4.4.10}$$

If no a priori assumptions are made about the relationship between the sources such that they are independent of each other and of equal variance then $C_S = I$, the identity matrix. In that case, the weighted solution reduces to the *classical* minimum norm solution in Eq. 4.4.8.

4.4.5 Regularisation

From Eq. 4.4.10 we can see that the calculation of the source strengths involves the inversion of the term $LC_S L^T$. Due to the highly underdetermined nature of the MEG inverse problem, we have too few measurements to find a unique set of source strengths which could explain the measured data. This feature of the inverse problem has the consequence that $LC_S L^T$ is *ill-conditioned*. Therefore, small changes in the measured data d can lead to large changes in the solution, s; i.e. the solution is numerically unstable (Golub & Van Loan, 1996) (see Appendix A, Section A.3, p. 252). This problem is caused by the fact that the estimated solution fits the specified data well but not data which is close to the specified data (Figure 4.5). Thus, small changes to the MEG data such as might be caused by noise in the sensor measurements could lead to very different solutions for the source strengths. To improve the stability of the estimated source strengths, *regularisation* is required. Regularisation can be thought of as the process of varying an a priori parameter which specifies the expected degree of smoothness of the estimated solution (Hansen, 1992). The result of this process is that changes in the estimated source strengths based on the time-varying MEG data are smooth and not highly dependent on the exact configuration of the data and the noise in the measurements.

A common approach to the regularisation of ill-posed linear problems is Tikhonov regularisation (Tikhonov & Arsenin, 1977), also referred to as damped least squares. This approach can be explained by considering the following problem:

$$\min_{s}\left\{\|Ls-d\|^{2}+\lambda\left(s^{T}C_{s}^{-1}s\right)\right\} \tag{4.4.11}$$

which is identical to the weighted minimum norm constraints (Eq. 4.4.9) with the inclusion of λ: the *regularisation parameter*. From the above formulation of the problem to be solved, it can be seen that the regularisation parameter balances the minimisation of the two terms: the fit of the model to the measured data and the minimisation of solution power. Increasing the parameter results in a spatially smoother solution with lower spatial resolution and increases the modelling error; i.e. the difference between the modelled and measured data (Phillips, Rugg, & Friston, 2002). When the problem in Eq. 4.4.11 is minimised, the solution is the *weighted*

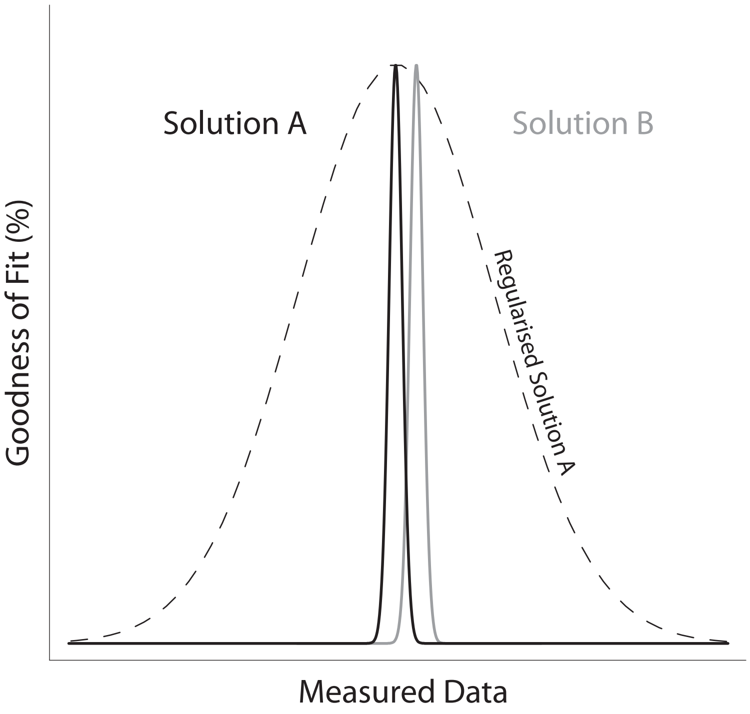

Figure 4.5. A schematic representation of the regularisation problem in MEG. The x axis represents a continuum of possible configurations of the recorded MEG data with data sets differentiated by small numerical changes being adjacent. For a certain set of measured data, solution A provides an excellent fit. However, if a small numerical change in the data is introduced, the goodness of fit for that solution decreases rapidly and a new solution, B, is required to explain the data. Thus, small numerical changes in the input data can lead to large changes in the best-fit solution of source strengths. Regularisation results in a spatially smoother estimate of the source strengths which is less numerically sensitive to small changes in the MEG data; i.e. it is a more stable solution.

minimum ℓ_2-norm estimate with Tikhonov regularisation:

$$s = C_s L^T (L C_s L^T + \lambda I)^{-1} d \tag{4.4.12}$$

where I is the identity matrix with dimensions equal to the number of sensors. The regularisation parameter, λ, can be estimated using several methods. These include generalised cross-validation (GCV) (Golub, Heath, & Wahba, 1979) and L-curve estimation (Hansen, 1992). For GCV, a value for the regularisation parameter is chosen which provides accurate estimates of missing data values. In the case of the L-curve, if the error between the model and the measured data is plotted against the norm of the solution, which are the two constraints to be minimised in Eq. 4.4.6, for different values of λ, an L-shaped curve is observed (Figure 4.6). The value of λ at the 'corner' of the curve represents a near optimal parameter choice, as both constraints are balanced simultaneously (Hansen, 1992).

Alternative methods to choose a suitable parameter can be based directly on the matrix which is to be inverted, LC_sL^T, in the form of its singular values. The smallest singular values, obtained through singular value decomposition (SVD) (Appendix A, Section A.2.4, p. 249), are often near zero when the problem is underdetermined and can have a large influence on the estimated solution. This can be seen by reformulating the computation of the Moore-Penrose Pseudoinverse in Eq. 4.4.7 in

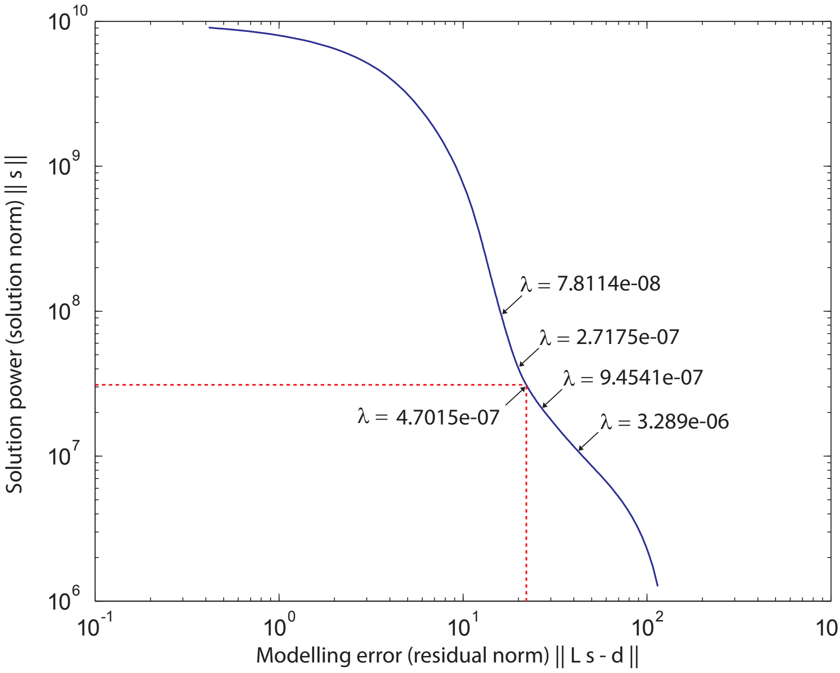

Figure 4.6. An example of an estimated L-curve. The norm of the solution and the modelling error (see Eq. 4.4.6) are plotted at various values of the regularisation parameter for Tikhonov regularisation (solid line). An optimum value is at the 'corner' of the curve where both of the constraints are balanced simultaneously.

terms of the decomposed parts of the matrix (Barnett, 1990):

$$LC_sL^T = U\Sigma V^T \qquad \text{(Singular Value Decomposition)}$$
$$\left(LC_sL^T\right)^+ = V\Sigma^+U^T \qquad \text{(Moore-Penrose Pseudoinverse)}$$

where Σ is a diagonal matrix containing the singular values, and Σ^+ is derived by taking the reciprocal of the singular values when they are non-zero. From this, it can be seen that singular values in Σ which are very close to zero become very large in Σ^+ and it is desirable to suppress such values (Press, Teukolsky, Vetterling, & Flannery, 2002). A tolerance factor is commonly employed to determine which values are non-zero and all other values are then set to zero. This cut-off point can be manipulated to remove a desired quantity of the weakest components. This process has the effect of regularising the solution, a technique referred to as truncated SVD or TSVD. Tikhonov regularisation has been shown to be equivalent to a smoothed truncation of these singular values (Hansen, 1992) but does not require the decomposition of the large matrix which can be computationally expensive.

A method of estimating the regularisation parameter for Tikhonov regularisation which avoids matrix decomposition is to use a small percentage (usually < 1%) of the trace of the matrix to be inverted, trace(LC_sL^T), to estimate the smallest singular values. This method can be used to estimate the smaller singular values as the trace of a square matrix is equivalent to the sum of its singular values. This value can then be used as the regularisation parameter (Figure 4.7). An alternative method has been suggested by Press et al. (2002) which gives equal weight to both of the constraints that the regularisation parameter is balancing:

$$\lambda = \frac{\text{trace}(LC_sL^T)}{\text{trace}(C_n)} \tag{4.4.13}$$

where trace() is the matrix trace operation and C_n is the estimated noise covariance across the sensor array. Lin, Witzel, Hämäläinen, Dale, Belliveau, & Stufflebeam (2004) proposed an extension of this method by deriving λ from an estimate of the signal-to-noise ratio (SNR) of the measured data, specific in decibels, as follows:

$$\lambda = \frac{\text{trace}(LC_sL^T)}{\text{trace}(C_n) * \text{SNR}^2} \tag{4.4.14}$$

In this way, the amount of regularisation increases as the SNR decreases, which is desirable (Phillips et al., 2002). For their study, Lin et al. (2004) used a fixed value of 5 for the SNR. In practice, the value could be estimated from recorded data by averaging the data based on the onset of a stimulus and taking the ratio between the variance pre- and post-stimulus as follows (Leonowicz, Karvanen, & Shishkin, 2005):

$$\mathrm{SNR} = 10\,\log_{10}\frac{\sigma^2_{poststim}}{\sigma^2_{prestim}}\,[\mathrm{dB}] \tag{4.4.15}$$

4.4.6 Noise Estimates

Up to this point, the solution to the *inverse problem* has been derived such that it explains the measured data as closely as possible. As the observed data are contaminated with some degree of noise in practical situations, it is usually more desirable to have the solution only explain that part of the measured data which relates to the signals of interest; i.e. the distributed cortical dipoles and not the sensor noise. We start from our original forward problem with the inclusion of some additive noise, ε:

$$d = Ls + \varepsilon \tag{4.4.16}$$

We then modify our least squares criteria of minimising the difference between the modelled and observed data so that instead of attempting to match all of the measured data as best as possible, we only explain part of it. Hauk (2004) expresses

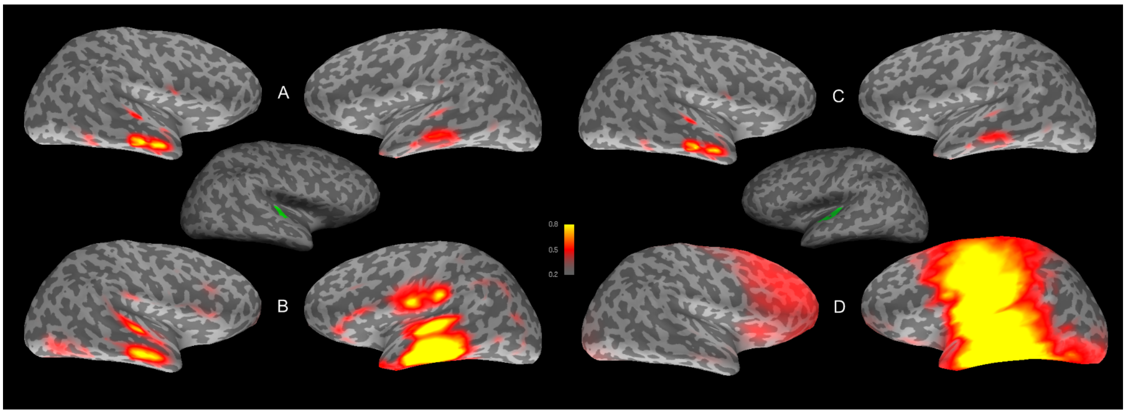

Figure 4.7. Effect of the choice of regularisation parameter on the minimum norm estimate of an auditory response 100 ms after the onset of a spoken phrase presented acoustically. Data are from a single participant and averaged over 90 presentations. The locations of primary auditory areas on Heschl's Gyrus are shown in the middle row marked in green. The magnitude of the solutions has been normalised to facilitate comparison. (A) and (B): regularisation based on a percentage of trace(LCL^T) (0.01% and 0.75% respectively); (C): regularisation parameter estimated from the L-curve; (D): artificially high regularisation parameter chosen manually. The amount of regularisation is too small in solution (A) as the peak activity is not close to the expected region of primary auditory areas. The regularisation in (B) is preferred compared to (A) as the data now show some focal peaks of activation in the expected regions. The parameter estimated from the L-curve (C) has produced an estimate similar to (A). The parameter used to compute (D) has resulted in an over-regularised solution, leading to a distributed and over-smoothed estimate of the source activity.

this as constraint as a modified version of the least squares constraint in Eq. 4.4.5:

$$(Ls - d)^T C_n^{-1} (Ls - d) = \varepsilon > 0 \tag{4.4.17}$$

where C_n is a weighting matrix representing the “reliability” of the sensors (Hauk, 2004) or the covariance of noise across the sensors (Phillips et al., 2002), and ε is the part of the data which is not accounted for by the solution. The expression for the solution in Eq. 4.4.10 is then expanded to become the regularised weighted minimum ℓ_2 norm estimate of the source distribution, s, with noise priors:

$$s = C_s L^T (L C_s L^T + \lambda C_n)^{-1} d \tag{4.4.18}$$

If $C_N = I$, all sensors are weighted equally and the solution reduces to the *weighted* minimum norm solution with Tikhonov regularisation in Eq. 4.4.12.

The formulation of the *minimum norm* solution in Eq. 4.4.18 has been widely used to estimate source activity in both MEG and EEG (Dale, Liu, Fischl, Buckner, Belliveau, Lewine, & Halgren, 2000; Dale & Sereno, 1993; Fuchs et al., 1999; Hauk, 2004; Lin, Belliveau, Dale, & Hämäläinen, 2003) as it provides the possibility of constraining the solution using estimates of sensor noise, a priori information about regions of interest identified in other modalities such as fMRI, estimates of source covariance, and for the correction of a bias towards more superficial sources. Bias correction is discussed later in this chapter.

4.4.7 Anatomical Constraints

The forward solution, or *lead field* matrix, is central to the calculation of the minimum norm estimates. The lead fields rely on information about the locations and orientations of a discrete number of sources; i.e. the source model. As the minimum norm solution does not provide information about the depth of the sources (Fuchs et al., 1999), that information must be provided as part of the source model.

The magnetic fields that are observed with MEG are thought to be generated by synaptic activity in the pyramidal neurons of the cerebral cortex (ignoring ‘external’ sources of noise such as cardiac and ocular artifacts) (Hämäläinen et al., 1993). A sensible set of source locations for a model would therefore be the cortical surface. In practice, calculations are made at each of a discrete set of points evenly spaced over the surface of the cortex. Typically, these points are separated by 5–10 mm. The exact surface used for this ‘shell’ based approach varies between the white-grey matter boundary, the grey-cerebro-spinal fluid (pial) boundary, or intermediate surfaces. These surfaces can be simple approximations of the underlying anatomy or based on detailed anatomical models extracted from high-resolution MRI scans (Dale & Sereno, 1993). As *minimum norm* estimates do not provide depth information, the result of a shell-based approach is a two-dimensional projection of the true 3-dimensional distribution of sources onto a surface (Hauk, 2004).

The accuracy of this projection in terms of source localisation depends on the error between the actual source distribution and the surface that is modelled. For example, if a simple approximation to the surface of the cortex, extracted from an average brain, is used for the analysis then the solution is capable only of providing information about which gross regions of the cortex are active. In contrast, if a model of an individual subject's cortical anatomy is used to reconstruct the MEG data of that participant, peaks of activity can be localised to identifiable gyrii and sulci (Lin, Witzel, Ahlfors, Stufflebeam, Belliveau, & Hämäläinen, 2006b). However, the localisation accuracy of the minimum norm approach is fundamentally limited by the underdetermined nature of the inverse problem in MEG and, like other linear estimation techniques, it tends to produce estimates which are spatially smeared (Dale et al., 2000).

Aside from their locations, the orientations of sources are also required to compute the lead fields. The choice of orientation is important as it can affect the accuracy of the minimum norm solution (Lin, Belliveau, Dale, & Hämäläinen, 2006a). From cortical anatomy, it has been shown that the orientation of current flow due to synaptic activity in pyramidal neurons is largely perpendicular to the cortical surface (Okada, Wu, & Kyuhou, 1997). Therefore, estimates of source orientations, like their locations, can be derived from cortical models extracted from MRI scans. The orientations can be estimated by calculating the direction normal to the surface at each of the source locations. However, small movements away from most cortical locations can be accompanied by large changes in the normal direction due to the corrugated morphology of the cortex. Therefore, a small modelling error in the location of a source, such as might be introduced through MEG/MRI co-registration errors, could be accompanied by a large modelling error in the source orientation.

A more robust approach, termed 'cortical patch statistics' (Lin et al., 2003), involves estimating the normal direction from an area around each source location thereby making the orientation estimation more robust to localisation errors. However, a more straightforward approach is often adopted due to limitations on processing capacity and/or time in which the lead fields are computed for three orthogonal components at each cortical location. The total power of source activity at a particular location can then be calculated by computing the vector sum of the three orthogonal components; i.e. the sum of the squared magnitude of each component.

4.4.8 Depth Weighting

Minimum norm estimates are biased towards more superficial sources (Jeffs, Leahy, & Singh, 1987). This is not an inherent limitation of this approach to solving the inverse problem. Rather, MEG is not equally sensitive to all source locations. With increasing source depth, the *lead fields* are increasingly attenuated (Figure 4.8), and the probability of source detection is reduced (Hillebrand & Barnes, 2002).

As a solution to this inherent bias in the lead fields, the weighting matrix in Eq. 4.4.18, C_s, can be used to apply a weighting factor to the sources based on their depth. This factor can be derived from the *forward fields* of each source (the columns of the forward solution, L) as the magnitudes of the *forward fields* are related to the depth of the source (Figure 4.8). One approach based on this decrease in magnitude of the forward fields with increasing depth was proposed by Lin et al. (2006b). The diagonal values of the source weighting matrix, C_s, corresponding to the three orthogonal components at the i^{th} source location are scaled by a factor derived from the forward fields for that location:

$$f_i = (L_{i,x}^T L_{i,x} + L_{i,y}^T L_{i,y} + L_{i,z}^T L_{i,z})^{-\gamma} \tag{4.4.19}$$

where $L_{i,n}$ contains the forward fields for the n component at the i^{th} source location, and γ is the depth weighting parameter. This equation has the effect of magnifying deeper sources compared to more superficial sources. The amount of

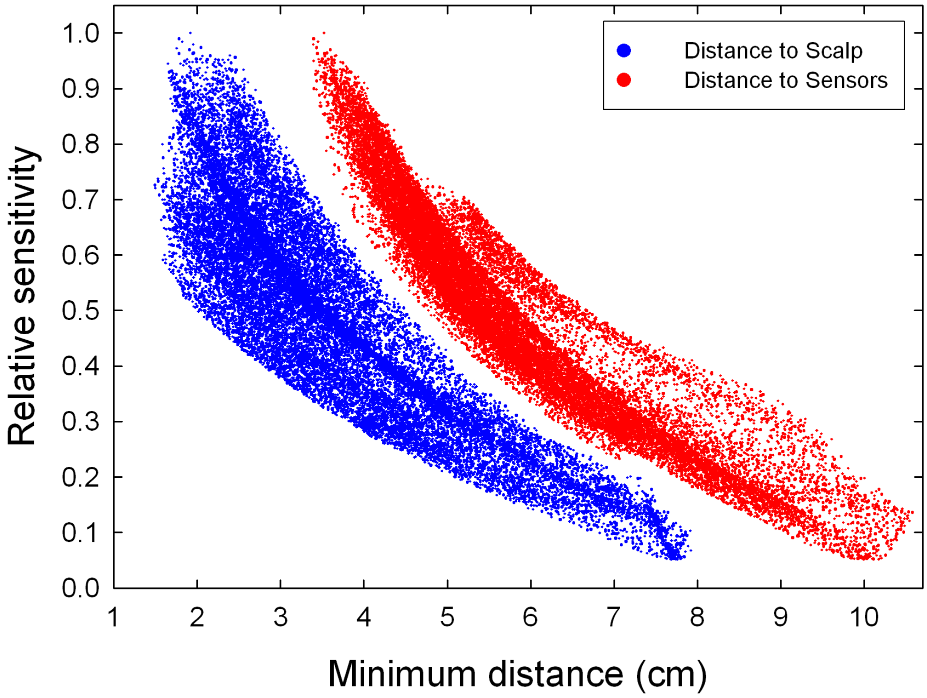

Figure 4.8. Sensitivity of MEG illustrated through the relative size of source lead fields with increasing source depth. Data are the magnitude of the source lead fields (ℓ_2-norm) computed for each of 20,484 locations on a model of a participant's cortical surface using 248 magnetometer sensors. Distances to the scalp (blue) and sensors (red) were calculated as the minimum distance between each source and the vertices of a scalp model and the sensor positions, respectively.

depth weighting is increased as the value of γ is increased. Setting $\gamma = 0$ means that no depth weighting is applied as for all i, $f_i = 1$.

An evaluation of this approach to depth weighting suggested that setting $\gamma = 0.5$ can improve cortical localisation of minimum norm estimates (Fuchs et al., 1999). However, a more detailed examination by Lin et al. (2006b) suggested that the actual amount of depth weighting required to correctly account for the superficial bias of MEG depends on the nature of the neural activity, with more focal activity (e.g. early auditory or somatasensory activation) requiring a higher parameter value (γ =0.7–0.8) for correct localisation. In particular, Lin et al. (2006b) warned that diffuse cortical activation patterns, as observed in some cognitive fMRI tasks, may require a specific degree of depth weighting; i.e. a different value of γ compared to that suitable for focal activity. The parametric nature of the above approach to depth weighting allows for the examination of several levels of bias correction to estimates of source strengths, and therefore for an appropriate value of γ to be estimated should the depths of sources be known a priori.

In situations in which the depths of the sources are not known, the objective choice of a suitable depth-weighting parameter may be difficult. An alternative, parameter-free, method of depth-weighting involves decomposing the lead fields of the three orthogonal components at each source location using singular value decomposition (SVD). The inverse of the maximum singular value for each location can then be used as a weighting factor for all three components (Fuchs et al., 1999):

$$C_{s,i} = \frac{1}{\max(\Sigma)} \text{ where } L_i = U\Sigma V^T \text{ and } \Sigma = \left[\sigma_x, \sigma_y, \sigma_z\right] \tag{4.4.20}$$

where $C_{s,i}$ is the diagonal values of the weighting matrix associated with the i^{th} source location, L_i contains the forward fields for that same location, and Σ contains the singular values of those forward fields. This method has the advantage of being parameter free and therefore straightforward to compute.

4.4.9 Summary

In this section, the minimum norm approach to solving the MEG inverse problem has been presented. The fundamental assumptions behind the technique, along with the mathematical foundations have been outlined. Extensions of the classical minimum norm solution have been discussed which provide more accurate estimates of the neural activity which produces the MEG data. These include improved localisation accuracy through depth-weighting and high-resolution anatomical constraints, and the inclusion of noise estimates with regularisation to increase the sensitivity and stability of the solution. Limitations of the technique have been discussed, including the need for a priori source models and low-resolution localisation due to the linear approach to solving the inverse problem.

4.5 Linearly Constrained Spatial Filtering

4.5.1 Introduction

Like minimum norm estimation, spatial filtering is a method of obtaining estimates of activity at locations within the brain. The technique involves weighting the contribution of each sensor such that signals from the source location of interest are neither attenuated nor amplified, and signals from all other spatial locations are suppressed. Thus, a spatial filter is formed which acts as a *pass-band* filter at the source location of interest, and a *stop-band* filter at all other locations (Huang, Shih, Lee, Harrington, Thoma, Weisend, Hanlon, Paulson, Li, Martin, Millers, & Canive, 2004). In reality, such an ideal filter design is not possible due to the small number channels employed in MEG systems (200–300) which provide insufficient degrees of freedom to ensure that the filter perfectly attenuates signals from all unwanted source locations (Figure 4.9). However, additional constraints can be introduced to construct an optimised filter for the reconstruction of localised neural activity. In this section, the fundamental theory behind the technique is introduced and its application to the analysis of MEG data is discussed along with its advantages and disadvantages.

4.5.2 Designing a Spatial Filter

The spatial filter is realised though a set of weights, w, which, when applied to the measured data, d, yield the estimated source strength, s, at the location of interest for

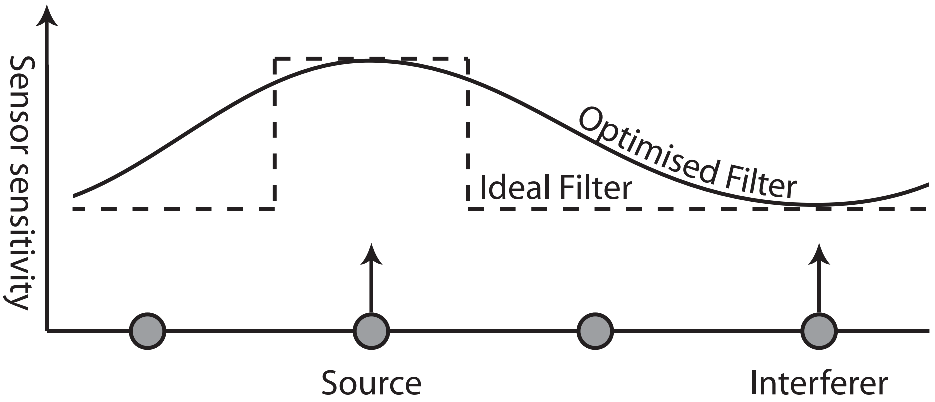

Figure 4.9. Graphical representation of the spatial filter focused on a source of interest. Spatial locations are on the horizontal axis. The theoretical spatial filter only permits signals from a location of interest to pass through ('Ideal filter'). The constraints of a pass-band at the source location of interest and minimum power have the effect of attenuating the interfering sources only. The power from the source of interest is not affected by the minimum power constraint because it is subject to the pass-band constraint; i.e. its power must remain unperturbed. The result of the constraints is that the filter is focused on the location of interest ('Optimised filter') (adapted from Hillebrand et al., 2005).

a particular source orientation:

$$s = w^T d \tag{4.5.1}$$

where w^T is the transpose of w making it a row vector which can then be multiplied by the data, a column vector, yielding a single value for the source strength, s. If the weights are applied to the measured data across multiple samples within a time window, d is replaced with D, an $m \times t$ matrix of m measurements over t samples. The result is the time-varying amplitude of source activity, s, at the location of interest for a given source orientation:

$$s = w^T D \tag{4.5.2}$$

This reconstructed source time-course, s, is sometimes referred to as a *virtual electrode*. The total power, Q, of this reconstructed source over the time window is defined as (Huang et al., 2004; Vrba, 2002):

$$Q = (s)^2 = (w^T D)^2 = w^T D D^T w = w^T C w \tag{4.5.3}$$

where C is the data covariance matrix containing the variances and covariances between the sensors calculated over the time window of interest. The calculation of the weights, w, for each source location and orientation is subject to two constraints: minimise the total power of the reconstructed source, Q, within the time-window of interest, and maintain a *pass-band* filter at the location of interest (Figure 4.9). Thus the variance of the output is minimised subject to a linear constraint; hence the technique is referred to as *linearly constrained minimum variance* (LCMV) spatial filtering. The two constraints which are applied to the spatial filter can be expressed as:

$$\min_{Q} \{Q\} \qquad \text{subject to:} \qquad w^T \ell = 1 \tag{4.5.4}$$

where Q is the source power and ℓ is the *lead field* for that source. A formulation of the filter weights, w, can then be derived such that both expressions in Eq. 4.5.4 are satisfied (Van Veen, van Drongelen, Yuchtman, & Suzuki, 1997) (Appendix B, Section B.2, p. 255):

$$w = \frac{C^{-1}\ell}{\ell^T C^{-1} \ell} \tag{4.5.5}$$

By substituting Eq. 4.5.5 into Eq. 4.5.3 we can arrive at a new expression for the power from a source at a location of interest with a particular orientation within a chosen time window, or the estimated *spatial spectrum* (Van Veen et al., 1997; Vrba & Robinson, 2001):

$$Q = \left(\ell^T C^{-1} \ell\right)^{-1} \tag{4.5.6}$$

Just as with minimum norm estimates, the LCMV problem can be regularised to obtain a more stable solution. Regularisation is achieved by replacing the covariance matrix C in Eqs. 4.5.5 and 4.5.6 with $(C + \lambda\Sigma)$ where λ is the regularisation parameter and Σ is the estimated noise covariance (Robinson & Vrba, 1999).

4.5.3 Factors which Influence the Spatial Filter

From Eqs. 4.5.5 and 4.5.6 it can be seen that the spatial filtering and the reconstructed source power depends on two factors. The first is the lead field, ℓ, of the source which is affected by the number, location, and sensitivity of the measurement sensors. Limitations on the sensitivity of MEG to radial source components (Hämäläinen et al., 1993) have obvious consequences for the ability of any source localisation technique to reconstruct such source activity, but more specific to this technique, the layout and density of the sensor array affects the spatial selectivity of the filter; i.e. its ability to attenuate signals originating from spatial locations other than the location of interest (Hillebrand et al., 2005; Van Veen et al., 1997; Vrba, Robinson, & McCubbin, 2004).

The second factor affecting the spatial filter is the data covariance matrix, C. The matrix depends principally on the time window over which it is calculated, and is influenced by the amount of correlation between the neural sources with give rise to the measured magnetic fields (Figure 4.10) (Van Veen et al., 1997). The dependence of this technique on deriving an accurate estimate of the data covariance highlights one of its main limitations: the necessity to select long data windows (at least several hundred milliseconds or a number of samples several times greater than the number of sensors) to derive a stable estimate of the covariance so that it is non-singular and may be inverted (see Eqs. 4.5.5 and 4.5.6), and to incorporate lower frequency information in the solution. To ensure that sufficient power is reconstructed, the experimental design must provide enough data to compute the covariance or the amount of regularisation must be increased, which has the effect of decreasing the spatial resolution of the technique (Brookes, Vrba, Robinson, Stevenson, Peters, Barnes, Hillebrand, & Morris, 2008). Instability in the underlying source configuration within the time window chosen for the covariance, such as may arise in cognitive tasks which are associated with a distributed array of cortical processes, manifests itself in the solution as a mixture of the time-varying spatial activity. Such fluctuations in the source configuration could affect the quality of the covariance estimate (Van Veen et al., 1997). The effects of source correlation on source reconstruction are discussed later in this chapter.

4.5.4 Depth Weighting

Unlike minimum-norm estimates, LCMV spatial filtering methods contain a correction for the bias towards more superficial sources that is inherent in MEG. Mosher et al. (2003) showed that spatial filtering could be described as a special case of a minimum-norm estimation, in which the source covariance matrix is calculated as the norm of the source lead fields, weighted by the data covariance matrix:

$$C_s = (L^T C^{-1} L)^{-1} \tag{4.5.7}$$

where C_s is the source covariance matrix which is a diagonal matrix due to the assumption that the activity of sources is independent, L is the lead field matrix, and C is the data covariance matrix. Eq. 4.5.7 shows that the lead fields are weighted when estimating the variances of the sources (Lin et al., 2006b). Thus, the spatial filtering approach contains a form of depth-weighting similar to the techniques used to correct for the depth bias in minimum-norm estimates (Section 4.4.8, p. 86). Therefore, the spatial filtering approach shows less bias towards superficial sources without any additional corrections compared to minimum norm estimates, and does not necessitate the incorporation of an explicit depth-weighting correction.

4.5.5 Characterising Distributed Neural Activity

The expression for the total source power at a certain location and orientation as shown in Eq. 4.5.6 can be used to determine the spatial distribution of activity within the cortex. Typical approaches involve computing the source power at a discrete number of points on a model of the cortical surface or on a regular grid of source locations which encompasses the cortex, a technique referred to as ‘source scanning’ (Vrba & Robinson, 2001). Although Eqs. 4.5.5 and 4.5.6 form the basis of most spatial filtering techniques which have been applied to the analysis of MEG data (Huang et al., 2004), in practice those equations alone are not sufficient to generate a spatial ‘map’ of cortical activity.

MEG measurements are more sensitive to sources closer to the sensor array than to deeper sources (Hillebrand & Barnes, 2002) (see Section 4.4.8, page 86). In contrast, regardless of source depth, the noise at the sensors remains at a relatively constant level. Therefore, the signal-to-noise ratio of the MEG measurements decreases with increasing source depth. This effect influences source power estimates determined using Eq. 4.5.6 such that the reconstructed power at locations closer to the centre of the head is largely determined, if not completely dominated, by noise rather than source activity (Van Veen et al., 1997; Vrba & Robinson, 2001).

As a solution to this problem, Van Veen et al. (1997) suggested normalising the reconstructed source power at each source location using estimates of the sensor noise projected to the same location, a technique which has also been applied to minimum norm estimates (Dale et al., 2000). The power estimate in Eq. 4.5.6 is calculated using the data covariance which in turn is derived from the non-averaged measured data—it contains both signal and noise. An estimate of the noise power alone, or the *noise spatial spectrum*, at a particular location and orientation can be calculated as (Van Veen et al., 1997):

$$v = \left(\ell^T \Sigma^{-1} \ell\right)^{-1} \tag{4.5.8}$$

where Σ is an estimate of the noise covariance across the sensor array and ℓ is the lead field of a source at the location and orientation of interest. The estimated source power, Q, can then be adjusted by normalising it in relation to the noise power, v, resulting in a reconstructed estimate, R, of source-to-noise variance:

$$R_{\text{NAI}} = \frac{Q}{v} = \frac{\left(\ell^T C^{-1} \ell\right)^{-1}}{\left(\ell^T \Sigma^{-1} \ell\right)^{-1}} \tag{4.5.9}$$

Van Veen et al. (1997) referred to this as the *neural activity index*. This calculation is performed for each source location within the source model, creating a map of neural activity.

So far only the case of a single source orientation has been discussed. In practice, the orientation of each source is either estimated from anatomical data (see Section 4.5.6, p. 93), found using non-linear search methods (Robinson & Vrba, 1999), or more frequently it is treated as unknown, such as when a grid is used as the source space. Just as for the minimum norm estimates, the spatial filter can be constructed for each of three orthogonal components at each source location. This information can then be combined to arrive at an estimate of combined source power at each location (Van Veen et al., 1997):

$$R_{\text{NAI(VanVeen)}} = \frac{(\ell_x^T C^{-1} \ell_x)^{-1} + (\ell_y^T C^{-1} \ell_y)^{-1} + (\ell_z^T C^{-1} \ell_z)^{-1}}{(\ell_x^T \Sigma^{-1} \ell_x)^{-1} + (\ell_y^T \Sigma^{-1} \ell_y)^{-1} + (\ell_z^T \Sigma^{-1} \ell_z)^{-1}} \tag{4.5.10}$$

Huang et al. (2004) pointed out that this approach has a potential weakness: the presence of one very weak component causes the estimate to be dominated by that component. This is particularly problematic in MEG, where the radial component is either very weak or near zero, depending on the head-model used to compute the lead fields. Therefore, Huang et al. (2004) presented a solution which involves noise-normalising the power of each component separately which removes this bias:

$$R_{\text{NAI(Huang)}} = \frac{(\ell_x^T C^{-1} \ell_x)^{-1}}{(\ell_x^T \Sigma^{-1} \ell_x)^{-1}} + \frac{(\ell_y^T C^{-1} \ell_y)^{-1}}{(\ell_y^T \Sigma^{-1} \ell_y)^{-1}} + \frac{(\ell_z^T C^{-1} \ell_z)^{-1}}{(\ell_z^T \Sigma^{-1} \ell_z)^{-1}} \tag{4.5.11}$$

4.5.6 Anatomical Constraints

One of the benefits of the spatial filtering approach over linear methods such as minimum norm is that a priori source models based on accurate anatomical information are not required. For minimum norm, the activity of a source whose location is not specified a priori is projected onto the specified source model. This is not the case with spatial filtering, as the weights are optimised for each source location independently. Thus, spatial filters can be calculated for a regular grid of locations which encompass the cortex, a property of all 'source scanning' techniques. Despite the absence of a requirement to constrain the distributed activity estimates to the cortical surface, it may still be useful to do so in order to make comparisons with the results of other methods, such as distributed dipole models including minimum norm estimates. The use of a cortical model as the source space of interest is also a valid constraint as it is in agreement with the knowledge of how and where the signals measured by MEG originate (Okada, 2003).

4.5.7 Correlated Sources

When applying the spatial filtering approach to MEG data, it is typically assumed that all sources are uncorrelated; i.e. the source covariance matrix is a diagonal matrix. If the time courses of two or more sources are correlated with each other, then the estimated variance for those sources is reduced (Figure 4.10) (Van Veen et al., 1997). This result arises because the spatial filtering method cancels the correlated part of the variance of the target source. The lead fields of the correlated sources sum to produce a lead field which does not match the expected lead field for the target location, and it is therefore minimised according to the minimisation of output power constraint (see Eq. 4.5.4) (Hillebrand et al., 2005; Van Veen et al., 1997). In other words, the activity is seen as originating from a source at a location other than that of the target location, and the output power of that location is attenuated due to the a priori constraint that the total power of the solution is minimised.

To examine the effects of source correlation on the ability of spatial filtering methods to reconstruct source power, Van Veen et al. (1997) computed the *neural activity index* for two sources at varying levels of correlation and relative distance. They found that the two sources could be resolved as peaks in the neural activity index when partially correlated at 50%, even when spaced close together. This outcome is consistent with the findings of Huang et al. (2004) who used a slightly higher level of correlation of 61%. When the sources were perfectly correlated, Van Veen et al. (1997) found that sources were merged when located close together and tended to cancel each other when they were more distant from each other.

While occurrences of perfectly correlated sources are not likely to occur very often, it is reasonable to presume that activity which is tightly time-locked to the onset of a stimulus, e.g. sensory input, is at least weakly correlated. There are however some cases, such as activity found bilaterally on the transverse temporal gyrii (primary and associated auditory cortices) in response to an acoustical stimulus, when sources are highly correlated and therefore would not be reconstructed properly when the spatial filtering approach is used without additional modifications (Hillebrand & Barnes, 2005). Necessary modifications might include changes to the lead fields (Brookes, Stevenson, Barnes, Hillebrand, Simpson, Francis, & Morris, 2007) or using only a subset of the available sensors which are positioned close to the target location (Hillebrand et al., 2005). While the first of these approaches does not require a priori knowledge of the location of the correlated sources, determining the locations is computationally prohibitive and is ultimately limited to the case of two correlated sources. Therefore, the use of spatial filtering methods when no a priori information on the number and extent of source correlations is available has the potential to provide inaccurate reconstructions of the distributed source activity should strong correlations exist between source time courses. For cases where the presence of highly-correlated source activity is known, the selection of sub-groups of sensors is currently the most convenient, and computationally-efficient, method to deal with correlated sources. For auditory data, this can be accomplished by selecting sensor groups for each hemisphere to reduce the effects of the correlated source activity in the contra-lateral hemisphere.

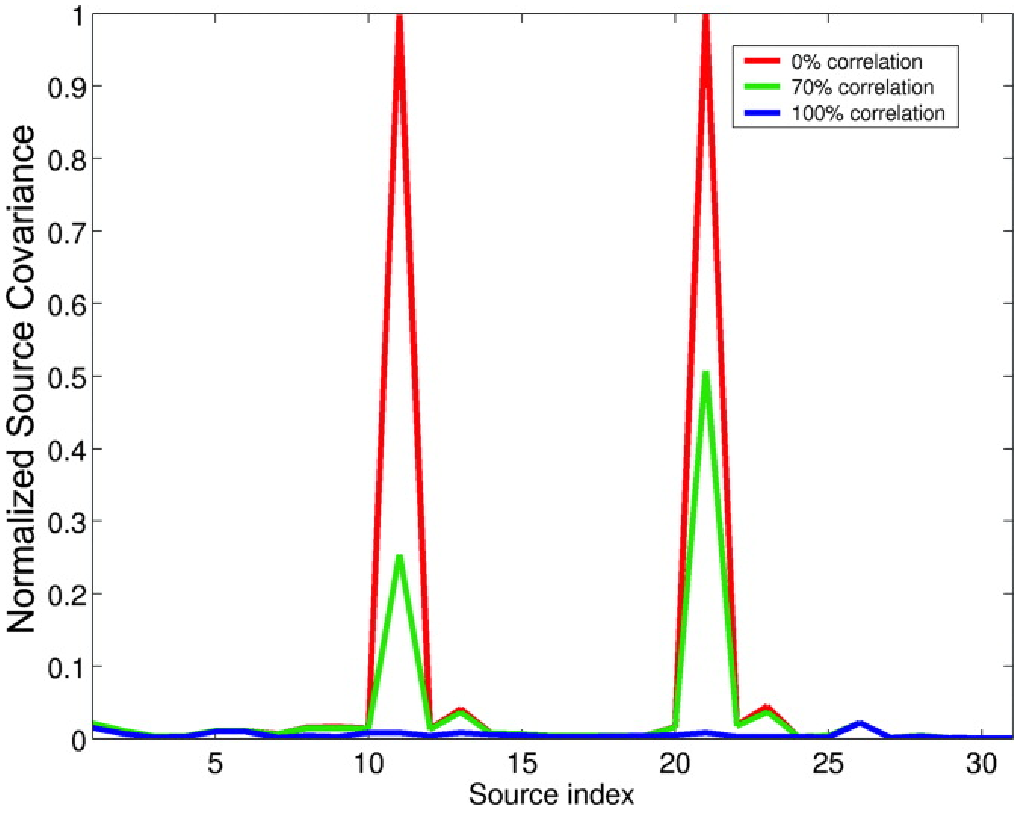

Figure 4.10. The graph shows the variance of individual sources, the diagonal elements of the source covariance matrix, for simulated data. Sources are arranged spatially along the x-axis with their variance plotted on the y-axis. The simulation comprised two sources with varying degrees of correlation between their activity over the time window used to compute the data covariance matrix. The results show that perfectly correlated sources are not reconstructed, and partially correlated sources are attenuated by the spatial filtering approach (after Hillebrand et al., 2005).

4.5.8 Summary

The reconstruction of source activity using LCMV spatial filtering has been presented in terms of its underlying assumptions and mathematical formulation. Factors influencing the success of source reconstruction have been outlined. These include the absence of strongly correlated source activity and deriving an accurate estimate of data covariance over a window of interest. Advantages of the technique have been discussed, such as the ability to scan for source activity over a regular grid of locations requiring no a priori anatomical information.

4.6 Comparison

A comparison of the two techniques which have been outlined so far can be formulated in terms of the constraints which are applied to the estimated solution of source activity. In the case of minimum norm, the power of the solution is minimised subject to the constraint that the modelled data produced by the estimated solution matches, or is close to, the measured data (Pascual-Marqui, 1999):

$$\min_{s}\{s^T W^{-1} s\} \qquad \text{subject to:} \qquad d = Ls \tag{4.6.1}$$

where s is a column vector of the estimated source strengths and W is a diagonal weighting matrix. In the case of classical minimum norm, $W = I$, and for the weighted approach W is usually a diagonal matrix used to correct for the inherent depth bias or to incorporate a priori information from other modalities such as fMRI. For spatial filtering, the power of the source activity is also minimised, but the constraint ensures a pass-band at the location of interest (Van Veen et al., 1997):

$$\min_{w}\{w^T C w\} \qquad \text{subject to:} \qquad w^T \ell = 1 \tag{4.6.2}$$

where w is the spatial filter weights, C is the estimated data covariance over the time window of interest, and ℓ is the forward field for that location. Both include a minimisation of power as the first constraint, but this serves a different purpose for each technique. For minimum norm, the constraint ensures that a unique solution to the linear inverse problem is found (Hämäläinen et al., 1993). For spatial filtering, the constraint has the effect of increasing the attenuation of source power from locations other than the location of interest (Van Veen et al., 1997). The difference in the second constraint determines the nature of the approach; i.e. a linear inverse or scanning method. For minimum norm, the solution must conform to the relationship defined in the forward problem; i.e. the difference between the modelled and measured data must be minimised. In contrast, no such constraint exists for spatial filtering, and instead the existence of a pass-band at the source location of interest is the critical constraint in the reconstruction of the source activity.

The minimum norm approach fits a model in terms of source power, at a discrete set of source locations and orientations specified a priori, to the measured data by minimising a cost function (Eq. 4.4.5); i.e. through a least-squares fit. Therefore, it is a true inverse solution as it provides a solution for the system of linear equations (Eq. 4.4.1). However, due to the underdetermined nature of the inverse problem in MEG, the resolution of minimum norm estimates, and indeed of most linear estimation methods, is quite low. As a result, estimates of focal source activity tend to be spread out over a far wider area than the original extent of the activation. This smearing can be reduced significantly by normalising the reconstructed source power by the estimated power due to noise (Dale et al., 2000) or through the use of iterative focussing techniques (Gorodnitsky et al., 1995; Grave de Peralta Menendez, Gonzalez Andino, Lantz, Michel, & Landis, 2001; Liu et al., 2005).

In contrast, the spatial filtering approach uses information about the location and orientation of each source to reconstruct its signal spectrum from the sensor data and does not attempt to match the entire source model against the measured data, such as through a least-squares fit. The absence of the least-squares constraint is common to all 'source scanning' approaches (Vrba & Robinson, 2001). This feature of the approach has the advantage of giving the technique excellent spatial resolution, which can be as high as the grid chosen for the source scanning (Barnes & Hillebrand, 2003; Sekihara et al., 2005). However, it also has the consequence that the reconstructed map of distributed activity does not necessarily produce data close to the measured data. This outcome results from the fact that the calculation of the weights for each location is independent of the weights at any other location, and that the spatial filter is not completely successful in attenuating source activity from locations other than the location of interest. As a result, the introduction of 'phantom', or false-positive, sources in the reconstructed source maps is possible, and is a function of the noise estimation approach, the signal-to-noise ratio of the data, and the extent of correlation between sources (Huang et al., 2004).

4.6.0.1 Source Models

One distinct advantage of the spatial filtering approach over linear modelling approaches such as minimum norm estimates is that a model of source locations, such as a cortical surface, does not necessarily need to be specified a priori. As the power at each location is reconstructed independently, this allows for activity to be scanned on a regular grid which encompasses the cortex. With standard minimum norm estimates, all of the activity, excluding a component assumed to be purely due to noise in the measurement system, is projected onto the set of source locations (Hauk, 2004). Therefore the choice of the a priori source model, particularly in terms of its anatomical accuracy, affects the localisation accuracy of the solution along with the correction for depth weighting. If multiple sources need to be reconstructed and they are located in close proximity to each other (within a few millimetres) then the spatial filtering approach is more likely to resolve them (Sekihara et al., 2005).

4.6.0.2 Source Correlation

Correlation between the sources can greatly affect the performance of spatial filtering techniques. This is independent of the distance between the sources, although the manner in which the source reconstructions are altered does depend on distance such that adjacent correlated sources are merged and the activity of spatially distance sources is smoothed along a line between them (Van Veen et al., 1997). Although partially correlated sources are still reconstructed, the activity is attenuated due to the removal of the correlated portions of the signal. This is not the case with minimum norm estimates which do not attenuate or remove the correlated components of source activity, despite the typical assumption when applying either technique that sources are uncorrelated (Figure 4.11).

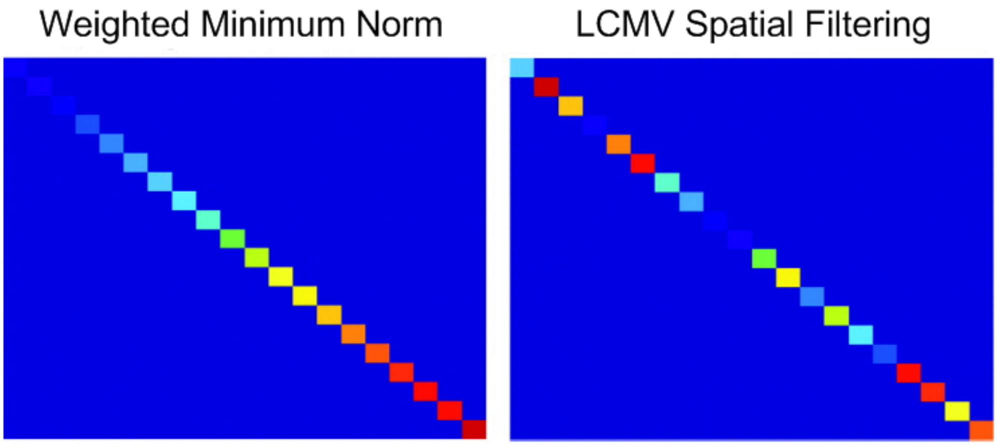

Figure 4.11. A graphical representation of the source covariance matrices for the weighted minimum norm (left) and spatial filtering (right) approaches. The most superficial sources are plotted in the top-left corner and the deepest sources in the bottom-right corner. In both cases, the values away from the main diagonal are zero (blue). The minimum norm approach typically assumes that the sources are uncorrelated (hence the diagonal matrix), with the values of the covariance matrix being used to weight the deeper sources (colours closer to red) to correct for the bias towards superficial sources. However, should sources be correlated, they are not removed from the solution. This behaviour is in contrast with spatial filtering which relies on the assumptions that sources are independent and that correlated activity is suppressed. Whereas the values for the source covariance are specified a priori in the form of weighting factors for the minimum norm approach, the values of the matrix in the case of spatial filtering represent the estimated variance of the sources (red is the highest variance) and are determined directly from the MEG data (adapted from Hillebrand et al., 2005).

4.6.0.3 Evoked and Induced Data

The two techniques can also be contrasted in terms of the form of the MEG data which is used to calculate the solutions. The minimum norm approach is usually applied to the average MEG data for a particular epoch. If the number of averages is substantial, over 80 trials, the signal-to-noise ratio is relatively high, reducing the need for noise covariance estimates to be used, or at least making the effects of including noise estimates less pronounced. The averaging process only retains phase-locked activity in the sensor data, so minimum norm estimates are fit to the *evoked* component of the measured data. For spatial filtering, the data covariance matrix is estimated using the non-averaged MEG data and therefore includes both phase-locked and non-phase-locked activity, or *evoked* and *induced* data, respectively. This ability to reconstruct induced activity is of considerable interest particularly in cognitive tasks where significant amounts of activity are likely to exhibit more variable latencies, or might be ongoing and oscillatory in nature, and consequently might not be phase-locked to the stimulus onset (Hillebrand et al., 2005). Induced activity may also be indicative of changes in functional connectivity (Pfurtscheller & Lopes da Silva, 1999). However, the reliance on non-averaged data means that the signal-to-noise ratio of each data window is poor when compared to average data, which in turn necessitates the use of large numbers of samples to derive stable estimates of data covariance (Brookes et al., 2008) and accurate estimates of sensor noise.

4.6.1 Summary

Table 4.1 provides a summary of the comparison between minimum-norm estimation and spatial filtering.

	Minimum Norm	**Spatial Filtering**
Cortical Surface	*Required*	*Not required*
Covariance estimation	*Not required*	*Required*
Spatial Resolution	*Low*	*High*
Least-squares Fit	*Yes*	*No*
Depth weighting	*Explicit*	*Implicit*
Correlated activity	*Unmodified*	*Removed*
Non-phaselocked activity	*Removed*	*Unmodified*

Table 4.1. Summary of comparison between Minimum norm and LCMV spatial filtering techniques.

4.7 Application to the Current Research

The ill-posed nature of the inverse problem in MEG means that no unique solution exists, and that constraints on the solution are necessary to estimate the spatial distribution and time-varying activity of sources within the cerebral cortex. The contrasting properties and performance of the minimum norm and LCMV spatial filtering approaches to estimating neural activity highlight the consequences of such an ill-posed system, and also indicate that no individual source analysis technique offers a ‘perfect’ solution for all types of cortical activity. Additionally, the underdetermined nature of the problem makes solutions numerically unstable, and imposes additional complications on the calculation of valid solutions.

A central goal of the research reported in this thesis is to characterise the location and time course of cortical activity that accompanies auditory attention to spatially presented speech against a background of other speech. From previous research in other modalities, such as fMRI (Nakai et al., 2005; Pugh et al., 1996) and PET (Tzourio, Massioui, Crivello, Joliot, Renault, & Mazoyer, 1997), we know that cortical activation associated with auditory attention involves a widely distributed network of regions. However, the extent of correlation, if any, between the sources of activity is unknown. Additionally, attention-related activation or modulation is often characterised by slow-wave activity (<10 Hz) (Picton, 1992) and constructing data windows to compute the data covariance which are long enough to include such low-frequency information is difficult when the stimuli are highly complex and non-stationary, as is the case when simulating multi-talker environments. These two issues make the application of spatial filtering to data from the current research challenging.

However, the LCMV spatial filtering approach offers significant advantages over linear inverse solutions such as minimum-norm. The advantages include high spatial resolution and the ability to reconstruct the induced data; i.e. the portion of the MEG data which is not tightly phase-locked to the onset of the stimuli or events with the stimuli. Therefore, this method is of considerable interest in analysing aspects of the MEG data which are not considered when the minimum-norm technique is used, such as changes in ongoing oscillatory activity at high frequencies. As discussed in Chapter 3 (Section 3.3.3, p. 53), oscillatory activity has been suggested to be involved in the coupling of different regions of the brain (Tallon-Baudry & Bertrand, 1999), the communication between different cortical processes (Bibbig et al., 2002), and a possible reflection of different attentional states (Gross et al., 2004).

Thus, both minimum norm and spatial filtering techniques were used to characterise the distributed cortical activity associated with the spatial listening task, introduced in Chapter 5. The application of the two techniques allows for a more thorough examination of the MEG data, compared to the use of either technique in isolation. To ensure that the highly-correlated bilateral sources of activity associated with auditory processing were not reconstructed incorrectly, a separate group of MEG sensors was used for spatial filtering analysis to reconstruct the source activity in each cortical hemisphere.

4.8 Summary

- Magneto-encephalography (MEG) measures magnetic fields around the head which result from the activity of large populations of neurons within the grey matter of the cerebral cortex.
- Two contrasting techniques for reconstructing the distribution and time-course of activity in the cortex are minimum-norm estimation and spatial filtering.
- Minimum-norm estimation provides a linear solution to the problem of estimating the strengths of a large number of sources, whose locations and orientations are known in advance. The accuracy of the solutions is dependent on the anatomical accuracy of the source locations provided in advance, and the signal-to-noise ratio of the MEG data.
- Minimum-norm solutions to MEG data are often numerically unstable, and require regularisation. This process has the effect of spatially-smoothing the solutions. A weighting has to be applied to deeper sources to correct for a bias towards more superficial sources.
- Spatial filtering reconstructs the activity at a known source location by focussing the MEG sensors through the application of a set of sensor weights.
- Unlike minimum-norm, the reconstruction of source activity in the brain with spatial filtering can be performed using a regular grid of source locations rather than a detailed anatomical model of the cortex.
- Application of the spatial filtering technique requires a stable estimate of the data covariance within a specified time window. It is not necessary to explicitly weight deeper sources, but estimates of source activity are often normalised using noise estimates, creating the neural activity index.
- The spatial resolution of spatial filtering is high, and the reconstructed source activity includes evoked and induced data.
- High levels of correlation between the activity of different sources results in that activity being suppressed or incorrectly localised by the spatial filtering technique. The effects of correlated source activity can be reduced through the selection of sub-groups of MEG sensors local to the source location of interest.

Chapter 5

Voluntary and Involuntary Attention in a Multi-talker Environment

This chapter reports four experiments which examined the ability of listeners to perceive speech in a demanding multi-talker listening environment. The experiments studied the benefits of providing prior information about the identity of who would speak a target phrase, where they would be located, and when they would speak. The experiments also examined the distracting effects of a new person starting to speak. Several phrases were presented in an overlapping sequence in which a new phrase started every 800 ms. Each phrase had the form "Ready CALL-SIGN, go to COLOUR NUMBER now." Participants had to attend to each new phrase, and determine whether it contained the call-sign keyword that had been allocated to them. The task was to report the colour and number keywords from that 'target' phrase. Experiments 1 and 2 examined the benefits of providing prior information about the phrase containing the target call-sign—*who* would speak it, *when* it would be spoken, and *where* it would be spoken from. In Experiment 1, phrases started one at a time. In Experiment 2, phrases started in pairs. The results suggest that the onset of a new talker captures attention automatically. In Experiment 1, this effect largely overcame uncertainty about *who*, *when*, and *where*. When pairs of talkers started speaking simultaneously (Experiment 2), attention was captured by the more intense talker and generally disrupted the ability to attend to the less intense talker. Experiments 3 and 4 examined the distracting effects of a phrase which started shortly before or after the onset of a target phrase. Compared to when the target and masker onsets were simultaneous, introducing a target-masker asynchrony improved performance for the condition in which uncertainty about when the target phrase would be spoken was minimised by placing it within a temporally-regular sequence of phrases (Experiment 4). No effect of asynchrony was found when the onset of the target phrase was unpredictable (Experiment 3). The results of the experiments suggested that attention, specifically the ability to adopt a diffuse attentional state and to take advantage of prior information, plays an important role in listening to what one person is saying while many other people are speaking at the same time.

5.1 Introduction

Listeners face considerable challenges in multi-talker environments. Talkers start speaking and stop speaking at unpredictable times. Sometimes they start speaking in sequence; sometimes they start speaking at the same time. In order to attend to the talker who is saying things that are relevant, a listener may need to monitor many talkers, divide attention among them, or alternate attention from one talker to another. Then, when a talker says something relevant, the listener needs to selectively attend to that talker, resisting distraction from competing talkers or *maskers*. Difficulties in dividing, alternating, and selecting attention are associated with auditory handicap (Gatehouse & Noble, 2004). The experiments reported in this chapter explored the ability of listeners to switch and sustain attention among talkers. The experiments measured the improvements in those abilities that arise from knowing who would speak relevant information, where they would be located, and when they would start to speak.

5.1.1 Benefits of Knowing 'Where' to Listen

Kidd et al. (2005) showed that knowledge about *where* a target talker will speak improves intelligibility when several spatially-separated talkers start to speak simultaneously. Performance at reporting the colour and number co-ordinates in CRM phrases improved significantly with increasing certainty about where a target phrase would be presented. This result occurred regardless of whether the identity of the target call-sign was provided before or after the stimuli. Thus, knowing where to direct attention provided an advantage to the listener in hearing out the target phrase when there were high levels of uncertainty about the vocal characteristics of the target and masker phrases. Mondor & Zatorre (1995) suggested that this selective auditory attention acts like a spatial filter, which attenuates signals based on their distance from the focal point.

The advantage from knowing where to listen was found when there was uncertainty about *who* would say the target phrase, as the talker who spoke the target phrase was selected at random on each trial. It is not clear whether knowledge about *where* a target phrase will be spoken from is also advantageous when the identity of the target talker is known to the listener.

5.1.2 Benefits of Knowing 'Who' to Listen to

Benefits from knowing *who* will speak a target phrase were reported by Brungart et al. (2001). By presenting a target phrase, identified by a unique keyword, and between 1 and 3 masker phrases simultaneously, Brungart et al. (2001) found that performance improved when listeners were provided with prior information about the identity of the target talker, compared to when the target was chosen randomly. This effect was largest when all maskers were of a different gender to the target talker, or when the maskers comprised both males and females. In Brungart et al.'s experiment, knowledge about *where* the target phrase was located was not used, as all phrases were presented diotically over headphones. It is not clear whether knowing *who* will speak would provide a benefit when a listener knows *where* the talker will be located.

5.1.3 Benefits of Knowing 'When' to Listen

In the experiments of Kidd et al. (2005) and Brungart et al. (2001), the advantages from knowing *where* to listen for a target phrase and from knowing *who* would speak it were identified in situations in which multiple talkers started speaking at the same time. Simultaneous phrase onsets have also been used in other studies of multi-talker environments (Allen et al., 2008; Arbogast et al., 2002; Humes et al., 2006; Lee & Humes, 2005; Shafiro & Gygi, 2007). Few studies have examined situations in which phrase onsets are asynchronous. In an example of one such study, Webster & Thompson (1954) recreated the listening demands of an aircraft control tower by presenting messages containing identification information for each plane, which participants had to repeat out loud. Overlapping pairs of messages were presented. The asynchrony between the messages (2, 4, and 6 sec) and the intensity difference (8 and 24 dB) within each pair were varied. Webster & Thompson (1954) found that identification performance was better for the more intense phrases, even when they lagged behind the other message in the pair. However, when the two overlapping messages were presented at the same intensity, the first of the two messages was reported with a higher accuracy compared to the lagging message. It is not clear whether smaller asynchronies would create similar effects, or whether asynchrony would be beneficial when messages are presented from different spatial locations.

Previous research has suggested that when a target is embedded within a temporally-regular sequence of stimuli, uncertainty about *when* the target will occur has little or no effect on the ability of participants to detect the target (Leibold et al., 2005). It is not clear whether this finding is applicable to speech stimuli, or whether the benefits of reducing uncertainty about *where* and *who* that are observed when phrase onsets are simultaneous (Brungart et al., 2001; Kidd et al., 2005) would arise when talkers start to speak at different times.

5.1.4 The Present Experiments

Against this background, four experiments were conducted. The task in each experiment was to identify the target phrase and report the key-words which it contained in a multi-talker spatial listening task. Experiments 1 and 2 examined the relative benefits of prior information about the target phrase. In Experiment 1, phrases overlapped in time but the onset of each new phrase was separated in time by a fixed duration from other onsets, creating a temporally-regular sequence of phrase onsets. The design of Experiment 2 was similar, except that phrases started in pairs. In both experiments, uncertainty about the location of a target phrase, the talker who spoke it, and its position in a sequence of maskers were manipulated. These experiments sought to establish whether knowledge about who would speak the target phrase and when it would be spoken would facilitate performance. If so, the relative advantages from knowledge of *who*, *when*, and *where*, would be assessed. The experiments would also establish whether the benefits are independent and hence additive, or whether they interact.

Experiments 3 and 4 examined the effect of target-masker onset asynchrony within the same multi-talker environment used in Experiment 1. In Experiment 3, the onset of the target phrase was varied relative to a paired masker phrase whose onset occurred at a fixed interval relative to the other phrases in the sequence. This arrangement resulted in local uncertainty about the onset time of the target phrase. The arrangement was reversed in Experiment 4 so that the onset of the target phrase was aligned with the temporal sequence of phrase onsets and was thus predictable, while the onset of the masker phrase was displaced in time. This arrangement resulted in local uncertainty about the onset of the masker phrase. The experiments examined whether the introduction of an onset asynchrony between the target and masker phrases leads to improved performance. The experiments also sought to ascertain whether or not there is an advantage when the target starts prior to the masker phrase compared to following the masker phrase by examining asynchronies in which the target phrase started before and after the masker phrase.

5.2 Experiment 1

5.2.1 Introduction

The first experiment examined the effects of knowing *who* would speak a target phrase, *where* it would be spoken from, and *when* it would be spoken. Unlike many previous tasks of speech perception, phrases were presented in an overlapping sequence. Participants were required to identify a target phrase and report the information within it.

Based on the findings of Kidd et al. (2005) and Brungart et al. (2001), it was expected that knowing *who* and *where* would allow information within the target phrase to be heard out at lower target-to-masker ratios (TMRs) compared to when no information was available. The presentation of phrases in an overlapping sequence meant that attention had to be focussed on each new phrase to determine whether it was the target phrase or not. It was possible that knowledge about *when* the phrase would be spoken would also provide an advantage, although the results of Leibold et al. (2005) predicated that any advantage arising from knowing *when* was likely to be smaller than the advantages arising from knowing *who* and *where*.

5.2.2 Methods

5.2.2.1 Participants

Eight paid listeners, one male and seven females, between the ages of 18–21 years (Mean age 19.4, $\sigma = 0.9$) participated in Experiment 1. All participants had lived in Britain or Ireland for at least 10 years, spoke English as their native language, and had pure-tone sensitivity better than 20 dB HL at octave frequencies from 250 to 8000 Hz, inclusive, tested in accordance with BS EN ISO 8253-1 (BSA, 2004). Responses to a questionnaire confirmed that participants had no history of hearing health problems (Appendix D).

5.2.2.2 Presentation of Stimuli

Stimuli were presented through an array of 24 loudspeakers (Bose Acoustimass 3 Series IV) spaced at 15° intervals around the perimeter of a circular stage with a diameter of 3.3 m. Only the front arc of 13 speakers was used, giving a range of spatial positions from −90° to +90° azimuth, where 0° was directly in front of the listener and positive azimuths were to the listener's right (Figure 5.1). The axes of the loudspeakers were 104 cm above the floor of the stage. The array was calibrated by measuring the intensity of an octave band of noise centred on 1 kHz presented from each loudspeaker at the centre of the array with a Brüel & Kjær 0.5-inch microphone (Type 4189) and sound level meter (Type 2260 Investigator). The output of individual loudspeakers was adjusted to give the same level, within ±0.1 dB.

The experiment was conducted in a 5.3 × 3.7 m single-walled IAC audiology test room located within a larger sound-proofed room. The ceiling consisted of a suspended grid of mineral tiles, the floor was carpeted, and the walls were lined with foam and fabric. The wall directly facing the listener contained an observation window approximately 2.5 × 1.0 m in size.

5.2.2.3 Stimuli

Stimuli were phrases from the Co-ordinate Response Measure (CRM) corpus (Moore, 1981). The phrases were similar to those described by Bolia et al. (2000) except that they had been spoken by native British-English talkers. Phrases had the form "Ready CALL-SIGN go to COLOUR NUMBER now," with eight call signs ("arrow," "baron," “charlie,” “eagle,” “hopper,” “laker,” “ringo,” and “tiger”), four colours (“blue,” “green,” “red,” and “white”) and the numbers from one to eight, inclusive, giving 256 different phrases. Phrases were spoken by four male and four female adult talkers, producing a corpus of 2048 phrases. Recordings were made in a carpeted sound-attenuated room using a Sennheiser K3N/ME40 microphone whose output was digitised at a sampling rate of 44.1 kHz with 16-bit amplitude quantisation using a LynxONE soundcard (Lynx Studio Technology Inc., CA, USA). The recording of each phrase was edited to remove leading and trailing silences. The average duration of the edited phrases was 2.5 s. The levels of the digitised phrases were subsequently normalised to the same total RMS power using the Praat software package (Boersma & Weenink, 2008). When presented from the loudspeaker at 0°, the variation in the peak A-weighted level among the phrases measured with the calibration equipment using a 1-sec integration time was ±2.5 dB. The gain in the system was set such that the average level of individual phrases was 62.5 dB (A).

5.2.2.4 Phrase Sequences

On each trial, phrases were presented in a sequence of 13 overlapping time slots and each phrase was presented from a different loudspeaker. The slots started at intervals of 800 ms (Figure 5.2). This interval was chosen so that the initial part of the phrases containing the call sign, “Ready CALL-SIGN...”, was not interrupted by the onset of a new phrase. Each sequence included one target phrase, containing the call-sign ‘Baron’, and several masker phrases containing other call-signs. Performance in identifying the colour and number key-words within the target phrase was measured

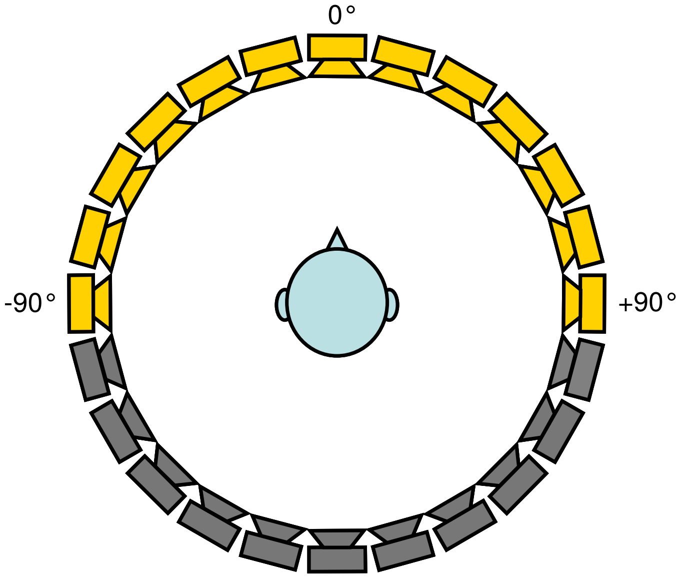

Figure 5.1. The loudspeaker array used in the experiments. Only the subset of loudspeakers in yellow was used.

as the level of the target phrase relative to the average level of individual masker phrases at which performance was 71% correct. This threshold was measured with a 2-down, 1-up adaptive procedure.

To reduce confusion between target and masker phrases, restrictions were imposed on the choice of the masker phrases in the six time slots which partly overlapped the target phrase. These masker phrases and the target phrase were spoken by different talkers, and included unique call-signs and colour-number co-ordinates.

5.2.2.5 Varying Uncertainty

Three parameters of the target phrase could either be fixed within a block of trials or could vary randomly from trial to trial: the talker who spoke the target phrase (*who*), the loudspeaker from which the phrase was presented (*where*), and the time slot occupied by the phrase in the sequence of slots (*when*).

For *who*, the identity of the target talker was either fixed within a block of trials or was varied randomly among the eight talkers from trial to trial. If one talker had been selected as the target talker in conditions where *who* was fixed, performance in those conditions would have been confounded with the intelligibility of the chosen talker. Instead, four of the eight possible talkers were used for the conditions in which *who* was fixed. Using the methods described in Appendix C (Section C.1), two male and two female talkers were selected whose intelligibility was closest to the average intelligibility of the eight talkers.

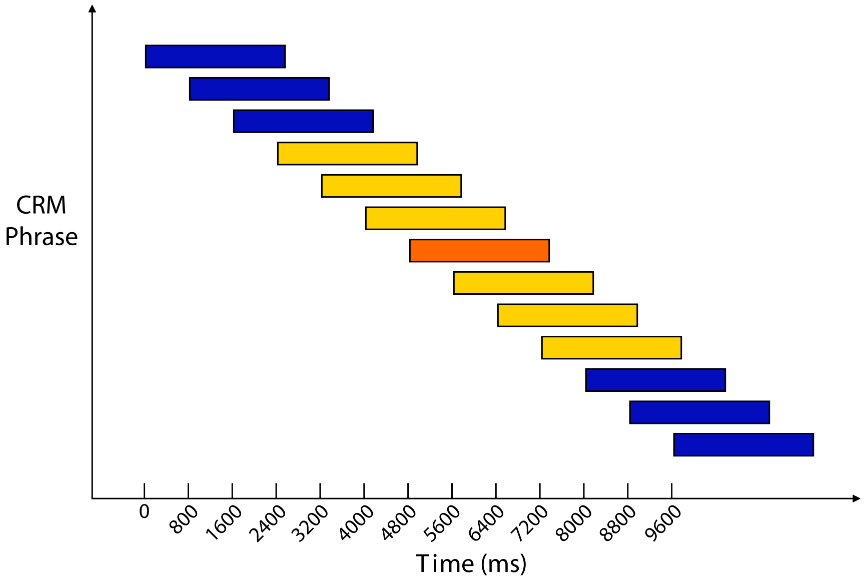

Figure 5.2. A schematic illustration of the overlapping sequence of phrases in Experiment 1. The target appeared either in the centre slot (orange) or in one of the surrounding slots (yellow). Maskers were positioned in the remaining slots. The seven central phrases had unique talkers, call-signs, colour, and number keywords.

When *where* was fixed, the target phrase was presented at 0°. Otherwise, the target was presented randomly from any of the 13 loudspeakers. When the location of the target phrase was fixed, it was possible that any benefits that arose from knowing *where* were confounded with effects directly related to the location of the target phrase. Therefore, the intelligibility of target phrases presented from different loudspeakers was examined using the methods described in Appendix C (Section C.2). The intelligibility of target phrases was greater when presented from ±90° than when presented at 0°. Therefore, constraining the target phrase to the location in front of the listener was unlikely to provide a greater advantage than would have been obtained at other fixed locations.

The position of the target phrase in the sequence of phrases, or *when* the target phrase would be spoken, was either fixed as the 7^{th} time slot or was assigned randomly to one of the 4^{th} to 10^{th} slots. In this way, the target phrase was always overlapped by at least three preceding and three following phrases.

5.2.2.6 Threshold Measurement

Performance in identifying the colour and number key-words within the target phrase was measured as the level of the target phrase relative to the average level of the individual masker phrases at which performance was 71% correct. This threshold was measured with a 2-down, 1-up adaptive procedure (Levitt, 1971). Blocks of trials started at a fixed TMR of +12 dB. The level of the target was reduced following two correct responses and increased following an incorrect response. A correct response was defined as one in which both colour and number key-words were identified correctly. The step size was 6 dB for the first three reversals and then 2 dB for the next twelve reversals. Data collected in the second phase of each block contained twelve runs, where a run was a sequence of changes in target level in one direction only. The threshold was estimated by averaging the mid-points of the even-numbered runs. A single threshold for each participant for each condition was obtained by averaging the four thresholds for that condition from the four sessions. These average thresholds are referred to as Speech-reception Thresholds (SRT).

5.2.2.7 Training

Participants completed four blocks of trials prior to starting the experiment. In two blocks, all three cues (*who, where,* and *when*) varied randomly from trial to trial. In the other two blocks, the values of the three cues were fixed. The order of the training blocks was counterbalanced across participants. In addition, one of each type of training block was completed as a 'warm-up' at the start of the subsequent sessions. In conditions where *who* was fixed, participants experienced a different target talker in each session in an order that was counterbalanced across participants.

5.2.2.8 Design

Eight conditions were defined by the factorial combination of three variables (*who*, *where*, and *when*) with two states (*fixed* or *randomised*). Each participant took part in four sessions. Each session comprised eight blocks of trials, one for each condition. The order of the conditions was counterbalanced across participants and sessions using a first-order Williams design (Williams, 1949), calculated using the R statistical computing environment (R Development Core Team, 2008). A first-order Williams design is a latin square which is constrained by two requirements: 1) each condition is preceded an equal number of times by each of the other conditions, and 2) each condition appears an equal number of times in each position. This design controls, in particular, for the residual effects of the condition immediately preceding the condition of interest.

5.2.2.9 Procedure

Participants sat in the middle of the array of loudspeakers and were instructed to face straight ahead for the duration of the experiment. Prior to each block of trials, the listener was informed which parameters of the target phrase were fixed, what their fixed values were, and which would vary from trial to trial. Responses were made using a touch-screen positioned directly in front of the participant at a comfortable height between 67 and 76 cm. The screen was divided vertically into two areas, one containing four buttons for the colours and the other containing eight buttons for the numbers. Participants were instructed to touch the buttons corresponding to the colour and number key-words in the target phrase on each trial. Feedback on the accuracy of responses was given by a change in the colour of the button: green for correct, red for incorrect. The pacing of the trials was determined by the participant. A 1-sec inter-trial interval began after the colour and number responses had been registered.

5.2.2.10 Analyses

To determine whether the cues (*who*, *where*, *when*) had significant individual effects on performance, a repeated-measures analysis of variance (ANOVA) was performed with four within-subjects factors: *who*, *where*, *when*, and *session*, each of which had two levels, *fixed* or *randomised*, using the general linear model in SPSS (SPSS Inc., 2006). The *session* factor was included to determine whether there were learning effects across the four sessions. Kolmogorov-Smirnov tests were used to confirm that the distribution of SRTs within each condition did not differ significantly from the normal distribution. Effect sizes were calculated by converting F-values to correlation values, denoted by r, as follows (Field, 2005):

$$r = \sqrt{\frac{F}{F + \mathrm{df}}} \tag{5.2.1}$$

where df is the residual degrees of freedom.

5.2.3 Results

Figure 5.3 shows group-mean and individual SRTs in each condition. A repeated-measures analysis of variance assessed the effects of the three cues (*Who, Where & When*) and learning (*Session*). The main effect of *who* was significant [$F(1,7) = 31.457$, $p < .01$, $r = .90$], but *where* [$F(1,7) = 3.515$, $p > .05$ ns, $r = .58$] and *when* [$F(1,7) = 0.001$, $p > .05$ ns, $r = .01$] were not. The blue bars in Figure 5.6 show the estimated mean benefit of each parameter. When *who* would speak was fixed, participants could hear out the colour-number co-ordinate at an SNR that was 1.9 dB less favourable than when *who* varied. The improvements from constraining where (1.0 dB) and when (0.01 dB) the target phrase would occur were not significant. None of the interactions was significant.

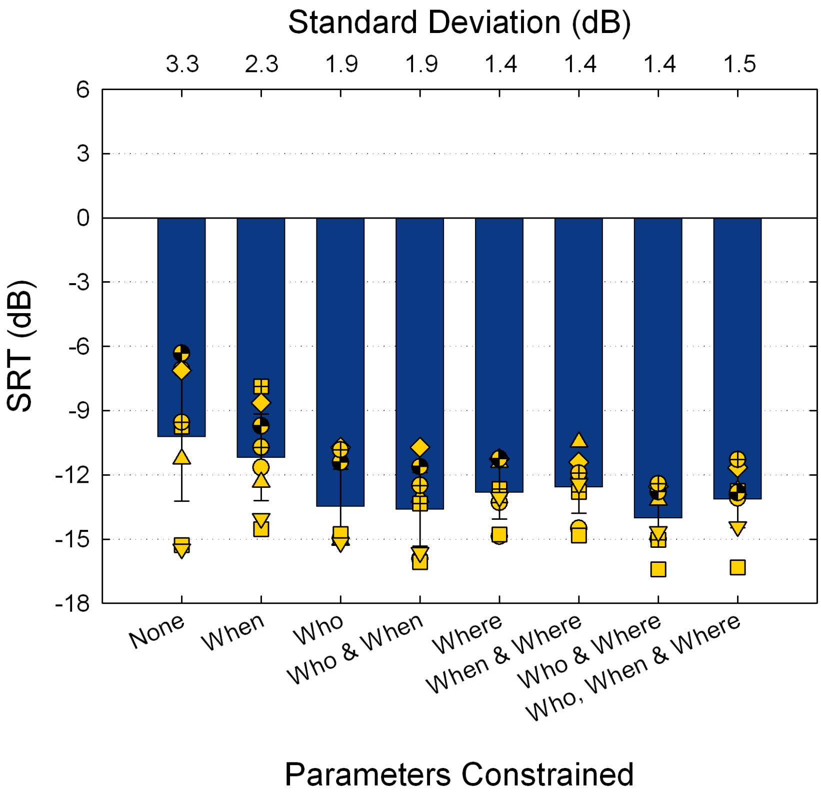

Figure 5.3. Group (bars) and individual (symbols) average speech reception thresholds (SRTs) for each condition in Experiment 1. Error bars show 95% confidence intervals.

5.2.4 Discussion

All participants could hear out information in target phrases that were less intense than the surrounding masker phrases. Individual SRTs ranged from -6 dB to -16 dB among participants and conditions. Group-mean SRTs ranged from -10 dB to -14 dB across conditions.

Fixing *who*, *where*, and *when* provided only small advantages over the condition in which the three parameters varied randomly. The results do not support the hypothesis that all three cues benefit listeners in hearing out information from the target phrase at lower SNRs. However, the results are compatible with the experience of participants. They reported that the onset of each phrase captured their attention, irrespective of who spoke, or where or when they spoke. This suggestion is congruent with previous visual research which observed that isolated abrupt onsets automatically capture attention and are allocated attentional resources (Yantis & Jonides, 1984). In the current task, the greater challenge was to sustain attention on the target talker when the target call-sign was detected and to resist distraction from subsequent masker phrases.

Possibly, the small significant advantage from constraining *who*, but not *where*, arose because knowledge of the identity of the target voice could be exploited more efficiently than knowledge of the location of the target voice in order to link the second part of the target phrase (‘go to COLOUR NUMBER now’) with the first part (‘Ready CALL-SIGN’). The absence of a benefit from knowing when the target phrase would be spoken supports the suggestion that the presentation of a target stimulus within a temporally-regular sequence of stimuli largely negates any effect of uncertainty about the target onset time (Leibold et al., 2005).

In summary, the results of Experiment 1 suggest that phrase onsets capture attention automatically by inducing an involuntary stimulus-driven shift in attentional focus. The strength of this effect may have rendered prior information about the target phrase unnecessary to the listener in performing the task.

5.3 Experiment 2

5.3.1 Introduction

If the onset of a voice captures attention, as was suggested by the subjective experience of participants in Experiment 1, then a masker phrase that starts at the same time as the target phrase should be particularly distracting. Knowledge of *who*, *where* and *when* might overcome that distraction. These hypotheses were tested in Experiment 2.

5.3.2 Methods

5.3.2.1 Participants

Eight paid listeners, one male and seven females, between the ages of 20–27 years (Mean age 23.3, $\sigma = 2.2$) participated in Experiment 2. All participants had lived in Britain or Ireland for at least 10 years and spoke English as their native language. Participants were confirmed to have hearing thresholds within normal ranges and no history of hearing health problems using the procedures outlined for Experiment 1.

5.3.2.2 Stimuli

The sequences of phrases presented in Experiment 2 were similar to those of Experiment 1, except that each time slot was occupied by a pair of phrases with simultaneous onsets (Figure 5.4). Each pair included one phrase spoken by a man and one spoken by a woman. The level of one phrase in the pair was 62.5 dB (A). In slots that did not contain the target phrase, the level of the other phrase was 52.5 dB (A). The target phrase occurred either in the 7^{th} slot, in conditions where *when* was fixed, or in one of the 4^{th} to 10^{th} slots, chosen randomly in conditions where *when* was varied. The masker phrase in the slot containing the target had a level of 62.5 dB (A).

Restrictions on the call-sign, talker, colour, and number keywords of the masker phrases which overlapped the target were applied as in Experiment 1. As the number of phrases in the sequence was greater than the number of loudspeakers available, it was not possible to present each phrase from a unique loudspeaker. Instead, each

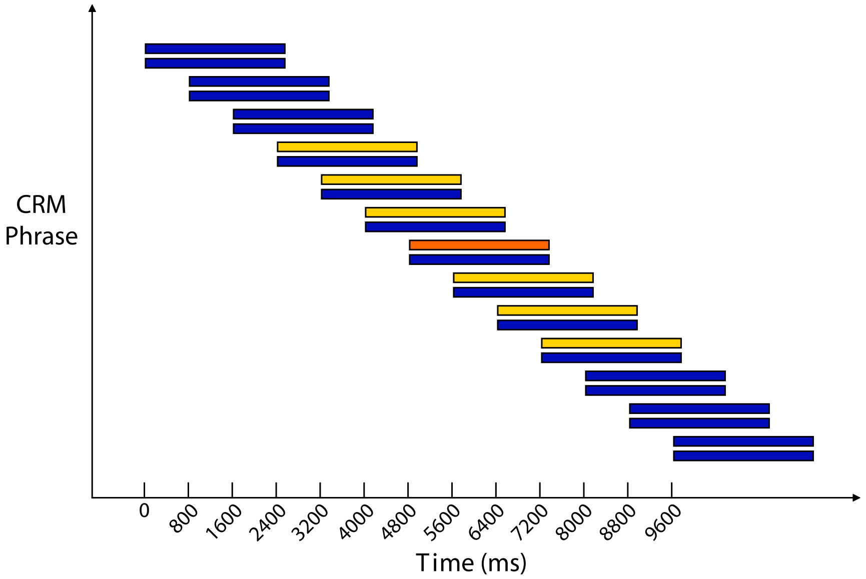

Figure 5.4. In Experiment 2, phrases were presented in pairs with simultaneous onsets and a 10-dB level difference within each pair that did not contain the target phrase. The target phrase appeared either in the centre slot (orange) or in one of the surrounding slots (yellow).

phrase within a pair was allocated a loudspeaker which was different from the other phrase in the pair, and from the phrases which immediately followed it.

5.3.2.3 Assessing the Increase in Energetic Masking

The addition of a second phrase to each time slot increased the amount of energetic masking of the target phrase as compared to Experiment 1. The increase was measured using a head and torso simulator (HATS) (Brüel & Kjær Type 4128C) with high-fidelity microphones in its ears (B&K Type 4158/9C). The HATS was placed in the centre of the array of loudspeakers, facing the loudspeaker at 0°. One hundred trials from Experiments 1 and 2 were presented with the target phrase omitted. The output of the microphones was digitally recorded at a sample rate of 44.1 kHz with 16-bit amplitude quantisation. Average RMS levels were measured for each ear in a 200-ms rectangular window centred on the moment when the target phrase would have started. An average for each ear across the 100 trials was calculated. The difference in this average between the two experiments was taken as an indication of the change in energetic masking at the onset time of the target phrase. Mean differences of 1.0 dB for the left ear and 1.3 dB for the right ear were found (Table 5.1).

5.3.2.4 Design & Procedure

The same 2 × 2 × 2 factorial design from Experiment 1 was used in Experiment 2. Participants completed four repetitions of each condition across four sessions. Each session was completed on a separate day and comprised 8 blocks of trials, one for each condition. The procedure for each trial was identical to that of Experiment 1.

5.3.2.5 Analyses

The data were subjected to a repeated-measures ANOVA with four within-subjects factors: *who*, *where*, *when*, and *session*. The analysis assessed the benefits of constraining each of the three cues, and whether learning effects were present across the four experimental sessions. To compare performance levels in Experiments 1

	Left Ear		**Right Ear**	
	Mean	σ	**Mean**	σ
Experiment 1	−25.9	2.6	−25.0	2.8
Experiment 2	−24.9	2.3	−23.7	2.3
Difference	1.0		1.3	

Table 5.1. RMS power measurements in dB (FS) for the phrase sequences from Experiments 1 and 2 obtained within a 200-ms window centred on the moment in time when the target phrase, which was omitted, would have started. Each value represents an average across 100 trials.

and 2, average SRTs for each participant for each condition were calculated across the four sessions. The average data was subjected to a repeated-measures ANOVA with within-group factors of *who*, *where*, and *when*, and the between-group factor of *experiment*. Interactions were assessed between each of the within-group factors and the between-group factor. This analysis examined whether there was a difference in the benefit that listeners received from prior information between Experiments 1 and 2. Effect sizes were calculated as for Experiment 1.

5.3.3 Results

Figure 5.5 shows group-mean and individual SRTs in each condition. To determine whether constraining the cues had significant effects on performance, and whether learning effects were present, a repeated-measures ANOVA was performed on the data with four within-subjects factors: *who*, *where*, *when*, and *session*. The main effects of *who* [$F(1,7) = 407.352$, $p < .001$, $r = .99$], *where* [$F(1,7) = 363.432$, $p < .001$, $r = .99$], and *when* [$F(1,7) = 31.121$, $p < .01$, $r = .90$] were significant. The effect of *session* was not significant [$F(3,21) = 0.561$, $p > .05$ ns, $r = .16$]. The interaction between *who* and *where* was also significant [$F(1,7) = 11.360$, $p < .05$, $r = .79$].

The yellow bars in Figure 5.6 show the estimated mean advantages from constraining each of the three parameters. The advantage from knowing when the target

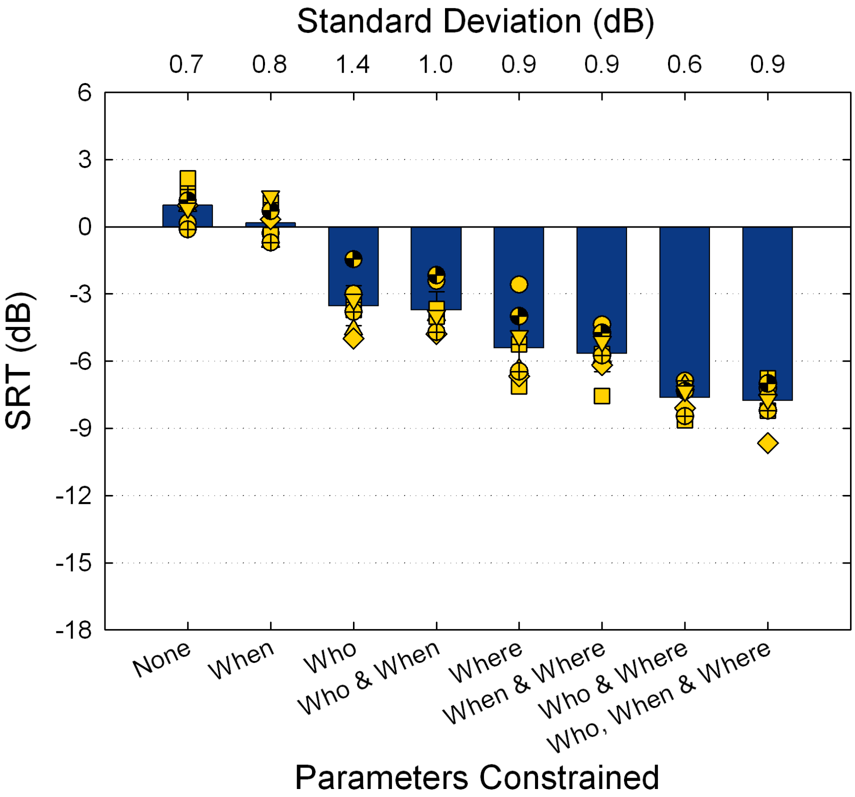

Figure 5.5. Group (bars) and individual (symbols) average speech reception thresholds (SRTs) for each condition in Experiment 2. Error bars show 95% confidence intervals.

phrase would occur was small (0.3 dB). Knowing who would speak the target phrase improved performance by 3.2 dB. Knowing the location of the target provided a larger benefit, amounting to an average improvement of 5.1 dB. The interaction between who and where arose because the benefit from constraining both cues together was less than the sum of the benefits from constraining the two cues individually.

A second ANOVA, with three within-subjects factors of *who*, *where*, *when*, and the between-group factor of *experiment* compared the results of Experiments 1 and 2. The main effects of *who* $[F(1,14) = 188.387,\ p < .001,\ r = .96]$, *where* $[F(1,14) = 56.786,\ p < .001,\ r = .90]$, *when* $[F(1,14) = 44.662,\ p < .001,\ r = .87]$, and *experiment* $[F(1,14) = 180.413,\ p < .001,\ r = .96]$ were significant. Significant interactions with *experiment* were observed for *who* $[F(1,14) = 12.976,\ p < .01,\ r = .69]$, *where* $[F(1,14) = 19.777,\ p < .01,\ r = .77]$, and *when* $[F(1,14) = 32.037,\ p < .001,\ r = .83]$.

5.3.4 Discussion

Performance improved when prior information was provided. These results support the hypothesis that knowing *who*, *where*, or *when* provided a benefit to the listener in overcoming the distracting effects of phrase onsets which were simultaneous with the target phrase. When listeners received no prior information about *who*, *where*, or *when*, target phrases had to be more intense than masker phrases for participants to identify the colour-number co-ordinate with an accuracy of 71%. The results support

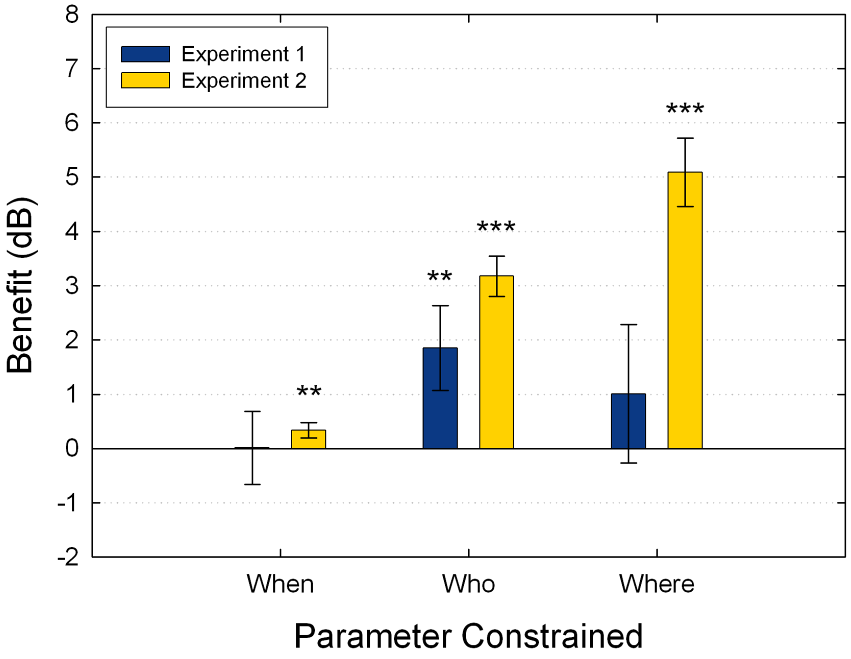

Figure 5.6. Estimated mean advantages from constraining *who*, *where*, and *when* in Experiments 1 and 2. Error bars show 95% confidence intervals (** $p < .01$, *** $p < .001$).

the hypothesis that a simultaneous phrase onset has a distracting effect and are compatible with the idea that attention is captured by the more intense of two voices that start at the same time. Indeed, the capture of attention was sufficiently complete to mean that participants contended that no target phrase had been presented on many trials when the target phrase was less intense than the masker phrase in the same time slot as the target.

The capture of attention by a concurrent masker was a sufficiently powerful effect to raise SRTs by 11 dB in the 'None' condition of Experiment 2 compared with the corresponding condition in Experiment 1. Two results show that this is an effect of attentional masking rather than energetic masking. First, the analysis of the energy in the stimuli showed that Experiment 2 involved only between 1 and 1.5 dB more energetic masking than Experiment 1. Second, the benefit from all three cues was significantly greater in Experiment 2 than in Experiment 1, as indicated by the significant interactions between each of the three cues and the *experiment* factor. While each new phrase onset captured attention in Experiment 1, largely eliminating the benefit from the three cues, knowledge of *who*, *when*, and *where* in Experiment 2 allowed listeners to control attention voluntarily and hence to overcome about 9 dB of the disadvantage.

The benefits that participants received from information about the target phrase differed between Experiments 1 and 2. This difference suggests that participants adopted different strategies in each experiment. In Experiment 1, it seemed to be most advantageous to use knowledge of the target talker's voice than knowledge of the talker's location to hear out the colour-number co-ordinate, indicated by the significant benefit from *who* but not *where* in Experiment 1. As discussed in Section 5.2.4, any benefit from knowing the location of the target phrase may have been largely negated by stimulus-driven shifts of attention induced by onset of each new phrase. In Experiment 2, the results suggest that it was most advantageous to set up a spatial filter to exploit knowledge of the target location than to set up a filter based on vocal characteristics to exploit knowledge of the target talker's voice, indicated by the larger benefit from *where* (5.1 dB) compared to *when* (3.2 dB) in Experiment 2.

The interaction between knowing who would speak and where they would speak in Experiment 2 possibly reflects the dominant use of one cue over the other. Knowing where the target would be provided a larger advantage than knowing who would speak the target when the cues were constrained individually. When both where and who were constrained, listeners may have allocated more attentional resources to focusing their attention at the known location of the target phrase, possibly because they found it easier to use information about where the target phrase was located as compared to who would say it, to focus their attention on the target phrase.

In Experiments 1 and 2, the main effect of *session* did not have a significant effect on performance, indicating that learning effects did not occur across sessions. This could be due to the training received prior to performing the task in the first session, and the warm-up trials at the start of the three subsequent sessions. Participants showed little difficulty in learning the task, and their performance levels had stabilised by the end of the training sessions.

5.4 Experiment 3

5.4.1 Introduction

The results of Experiment 2 raise a question about whether a masking phrase captures attention only when its onset is precisely synchronised with the onset of an accompanying target phrase, or whether there is a broader time window within which attention can be captured. Experiment 3 examined the extent of the distracting effect of the masker phrase by introducing an asynchrony between the target and paired masker phrases.

It is possible that the introduction of an asynchrony would provide a benefit to the listener in focussing their attention on the target phrase. Shafiro & Gygi (2007) presented multiple CRM phrases simultaneously, and found that increasing the number of target phrases that had to be monitored reduced the ability of participants to detect the target phrases and report the information they contained. This decrease in performance was attributed to an increase in attentional load. This result suggests that by reducing the number of phrases that have to be monitored simultaneously, attentional load is also reduced. The first hypothesis was therefore that introducing an asynchrony between the target and masker phrases would improve performance by reducing the attentional demands on the listener, and providing a possibility for the listener to overcome the distracting effects of a simultaneous phrase onset.

Experiment 3 examined asynchronies in which the target onset occurred before and after the masker onset, and sought to ascertain whether or not there was an advantage to the target occurring prior to the masker phrase even at small asynchronies. In the case that the target phrase started before the masker phrase, the results of Experiment 1 suggest that attention would be captured initially by the target phrase. Possibly, this initial capture would provide a benefit to the listener in identifying the target phrase, and in focussing attention on that phrase. In addition, asynchronies at which the target phrase would lead the masker phrase would provide listeners with a ‘glimpse’ at the target phrase during which it would not be subject to additional masking due to the paired masker phrase. Therefore, the second hypothesis was that performance would improve by a greater extent when the onset of the target phrase preceded the paired masker phrase.

5.4.2 Methods

5.4.2.1 Participants

Eight paid listeners, three males and five females, between the ages of 19–22 (Mean age 20.1, $\sigma = 1.0$) participated in Experiment 3. All participants had lived in Britain or Ireland for at least 10 years and spoke English as their native language. Participants were confirmed to have hearing thresholds within normal ranges and no history of hearing health problems using the procedures outlined for Experiment 1.

5.4.2.2 Stimuli

The conditions of Experiment 3 were derived from the 'None' condition of Experiment 1 in which each time slot was occupied by a single phrase and each of the three cues (*who*, *when*, and *where*) was varied randomly from trial to trial. In Experiment 3, a masker phrase, at the same fixed level of 62.5 dB (A) as the other masker phrases, was added to the time slot containing the target phrase. This masker phrase always started 800 ms after the onset of the previous masker phrase.

The onset of the target phrase was varied relative to the onset of this masker phrase to create nine asynchronies relative to the onset of the masker phrase: -320, -160, -80, -40, 0, +40, +80, +160, and +320 ms (Figure 5.7). Negative values indicate that the target preceded the masker. Asynchronies outside this range were not sampled because a negative asynchrony that is x ms greater than 400 ms is equivalent to a positive asynchrony of $400 - x$ ms, and *vice versa*. An additional condition in

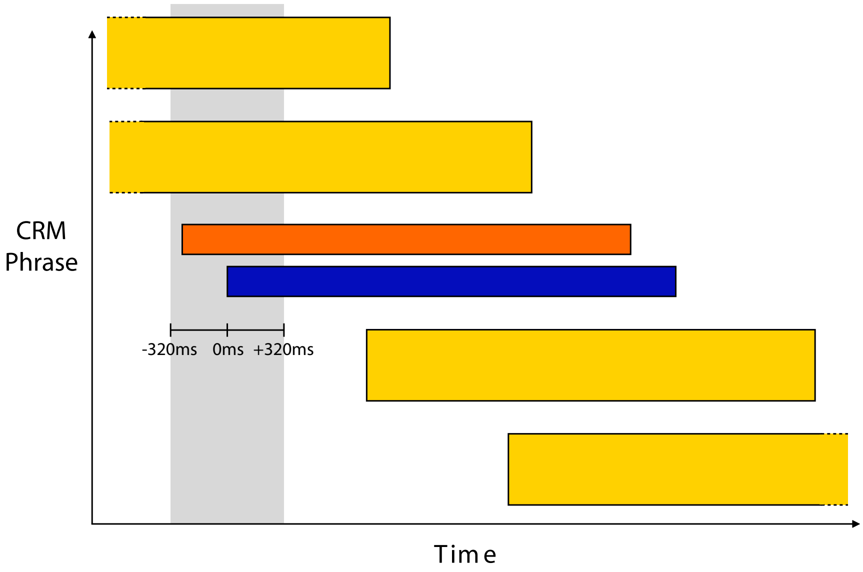

Figure 5.7. An expanded section of the sequence of phrases similar to that used in Experiment 1. The target phrase (orange) was paired with a masker phrase (blue) within a sequence of other masker phrases (yellow). The onset of the paired masker was always 800 ms after the previous masker in the sequence. The onset of the target phrase was varied so that it preceded or followed the masker onset by up to 320 ms.

which the paired masker phrase was omitted (equivalent to the 'None' condition in Experiment 1) was also included.

5.4.2.3 Training

Participants completed a single training block of the 0-ms asynchrony condition; i.e. when the target and paired masker phrase had simultaneous onsets at the start of each session.

5.4.2.4 Design

The 10 conditions were presented in an order that was partially counterbalanced across the eight participants by ensuring that each participant completed the conditions in a unique order. Each participant completed two repetitions of each condition across two sessions.

5.4.2.5 Procedure

The procedure was similar to that of Experiments 1 and 2. Participants were reminded of the experimental task after each block of trials. However, they were not informed as to whether the target phrase would occur before, simultaneously with, after the paired masker phrase, or alone.

5.4.2.6 Analyses

SRTs were calculated by averaging the two thresholds for each condition from the two sessions. The data were subjected to a repeated-measures ANOVA with a single within-subjects factor of *asynchrony*. This analysis did not include data from the condition in which the paired masker was omitted. The analysis examined whether introducing an asynchrony between target and masker phrase onsets would reduce the distracting effect of the masker and therefore lead to improved performance. Planned contrasts were performed between the 0 ms asynchrony condition and each of the other 8 levels of asynchrony. The contrasts assessed whether a larger benefit from asynchrony was received when the target phrase preceded the paired masker, or vice versa. Mauchly's test was used to confirm that the assumption of sphericity had been met. Omega squared (ω^2) was used to estimate the effect size for the *asynchrony* factor, and was calculated as (Field, 2005):

$$\omega^2 = \frac{\frac{k-1}{nk}\left(\mathrm{MS}_M - \mathrm{MS}_R\right)}{\mathrm{MS}_R + \frac{\mathrm{MS}_{BG}-\mathrm{MS}_R}{k} + \frac{k-1}{nk}\left(\mathrm{MS}_M - \mathrm{MS}_R\right)} \tag{5.4.1}$$

where k is the number of conditions in the experiment, n is the sample size, and MS_M, MS_R, and MS_{BG} are the mean squares for the model, the residual, and the between-group variance respectively. The effect sizes for each of the planned contrasts were calculated by converting the F-values in to correlation values, as described in Section 5.2.2.10.

A one-sample t-test was performed to test whether SRTs from the condition in which the paired masker phrase was omitted were significantly lower than 0 dB. This analysis examined whether participants could hear out information from the target phrase when it was less intense than the individual masker phrases which surrounded it in time. Independent-samples t-tests were performed to test 1) whether SRTs in that condition were not significantly different to the equivalent condition, 'None', in Experiment 1, and 2) SRTs in the 0 ms asynchrony condition were not significantly different from the equivalent condition, 'None', in Experiment 2. Effect sizes for the t-tests were calculated by converting the t-values to correlation values, r, as follows (Field, 2005):

$$r = \sqrt{\frac{t^2}{t^2 + \mathrm{df}}} \tag{5.4.2}$$

where df is the degrees of freedom. Bonferroni corrections were used to control for the inflation of the family-wise error rate due to the calculation of multiple comparisons based on the data from the condition in which the paired masker phrase was omitted, and are denoted by p_{bf}.

5.4.3 Results

Figure 5.8 shows group-mean and individual SRTs in each condition. In the asynchrony conditions, the average SRT was +0.5 dB and varied by less than ±0.6 dB as a function of the size of the asynchrony. A repeated-measured ANOVA with a single within-subjects factor of *asynchrony* showed that main effect of asynchrony was not significant [$F(8,56) = .605$, $p > .05$, $\omega^2 = 0$]. As the main effect was not significant, the planned contrasts were not performed.

Thresholds in the 0 ms asynchrony condition (Mean SRT 0.7 dB) did not differ significantly from the 'None' condition in Experiment 2 (Mean SRT 1.0 dB) [$t(14) = .584$, $p > .05$ ns, $r = .15$]. The high thresholds did not arise because listeners were unable to hear out target phrases at negative target-to-masker ratios: the group-mean SRT was significantly lower than 0 dB, at -6.6 dB, when the additional masker was omitted in the condition labelled 'alone' in Figure 5.8 [$t(7) = -5.294$, $p_{bf} < .01, r = .89$]. In the 'alone' condition, performance did not differ significantly from the 'None' condition of Experiment 1 (Mean SRT -10.2 dB) [$t(14) = -2.041$, $p_{bf} > .05$ ns, $r = .48$].

5.4.4 Discussion

The hypothesis that introducing an asynchrony between a target and a paired masker phrase would improve performance by reducing attentional load was not supported by the results. The failure of asynchronies as great as ±320 ms to reduce thresholds suggests either that the time window over which the onset of a masker can capture attention is broad, or that a second effect counteracted a beneficial effect of asynchrony.

Subjectively, the asynchrony conditions of Experiment 3 were challenging for listeners because the targets started at times that were out of synchrony with the regular rhythm of phrase onsets. Given that the target could occur in any of the 4^{th} to the 10^{th} time slots, listeners did not know when to listen for the target. Possibly, the disadvantage of not knowing when to listen counteracted the advantage of asynchrony. If that explanation is correct, then an advantage for asynchrony would be found if the timing of targets and paired maskers was reversed, with the target starting in synchrony with the regular rhythm of phrase onsets and the onset of the masker being varied in relation to the onset of the target. This hypothesis was tested in Experiment 4.

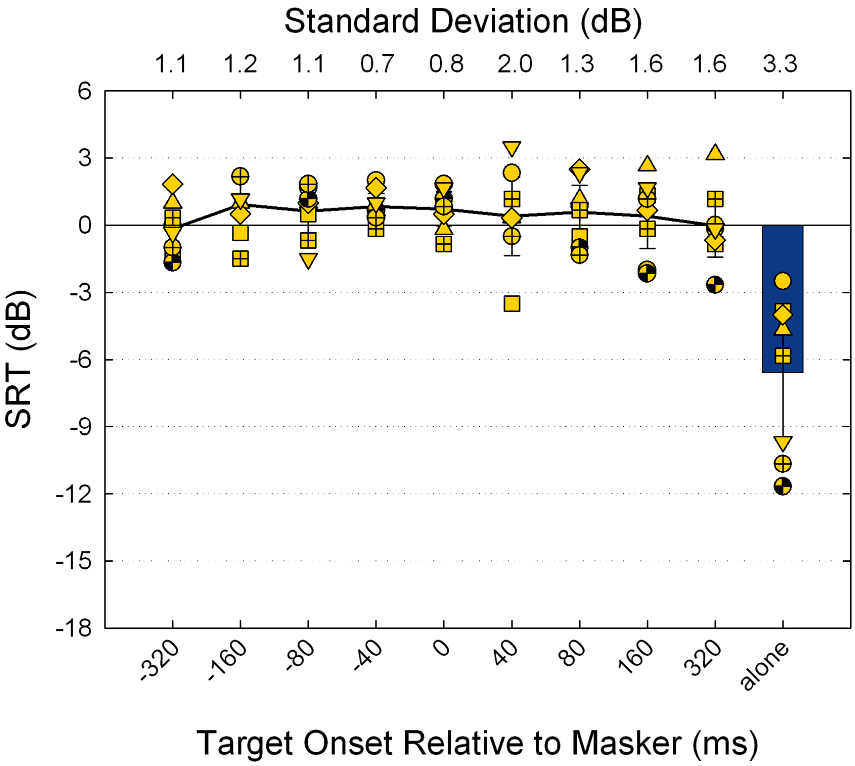

Figure 5.8. Individual (symbols) and group (solid line) SRTs for a range of target-masker asynchronies. Negative values indicate the target onset occurred before the masker phrase onset. In the ‘alone’ condition (bar) the masker was omitted making it equivalent to the ‘None’ condition of Experiment 1. Error bars show 95% confidence intervals.

5.5 Experiment 4

5.5.1 Introduction

The results of Experiment 3 suggested that phrases which were not synchronised with the regular rhythm of phrase onsets were less successful at capturing attention. In Experiment 4, an asynchrony was introduced between the target and masker phrases while keeping the target phrase in synchrony with the regular sequence of phrase onsets. The first hypothesis was that the distracting effect of the masker phrase would be reduced by moving the masker phrase out of synchrony with the sequence of phrases.

If the target-masker asynchrony resulted in the masker phrase following the target phrase, a larger benefit may be experienced by the listener as attention would be captured by the target phrase. The second hypothesis was therefore that a larger benefit from a target-masker asynchrony would be observed when the masker phrase followed the target phrase, compared to when the masker phrase preceded the target phrase.

5.5.2 Methods

5.5.2.1 Participants

Nine paid listeners, two males and seven females, between the ages of 19–25 (Mean age 21.2, $\sigma = 2.2$) participated in Experiment 4. All participants had lived in Britain or Ireland for at least 10 years and spoke English as their native language. Participants were confirmed to have hearing thresholds within normal ranges and no history of hearing health problems using the procedures outlined for Experiment 1. One of the participants had previously taken part in Experiment 2. Two other participants had prior exposure to the stimuli.

5.5.2.2 Stimuli

The design was similar to Experiment 3, except that an asynchrony between the onsets of target phrases and paired masker phrases was introduced by displacing the masker phrases in time (Figure 5.9). Therefore, the target phrase maintained its temporal position in the sequence of phrases, 800 ms after the onset of the previous phrase in the sequence. Nine conditions with asynchronies of -640, -320, -160, 0, 160, 320, 480, and 640 ms were created. An additional condition in which the paired masker phrase was omitted (equivalent to the 'None' condition in Experiment 1) was also included.

5.5.2.3 Training

Participants completed a single block of trials from the 0 ms asynchrony condition prior to the experimental conditions.

5.5.2.4 Design

The 10 conditions were presented within a single session in an order that was partially counterbalanced across the 9 participants by ensuring that each participant completed the conditions in a unique order. Each participant completed a single block in each condition.

5.5.2.5 Procedure

The procedure was identical to that of Experiment 3. Participants were reminded of the experimental task after each block of trials but not provided with information about the target-masker asynchrony or whether the paired masker would be omitted.

5.5.2.6 Analyses

A repeated-measures ANOVA with a single within-subjects factor of *asynchrony* was performed. The condition in which the paired masker was omitted was not included in this analysis. Mauchly's test was used to confirm that the assumption of sphericity had been met. The analysis examined the effect of introducing an

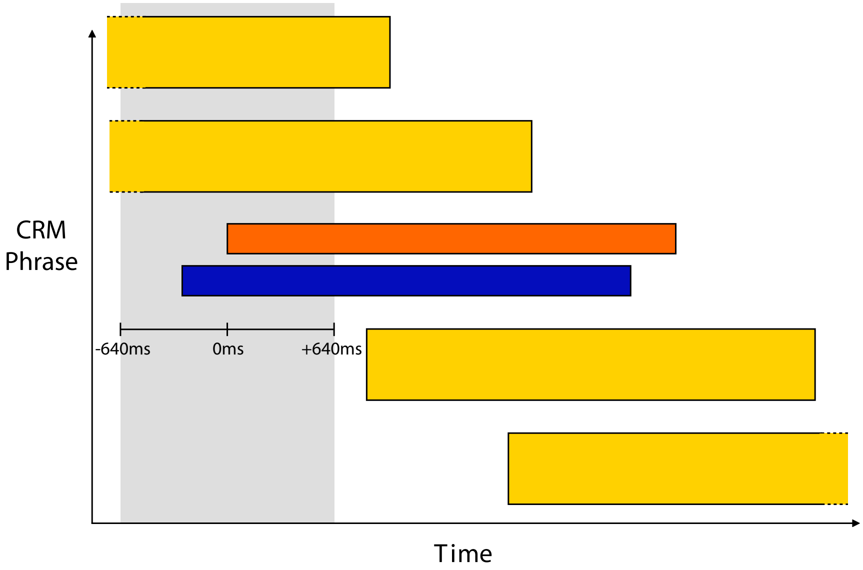

Figure 5.9. As in Experiment 3, the target phrase (orange) and a paired masker phrase (blue) are both allocated to the same time slot within a sequence containing 21 other phrases (yellow). In Experiment 4, the onset of the target phrase was aligned with the start of the time slot and the onset of the masker was varied. The maximum onset asynchrony between the target and paired masker phrases was 640 ms.

asynchrony between the target and paired masker phrase on SRTs. Planned contrasts were performed between the 0 ms asynchrony condition and each of the other 8 levels of asynchrony. The contrasts indicated which levels of asynchrony provided a significant reduction in the distracting effect of the paired masker relative to when it was simultaneous with the target phrase, and examined whether listeners received a larger benefit from asynchrony when it resulted in the target phrase preceding the paired masker, compared to when it followed it. Effect sizes for the main AVNOVA and the contrasts were calculated as described for Experiment 3 (Section 5.4.2.6).

Independent-samples t-tests were used to compare 1) performance in the 0 ms asynchrony condition to the equivalent conditions in Experiments 2 and 3, and 2) performance when the paired masker phrase was omitted to the equivalent conditions in Experiments 1 and 3. This analysis confirmed that performance was consistent across experiments when the task was identical. Independent-samples t-tests were also used to compare performance in the +320 ms asynchrony condition in the current experiment and the equivalent condition (-320 ms) in Experiment 3. This analysis directly contrasted the benefit from the target phrase preceding the masker phrase by 320 ms when the target was in synchrony with the sequence of phrase onsets (Experiment 4) and when it was out of synchrony (Experiment 3).

Bonferroni corrections were used where multiple comparisons were performed based on the same performance data, and are denoted by p_{bf}. Welsh's t-test was used for cases where equality of variance could not be assumed, denoted by non-integer degrees of freedom. Effect sizes for the t-tests were calculated by converting the t-values to correlation values using the same method described for Experiment 3.

5.5.3 Results

Figure 5.10 shows group-mean and individual SRTs. A repeated measures ANOVA was performed on the conditions containing the paired masker with a single within-subjects factor of asynchrony. Mauchly's test indicated that the assumption of sphericity had been violated [$\chi^2(35) = 58.068,\ p < .05$] therefore Greenhouse-Geisser estimates of sphericity were used to correct the degrees of freedom ($\hat{\varepsilon} = .32$). Asynchrony had a significant effect on performance [$F(2.53, 20.26) = 10.061,\ p < .001,\ \omega^2 = .30$]. Planned contrasts between the 0 ms asynchrony condition and each of the other asynchrony conditions revealed that performance improved significantly when the masker preceded the target phrase by 480 ms [$F(1,8) = 6.646,\ p < .05,\ r = .67$] or 640 ms [$F(1,8) = 19.890,\ p < .01,\ r = .84$]. When the masker onset followed the target onset performance improved significantly with asynchronies of 160 ms [$F(1,8) = 9.140,\ p < .05,\ r = .73$], 320 ms [$F(1,8) = 25.360,\ p < .01,\ r = .87$], 480 ms [$F(1,8) = 31.705,\ p < .001,\ r = .89$], and 640 ms [$F(1,8) = 50.000,\ p < .001,\ r = .93$].

The 0-ms condition produced the highest SRT (+0.4 dB) which was similar to the corresponding condition in Experiment 3 (+0.7 dB) [$t(15) = -0.470$, $p_{bf} > .05$ *ns*, $r = .12$] and to the 'None' condition in Experiment 2 (+1.0 dB) [$t(11.1) = -0.883$, $p_{bf} > .05$ *ns*, $r = .26$]. When the paired masker was omitted, performance was similar to the equivalent conditions in Experiment 3 ('alone' in Figure 5.8) [$t(15) = -0.860$, $p_{bf} > 0.05$ *ns*, $r = .22$] and Experiment 1 ('None' in Figure 5.3) [$t(15) = 0.649$, $p_{bf} > 0.05$ *ns*, $r = .17$] (Figure 5.11).

The benefit from the target phrase preceding the paired masker phrase by 320 ms was significantly larger in the current experiment (3.7 dB) compared to the equivalent condition in Experiment 3 (0.9 dB) [$t(15) = -3.341$, $p < .05$, $r = .65$].

5.5.4 Discussion

In contrast to Experiment 3, participants benefited from the introduction of an asynchrony between the onsets of the target and paired masker phrases. This result confirmed the hypothesis that moving the masker phrase out of synchrony with the sequence of phrase onsets would provide a benefit to the listener in focussing attention on the target phrase. The difference between the experiments was evident at the asynchrony common to both experiments (320 ms). A larger benefit was provided

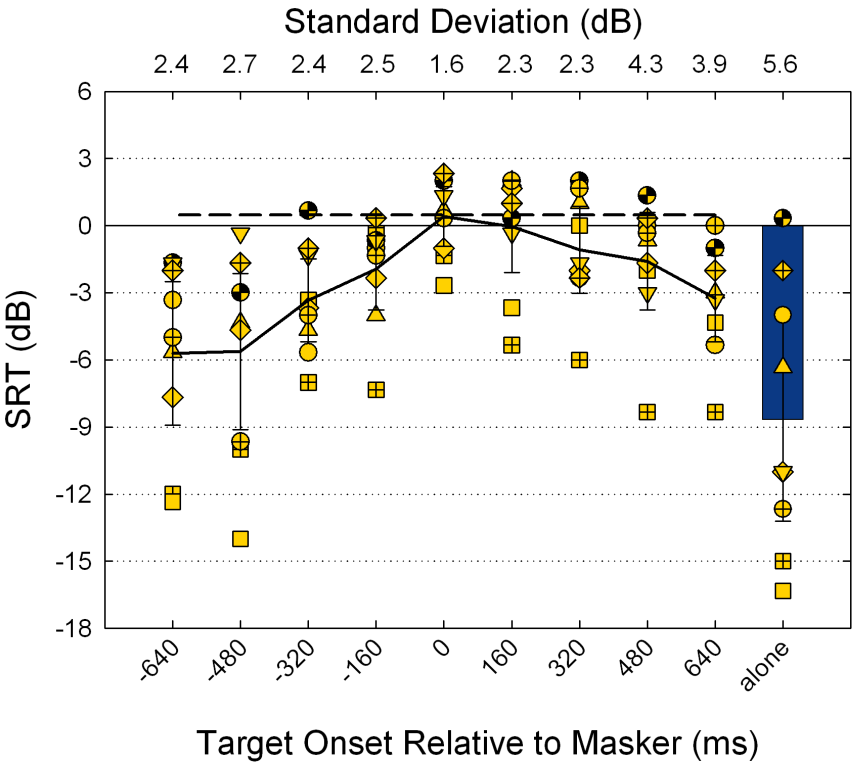

Figure 5.10. Individual (symbols) and group (solid line) target-masker ratios for a range of masker-target asynchronies. Negative values indicate the masker onset occurred before the target phrase onset. The dotted line shows mean performance across the asynchrony conditions in Experiment 3. In the 'alone' condition (bar) the masker was omitted making it equivalent to the 'None' condition of Experiment 1. Error bars show 95% confidence intervals.

on average in Experiment 4 (3.1 dB) when the target preceded the masker phrase by 320 ms (+320 ms condition) compared to the equivalent condition (-320 ms) of Experiment 3 (0.9 dB). The only difference between these two conditions was that the masker was synchronised with the phrase sequence in Experiment 3 while the target phrase was synchronised in Experiment 4. The results suggests that the contrast between Experiments 3 and 4 was due, at least in part, to differences in the predictability of the onset time of the target phrase.

The results suggested that the distracting effect of the masker phrase varied based on whether it preceded or followed the target phrase. Delaying the masker phrase by 160 ms was sufficient to improve performance significantly (2.3 dB). When the masker preceded the target phrase, an asynchrony of 480 ms was required to improve performance (2.0 dB). This result supported the hypothesis that a larger benefit would arise in conditions where the target phrase preceded the masker phrase.

This difference may have arisen for two reasons. First, it is possible that the process of focusing attention on the target phrase could have been completed more rapidly than the process of disengaging attention from the masker phrase, when it

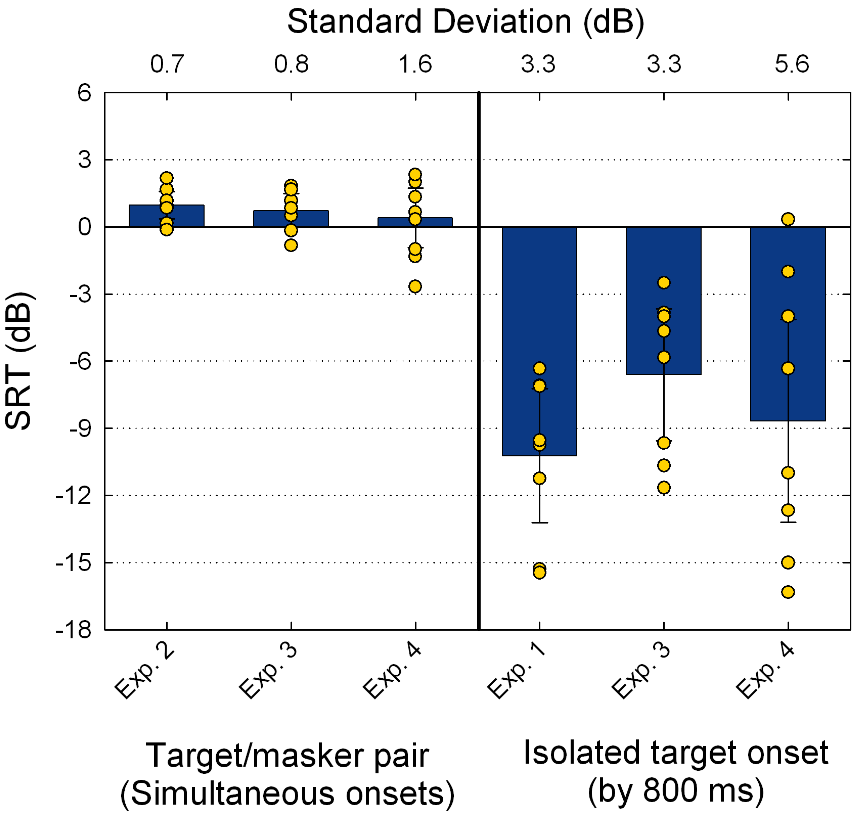

Figure 5.11. Group-mean (blue bars) and individual (yellow symbols) SRTs for the three conditions in which the onsets of the target and a masker phrase were simultaneous (left half) and conditions in which the target phrase was maximally isolated in time from the onset of a masker phrase (right half). The three bars on the left correspond to the 'None' condition in Experiment 2 and the '0 ms' conditions in Experiments 3 and 4. The three bars on the right refer to the 'None' condition in Experiment 1 and the 'Alone' conditions in Experiments 3 and 4. Error bars show 95% confidence intervals.

preceded the target. The results of Experiment 1 suggested that the onset of a new phrase captures attention involuntarily. In the visual domain, Wolfe et al. (2000) have found that voluntary attentional shifts can be up to an order of magnitude slower than involuntary shifts, and Lachter et al. (2004) have suggested that involuntary shifts of attention can take less than 100 ms whereas voluntary shifts can require between 150-500 ms. These findings suggest that the voluntary process of disengaging attention from one talker and refocusing it on another would require a much longer time than an involuntary stimulus-driven shift in attentional focus towards a new phrase. Thus, the absence of a significant benefit from the masker phrase preceding the target phrase at shorter asynchronies (160 and 320 ms) may suggest that a longer period of time was necessary for the listener to disengage attention from the masker phrase and subsequently attempt to refocus attention on the target phrase, compared to their attention being captured by the onset of the target phrase. This difference in the time required to attend to the target phrase may have contributed to the asymmetry in the distracting effect of the masker phrase.

Another possible explanation for the asymmetry is that the length of time necessary for participants to determine whether a new phrase was the target or a masker meant that their attention was still focussed on the masker phrase while the target call-sign was being spoken. Thus, even if they disengaged attention from the masker phrase, the identification of the target phrase may have been impossible as the target call-sign was no longer audible.

The results of Experiment 4 suggest that even a brief period of exposure to a target phrase (160 ms) provides a benefit to the listener in focusing and sustaining their attention on that phrase. This finding was similar to the effect observed in the spatial domain by Allen et al. (2008) where an initial exposure to a target phrase at a spatially distinct location was beneficial in hearing out information within the target phrase, even when it was subsequently collocated with masker phrases. Both effects can be accounted for by a more general effect of selective attention mechanisms as outlined by Allen et al. (2008); i.e. that the opportunity to extract cues or features from a target phrase prior to interference from a masker phrase (within the time domain in the case of the current study) provides a benefit in selectively attending to the target phrase even in the presence of the masker phrase.

The introduction of the single paired masker (0 ms condition) produced performance levels similar to that of Experiment 2 (‘None’ condition), despite the higher number of simultaneous talkers present in that experiment. The dramatic effect of a single masker provides further evidence that the difference in performance levels of 11.2 dB between the ‘None’ conditions of Experiments 1 and 2 was not the product of an increase in the level of energetic masking, but rather due to the attentional effects of the paired masker. In Experiments 3 and 4, this performance difference could not be completely eliminated by introducing an asynchrony between the onsets of the target and paired masker phrases, implying that simultaneous onsets is not a 'special' case in multi-talker listening but rather reflects a more general distracting effect of temporally-adjacent onsets.

5.6 General Discussion

5.6.1 Attentional Effects of Phrase Onsets

The results from Experiment 1 show that, regardless of what information about the target phrase was available, it was possible for participants to identify and hear the target phrase when it was presented at a level below that of the masker phrases. Studies in vision have shown that isolated abrupt stimulus onsets can capture attention unavoidably (Yantis & Jonides, 1984) but, if sufficiently spaced in time, can allow attention to be released and grabbed by each succeeding onset allowing for the identification of a target stimulus (Kahneman et al., 1983). In contrast, when a visual target and irrelevant distractor start at the same time, attention can be misdirected and then must be switched between the sources to identify the target. The need to switch attention introduces a 'filtering cost', which can significantly increase the time required to recognise the target (Kahneman et al., 1983). Yantis & Jonides (1984) demonstrated that the time necessary to identify a target letter within a group of letters is greater when the targets have gradual onsets, in comparison to when they have abrupt onsets, regardless of how many distractor stimuli were presented at the same time as the target. From this result, they suggested that isolated abrupt onsets automatically capture attention and are allocated attentional resources. These results from visual research provided further evidence for the assertion that attention was captured by the abrupt onset of new phrases. In Experiment 1, this stimulus-driven attentional shift largely invalidated the need to use information about *who*, *where*, and *when*, to focus attention on the target phrase.

Variability in performance was higher in Experiment 1 in all but one of the conditions than in the corresponding conditions from Experiment 2. In Experiment 2, listeners used prior information about the target phrase to focus their attention on the correct location, the correct voice, at the correct time, or a combination of the three to hear out the target phrase. This strategy led to improvements in performance over the condition in which no information was provided, and involved top-down control over attentional focus. In contrast, performance in Experiment 1 was suggested as being based on the ability of phrase onsets to capture attention involuntarily and the ability of participants resist distraction once attention was focussed on the target phrase. Thus, better performance may reflect the ability of participants to adopt 1) a diffuse, or non-focussed, attentional state, and 2) a highly-focussed attentional state.

In the visual literature, the ability of an abrupt onset to capture attention has previously been found to depend on the attentional state of the listener, and can be diminished if attention is focused on a different location than that of the onset. Yantis & Jonides (1990) cued participants to the location of a target letter prior to presenting the target and several distractor letters simultaneously. By presenting a perfectly valid cue before the onset of the stimuli, Yantis & Jonides (1990) found that the effects of abrupt onsets can be decreased through the use of highly focused attention on a location other than that of the onset. In other words, focused attention increases resistance to the attention-capturing effects of abrupt onsets. Possibly, the large individual differences that arose when no information was provided in Experiment 1 could indicate a failure of some participants to adopt an attentional state in which their attention could be captured automatically by new phrase onsets.

The results of Experiments 2 supported the finding from visual research that focussed attention affects the ability of abrupt onsets to influence the attentional state of the listener. When prior information was provided about the target phrase, participants were able to hear out the target phrase at lower SNRs compared to when no information was provided. The benefits from knowing information about the target phrase suggest that the information was used to focus attention on the target phrase, and to overcome distraction from the onset of masker phrases.

5.6.2 Temporal Uncertainty

Information about *when* a target phrase would occur provided little benefit in hearing out the colour-number co-ordinate in Experiments 1 and 2. In both experiments, there was a high degree of predictability about the onset of each phrase in the sequence relative to the previous phrase; i.e. that each phrase would occur 800 ms after the preceding phrase. Uncertainty about the position of the phrase within the sequence may not have affected performance due to the strong effect of the temporal regularity of the sequence of phrases. Leibold et al. (2005) examined the ability of participants to detect a target tone within a temporally-regular sequence of maskers which where either tones or noise bursts. No advantage was observed as arising from reducing uncertainty about *when* the target would appear within the sequence. The findings of Leibold et al. (2005) and the results of Experiment 1 suggest that the regularity of the sequence of stimuli was beneficial to the listener in overcoming any detrimental effect of uncertainty about when the stimulus would occur.

The contrast between Experiments 3 and 4 provides some evidence for the effects of temporal uncertainty. In Experiment 3, there was greater uncertainty about the onset time of the target as it was not synchronised to the regular sequence of phrase onsets. In Experiment 4, the target phrase was positioned within the regular sequence of phrases and it was the masker, paired with the target phrase, which was temporally-displaced from the sequence. Green & Swets (1966) suggested that the introduction of even a small amount of temporal uncertainty can result in a 3 dB decrease in detection performance for tones in noise. The benefit from reducing temporal uncertainty about the onset of the target phrase, obtained by directly contrasting conditions with the same degree of target-masker asynchrony in Experiments 3 and 4 (target leading by 320 ms), was found to be 3.1 dB. This benefit may suggest that the regular sequence of onsets, or perhaps the act of reducing uncertainty about when the target phrase would occur, provided a cue which aided the reorienting or reconfiguring of attention so that the attentional state of the listener was optimised for the onset of a new phrase. However, the use of different degrees of asynchrony in Experiments 3 and 4 may have contributed to the observed differences in SRTs. Possibly, the inclusion of larger asynchronies in Experiment 4 alerted participants to the presence of a target-masker asynchrony and provided a benefit to the listener in exploiting smaller degrees of asynchrony.

5.6.3 Energetic Masking

It is possible that the introduction of additional maskers in Experiment 2 increased the amount of energetic masking of target phrases. If so, the higher SRTs in Experiment 2 compared to Experiment 1 might be attributed to the fact that participants found it harder to hear the target phrase, rather than any difficulties related to focusing attention on it. However, while the average RMS level of the maskers at the onset of the target phrase differed by only 1dB between experiments, thresholds were lower on average by 12.6 dB in Experiment 2 compared to Experiment 1. More specifically, performance was 11.2 dB better in Experiment 2 in the condition where no prior information was provided about the target compared to the equivalent condition in Experiment 1. Additionally, performance levels in the conditions with a 0-ms asynchrony in Experiments 3 and 4 were similar to the 'None' condition of Experiment 2, despite the fact that Experiments 3 and 4 only differed from Experiment 1 by the addition of the single paired masker phrase. Thus, it is more likely that the differences in performance observed between experiments were caused by the differing attentional demands of the tasks, rather than any differences in energetic masking.

5.7 Conclusions

Knowledge of *where* a talker is located is beneficial when talkers start to speak at the same moment. It is largely unnecessary when talkers start to speak sequentially, as the benefit is outweighed by the attention-capturing effect of phrase onsets, providing those onsets are sufficiently isolated from each other in time. Familiarity with the target voice, or perhaps knowing the sex of the talker, is helpful in identifying it among other talkers, regardless of whether the onsets are simultaneous or not. The strong distracting effect of a phrase onset which is temporally adjacent to a target phrase is dependent on the degree of uncertainty about when the target phrase will be spoken. The benefits derived from information about a target phrase, and the presence of strong attentional effects, particularly in relation to temporally-adjacent phrase onsets, suggest that mechanisms related to attention play an important role when listening in a multi-talker environment.

5.8 Summary

- A multi-talker spatial listening task was developed to assess the benefits from knowing *who* would speak a target phrase, *where* it would be spoken from, and *when* it would be spoken.
- In Experiment 1, regardless of the prior information provided, participants were able to identify and hear out information in the target phrase even when it was presented at a less intense level than the maskers.
- When pairs of talkers started speaking simultaneously in Experiment 2, attention was captured by the more intense talker, requiring the target phrase to exceed the level of the masker phrases to receive attention.
- Three results show that this effect is largely one of 'attentional' masking, rather than energetic masking.
    1. First, the maximum difference in average thresholds between Experiments 1 and 2 was 11.2 dB, but was accompanied by a difference of only 1dB in the average level of the masker phrases.
    2. Secondly, thresholds improved significantly when uncertainty about the target phrase was reduced, by providing the listener with prior information about the target phrase.
    3. Finally, performance levels in Experiments 3 and 4, which contained 7 masker phrases, were comparable to those of Experiment 2, which contained 25 masker phrases.
- These results are compatible with the subjective experience of participants who reported that each new phrase onset captured attention automatically. The greater challenge was to resist distraction from the onset of the masker phrase that followed the onset of the target.
- The release from the attentional effects of a distracting masker phrase was found to be dependent on uncertainty about *when* the target phrase would be spoken.

- When uncertainty about the onset time of the target phrase was reduced by placing it within a temporally-regular sequence of phrase onsets, there was a significant improvement in performance when the target and masker onsets were separated by a small amount (< 200 ms). The window in which the masker phrase had a distracting effect, defined by those asynchronies which produced a 3 dB decrease in SRTs, was -640 ms to $+320$ ms relative the onset of the target phrase.
- When there was a high degree of uncertainty about the onset of the target phrase, obtained by desynchronising it from the regular sequence of stimuli, the introduction of a target-masker asynchrony did not provide a benefit in attending to the target phrase.

Chapter 6

Relationships among Age, Hearing Level, Attentional Abilities, and Performance on a Spatial Listening Task

The experiment reported in this chapter investigated the relationship between attentional ability as measured using an attentional test battery, hearing sensitivity, performance on a task of spatial listening involving multiple talkers, and self-reported difficulties with listening in everyday situations. Younger and older groups of normally-hearing adults were compared to expose differences in their ability to cope with speech-in-speech listening situations. Self-reported difficulties experienced in everyday listening situations were assessed using a questionnaire, and examined in relation to performance on a laboratory task of spatial listening. Greater self-reported difficulties in everyday situations were associated with poorer performance on the spatial listening task. This result suggests that the task was successful in placing demands on the participant similar to difficult listening situations in which older adults commonly experience difficulties in everyday life. Poorer performance on the listening task was related to poorer hearing sensitivity, and was observed among participants who scored less well on several subtests of the attentional battery. This included purely visual, purely auditory, and audio-visual tests of attention. Two models which which could account for the pattern of results are discussed.

6.1 Introduction

The results of Experiments 1 and 2 (Chapter 5) are compatible with the idea that attention can play an important role in the perception of speech when many people are talking at the same time. Speech-reception thresholds (SRTs) improved by up to

8.7 dB when purely informational cues were provided about the target phrase that participants were listening for within a sequence of similar phrases. This performance benefit was compatible with the notion that listeners can use prior information to focus or direct their attention in a more efficient or task-relevant manner. In the spatial listening tasks, the results indicated that the informational cues had two main effects: 1) to overcome the distracting effect of phrase onsets which are temporally-adjacent to the target phrase, and 2) to set up attentional filters based on knowledge of the voice characteristics of the talker who spoke the target phrase. The results of these attentional modulations were improved SRTs even in the most challenging of the listening tasks.

Older adults report difficulties in understanding what one person is saying when many other people are speaking at the same time (Gatehouse & Noble, 2004). Many studies have examined age-related differences in performance on speech-in-speech and speech-in-noise tasks (van Rooij & Plomp, 1992; Dubno et al., 2002; George et al., 2007; Humes, 1996). Studies which have examined large samples of the general population have identified age-related deficits in the ability to perceive speech in noise (Wiley et al., 1998), and that a large proportion of older adults ($\sim$ 40%) report experiencing difficulties with speech perception in the presence of noise (Davis, 1989). While many studies have linked poor speech perception in noise with decreased hearing sensitivity (e.g. Festen & Plomp, 1983; Helfer & Wilber, 1990; Jerger et al., 1989; van Rooij & Plomp, 1992), other studies have observed deficits in speech perception which could not be fully accounted for by hearing loss. Dubno et al. (1984), for example, reported age-related decreases in performance in speech-in-noise tasks which were independent of hearing loss, and suggested that the ability of adults to perceive speech against a background of speech can only be accounted for by considering both age and hearing sensitivity. George et al. (2007) reported that among normally-hearing listeners, performance on a purely visual task which measured the ability of participants to perceive text obscured by a grating explained a significant amount of variance in SRTs. Deficits in speech perception have also been associated with decreased working memory capacity (Humes et al., 2006; Pichora-Fuller et al., 1995).

Zekveld et al. (2007) reported that hearing loss was not a predictor of deficits in sustained attention or working memory, cognitive variables which have been associated with the ability to cope with complex listening tasks (Dubno et al., 1984; Gatehouse et al., 2006). This result provides further evidence that age-related decreases in performance on cognitively-demanding listening tasks cannot be fully accounted for by decreased peripheral sensitivity. A hypothesis arising from the results of these studies is that cognitive abilities play a significant role in the ability to perceive speech in noise. However, few studies have directly related several different cognitive variables to performance on speech-in-speech or speech-in-noise tasks.

The present experiment examined the relationships between attentional abilities, hearing sensitivity, and performance on a multi-talker spatial listening task. Two groups of participants, young and older adults, were recruited to examine age-related effects across these measures. Participants were selected to ensure that hearing levels in the older adult group were not outside the normal range, removing any large differences due to sensori-neural hearing loss. To assess the attentional factors that have been implicated in challenging listening tasks, a standardised attentional test battery was used. The *Test of Everyday Attention* (TEA) (Robertson et al., 1996) contains a wide range of attentional measures which examine the ability of participants to divide, switch, and selectively direct their attention. The battery includes purely visual tasks, purely auditory tasks, and cross-modal tasks. To examine the difficulties that older adult listeners report experiencing in everyday situations, participants also completed the *Speech, Spatial, and Qualities of Hearing scale* (SSQ) (Gatehouse & Noble, 2004).

The first hypothesis was that the older adults would experience more difficulty than the young adults in hearing out what one person is saying while many people are speaking at the same time. As a consequence, performance on the spatial listening task should be poorer amongst the older group. The results of Experiments 1–4 have shown that the task is successful at placing high attentional demands on the listener. Although the older group was selected to have hearing thresholds within normal ranges, the findings of previous studies that examined normally-hearing older adults (e.g. Dubno et al., 1984; Helfer & Freyman, 2008) predicted that they would experience more difficulty with the perception of speech in a multi-talker task. Specifically, two factors of the spatial listening task were likely to induce poorer performance in the older group compared to the young group. First, target and masker phrases comprise intelligible speech. Previous research has suggested older adults have greater difficulty in ignoring speech which contains meaningful words compared to background speech which is unintelligible (Tun et al., 2002). Second, masker phrases would include phrases spoken by talkers of the opposite gender as the talker that would speak the target phrase. An odd-sex distraction effect has been observed in older listeners, in which greater difficulty is experienced in ignoring information from talkers of a different gender to that of a target phrase (Helfer & Freyman, 2008).

The second hypothesis was that older adults experience cognitive deficits compared to the young adults, and would therefore perform more poorly on many of the tasks of attention compared to the young group. This hypothesis was based on three aspects of the TEA test battery. Firstly, several of the measures from the TEA are subject to time limits, or are scored based on the speed at which the participant can complete the task (Robertson et al., 1996). The general finding from ageing studies of an age-related decrease in the speed of processing (Kok, 2000; Verhaeghen & Cerella, 2002) and the finding that decreased processing speed contributes to poor performance on cognitive tasks (Salthouse, 1996) predicted that the ability of the older group to cope with tasks which involved a speed component would be compromised compared to the young group. Secondly, two of the tasks in the TEA are designed to assess the ability to switch attention, both of which also include a speed component. There is evidence that older adults exhibit a reliance on top-down attentional control (Madden, 1990). When switching attention, voluntary shifts initiated by top-down control have been found to be relatively slow, requiring 150–500 ms compared to involuntary shifts which have been found to take less than 100 ms (Lachter et al., 2004). Together, these results predict that the performance of older adults would be poorer than that of the young adults on tasks which require information to be processed quickly or require that the focus of attention be switched rapidly between different sources of information. Lastly, many of the TEA tasks are designed to place considerable demands on participants by requiring the use of numerous cognitive functions, including attention and working memory. Another general finding of ageing studies, referred to as the “complexity effect”, suggests that age-related deficits in performance are more prevalent in tasks which impose high cognitive demands (Kok, 2000). This “complexity effect” predicts that the older group will show deficits in those tasks which impose high cognitive demands.

The third hypothesis was that performance on the spatial listening task would be related to both hearing level and attentional abilities as measured by the TEA. Several studies have observed poorer speech reception ability with increased sensori-neural hearing loss (George et al., 2007; Peters & Moore, 1992). However, even amongst older adults with hearing loss, correlations have been found between individual differences in performance on an attention-demanding listening task and memory capacity (Humes et al., 2006). Both results predict that performance on a complex listening task, which places high cognitive demands on the listener, will be poorer in the older group compared to the young group. The present experiment explored whether a more diverse range of cognitive abilities, as measured by the TEA, would be related to performance on the spatial listening task.

6.1.1 Summary of Hypotheses

1. Normally-hearing older adults are poorer at hearing out what one person is saying when many other people are speaking at the same time.

2. Older adults would exhibit poorer performance on tasks designed to assess attentional ability which:

    - require the rapid processing of information
    - involve frequent switching of attention focus
    - make high cognitive demands

3. Deficits in the performance of older adults compared to young adults on a complex speech-perception task would be associated with individual differences in:
    - peripheral sensitivity
    - attentional ability

6.2 Methods

6.2.1 Participants

Twenty-four paid listeners participated, 12 young adults and 12 older adults (Table 6.1). The young adult group comprised 2 males and 10 females aged between 19–30 years (Mean age 22.1 years, $\sigma = 3.3$). The older adult group comprised 5 males and 7 females aged between 57–71 years (Mean age 62.7 years, $\sigma = 4.6$). All had lived in Britain or Ireland for at least 10 years, and spoke English as their native language. Participants had better-ear average (BEA) pure-tone sensitivity thresholds ≤ 20 dB HL and had a normal hearing health history (Appendix D). None of the participants had prior experience with auditory experiments or the stimuli employed in the current experiments.

6.2.2 Spatial Listening Tasks

A subset of the spatial listening tasks that were presented in Experiments 1 and 2 (Chapter 5) was included in the current experiment. To address whether older adult listeners would be susceptible to the distracting effects of a competing voice which

Young adults			Older adults		
Age	Gender	BEA	Age	Gender	BEA
30	F	1.25	67	M	18.13
19	M	3.75	61	F	10.00
22	F	0.00	65	F	16.25
21	F	−1.25	59	F	13.75
24	F	1.25	71	F	15.00
19	M	5.00	60	M	3.75
23	F	3.13	70	M	16.25
20	F	8.75	58	M	12.50
20	F	1.25	61	F	13.13
26	F	5.00	63	F	18.75
19	F	2.50	57	M	16.25
22	F	10.00	60	F	15.63

Table 6.1. Age, gender, and hearing sensitivity (BEA) values for each participant in the young and older adult groups.

started simultaneously with a target phrase ("2 by 2") as compared to when the target phrase started on its own ("1 by 1"), the 'None' conditions of Experiments 1 and 2 were included. In the 'None' conditions, the target phrase was identifiable only by its call-sign. To examine the size of attentional masking release that older adult listeners would gain from the use of information cues about the target phrase, the 'All' conditions of Experiments 1 and 2 were included. In the 'All' conditions, information about the location, talker, and position of the target phrase in the sequence of phrases was provided in addition to the target call-sign. The four selected conditions created a "2 by 2" factorial design with the factors *Paired* and *Cued*.

An adaptive procedure was used in all four conditions and was similar to that used in previous experiments (Table 6.2). The number of reversals in the second phase of the routine was reduced from 12 to 6 to reduce the time necessary to complete the listening tasks. Also, as the previous experiments had suggested that thresholds stabilised after 3–5 reversals. The apparatus used to present the stimuli as well as the method of collecting responses was identical to that of Experiments 1 and 2.

The 12 participants in each age group were divided into three subgroups and each subgroup was assigned a different target call-sign: 'Baron', 'Tiger', or 'Ringo'. Each participant was allocated one target call-sign which did not change across the four conditions. The same counterbalancing method was applied to each subgroup of 4 participants for the 4 experimental conditions. The order of the four conditions was counterbalanced across subjects using a first-order Williams design (Williams, 1949) calculated using the R statistical computing environment (R Development Core Team, 2008).

6.2.2.1 Training

To familiarise participants with the stimuli and the task, two training conditions were completed before the four experimental conditions. In one of the training conditions, no informational cues were provided, and in the other condition, cues about the location, talker, and temporal position in the sequence of the target phrase were provided. In addition, to familiarise the participants with the difference between conditions in which phrases started one at a time or when they started in pairs, one of the conditions was taken from the "1 by 1" task (Experiment 1) and the other from the

Rule	Adaptive Routine	Step size (dB)	Reversals
1	2-down, 1-up	6	3
2	2-down, 1-up	2	6

Table 6.2. Details of the adaptive procedure employed to estimate speech reception thresholds (SRTs) in the spatial listening task.

"2 by 2" task (Experiment 2). An example choice of training conditions was therefore: "1 by 1" 'All' and "2 by 2" 'None'. The order of the conditions was counterbalanced across participants.

6.2.3 Attentional Test Battery

The attentional ability of the participants was assessed using the *Test of Everyday Attention* (TEA) (Robertson et al., 1996). The TEA was designed to examine different aspects of attention: selective attention, sustained attention, attentional switching, and divided attention. The test battery comprises 8 subtests: 3 visual tests, 4 auditory tests, and 1 audio-visual test. All of the subtests are presented in real-world contexts and involve common materials, such as maps and telephone directories.

The *Map search* subtest (subtest 1) (Figure 6.1) is a visual search task in which participants are instructed to circle as many target symbols (see inset) as possible within a fixed time limit. This test yields 1 and 2 minute scores. The three *Elevator counting* tests (subtests 2, 3, & 5) involve tracking which floor an elevator is at by counting a series of acoustical tone bursts either in silence, in the presence of distracting tones of a different frequency, or with the aid of high- and low-frequency tones which indicate movement up and down respectively. Scores on the three tasks represent the number of sequences in which the participant correctly identified the final position of the elevator. The *Visual elevator* task (subtest 4) is a visual analogue of the former task, using symbols instead of tones to indicate movement. This task

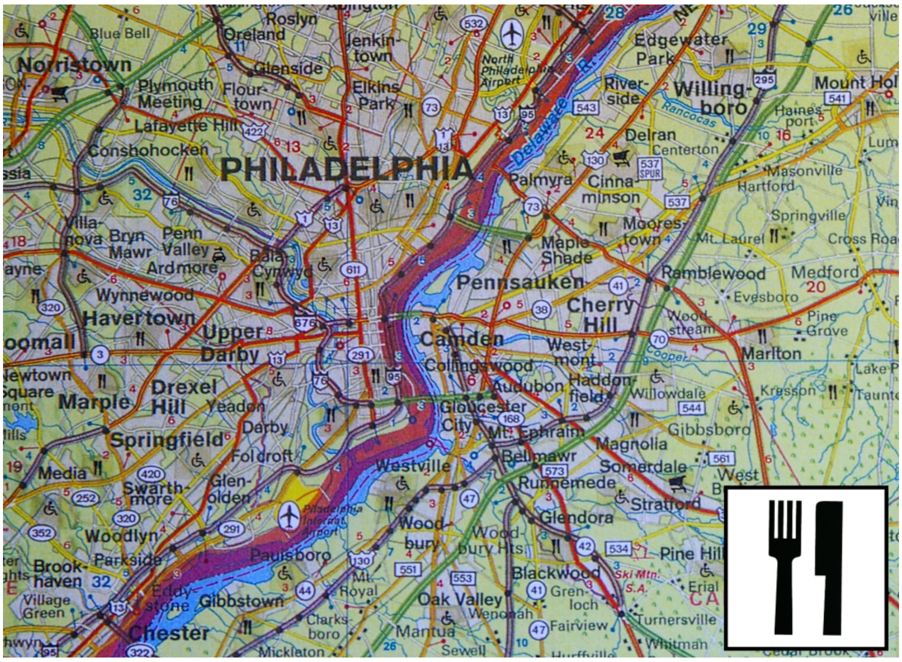

Figure 6.1. A subsection of the map used as the basis of the map search subtest of the TEA. Participants are instructed to circle as many target symbols as possible within 2 minutes. Inset: the target symbol for the version of the TEA that was administered. (Robertson et al., 1996)

yields the number of correct responses and the average time required for a participant to switch counting direction. Robertson et al. (1996) interpreted this switching cost as the ability to rapidly change attention from one task (counting upwards) to another (counting backwards).

In the *Telephone search* tasks (subtests 6 & 7), participants are presented with a page of a telephone directory in which each entry is accompanied by a pair of symbols, e.g. circles, squares, and stars. Their task is to search the page as fast as possible to identify pairs of matching symbols. The task is performed twice. First, in quiet, yielding an average time per item. Then, while doing a concurrent auditory counting task, which yields a dual-task cost in terms of a change in the average time per item compared to performance in quiet. Finally, the *Lottery* task (subtest 8) presents participants with a 10 minute radio broadcast of lottery ticket numbers. Their task is to identify the first two letters of winning tickets which they have been told end in the digits ‘55’. Scores on this task are expressed as the number of tickets for which at least one letter was identified in the correct position on the ticket. A summary of the subtests of the TEA is shown in Table 6.3.

Table 6.4 shows an alternative summary of the 10 measures from the 8 subtests of the TEA in terms of four cognitive variables: processing speed, working memory, attention switching, and resisting distraction. A task was categorised as involving *processing speed* if it was subject to a time limit, if participants were told to perform the task “as fast as possible”, or it was scored based on the speed at which a task could be performed; e.g. find as many symbols matching a target symbol in 1 minute (subtest 1A). A task was categorised as involving *working memory* if it required the participant to store information internally and update that information over the course of a task, e.g. tracking what floor an elevator was on as it moved upwards or downwards (subtests 4A/B and 5), or alternatively it required the simultaneous performance of two tasks (subtest 7). A task was categorised as involving *attention*

Subtest	**Modality**	**Measure**
Map Search	V	1A: Items found - 1 min
		1B: Items found - 2 min
Elevator Counting	A	2: Correct responses
Elevator Counting with Distraction	A	3: Correct responses
Visual Elevator	V	4A: Correct responses
		4B: Time to switch attention
Elevator Counting with Reversal	A	5: Correct responses
Telephone Search	V	6: Symbol pairs found
Telephone Search with Counting	A/V	7: Dual-task cost
Lottery	A	8: Lottery tickets identified

Table 6.3. The 8 subtests of the TEA. Tasks were either purely visual (V), purely auditory (A), or audio-visual (A/V).

switching if it required participants to switch attention, either between two streams of information or between two task sets; e.g. tracking the elevator movement *upwards* or *downwards* (subtest 5). Finally, a task was categorised as involving *resisting distraction* if it required the participant to actively ignore irrelevant information; e.g. irrelevant symbols and markings on a map (subtest 1A/B) or irrelevant high-frequency tones in an auditory counting task (subtest 3).

Three versions of the TEA subtests are available: A, B, and C. Multiple versions can be employed to avoid learning effects across repeated administrations. The test battery was administered once for each participant and version A was used exclusively.

6.2.4 Self-reported Difficulties

Self-reported difficulties with listening in everyday situations were elicited from participants using the *Speech, Spatial, and Qualities of Hearing Scale* (SSQ) (Gatehouse & Noble, 2004). The questionnaire was designed to assess hearing difficulties that may arise in a variety of listening situations, specifically those situations which are commonly experienced in everyday life. The SSQ comprises 52 questions divided into three sections, each focusing on a different domain of hearing: speech, spatial, and qualities of hearing. The speech section includes situations with a variable number of background talkers, types of background noise (e.g. environmental sounds, TV sounds, and speech), and contexts (restaurants, telephone conversations, quiet room). The spatial questions examine difficulties associated with the perception of distance, localisation, and movement of sound sources, e.g. dog barks, vehicles, and other people. Questions in the qualities of hearing section focus on the ability of participants to separate sources of sounds (e.g. voices and music) and recognise familiar sounds (e.g. voices). Questions also elicit responses about the amount of effort the listener expends when listening, and the naturalness and clarity of sounds such as music and everyday sounds.

Each SSQ question is accompanied by a visual-analogue scale ranging from 0–10. The extremes of each scale are labelled; e.g. ‘Not at all’ at 0 and ‘Perfectly’ at 10. Participants indicate their response by making a vertical line through the scale at any

	1A	**1B**	**2**	**3**	**4A**	**4B**	**5**	**6**	**7**	**8**
Processing speed	Y	Y	–	–	Y	Y	–	Y	Y	–
Working memory	–	–	–	–	Y	–	Y	–	Y	–
Attention switching	–	–	–	–	Y	Y	Y	–	Y	–
Resisting distraction	Y	Y	–	Y	–	–	–	–	–	–

Table 6.4. Summary of the 10 measures of the TEA described in terms of different cognitive variables which are likely to be involved (Y, yes; -, no).

point. An example question is shown in Figure 6.2 and the complete list of questions is listed in Appendix E.

Gatehouse & Akeroyd (2006) arranged the questions from the three sections of the SSQ into 10 sub-scales, each containing questions which examine similar listening situations. This approach was based on informed consideration of the questions rather than an objective technique such as factor analysis, and serves to reduce the total number of measures for each participant. Table 6.5 lists the questions which were allocated to each sub-scale. Scores on each of the sub-scales are calculated as the mean of the scores across the questions within the sub-scale.

6.2.5 Design

Participants were divided into young (≤ 30) and older adult (≥ 55) age groups with 12 participants in each group.

6.2.6 Procedure

Each participant completed the four parts of the experiment within a single session. The parts were always presented in the same order: audiogram, SSQ questionnaire, TEA battery, and the spatial listening task.

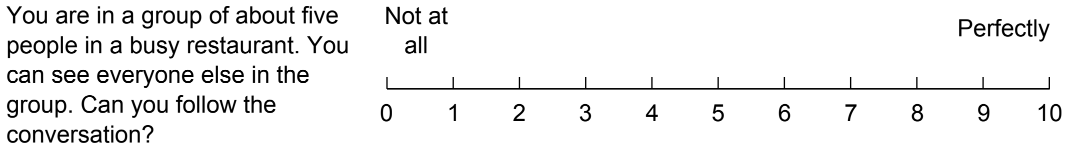

Figure 6.2. An example question from the speech section of the SSQ. Participants were instructed to make a vertical line through the scale to indicate their subjective experience.

	Sub-scale	Questions
1	Speech in quiet	Speech: 2, 3
2	Speech in noise	Speech: 1, 4, 5, 6
3	Speech in speech contexts	Speech: 7, 8, 9, 11
4	Multiple speech-stream processing and switching	Speech: 10, 12, 14
5	Localisation	Spatial: 1, 2, 3, 4, 5, 6
6	Distance and movement	Spatial: 7, 8, 9, 10, 11, 12, 13 ,15, 16
7	Sound quality and naturalness	Qualities: 8, 9, 10, 11, 12
8	Identification of sound and objects	Qualities: 4, 5, 6, 7, 13
9	Segregation of sounds	Qualities: 1, 2, 3
10	Listening effort	Qualities: 14, 18, 19

Table 6.5. The 10 sub-scales of the Speech, Spatial, and Qualities of Hearing scale.

6.2.6.1 Audiogram

An audiometric examination was performed in a sound-attenuated booth in accordance with BS EN ISO 8253-1 (BSA, 2004). Pure-tone thresholds were measured at octave frequencies from 250 to 8000 Hz inclusive for each ear.

6.2.6.2 SSQ

Participants completed the questionnaire in a quiet room and were given as much time as needed. Participants were encouraged to ask questions if the meaning or context of the questions were not clear to them.

6.2.6.3 TEA

The TEA was administered in the same sound-attenuated booth as the spatial listening tasks. The participant sat in front of a table containing the tape-recorder and the test materials. Instructions were given as recommended in the TEA test manual. All 8 subtests were completed in a single session.

6.2.6.4 Spatial Listening Task

The task was explained verbally to the participant prior to the presentation of the training conditions. After each training condition, participants were encouraged to ask questions if the task was not clear to them. If a participant had misunderstood any of the task instructions, the training conditions were repeated until they were comfortable with the task. Following training, the four experimental conditions were presented. If appropriate, the participant was provided with informational cues about the target phrase prior to the start of the condition. The participant was not told whether a “1 by 1” or “2 by 2” condition would be presented. Feedback was provided on each trial for each component of the response; i.e. both colour and number. The training conditions and the four experimental conditions were completed in a single session.

6.2.7 Analyses

6.2.7.1 Hearing Sensitivity

Better-ear averages (BEAs) were calculated for each participant from the results of the audiometric examination. Four-frequency averages (FFA) were calculated for each ear as the mean threshold level across octave frequencies from 500 to 4000 Hz inclusive. The lower of the two FFAs was designated as the BEA.

Age-related differences in hearing levels were assessed using independent-samples t-tests. The Kolmogorov-Smirnov (K-S) test was used to confirm that the distribution of hearing levels in both age groups was not significantly different from the normal distribution. Levene's test was used to check the assumption of equality of variance. Effect sizes were calculated by converting t-values into correlation values, denoted by r (Chapter 5, Section 5.4.2.6, p. 120).

6.2.7.2 Spatial Listening Task

Speech reception thresholds (SRTs) were estimated for each of the spatial listening tasks using the method employed in Experiments 1 and 2 (Chapter 5). The thresholds were calculated from the data collected in the second phase of the adaptive routine. The 6 reversals in the direction of change in the target level produced 6 'mid-run' estimates; 3 odd and 3 even. The odd mid-run estimates were averaged to arrive at an estimate of the threshold for each participant (Levitt, 1971). To examine consistency of performance, the standard deviations of the mid-run estimates were also calculated. The distribution of thresholds within each condition of the spatial listening task were confirmed to not be significantly different from the normal distribution using K-S tests.

The resulting SRTs were analysed using a repeated-measures ANOVA. The analysis addressed three questions. The first question was whether participants could use prior information about the target phrase to 'hear out' the target key-words at lower SNRs compared to situations in which they had no information about the target phrase apart from its call-sign (main effect of *Cueing*).

Secondly, the analysis sought to examine whether there was a detrimental effect of talkers starting to speak in pairs rather than one at a time (main effect of *Pairing*). The analysis examined whether the attentional effects observed between Experiments 1 and 2 were also observed in a repeated-measures design.

The third aim of the analysis was to ascertain whether there was difference between young and older adult groups. The ANOVA was used to examine whether there was a general effect of ageing (between-subjects factor of *Group*), age-related differences in susceptibility to distraction by a simultaneous phrase onset (*Group* × *Pairing* interaction), and in the older adult participants' ability to take advantage of information cues (*Group* × *Cueing* interaction).

Effect sizes for the ANOVA were calculated by converting F-values into correlation values, denoted by r (Chapter 5, Section 5.2.2.10, p. 110).

To examine the relationships between performance on the spatial listening task and scores on the SSQ and TEA, it was desirable to reduce the performance scores across the four conditions to a single measure. The results of the previous experiments suggested that performance across the different listening tasks was strongly related; i.e. participants who performed poorly in one condition generally performed poorly in other conditions. Therefore, the behavioural data from the four spatial listening tasks were subjected to an exploratory factor analysis to evaluate whether the data could be reduced to a single latent factor.

Before carrying out the factor analysis, the data were examined to assess whether all assumptions of the analysis were satisfied. K-S tests were used to test whether the data for each task were not significantly different from the normal distribution, and box-plots were used to identify outliers or extreme data points. Scatter plots between each of the performance measures were created to examine whether there were linear relationships between the variables. Principal Component Analysis (PCA) was used to extract the factors and varimax rotation was requested should more than a single factor be extracted. Only those factors with eigenvalues greater than 1 were accepted.

6.2.7.3 TEA

Norm-referenced scores are available for the 8 subtests of the TEA based on a sample of 154 normal volunteers broken down into four age groups (Robertson et al., 1996). Scores from the young adult group (19–30) were compared to the normative sample between the ages of 18–34, and the scores of the older adult group (57–71) were compared to the normative sample between 50–64 years of age. Normative scores at the 50% percentile were selected in both age ranges.

Visual inspection of boxplots was used to identify possible outliers in the data. If a data point was suspected as being an outlier, the score on the relevant measure was standardised and a z-score threshold of 3, corresponding to the two-tailed 99.9% percentile of the normal distribution of scores, was used to confirm the presence of an outlier. Data points which exceeded the threshold were excluded from all analyses.

To assess whether there were differences between the young and older adult groups in their scores on the TEA, the Mann-Whitney U statistic was chosen due to the non-normal distribution of scores on many of the subtests. The data were examined to see if the assumptions of homogeneity of variance and similarity in the shapes of the distribution of scores between the variables to be compared had been violated (Sheskin, 2003). Levene's test based on the median was used to assess homogeneity of variance across the two age groups for each of the measures. K-S Z was used to test the assumption that the distributions of scores within the two age groups to be compared did not differ significantly from each other. Because the Z statistic is sensitive to differences in the location of variables, the scores on all tests were first centred by subtracting the mean score for each group from the raw scores within that group. If assumptions had not been violated, the U statistic was then calculated on the original data.

If the homogeneity of variance assumption was violated for a measure but the shape of the distribution of scores in each age group did not differ significantly as revealed by the K-S Z test, the Westenberg-Mood median test was used to assess whether the medians of the samples for each age group differed significantly. Fisher's exact procedure was used as the test statistic instead of χ^2 due to the small number of observations in each group (Sheskin, 2003). For both the Mann-Whitney *U* and Fisher's test, significance was assessed at the two-tailed level using exact tests.

A linear regression analysis was performed to test the hypothesis that both hearing level and cognitive ability contribute to speech perception in a multi-talker environment. To ensure that the cognitive measures were independent of hearing level, only measures from purely visual tasks were included in the analysis: subtests 1A/B, 4A/B, and 6. The K-S test was used to ensure that each measure did not violate the assumption of being normally-distributed. Scores on the *Map search* task (subtest 1B) were transformed to ensure that the distribution of the data approximated the normal distribution. This was achieved by converting the data from 'items found' to 'items not found' and taking the square root of the scores. A K-S test on the transformed data confirmed that the data distribution did not differ significantly from the normal distribution. The scores from subtest 4A were not included in the analysis as they were heavily negatively skewed and could not be transformed to approximate the normal distribution. The regression analysis was performed twice. In the first analysis, the TEA measure was entered into the analysis followed by the BEA levels. This analysis examined whether the TEA measures accounted for a significant proportion of the variance in SLT performance, and whether BEA levels explained a significant amount of the residual variance. In the second, the order in which the variables were entered was reversed, so that the analysis examined whether the TEA measures explained a significant proportion of the variance not accounted for by the BEA levels.

6.2.7.4 SSQ

The data from the individual SSQ questions and the mean scores for each of the 10 sub-scales of the SSQ were used to assess self-reported hearing difficulties. Age-related differences in scores on the sub-scales were assessed using the same methods employed for the analysis of group differences in hearing levels. In addition to the sub-scales, the mean scores on individual questions from the SSQ directly related to listening in multi-talker environments were also analysed—questions 3, 4, & 12 from the Speech section of the SSQ. For cases where equality of variance could not be assumed, Welsh's t-test was used, denoted by non-integer degrees of freedom. If the K-S tests indicated that the distribution of scores in either age group was significantly different from the normal distribution, tests for age-related differences were performed using non-parametric statistics as outlined in Section 6.2.7.3.

6.2.7.5 Individual Differences

Correlations were used to assess the relationships between performance on the spatial listening task, hearing sensitivity, attentional measures from the TEA, and self-reported difficulties using the SSQ. Unless stated otherwise, relationships between variables were calculated using rank order correlations in the form of Kendall's τ_b. If tied pairs exist between two variables, Kendall's τ_b adjusts the total number of pairings between items to take into account the number of ties. If no ties exist, τ_b is equivalent to Kendall's τ. Correlations were performed separately for the young and older adult groups, and for all participants which represented a normally-hearing sample between 19–71 years of age. Significance was assessed at the two-tailed level by converting the τ_b value into a z-score using the SPSS software (SPSS Inc., 2006).

The analysis assessed relationships between participants' ability to hear out information in a multi-talker listening task, their ability to divide, switch, and sustain their attention, their hearing sensitivity, and the level of difficulty that they reported experiencing in everyday life.

Adjusting for Multiple Comparisons

The analysis of individual differences involved the calculation of multiple rank order correlations, often involving the same variable. The calculation of a probability value for each correlation did not take into account the inflation of the family-wise error rate (FWER), the probability of making a Type I error (Field, 2005), due to repeated comparisons with the same variable.

An analysis was devised to determine whether a group of observed significant correlations involving a common variable should be accepted as being statistically significant. The procedure took into account the number of multiple comparisons that were performed. The null hypothesis, H_0, was that the group of significant correlations had occurred by chance. An approach based on Monte Carlo methods was used.

Each time a single variable was compared to a set of related measures (e.g. BEA thresholds vs. the 10 measures of the TEA), the number of observed significant correlations, $S_{observed}$, the number of comparisons involving the same variable, C, and the number of observations in each variable, N, were recorded. The comparisons were then modelled with random variables. The Box-Muller transform (Box & Muller, 1958) was used to generate gaussian-distributed random variables from uniformly-distributed random variables generated with a random number generator (Press et al., 2002). The multiple-comparisons were modelled using the following steps:

- Generate a normally-distributed random variable with N observations. This was termed the *static* variable.
- Generate C normally-distributed random variables also with N observations. These were termed the *paired* variables.
- Calculate rank order correlations (τ_b) between the *static* variable and each of the *paired* variables.

- Record the number of significant correlations ($p < .05$, two-tailed), S_{random}.

This sequence of steps was termed a *random sample* and was repeated 10,000 times. This process created a distribution of the number of significant correlations, S_{random}, observed when C correlations are calculated against the same random variable. The number of *random samples*, R, which resulted in at least $S_{observed}$ significant correlations was determined from the distribution. The probability of $S_{observed}$ (uncorrected) significant correlations occurring when C correlations are computed against the same variable, given that H_0 was true, was calculated as a proportion of the number of random samples performed: $p_s = \frac{R}{10000}$. If $p_s < .05$ then H_0 was rejected and the group of significant correlations in the observed data were deemed not to have occurred by chance. Significance values which have been corrected for multiple comparisons are denoted by p_{group}, and uncorrected significance values by p_{uncorr}.

When a small number of comparisons are made (< 5), a bonferroni correction was used to control for the inflated FWER. In cases where bonferroni correction was used, significance values are denoted by p_{bf}.

6.3 Results

6.3.1 Hearing levels

Figure 6.3 shows the pure-tone sensitivity thresholds and better-ear average (BEA) hearing levels for the young and older adult groups. Pure-tone thresholds ranged from −5 to 70 dB HL and BEAs from −1.3 to 18.75. Analysis of the BEAs assessed whether there were age-related differences between the two groups of participants. A t-test revealed that BEA hearing levels were significantly poorer in the older adult group (Mean 14.1 dB HL, $\sigma = 4.1$) than in the young adult group (Mean 3.4 dB HL, $\sigma = 3.4$) [$t(22) = -7.03,\ p < .001,\ r = .83$].

6.3.2 Spatial Listening Task

6.3.2.1 Performance

Figure 6.4 shows performance for the four conditions of the spatial listening task by the young and older adult groups. The average SRT was −8.8 dB (σ = 5.1 dB) in the “1 by 1” tasks and −2.3 dB (σ = 4.0 dB) in the “2 by 2” tasks. The widest range of SRTs across participants was observed in the “1 by 1” ‘None’ task, with thresholds ranging from −15.0 to 2.3 dB.

To determine whether providing cues for the location, onset time, and vocal characteristics of the target phrase (‘All’ vs. ‘None’) and whether pairing phrase onsets (“1 by 1” vs. “2 by 2”) significantly affected performance levels, the data were

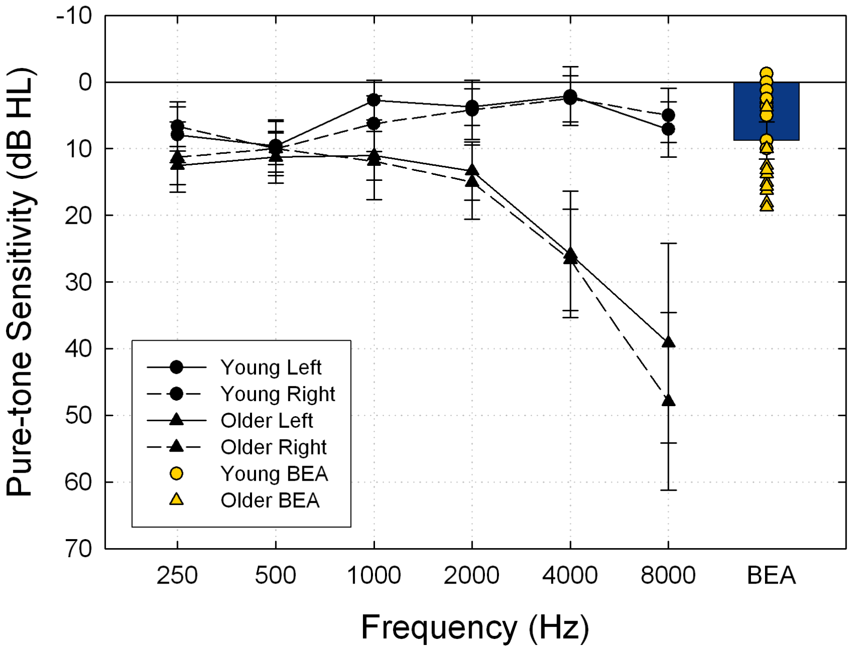

Figure 6.3. Pure-tone sensitivity thresholds for the young and older participants (circles and triangles) for left and right ears (solid and dashed lines). Individual (yellow symbols) better-ear averages (BEAs) are shown for both groups together with the mean BEA (blue bar). Error bars show 95% confidence intervals.

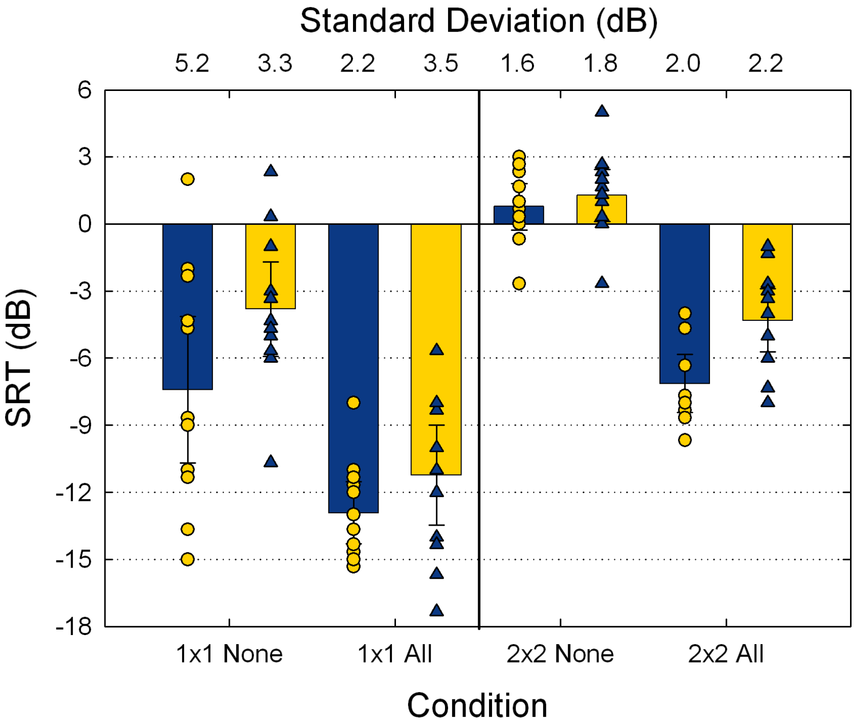

Figure 6.4. Mean (bars; young (blue) and older (yellow)) and individual (symbols; young (circles) and older (triangles)) SRTs when phrases started one by one (1x1) or in pairs (2x2). Participants were either provided with additional cues as to the target phrases location, talker, and onset time (all) or not (none). Error bars show 95% confidence intervals.

subjected to a repeated-measures ANOVA with two within-subjects factors, *Cued* and *Paired*, and a single between-subject factor of *Group*. Providing prior information about the target phrase significantly improved performance relative to when no cues were provided [$F(1,22) = 244.46,\ p < .001,\ r = .96$]. When phrases started in pairs, performance decreased significantly compared to when they started one at a time [$F(1,22) = 160.83,\ p < .001,\ r = .94$]. The main effect of group was also significant [$F(1,22) = 6.30,\ p < .05,\ r = .47$]. None of the interaction terms was significant.

6.3.2.2 Factor Analysis

Table 6.6 shows the correlation matrix between the SRTs in the four conditions of the spatial listening task. In agreement with Experiments 1 and 2, strong relationships were observed between performance levels in different conditions. This result suggests that a single variable may underlie performance across the four conditions.

Table 6.7 shows the results from the factor analysis. Of the four factors in the initial solution, one had an eigenvalue above 1; therefore only one factor was retained and no rotation was performed. The single factor accounted for 57.86% of the variance.

A threshold of .7 was used to select those tasks which had a high load on the factor. All of the tasks except the “2 by 2” ‘None’ loaded highly on the factor. The observed weak correlations involving that task and its lower loading on the extracted factor may be due to low variability in the threshold values. An examination of the standard deviations of each of the conditions revealed that SRTs on the the ‘2x2 None’ task had the lowest variability (σ = 1.7 dB). This low variability was also observed in the

	1x1 All	**1x1 None**	**2x2 All**	**2x2 None**
1x1 ‘All’	—	.49**	.48**	.16
1x1 ‘None’		—	.74***	.30
2x2 ‘All’			—	.32
2x2 ‘None’				—

Table 6.6. Correlation matrix (Pearson’s r) for the four conditions of the spatial listening task (SLT) (** $p < .01$, *** $p < .001$, two-tailed).

	Loading
1x1 ‘All’	.72
1x1 ‘None’	.88
2x2 ‘All’	.88
2x2 ‘None’	.51
Sum of squared loadings	2.31
% of variance explained	57.86

Table 6.7. Loadings for the single factor with an eigenvalue > 1. Principal Component Analysis was used as the extraction method.

analogous conditions of Experiments 2, 3, and 4 (Chapter 5).

A post-hoc analysis was performed to compare performance in the “2 by 2” ‘None’ condition to the equivalent conditions in Experiments 2–4 using independent-samples t-tests. Bonferroni correction was used to adjust the significance criteria based on the calculation of multiple comparisons involving the “2 by 2” ‘None’ data. The analysis confirmed that performance in “2 by 2” ‘None’ in the current experiment did not differ significantly from the equivalent conditions in Experiment 2 [$t(30) = 0.093$, $p_{bf} > .05$ ns, $r = .02$], 3 [$t(30) = 0.467$, $p_{bf} > .05$ ns, $r = .08$], or 4 [$t(31) = 0.921$, $p_{bf} > .05$ ns, $r = .16$]. The single extracted factor was used to represent performance on the spatial listening task (SLT) in all subsequent analyses.

6.3.3 Attentional Measures

Boxplots of the data from the 10 measures of the TEA revealed the presence of an outlier in the older adult group on the *Telephone search with counting* task (subtest 7). The z-score of the data point was found to exceed the threshold of 3. The outlier was removed from the data for that measure for all analyses.

Table 6.8 shows the mean and normative scores for the young and older adult groups across the 10 measures of the TEA. Close correspondence between the participants in the current experiment and the normative sample was observed on 8 of the 10 measures. The time required to switch attention on the *Visual elevator* (subtest 4B) was shorter (Mean difference .47 secs, mean % decrease 12.79) and accuracy scores on the *Elevator counting with reversal* (subtest 5) tasks were higher (Mean difference 2 items, mean % increase 35.42) on average in both groups relative to the normative data.

	Mean		**Norm (50%ile)**		
	Young	**Older**	**Young**	**Older**	**Group differences**
1A	48.58	36.33	49	33	$U = 25.50^{**}$
1B	77.08	68.42	77	65	FET**
2	6.92	7.00	7	7	n/a
3	9.25	8.17	9	9	n/a
4A	8.58	8.75	8	9	$U = 58.50$
4B	2.93	3.53	3.4	4.0	$U = 25.00^{**}$
5	8.92	7.17	7	5	$U = 25.00^{**}$
6	2.45	3.40	2.6	3.2	$U = 13.50^{***}$
7	.18	1.25	0.5	1.0	$U = 24.00^{**}$
8	9.25	9.50	9	10	n/a

Table 6.8. Mean scores for the young and older adult groups, the values at the 50^{th} percentile from the normative sample of 154 participants (Robertson et al., 1996), and the results of the between-group analysis (** $p < .01$, *** $p < .001$, two-tailed; FET=Fisher’s exact test; n/a=could not be computed).

The data from the 10 measures were assessed for age-related differences. Performance on the *Map Search* task (subtest 1A) after 1 minute was significantly lower in the older group (Mdn = 34.00) than in the young group (Mdn = 50.50) [$U = 25.50$, $p < .01$, $r = -.55$]. The number of participants with poor 2 minute scores (subtest 1B) (classified as being less than or equal to the median of all scores) was 11 out of 12 in the older group (91.7%, Older Mdn = 69.00) compared to 2 out of 12 in the young group (16.7%, Young Mdn = 77.50). This difference was significant [Overall Mdn = 74.00, $p < .01$, Fisher's exact test].

While no effect of age group was was found for performance on the *Visual Elevator* task (subtest 4A) [$U = 58.50$, $p = .43$ *ns*, $r = -.17$], the measure of attentional switching speed from same the task (subtest 4B) was found to be significantly longer in the older adults (Mdn = 3.55) compared to the young adults (Mdn = 2.90) [$U = 25.00$, $p < .01$, $r = -.55$]. Performance on the *Elevator counting with reversal* (subtest 5), an auditory analogue of the visual elevator task, showed poorer performance amongst the older group (Mdn = 7.00) compared to the young group (Mdn = 9.00) [$U = 25.00$, $p < .01$, $r = -.57$]. In the *Telephone search* task (subtest 6), older adults required longer to correctly identify matching symbol pairs (Mdn = 3.20) relative to the young adults (Mdn = 2.45) [$U = 13.50$, $p < .001$, $r = -.69$]. On the *Telephone search with counting* (subtest 7), older adults exhibited a significantly larger dual-task cost (N = 11, Mdn = 1.00) as compared to the young adults (N = 12, Mdn = .40) [$U = 24.00$, $p < .01$, $r = -.54$].

The shape of the distribution of scores was found to be significantly different between the age groups on 3 of the 10 measures: *Elevator counting*, *Elevator counting with distraction*, and *Lottery* (subtests 2, 3, and 8 respectively). Upon further examination of the data from those measures, ceiling effects were observed within both age groups. A comparison with the scores from related age groups in the normative sample showed a similar pattern of results (Robertson et al., 1996). No further analysis for group differences were performed on the data.

6.3.3.1 Regression Analysis

The results of the linear regression analysis for all 24 participants are shown in Table 6.9. For three of the four TEA measures, linear regression showed that attentional ability assessed using purely-visual tasks was a significant predictor of SLT performance, and that BEA level predicted a significant proportion of the residual variance. When BEA level was entered first, it explained a significant amount of the variance in SLT performance (67%), and none of the TEA measures explained a significant amount of the residual variance.

6.3.4 Self-reported Difficulties

Table 6.10 shows the ranges, means, and standard deviations for scores on the 10 sub-scales of the SSQ. On average, participants rated themselves highly on all aspects of their own ability to deal with listening in everyday environments (overall mean score 8.1) with more than half of participants providing a rating ≥ 7 on all sub-scales. Low variability was also observed across all sub-scales (overall $\sigma = 1.4$, maximum $\sigma = 1.5$ (sub-scale 3)).

Sub-scales which contained ratings below 7 tended to contain questions related to speech perception in noisy environments. These sub-scales included speech-in-noise and speech-in-speech (sub-scales 2 & 3, 5 participants each), localisation (sub-scale 5, 7 participants), and listening effort (sub-scale 10, 6 participants). The *Multiple speech-stream processing and switching* sub-scale (4) contained the highest number of ratings below 7 (11 participants). This sub-scale also had the largest number of responses of all sub-scales with a rating of less than 5 (3 participants).

Age-related differences in the level of self-reported difficulties were examined across 9 of the sub-scales. The first sub-scale, speech in quiet, did not meet the assumptions of the Mann-Whitney *U* or median tests. An examination of the data for this sub-scale revealed that only two participants reported a rating below 9 (8.3 & 7.5). Both participants were members of the older adult group.

Older adults reported experiencing more difficulty with identifying sounds (*Identification of sounds and objects*, sub-scale 8) (Mean 8.98, $\sigma = .76$) compared to the younger adults (Mean 8.23, $\sigma = .86$) [$t(22) = -2.26$, $p < .05$, $r = .43$]. The older group also reported having to work harder to understand speech or when focusing on other

	TEA First			**BEA First**	
	β	**R^2**		**β**	**R^2**
Map search (1A)					
TEA	.50	.25*	BEA	−.82	.67***
BEA	−.60	.36**	TEA	.03	.00
Map search (1B)					
TEA	−.58	.34**	BEA	−.82	.67***
BEA	−.47	.22*	TEA	.05	.00
Visual elevator (4B)					
TEA	−.46	.21*	BEA	−.82	.67***
BEA	−.66	.44***	TEA	−.08	.01
Telephone search (6)					
TEA	−.39	.15	BEA	−.82	.67***
BEA	−.62	.38**	TEA	.22	.05

Table 6.9. Results of the regression analysis which assessed whether both measures from purely-visual TEA tasks and BEA levels contributed to SLT performance. The analysis was performed across all 24 participants (* $p < .05$, ** $p < .01$, *** $p < .001$, two-tailed).

sounds (*Listening effort*, sub-scale 10) (Mean 8.16, $\sigma = 1.30$) relative to the younger group (Mean 6.93, $\sigma = 1.27$) [$t(22) = -2.36$, $p < .05$, $r = .45$]. No other significant age-related differences were observed.

The constituent questions of the SSQ for the two sub-scales which showed age-related differences were examined to assess which questions were giving rise to the observed effect. For the *Identification of sounds and objects* sub-scale (8), an age-related difference was found only for question 7 from the Quality section of the SSQ [$U = 27.50$, $p < .01$, $r = -.53$]: *"When you listen to music, can you make out which instruments are playing?"*. For the *Listening effort* sub-scale (10), only Quality question 14 was found to show a significant effect of group [$U = 30.50$, $p < .05$, $r = -.49$]: *"Do you have to concentrate very much when listening to someone or something?"*.

Individual questions which were directly related to listening to what one person is saying when other people are speaking at the same time were also examined. No significant age-related differences were found for question 3 [$U = 45.50$, $p > .05$ *ns*, $r = -.32$], 4 [$U = 58.00$, $p > .05$ *ns*, $r = -.17$], or 12 [$U = 52.50$, $p > .05$ *ns*, $r = -.23$] from the Speech section of the SSQ.

6.3.5 Individual Differences

6.3.5.1 Hearing Sensitivity

The first analysis assessed whether the hearing sensitivity of participants was associated with their performance on the SLT. A significant relationship was observed across all participants between SLT performance and BEA hearing level [$\tau_b = -.67$, $p < .001$]. This relationship was also observed within the young adult [$\tau_b =$

	Overall		Group means		
Sub-scale	**Range**	**Mean**	**Young**	**Older**	**Difference**
1	7.50–10.0	9.48	9.74	9.23	n/a
2	5.50–9.80	7.89	8.24	7.54	$t(16.43) = 1.39$
3	4.00–10.0	7.69	7.91	7.47	$t(22) = .73$
4	3.70–9.50	6.80	6.73	6.88	$t(22) = -.26$
5	4.00–9.90	7.75	7.75	7.75	$t(22) = -.02$
6	4.30–9.50	7.61	7.42	7.80	$U = 50.00$
7	6.70–10.0	8.96	8.58	9.35	$t(22) = -1.93$
8	6.60–9.70	8.61	8.23	8.98	$t(22) = -2.26^*$
9	6.70–10.0	8.90	8.75	9.06	$t(22) = -.82$
10	4.90–9.70	7.55	6.93	8.16	$t(22) = -2.36^*$

Table 6.10. Descriptives of self-reported scores on the 10 sub-scales of the SSQ and the results of the between-groups analysis. All questions were rated between 0 and 10 (* $p < .05$, two-tailed).

$-.63,\ p < .01$] and older adult [$\tau_b = -.73,\ p < .01$] groups, separately.

6.3.5.2 Attentional Abilities

The second analysis sought to examine relationships between SLT performance, participants' attentional abilities, and BEA levels. Table 6.11 lists the correlations between performance on the SLT and subtests of the TEA across all 24 participants. SLT performance was significantly related to 5 subtests after correcting for multiple comparisons. The 5 subtests with which the significant relationships were found included auditory, visual, and audio-visual tasks.

The strongest of these relationships involved the 2-minute score from the visual *Map Search* task (subtest 1B). Figure 6.5 shows a scatterplot of the data from the *Map Search* task and the single factor extracted from the four conditions of the spatial listening task. The distribution of the data suggests an approximately linear relationship and shows the effect of group previously identified for both variables.

An examination of the relationships between the TEA measures and performance on the spatial listening task within the age groups revealed a significant (uncorrected) correlation involving performance on the *Lottery* task (subtest 8) for the older adult group [$\tau_b = .62,\ p_{uncorr} < .05$]. However, after correcting for multiple comparisons using the techniques described in Section 6.2.7.5, no significant relationships were found between SLT performance and the TEA measures within the young or older adult groups.

Across all 24 participants, hearing sensitivity was found to be significantly related to 6 of the 10 measures of the TEA (Table 6.11). The significant relationships included purely visual subtests (1A, 1B, 4B, & 6) as well as auditory subtests (5 & 7). No

Subtest	Modality	Correlation with SLT	Correlation with BEA
1A	V	$.36^{*}$	$-.47^{**}$
1B	V	$.44^{**}$	$-.58^{***}$
2	A	.21	−.03
3	A	.24	−.22
4A	V	.11	−.02
4B	V	$-.36^{*}$	$.34^{*}$
5	A	$.42^{**}$	$-.48^{**}$
6	V	−.28	$.43^{**}$
7†	A/V	$-.32^{*}$	$.32^{*}$
8	A	.23	−.13

Table 6.11. Rank order correlations (Kendall's τ_b) between subtests of the TEA, performance on the SLT, and BEA levels across all 24 participants ($^{*}\ p_{uncorr} < .05$, $^{**}\ p_{uncorr} < .01$, $^{***}\ p_{uncorr} < .001$, two-tailed; $p_{group} < .05$ for all significant (uncorrected) correlations after correcting for multiple comparisons). †An outlier from the older adult group was removed for the correlations with subtest 7 of the TEA.

significant relationships were found between the TEA measures and BEAs within either age group.

6.3.5.3 Self-reported Difficulties

The third individual differences analysis sought to identify whether participants' performance on the spatial listening task, hearing sensitivity, and attentional abilities were related to difficulties that they reported experiencing in everyday life. Relationships between the SSQ and BEA levels were first assessed at the individual question level. After correcting for multiple comparisons, no significant correlations were found between the individual SSQ scores and BEA thresholds across all participants or within either age group.

Relationships between the SSQ, performance on the spatial listening task, attentional abilities, and BEA levels were then examined using the sub-scales of the SSQ. Using the sub-scales to summarise self-reported difficulties reduced the number of measures from 52 to 10. After correcting for multiple comparisons, no significant relationships were found within the young and older adult age groups or across all participants.

Finally, specific questions which were related to difficulties experienced while listening to speech in multi-talker environments were examined (speech questions

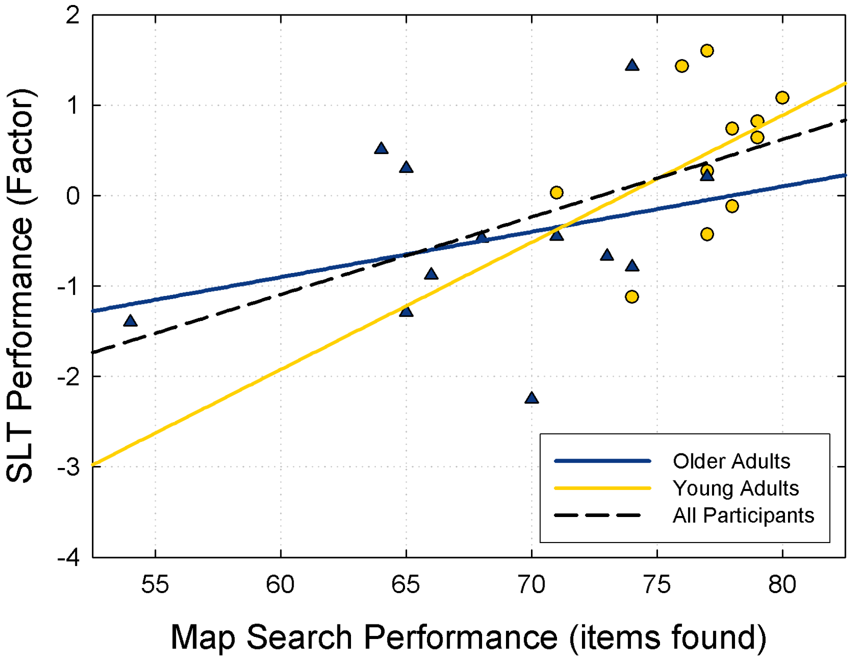

Figure 6.5. Scatterplot of the single PCA factor extracted from the spatial listening task (SLT) and the purely visual *Map search* task from the TEA for the young (circles) and older (triangles) adult groups. The two variables were found to be significantly related [$\tau_b = .44$, $p < .01$]. Regression lines are shown for illustration purposes for the data from all participants (dashed), and for the young (yellow) and older (blue) adults.

3, 4, & 12). A significant relationship was observed between performance on the spatial listening task and the self-reported difficulties with following a conversation in a group of people [$\tau_b = .37,\ p_{bf} < .05$]. Difficulty with hearing what people are saying when a conversation switches between different talkers was marginally related to the time required to switch attention (TEA subtest 4B) after correcting for multiple comparisons [$\tau_b = -.31,\ p_{uncorr} = .045,\ p_{bf} > .05\ ns$]. No significant correlations involving the three questions were found within either age group.

6.4 Discussion

This chapter has presented the performance of young and older adults on a task of spatial listening for speech which placed high attentional demands on the participants. The cognitive abilities of both age groups were assessed using an attentional test battery, and have been compared to performance on the listening task, hearing sensitivity, and self-reported difficulties in everyday listening situations.

Older adults exhibited poorer performance on all conditions of the spatial listening task relative to the young adult group. This result confirmed the hypothesis that older adults are poorer at listening to what one person is saying when many other people are speaking at the same time.

Task performance was related to self-reported difficulties in following a conversation in a group of people, suggesting that the task placed similar cognitive demands on the listener compared to difficult everyday listening situations.

Group differences were observed on 5 of the 8 tasks from the attentional test battery. The tests were identified as requiring the rapid processing of information, as loading working memory, and as involving the switching of attentional focus. These results confirmed the hypothesis that older adults experience cognitive deficits resulting from factors which may include a decrease in processing speed, a reliance on the slower top-down volitional attentional control compared to bottom-up stimulus-driven attentional control, and a reduced processing capacity.

An analysis of individual differences showed that performance on the spatial listening task was related to several of the cognitive measures from the attentional test battery and to hearing sensitivity. It was hypothesised that poorer hearing sensitivity and deficits in attentional abilities would both contribute to poorer spatial listening performance. This hypothesis could not be confirmed based on the results of a regression analysis. Two models are proposed in the next section which attempt to account for the observed patterns in the data.

In the following sections, the results from each aspect of the experiment will be discussed in detail.

6.4.1 The Roles of Hearing Loss and Attentional Abilities in Determining Spatial Listening Performance

Three of the measures in the current experiment were significantly inter-related; i.e. each measure was correlated with the other two. The measures were performance on the spatial listening task, hearing sensitivity as indexed by the BEA, and 5 of the 10 measures from the TEA. The regression analysis sought to determine whether both TEA and BEA measures were significant predictors of performance on the spatial listening task, and whether they made independent contributions to explaining variance in performance. When measures from the TEA were entered first into the analysis, a significant proportion of the variance in performance scores was explained and BEA levels accounted for a significant proportion of the residual variance. The highest proportion of variance in performance scores explained by the two variables was 65% (*Visual elevator* (Subtest 4B), 21%; BEA, 44%). However, when the BEA thresholds were entered first into the regression analysis, they explained 67% of the variance in performance scores and none of the TEA measures explained a significant proportion of the residual variance. The pattern of correlations and the results of the regression analysis will be discussed in terms of three models which attempt to formalise the causal and associative relationships between the three measures.

Based on the findings of previous studies (George et al., 2007; Helfer & Wilber, 1990; Helfer & Freyman, 2008; Humes et al., 2006; Pichora-Fuller et al., 1995; Tun et al., 2002; Zekveld et al., 2007), it was hypothesised that poorer hearing sensitivity and cognitive deficits would make independent contributions to performance on the spatial listening task (Figure 6.6, Model A). Thus, it was expected that measures of these underlying influences would correlate with SLT performance. As both variables have been associated with the natural ageing process, the measures would therefore be expected to show an inter-correlation. The data from the current experiment agreed with these two predictions. However, the results of the regression analysis did not support the hypothesis that the factors have independent effects on SLT performance (Table 6.9), as TEA measures did not explain a significant proportion of variance in SLT scores once variance associated with BEA levels had been partialled out. Therefore, Model A did not fit the data and was not considered further.

An alternative model embodies the idea that both hearing sensitivity and cognitive abilities make independent contributions to SLT performance, but that measures of each variable are not independent of each other (Figure 6.6, Model B). It is possible that the measurement of pure-tone thresholds was confounded with the ability of participants to sustain attention, as the measurement procedure required participants to maintain vigilance to detect pure tones at levels close to threshold over a period of up to 10 minutes. In the regression analysis, the reduction in the proportion of variance explained by the BEA thresholds when they were entered after the TEA variables compared to when they were entered first may therefore suggest that a latent factor of attention underlay both measures. This hypothesis may also explain the observed relationship between hearing level (BEA) and performance on the spatial listening task even within the young adult group.

While Model B suggests that both hearing level and attentional abilities contributed to spatial listening performance, and that BEA levels may have been confounded with the attentional abilities of participants, an alternative hypothesis is that hearing sensitivity alone contributed to SLT performance (Figure 6.6, Model C). The observed relationships between BEA levels and SLT performance within the young adult group may signal that the spatial listening task may have placed a sufficiently high load on the peripheral auditory system so that even small decreases

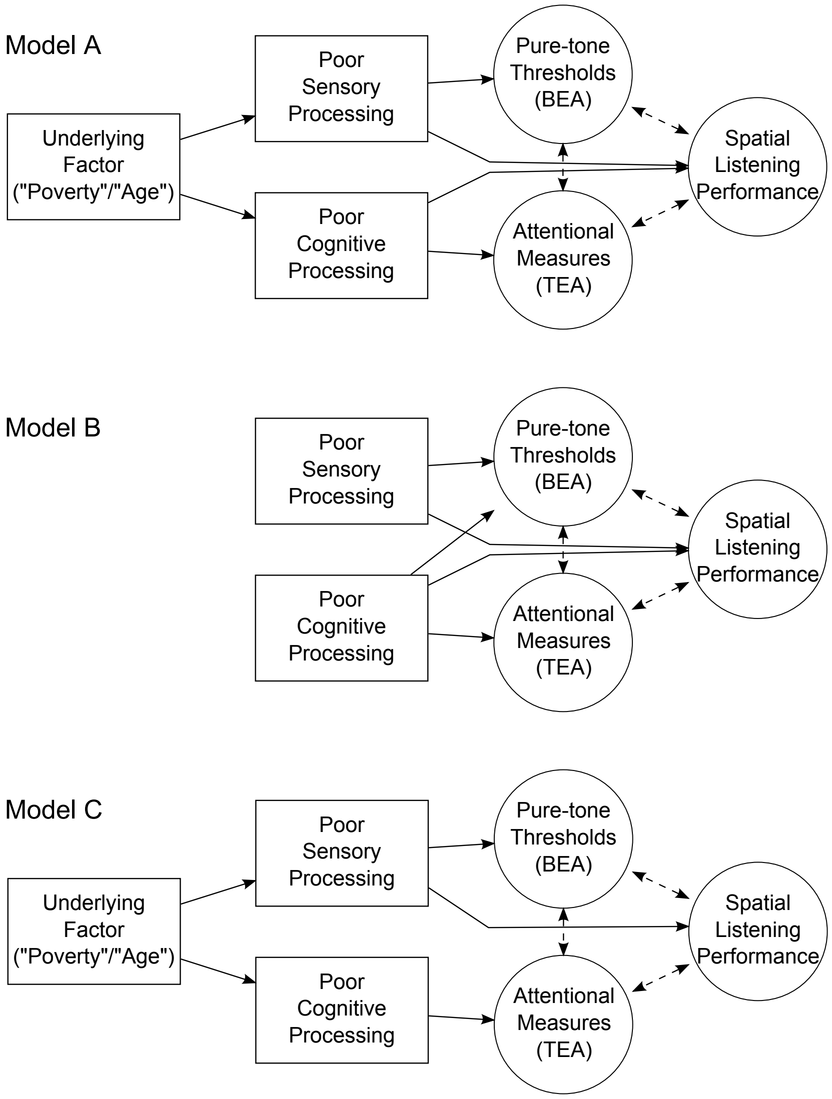

Figure 6.6. Three alternative models which attempt to account for spatial listening task performance in terms of hearing sensitivity and cognitive abilities. Causal relationships are indicated by solid arrows, and associative relationships indicated by dashed arrows.

in hearing sensitivity would be accompanied by increased SRTs. The implication is that complex listening situations which more closely model everyday environments compared to non-spatial headphone listening tasks are more likely to expose the effects of mild hearing loss in younger adults on speech-in-noise and speech-in-speech performance. In addition, the observed correlations between TEA measures and BEA levels may have arisen due to an underlying factor such as socio-economic status or poverty. The results of the regression analysis could thus be interpreted as confirming the ability of poorer hearing sensitivity to affect poorer spatial listening performance.

If attentional abilities did affect the pure-tone threshold measurement process (Model B), it would be expected that all pure-tone thresholds would reflect this, and therefore that thresholds at both low and high frequencies would correlate with SLT performance. Alternatively, if the measurement of pure-tone thresholds was not influenced by attentional ability and only hearing sensitivity contributed to poorer spatial listening performance (Model C), high-frequency loss would be a likely predictor of SLT performance. Therefore, only high-frequency thresholds (> 2 kHz) would be expected to correlate with SLT performance. A post-hoc analysis revealed that thresholds at low and high frequencies were significantly correlated with SLT performance (Table 6.12). These results provide some evidence for the suggestion that BEA levels were confounded with the attentional abilities of the participants (Model B). However, Models B and C can not be distinguished unambiguously based on the findings of the current experiment. For future research which examines the relative contributions of hearing sensitivity and attention, a measure of hearing sensitivity that is less likely to be influenced by the attentional abilities of participants may be preferable. Candidates would be the methods that have been proposed for predicting peripheral sensitivity based on evoked responses to tones (Johnson & Brown, 2005) or speech (Dajani, Purcell, Wong, Kunov, & Picton, 2005) measured in the brainstem using EEG.

Frequency	**Correlation with SLT**	
	Left ear	**Right ear**
250 Hz	$-.528^{**}$	$-.367^{*}$
500 Hz	$-.214$	$-.234$
1000 Hz	$-.359^{*}$	$-.329^{*}$
2000 Hz	$-.687^{**}$	$-.512^{**}$
4000 Hz	$-.525^{**}$	$-.548^{**}$
8000 Hz	$-.514^{**}$	$-.409^{**}$

Table 6.12. Rank order correlations (Kendall's τ_b) between pure-tone thresholds and performance on the SLT across all 24 participants (* $p_{uncorr} < .05$, ** $p_{uncorr} < .01$, two-tailed).

6.4.2 Multi-talker Listening

An overall decrement in the ability of the older adult group to perceive speech when presented against a background of other talkers was observed across all conditions of the spatial listening task. On average, speech reception thresholds were higher by 2.2 dB for the older adults, with the largest age-related difference arising in the “1 by 1” task when no prior information about the target phrase was provided. These results are compatible with the findings of previous studies that have identified an effect of ageing in speech-in-speech tasks (Dubno et al., 1984; Helfer & Freyman, 2008; Tun et al., 2002; Wiley et al., 1998). The inability of older adults to ignore background speech which is intelligible (Tun et al., 2002) or spoken by talkers of a different gender to that of the target phrase (Helfer & Freyman, 2008) may have contributed to the observed group differences.

In contrast, the ability of the two age groups to take advantage of informational cues about the target phrase did not differ significantly, as noted by the absence of a significant interaction between the *group* and *cued* factors. Thresholds improved by 6.6 dB on average when cues were provided compared to conditions in which they were not available to the participants. This result is compatible with the finding that older adults are able to take advantage of information about the voice of a talker to focus attention on what they say against a background of noise (Johnsrude et al., 2008). Yonan & Sommers (2000) found that older adults received a larger benefit from prior information about who would speak a target phrase, compared to younger listeners. Thus, it would seem that the ability of older adults to take advantage of information to focus their attention is not significantly impaired by the ageing process.

There were no significant age-related differences in susceptibility to the distracting effects of a phrase onset which was simultaneous with the onset of the target phrase, which raised thresholds on average by 6.5 dB compared to when phrases started one at a time. The results of Experiments 1–4 suggest that phrase onsets capture attention involuntarily, and that the occurrence of two simultaneous phrases undermines a listener’s ability to reliably focus attention on the less intense of the two phrases. The absence of a greater susceptibility to the distracting effects of phrase onsets among the older adults suggests that their attention was automatically captured by the onsets, regardless of age. This result provides evidence to support the findings of previous research which has suggested that the capacity to make voluntary and involuntary shifts in attention is preserved with age (Folk & Hoyer, 1992; Madden, 1990).

6.4.3 Effects of Hearing Sensitivity

Although the older adult group had BEA thresholds within the range of sensitivities commonly attributed to normally-hearing individuals, hearing sensitivity was found to be significantly poorer in the older adult group. The finding underlines the difficulty in disassociating the normal age-related decline in hearing levels from other age-related factors. The BEA thresholds were found to be significantly related to performance on the spatial listening task. This was also found to be the case within both the young and older adult groups. Previous research has suggested that the hearing level of normally-hearing older adults is not related to their performance on a speech-in-noise task. George et al. (2007) proposed that the absence of a relationship between hearing sensitivity and SRTs within a group of normally-hearing adults may have been due to the small range of hearing thresholds in that group (Mean 6.7 dB HL, $\sigma = 3$). In contrast, the thresholds of the older adult group in the current experiment were poorer (Mean 14.1 dB HL) and more varied ($\sigma = 4$). These factors may have contributed to the observed significant relationship. Helfer & Freyman (2008) also found significant correlations amongst older adults between hearing level and performance on a speech-in-speech task, but the group of participants included individuals with more pronounced levels of presbycusis than were tested in the current experiment.

Asymmetry in the degree of hearing loss between the ears has been associated with spatial listening difficulties. Using the SSQ, Noble & Gatehouse (2004) found that older adults with asymmetric hearing loss reported a greater level of difficulty with spatial listening in everyday situations compared to a group of older adults with symmetric hearing loss. An asymmetric loss was defined as a difference of more than 10 dB between the average thresholds, calculated from pure-tone thresholds at 0.5, 1, 2, and 4 kHz. None of the older adult participants in the current experiment exhibited an equivalent asymmetry. It is therefore unlikely that hearing asymmetry contributed to the significant relationship between BEA thresholds and SLT performance levels.

6.4.4 Attentional Factors

Age-related differences were identified on 6 of the 10 measures of the TEA. They included purely auditory, purely visual, and audio-visual tasks of attention. From the summary of the TEA measures in terms of four cognitive variables (Table 6.4), processing speed is associated with 5 of the measures which exhibited age-related differences. Of those 5 measures, 4 were also found to be significantly related to performance on the spatial listening task. This pattern is compatible with the idea that a decrement in the speed of processing amongst the older adult group underlay their difficulties in coping with the complex listening task. Processing speed has been identified as a general effect of the ageing process (Kok, 2000), and a major contributor to deficits in performance on cognitively-demanding tasks (Salthouse, 1996).

From the tasks which exhibited age-related differences, two other cognitive variables were associated with deficits in performance: working memory and attentional switching. Increasing age has been associated with a decrease in working memory performance on purely-visual tasks (Zekveld et al., 2007) and speech-based tasks (Pichora-Fuller et al., 1995). Furthermore, speech perception performance in noise has been associated with working memory capacity (Humes et al., 2006). In relation to attentional switching, it has been suggested that voluntary and involuntary shifts are preserved with age (Folk & Hoyer, 1992). Possibly, the age-related deficits in tasks which involved a switching component were a result of a greater reliance on top-down attentional control in the older group, resulting in slower and more error-prone performance. Madden et al. (2007) found evidence for such a reliance using a visual task of attention, and Lachter et al. (2004) suggested that top-down shifts are slower compared to involuntary, or stimulus-driven, attentional shifts. Poorer performance on tasks which load working memory or require shifts of attention is therefore congruent with previous research.

6.4.5 Self-reported Difficulties

Self-reported difficulties, as measured with the SSQ, revealed age-related differences on 2 of the 10 sub-scales: *Identification of sounds and objects* and *Listening effort*. Further analysis revealed that the source of these differences were two specific questions from the ‘Quality’ section of the SSQ. For the *Identification of sounds and objects* sub-scale, the question was related to the ability of participants to hear what instruments are playing when listening to music. The result that older adults reported more difficulty in this situation compared to the young adults may be related to deficits in frequency resolution arising from a broadening of the auditory filters with increasing age (Patterson et al., 1982; Peters & Moore, 1992) which could impair the ability of the older group to fully separate the different acoustical streams, or deficits in selective attention to the output of those filters arising from an inability to ignore irrelevant aspects of the signal such as variations in amplitude (Sommers, 1997).

For the *Listening effort* sub-scale, the question which exhibited an age-related difference queried the extent to which participants must concentrate when selectively listening to someone or something. Informally, many of the older adult listeners commented that they had to expend a substantial degree of effort across the range of tasks in the TEA and also in the spatial listening task. Previous research has identified a bias towards top-down control of attention in older adults. In a study of visual search performance, Madden (1990) found that while older adults did not show a deficit in top-down control of attention compared to young adults, they were slower at performing the search task and also had a greater tendency to rely on that top-down control compared to the young adults who could take advantage of stimulus-driven shifts in attention. In the auditory domain, Pichora-Fuller et al. (1995) compared young and older adult listeners on a combined speech-in-noise and working memory task to assess the relative contributions of auditory and cognitive factors to speech reception performance in noise. While the results did not support a general effect of age-related cognitive decline, the authors suggested that the older adults were engaging in more "effortful listening" which affected their ability to correctly identify words against a background of speech babble. This mode of listening was described in terms of the reallocation of resources from cognitive processes, such as working memory, to the process of decoding of auditory information. If such a reallocation involved the addition of extra processing stages, then older adults would be more likely to exhibit deficits in the speed at which the information could be processed compared to young adults (Verhaeghen & Cerella, 2002).

Gatehouse & Noble (2004) used the SSQ to assess the self-reported difficulties of 153 individuals who attended an audiology clinic. They observed significant correlations between hearing sensitivity and scores on 46 of the 52 questions. In the present experiment, no significant correlations were observed between the same two measures. This result may be a reflection of the small range of hearing thresholds both across and within groups. An examination of the average scores reported by Gatehouse & Noble (2004) reveals lower ratings amongst their sample of older adults (Mean age 71, mean score 5.5) compared to the older adult group in current study (Mean age 63, mean score 8.2), although the amount of variability was similar (Gatehouse et al. $\sigma = 1.9$, current study $\sigma = 1.5$). The average high ratings reported by participants suggest that the questionnaire may be more sensitive to the difficulties experienced by older adults with increased levels of cognitive decline or more pronounced levels of age-related hearing loss.

The present experiment also related self-reported listening difficulties to participants' ability to cope with multi-talker situations. A significant correlation was observed between a question related to following a conversation in a group of people and SRTs. The spatial listening task was therefore successful in recreating some of the listening contexts in which older adults report experiencing difficulties.

6.5 Conclusions

Poorer performance on a task of spatial listening amongst older adults was found to be related to poorer hearing sensitivity and slower cognitive function. The results of the present experiment suggest that the use of a spatial listening task which places a range of attentional demands (divided, switching, and sustained attention) on the listener is a valuable method of understanding the difficulties that older adults experience in everyday life.

6.6 Summary

- Speech perception ability, hearing sensitivity, and cognitive ability were assessed for groups of younger and older normally-hearing adults.
- Older adults performed more poorly on all conditions of the spatial listening task compared to young adults. This result supported the hypothesis that older adults experience greater difficulties with hearing what one person is saying when many people are speaking at the same time, but did not dissociate the relative contributions of hearing sensitivity and attentional abilities to spatial listening performance.
- No age-related differences were found in the ability of listeners to take advantage of prior information about a target phrase to hear out information within the target phrase at lower SRTs, or in susceptibility to the distracting effect of a masker phrase whose onset is simultaneous with a target phrase.
- Hearing sensitivity was found to be significantly related to performance on the spatial listening task across all participants and within the young and older adult groups.
- The group of older adults showed a decrement in performance on 5 of the 8 subtests of the TEA compared to the younger adults. An analysis of the requirements of those tasks indicated that good performance on four of the tasks required fast speed of processing. This analysis supported the hypothesis that older adults exhibit deficits in cognitive tasks which require the rapid processing of information, frequent switching of attentional focus, and impose high cognitive demands.
- Of those TEA tasks which exposed age-related differences, 4 were found to be significantly related to performance on the spatial listening task.
- The results of a regression analysis did not unequivocally support the hypothesis that both peripheral and central deficits are implicated in age-related deficits in speech perception in noise. Two models to account for the observed pattern of relationships were proposed.
- Responses on the SSQ revealed that older adults reported more difficulty in identifying instruments when listening to music and having to expend more effort when listening to someone or something.

- Self-reported difficulties with following a conversation in a group of people were found to be significantly related to performance on the spatial listening task. This result suggests that the spatial listening task was successful in recreating the everyday environments which best distinguish better from poorer listeners.

Chapter 7

Cortical Activation Patterns During a Spatial Listening Task: Evidence from Young and Older Normally-Hearing Adults

In this chapter, the neural bases of focussing attention and resisting distraction in a multi-talker environment were examined using Magneto-encephalography (MEG). The spatial listening task developed in Experiment 1 was adapted so that the task could be performed while participants were lying supine in the MEG scanner. Data were collected from the same groups of young and older adults who participated in Experiment 5. Neural activity was examined at key moments in the spatial listening task: when participants had to discriminate between target and non-target call-signs, when attention had to be sustained on a target phrase, and when the onset of the phrase following the target phrase had to be ignored. Differences in activation at these key moments were localised to regions of the brain previously implicated in attentional processing, in both the auditory and visual domains. The regions were distributed across temporal, parietal, and frontal cortices. Relationships were observed between differences in MEG power at the moment at which a new phrase onset had to be ignored and the accuracy with which the task was performed. The results suggest that complex tasks of spatial listening, such as are experienced in everyday situations, place high attentional demands on the listener. These demands elicit the activation of a wide range of cortical processes which are not limited to auditory-specific processes. Difficulties in sustaining attention and/or resisting distraction were reflected in differences in cortical activation, demonstrating that deficits in attentional control contribute to difficulties in coping with challenging 'cocktail-party' listening environments.

7.1 Introduction

Experiment 5 confirmed the suggestion arising from Experiments 1–4 that the multi-talker spatial listening task described in Chapter 5 taps into the attentional abilities of participants. The relationships observed in Experiment 5 between performance on the task and both purely visual and purely auditory measures of attention suggest that the extent of the cognitive functions involved is broad—encompassing processes which are independent of the stimulus modality. The current experiment examined the patterns of cortical activation which accompanied successful performance of the spatial listening task. One goal was to understand the nature of the cognitive processes which underpin the ability to cope with difficult listening environments.

There are only a limited number of studies which have examined the nature and patterns of cortical activity associated with listening to speech in the presence of competing sounds; e.g. random noise or competing speech. ERP studies have examined listening to speech in noise using dichotic listening paradigms (Hink & Hillyard, 1976; Woods et al., 1984) or virtual spatial presentation of multiple speech streams (Nager et al., 2008). A common finding of these studies is that attention to speech and ignoring background acoustical information produces low-frequency, or slow-wave, differences in the activity measured on the scalp which was evoked by attended compared to unattended events. Such studies provide information about the time-course of the effects of attention to speech on neural activity but do not provide evidence of the cortical structures which underlie such effects.

Evidence of the cortical representations of attention to speech in noise can be found in a small number of studies which have used fMRI and PET to examine changes in metabolic activity and blood flow during complex listening tasks. The tasks have examined the response to brief segments of synthetic speech (Hashimoto et al., 2000), phrases extracted from a continuous speech stream (Nakai et al., 2005), and continuous speech against multi-talker babble (Salvi et al., 2002). While the studies have consistently found activity within the temporal lobes, associated with the selective processing of acoustical information, they have also identified a wide range of cortical structures which are activated in complex listening tasks. These structures encompass parietal, frontal, and cingulate cortices. This distributed network of regions has much in common with the "multiple demand" network of cortical areas (Duncan, 2006), identified in a range of complex cognitive tasks, and with the "posterior parietal network" of attention identified in previous studies of visual attention (Posner & Petersen, 1990). Thus, even simple tasks of speech perception in the presence of noise or other speech which do not include elements common to difficult listening environments, such as multiple spatially-separated talkers, elicit the activation of a broad range of cognitive functions. However, in contrast with ERP studies, the PET and fMRI methods do not provide information about the evolution of that activity over time.

This chapter presents an experiment designed to examine the nature of cortical activity during an adapted version of the complex task of spatial listening used in Experiments 1–5. Compared to previous neuroimaging studies of 'cocktail-party' listening, this task provided a closer approximation to everyday situations in which listeners report difficulties. Magneto-encephalography (MEG) was used due to its ability to study neural activity in both the spatial and temporal domains at a high resolution (Chapter 4). Activation was examined at several 'key moments' during the spatial listening task.

The first key moment was the onset of each new phrase in the sequence of phrases. The results of Experiments 1–4 suggested that a new phrase onset 'grabs' attention—onsets invoke a stimulus-driven shift of attentional focus. The first hypothesis was that the onset of a new phrase induces a pattern of cortical activity independent of the target or non-target call-sign which follows it.

The second key moment was the onset of the target and non-target call-sign keywords. Successful performance on the task requires the listener to focus attention on the call-sign and examine whether or not it matches their target call-sign. The second hypothesis was that there should be distinct differences between the attentional processes elicited by the recognition of the target call-sign compared to the non-target call-sign. In the case of the target, attention would have to be sustained on the talker who spoke the call-sign and/or the spatial location of that talker. For the non-target call-sign, the listener would first have to detach their attention from that talker and/or spatial location. This detachment may involve the cessation of focussed attention on the current phrase, the deliberate disconnection of attention from the phrase, or both. The listener must then successfully adopt an attentional state which would allow them to detect the onset of the next phrase in the sequence. As a consequence, there should be differences in cortical activation at this key moment and those differences should be associated with cognitive processes related to selective focussing of attention, sustaining attention, and imposing top-down control on attentional focus.

The third hypothesis was that there would be two phases after the onset of the target call-sign in which cortical activity would differ compared to the activity which follows non-target call-signs. Subjective experience from performing the spatial listening task suggests that there are two distinct challenges when attending to the target phrase and hearing out the colour and number information within it. First, one must focus attention on the talker who spoke the phrase, requiring a shift of attentional focus, and analyse the call-sign keyword. If it is the target call-sign, attention must be sustained on the phrase; if it is a non-target call-sign, attention must be detached from the phrase or allowed to be 'captured' by the next phrase in the sequence of phrases. Second, one must resist distraction from the phrase onset which follows the target call-sign. It was therefore hypothesised that differences in the responses to target and non-target call-signs would be observed shortly after the onset of the call-signs and several hundred milliseconds later around the time of the onset of the phrase which follows the call-sign.

The fourth hypothesis was that these two 'phases' of processing would exhibit different patterns of cortical activity. It was expected that the differences in activity within the first phase should be related to the recognition of the target call-sign, and sustaining attention on the target phrase. The second phase related to resisting distraction from the phrase which follows the target call-sign. When compared to a phrase onset to which listeners must attend, i.e. one which follows a non-target call-sign, cortical activation related in time to the onset they must resist distraction from should be localised to regions associated with shifts of attention or the inhibition of such processes. This hypothesis was examined by an analysis of the activation at a third key moment—the onset of the phrase which followed the target call-sign.

Hypotheses 1–4 implied that the recruitment of a range of attentional processes at key moments would be critical for successful performance on the spatial listening task. The last hypothesis was therefore that indices from the MEG data, such as the magnitude or latency of cortical activation, extracted at key moments in the task would predict performance, and that such correlates would be localised to cortical regions implicated in attentional processing.

7.1.1 Summary of Hypotheses

1. The onset of a new phrase induces a pattern of cortical activity independent of whether it is followed by a target or non-target call-sign.
2. Different patterns of cortical activity are induced by target call-signs compared to non-target call-signs.
3. These differences in activation are divided into two 'phases'.
4. These processing phases exhibit different foci of activation related to:
   - recognition of and focussing attention on the call-sign (First phase)
   - resisting distraction from the following phrase onset (Second phase)
5. Parameters of the MEG data, such as the amplitudes or latencies of peak differences in activation between conditions, are correlated with task performance.

7.2 Methods

7.2.1 Participants

The participants were the same as those described in the previous chapter. The study was approved by the Ethics Committees of the Department of Psychology of the University of York and the York Neuroimaging Centre. All volunteers signed an informed consent according to the guidelines of both ethics committees and all other documents that were a required part of the procedures as dictated by the guidelines of the York Neuroimaging Centre.

7.2.2 Spatial Listening Task

To maximally expose individual differences in performance, the “1 by 1” ‘None’ task from Experiment 1 was chosen as the spatial listening task to be performed during MEG imaging. This condition had given rise to the largest variability in SRTs across participants in Experiments, 1, 3, 4, and 5.

Several aspects of the task were modified so that it could be performed while participants were in the MEG scanner. First, to increase the number of phrase sequences that could be presented within a short time period, the number of phrases in each sequence was reduced from 13 to 7. The target phase occurred in one of the central three slots in the sequence (Figure 7.1).

Second, due to the limited number of response buttons available within the scanner, only CRM phrases containing the numbers from one to four were used for the MEG listening task. The number of possible responses that participants could

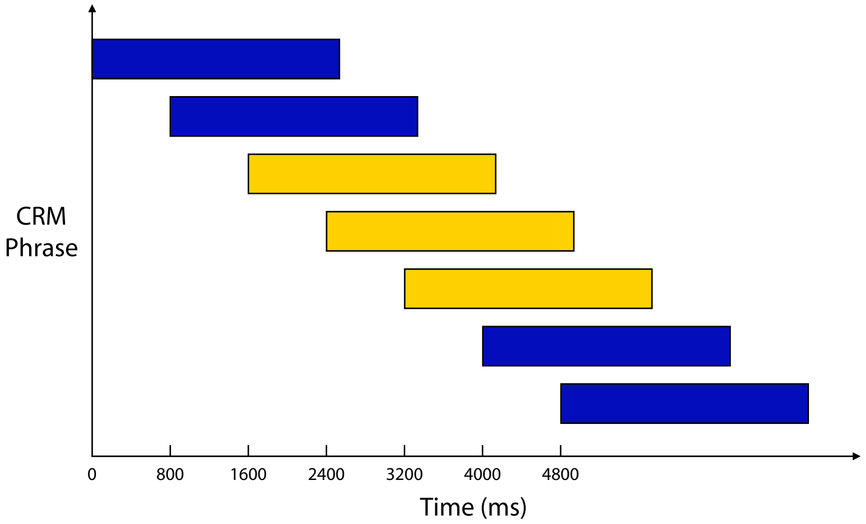

Figure 7.1. A schematic illustration of the overlapping sequence of phrases in Experiment 6. A new phrase started every 800 ms. The target appeared in one of the three central slots (yellow). Maskers were positioned in the remaining slots. The three central phrases had unique talkers, call-signs, colour, and number keywords.

choose from was therefore equal for both parts of the response; i.e. 4 colours and 4 numbers.

Third, due to the restrictions on the length of time that the participants could be in the scanner, an adaptive procedure was not used. Instead, the target phrase was always presented at a fixed level of −3 dB relative to the non-target phrases. This level difference was chosen during pilot testing to make the task challenging, without being impossibly difficult, and thus encouraged the participant to maintain a high-degree of attentional focus.

In all other respects, the task was identical to the "1 by 1" 'None' task from Experiment 1. Participants were allocated the same target call-sign that had been allocated to them when they performed the spatial listening task in the laboratory; i.e. 'Baron', 'Tiger', or 'Ringo' (Chapter 6, Section 6.2.2, p. 138).

7.2.2.1 Training

Due to the sensitivity of the MEG to metallic objects, silicone tube-phones (ER30 insert earphones, Etymotic Research, Illinois, USA) were used to present the stimuli. To familiarise participants with the tube-phones and the sensation of listening to the binaurally-recorded stimuli, they performed the new version of the spatial listening task in a sound-attenuated booth prior to the MEG session. Feedback on the accuracy of responses was provided on a computer screen, and responses were collected using a keyboard.

7.2.3 Stimuli & Recordings

The stimuli were created by presenting phrases through the circular loudspeaker array used in the previous experiments. Seven loudspeakers were used, equally spaced at 30° intervals (Figure 7.2). This arrangement produced a range of locations from −90° to +90°.

Binaural recordings of the 7-phrase sequences were made using a Brüel & Kjær Head and Torso simulator (HATS). The HATS was placed in the middle of the loudspeaker array so that its in-ear microphones were approximately level with the loudspeakers (Figure 7.3). The non-target phrases were presented at a fixed level of 70 dB SPL. The target phrase was presented at a fixed level of −3 dB relative to the non-target phrases. Recordings were made at a sampling rate of 44.1 kHz with 16-bit amplitude quantisation using a portable recording device (Marantz PMD670). A group of 96 7-phrase sequences were recorded for each of the three target call-signs ('Baron', 'Tiger', or 'Ringo') producing a total of 288 stimuli. Each group of 96 sequences comprised 3 sub-groups of 32 stimuli. In each sub-group, the target call-sign occurred in either the 3^{rd}, 4^{th} , or 5^{th} slot in the sequence.

7.2.3.1 Tube-phone Compensation

It was necessary to compensate for the frequency response of the tube-phones. The frequency response of the ER30 earphones declines at approximately 10 dB/octave above 1 kHz. Therefore, the stimuli were pre-emphasised to approximate a flat response when presented through the tube-phones. Prior to pre-emphasis, recordings were made of a complex tone with a fundamental frequency of 100 Hz and 59 harmonics presented through the tube-phones using a Brüel & Kjær ear canal coupler (Coupler Type 4157, Ear canal extension Type DB2012) and sound level meter (Type

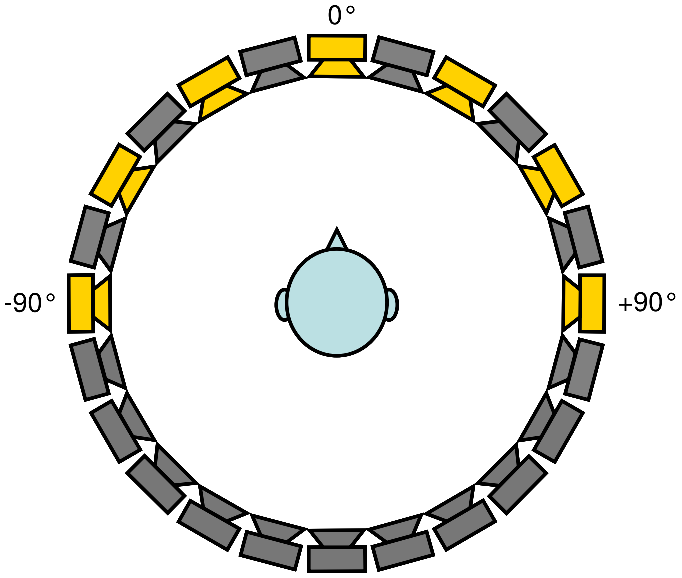

Figure 7.2. A schematic plan of the loudspeaker array used to create the binaural stimuli for the experiment. Only the subset of loudspeakers in yellow was used.

Figure 7.3. The head and torso simulator placed in the centre of the loudspeaker array. The height of its in-ear microphones and the loudspeakers were approximately equal.

2260 Investigator). Using the FFT of the recorded tone, an attenuation value for each of the harmonics was calculated relative to the frequency with maximum output level (Figure 7.4). These values were then inverted and applied as coefficients to a digital filter which pre-emphasised the recordings (Cusack & Long, 2006). The resulting stimuli were low-pass filtered at 5 kHz and all extraneous silence at the beginning and end of the stimuli was removed using digital editing software (Cool Edit 2000, Syntrillium Software).

7.2.3.2 Keyword Onsets

The onset times of key-words were located in the digital recordings of the stimuli. For each stimulus, the onset of each phrase in the sequence of phrases and the onset of the call-sign in each phrase were determined manually relative to the start of the stimulus (Figure 7.5). These processes yielded at total of 4032 keyword onset latencies. The timing information facilitated the analysis of the MEG data relative to the keyword onsets.

7.2.4 MEG & EEG Recordings

Auditory-evoked and -induced magnetic fields were recorded by means of a 248-channel magnetometer-based whole-head MEG system (Magnes 3600 WH, 4-D Neuroimaging, California, USA). Bipolar vertical and horizontal electro-oculograms

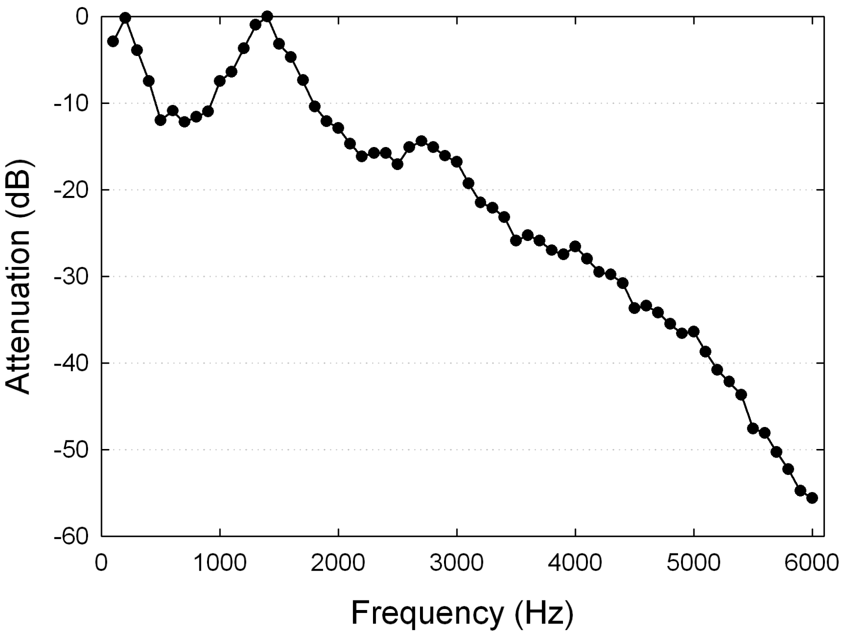

Figure 7.4. Normalised attenuation values for each of the 59 harmonics of a complex tone recorded through the ER30 tube-phones. The values were used to estimate the frequency response of the tube-phones and to pre-emphasise the stimuli to compensate for attenuation at high frequencies.

(EOG) were recorded from four electrodes attached to the left and right outer canthus and above and below the right eye. An electro-cardiogram (ECG) was recorded via two electrodes, one on each forearm. EOG and ECG data were acquired using high-speed bipolar EEG electrodes (SynAmps system, NeuroScan, Texas, USA).

Auditory stimuli were delivered by a stimulus computer using custom software written in the Python programming language (Python Software Foundation, 2007) and were presented through a soundcard (MOTU PCI 2048). A separate channel of the soundcard was used to send signals marking the onset time of each stimulus to the MEG scanner, which encoded the markers on a dedicated channel within the MEG data. The timing of the markers was adjusted to account for the time required for the stimuli to reach the ear canals of the participant (0.0104 sec). Stimuli were delivered via Etymotic ER30 tube-phones (Etymotic Research, Illinois, USA). The silicone tubes (20 ft, 4 mm internal diameter (ID)) were terminated with foam ear-tips (Etymotic ER13-14, 3 mm ID) via 90°-angled plastic tubes (2 mm ID). Responses were recorded using two 5-button Lumitouch response pads (Photon Control Inc., Vancouver, Canada). The MEG signal was DC-coupled and was digitised at a sampling rate of 678.17 kHz with 16-bit amplitude quantisation. The data were low-pass filtered online at 200 Hz. All EEG recordings were low-pass filtered at 50 Hz. The data were collected in one continuous acquisition.

To enable the functional data collected during the MEG scan to be co-registered with the participant's structural MRI scan, fiducial points (left/right pre-auricular and

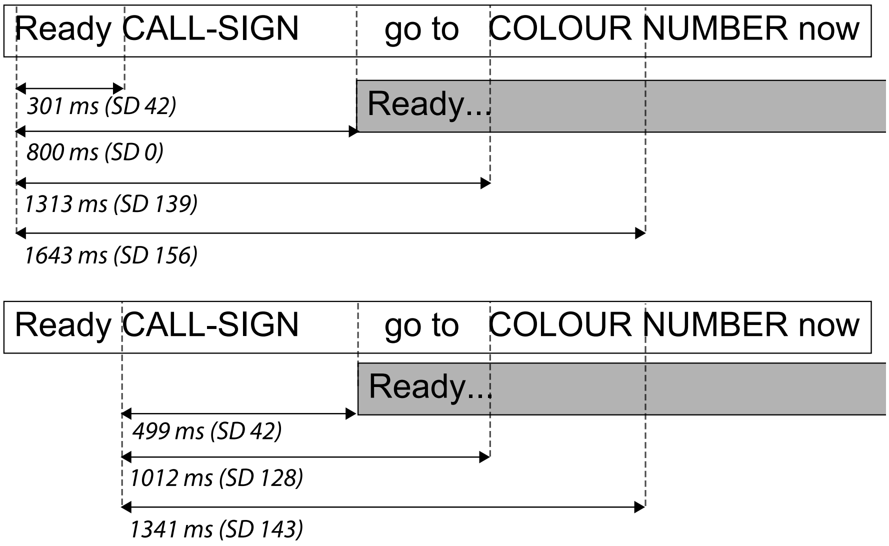

Figure 7.5. The mean and standard deviation onset times of the keywords within each phrase (white box) as measured manually across all 288 phrase sequences. The onset times are shown relative to the start of the phrase (top) and the onset of the call-sign keyword (bottom). The grey boxes indicate the relative position of the next phrase in the sequence of phrases. The time between each phrase onset ('Ready' keyword) was exactly 800 ms (top). Relative to the onset of the call-sign keyword, the next phrase started at 499 ms on average.

nasion points) were located using a 3-dimensional motion tracking system (Polhemus Fastrak, Polhemus, Colchester, VT, USA). A recording of the participant's scalp shape was also acquired using the same system while the participant was in a supine position.

7.2.5 MRI Acquisition

Structural whole-head MRI scans were recorded for each participant. T_1-weighted images were acquired using a 3 Tesla Signa Excite HDx (GE Healthcare). A Sagittal Isotropic 3D Fast Spoiled Gradient Recall Echo sequence was used with the following parameters: flip angle = 20°, TE = 3.07 ms, TR = 8.03 ms, FOV = $290 \times 290 \times 176$, matrix size = $256 \times 256 \times 176$, voxel size = $1.13 \times 1.13 \times 1.0$ mm^3.

7.2.6 Co-registration & Headmodel

To compute solutions to the MEG data at a cortical level, the positional information from MEG, including the location of the sensors and fiducial points, was co-registered with the anatomical MRI of each participant. This was achieved in a four-stage process. First, mesh models of the participant's scalp, inner, and outer skull were extracted from their structural MRI with a watershed algorithm using the Freesurfer image analysis suite (Freesurfer Development Team, 2007). The positions of the MEG sensors, fiducials, and the headshape information recorded using the Polhemus motion tracker were imported into the MATLAB ® computing environment (The MathWorks, 2007). Next, the approximate positions of fiducial points on the scalp mesh extracted from the MRI were specified manually. The MEG and MRI fiducial points were then aligned using a rigid registration algorithm (Press et al., 2002). This registration served as an initial solution for the final alignment process. The final alignment was based on the headshape acquired prior to MEG scanning and the scalp surface mesh extracted from the MRI. This alignment was performed using an iterative closest point (ICP) registration algorithm (Zhang, 1994). The resulting transform was applied to the sensor array, and the co-registration of the MEG and MRI information was visually inspected to ensure that the ICP algorithm converged on a valid solution.

Prior to the computation of the leadfields (see Chapter 4, Section 4.3, p. 73), a headmodel was computed. A multi-sphere head model (Huang et al., 1999) was created using MATLAB. The process involved computing the centre and radius of the sphere which best approximated the curvature of the inner skull below each sensor in the MEG sensor array. The location and the fit of each sphere to the inner skull model was visually inspected to ensure that the model was valid.

7.2.7 Design

All trials were presented in a single block. There were 32 stimuli for each position of the target phrase; i.e. in the 3^{rd}, 4^{th} , or 5^{th} slot within the sequence of phrases. The order of the stimuli was randomised. A total of 96 trials were presented resulting in a total scanning time of approximately 16 minutes. MEG data were collected on either the same day, or a different day, from the training data.

7.2.8 Procedure

7.2.8.1 MEG Recordings

Participants were familiarised with the magnetically-shielded room (MSR) prior to scanning. Before entering the MSR, the EOG/ECG electrodes and 5 fiducial marker coils were attached to the participant. The coils were used to locate the position of the participant's head relative to the MEG sensor array. Participants were then asked to lie on an adjustable bed in front of the scanner in the MSR. The Polhemus tracker was used to obtain the positions of the fiducial points, the coils, and the shape of the participant's head. Participants inserted the tube-phones into their ear canals with the aid of foam ear-tips. A sequence of test sounds was used to confirm that acoustic stimuli were audible in both ears individually and that a diotically-presented sound was located in the centre of the participant's head. A video camera installed inside the MSR allowed for the participant's behaviour to be monitored throughout the experiment.

7.2.8.2 Spatial Listening Task

The 96 stimuli were presented in random order. A 1-second inter-stimulus interval began at the end of each trial, which was determined either by the participant's response or the end of the stimulus—whichever came second. Participants indicated their colour choice using the four buttons on the left-hand response pad, and their number choice using the buttons on the right-hand response pad. Responses could be made at any time during a stimulus presentation. There was no time limit in which a response had to be made. Participants were instructed to maintain their fixation on a visually-present cross, displayed on a screen using a projector and a mirror (Figure 7.6). Feedback on the accuracy of each part of the response (colour and number) was also provided on this screen.

7.2.9 Preparation of MEG Data

After acquisition, the unprocessed MEG data for each participant were imported into the BESA ® software package (MEGIS Software GmbH, 2008). The data were subjected to a series of post-acquisition processing steps in preparation for all subsequent analyses. Figure 7.7 shows the steps in the order that they were performed. The following paragraphs describe each processing stage.

7.2.9.1 Post-Processing of Events

To identify and analyse the MEG data relative to the onset of key-words within the stimuli it was necessary to add the keyword onset information to the MEG data. Figure 7.8 shows the stages involved in processing the MEG events, which included stimulus onset latencies and response events. First, it was first necessary to extract the exact timing of stimulus presentations during the MEG acquisition. This information was encoded on additional channels within the MEG data during data acquisition and was then extracted using the BESA software. The keyword onset information, which had been extracted from the individual stimuli, was used to create additional markers on further channels in the MEG data. For each stimulus that had been presented, these markers encoded the exact onset times of the 7 phrases, the onset of call-sign keywords, and whether each phrase contained a target or non-target call-sign. The resulting events were imported into BESA and integrated with the MEG data.

7.2.9.2 Defining Analysis Windows

Five analysis windows were defined using the phrase and keyword onset markers that had been inserted into the MEG data. Figure 7.9 shows a schematic representation of

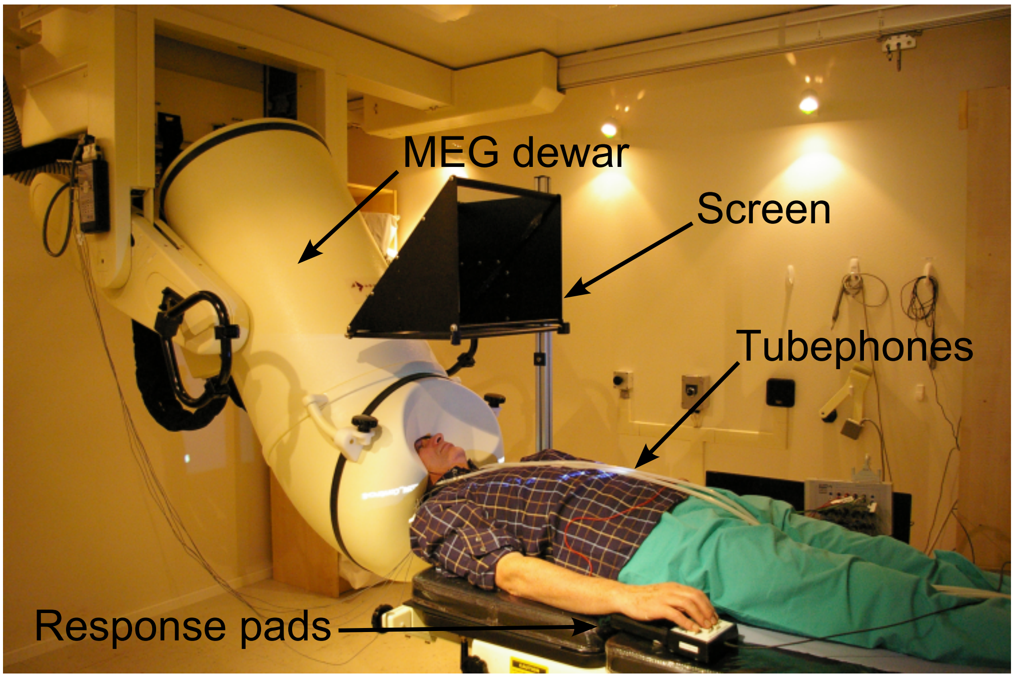

Figure 7.6. The magnetically-shielded room containing the MEG scanner. The participant lay supine with their head inside the dewar of the scanner and made their responses using two response pads. Stimuli were delivered via tube-phones and feedback based on responses was provided on a screen suspended above the participant.

the windows of interest. Windows were selected around the onsets (the word 'Ready') and the call-signs of target phrases, and of non-target phrases which appeared before the target phrase in the 3^{rd} or 4^{th} slots. For example, if the target appeared in the 5^{th} slot, non-target data windows were extracted from the phrases in the 3^{rd} and 4^{th} slots (Figure 7.10). Windows were not extracted from phrases in the 1^{st} or 2^{nd} slots because those phrases were overlapped by fewer phrases compared to phrases in slots 3–5. The onset, or 'Ready', windows were 250 ms in length, and the call-sign windows lasted 1000 ms.

A total of 96 phrase onset and call-sign windows were defined for target and non-target phrases. Windows were defined for the 32 target phrases which appeared in the 3^{rd}, 4^{th}, and 5^{th} slots. For the non-target phrases, windows were defined for 64 phrases which appeared in the 3^{rd} slot in the sequence of phrases and for 32 phrases which appeared in the 4^{th} slot. This resulted in an equal number, 96, of target and non-target windows for the onset and call-sign windows.

The windows were arranged into three pairs, as indicated by the window numbers

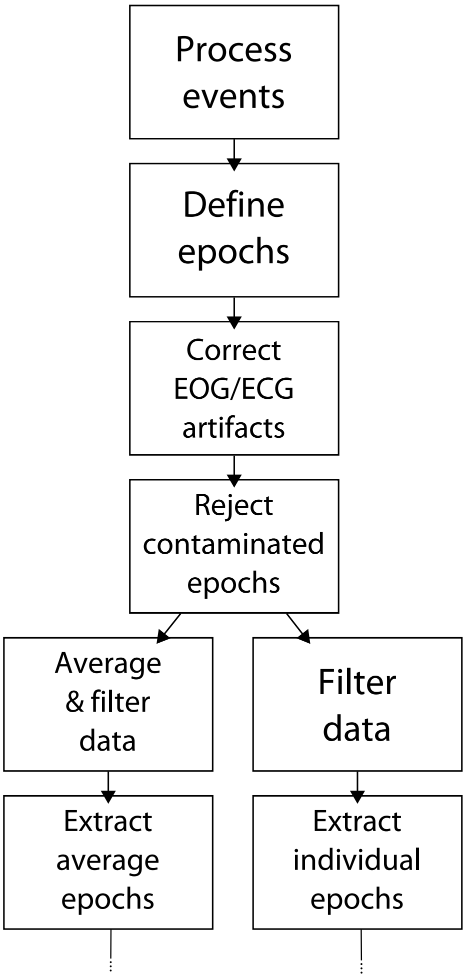

Figure 7.7. The sequence of processing steps to which all MEG data were subjected before sensor- or source-space analyses.

in Figure 7.9. The first epoch pair comprised the response to the word ‘Ready’ when it immediately preceded a target call-sign and when it preceded a non-target call-sign (1 vs. 2). This pair was termed “Onset processing”. The second pair comprised the response to the onset of the phrase which followed a target call-sign and the onset which preceded a target call-sign; i.e. the onset of the target phrase (3 vs. 2). This pair was termed “Onset attention”. The third pair involved the response to a target call-sign and to a non-target call-sign (4 vs. 5). This pair was termed “Call-sign processing”. Subsequent analyses were focussed on these three pairs of windows.

7.2.9.3 Artifacts

EOG and ECG Artifact Correction
The BESA software package provides the facility to remove artifacts from MEG data which result from activity of the ocular and cardiac muscles. The method employed is based on a surrogate model of brain activity (Berg & Scherg, 1994; Ille, Berg, & Scherg, 1997, 2002) and involves the following processing stages (MEGIS Software GmbH, 2008):

1. Identify the sensor topography associated with the artifact.
2. Model the spatial topography associated with the artifact.
3. Reconstruct the activity of the artifact at the sensor level using a dipole model

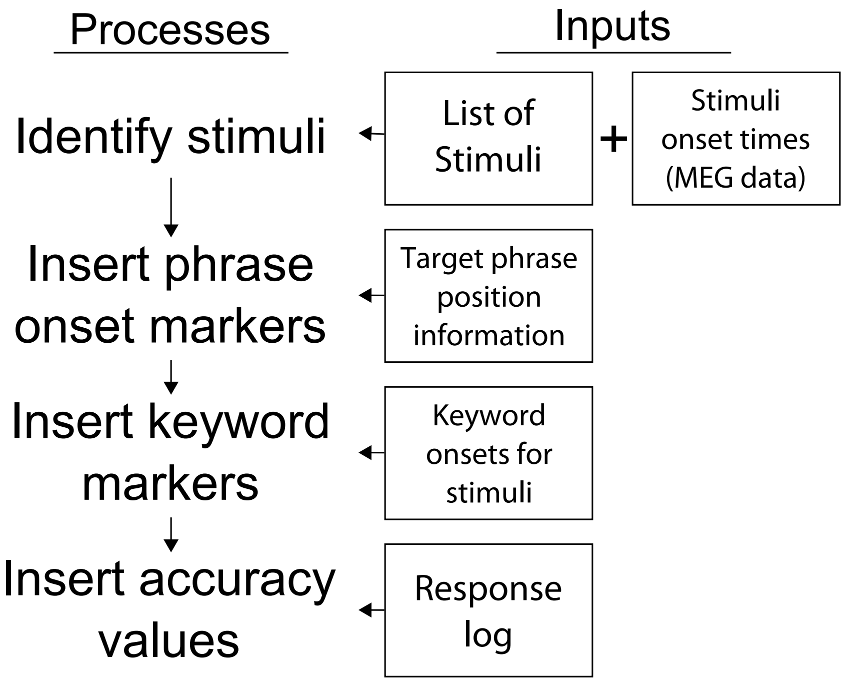

Figure 7.8. The steps involved in adding markers into the MEG data to identify keyword and phrase onsets within the individual stimuli. Information was also encoded in the MEG data which indicated whether the responses were correct or incorrect.

of the underlying brain activity to estimate the artifact and brain signal subspaces.

4. Remove (subtract) the artifact signals from the measured MEG signals.

The EOG and ECG data acquired simultaneously with the MEG data were used to identify the sensor topography of eye and cardiac artifacts. For eye blinks, the MEG and EOG data were digitally filtered from 0.5–8 Hz. First, an epoch from −100 to +400 ms relative to the onset of a specimen eye blink was manually selected using the EOG data as a guide. The search function in BESA was then used to identify subsequent artifacts based on the continuous EOG data. Each exemplar blink was inspected manually to check that it did not coincide with another artifact, such as a low-frequency drift or movement caused by swallowing. Up to 50 blink topographies were averaged. The total number was limited by the number of artifacts in the data.

A similar procedure was carried out to extract an average topography for the cardiac cycle. The data were digitally filtered from 5–12Hz. An epoch from −200 to +500 ms relative to the negative peak corresponding to the *Q* wave of a specimen cardiac cycle was selected using the ECG data as a guide. The *Q* wave marks the start of the *QRS* complex of the human ECG which signals the depolarisation of the

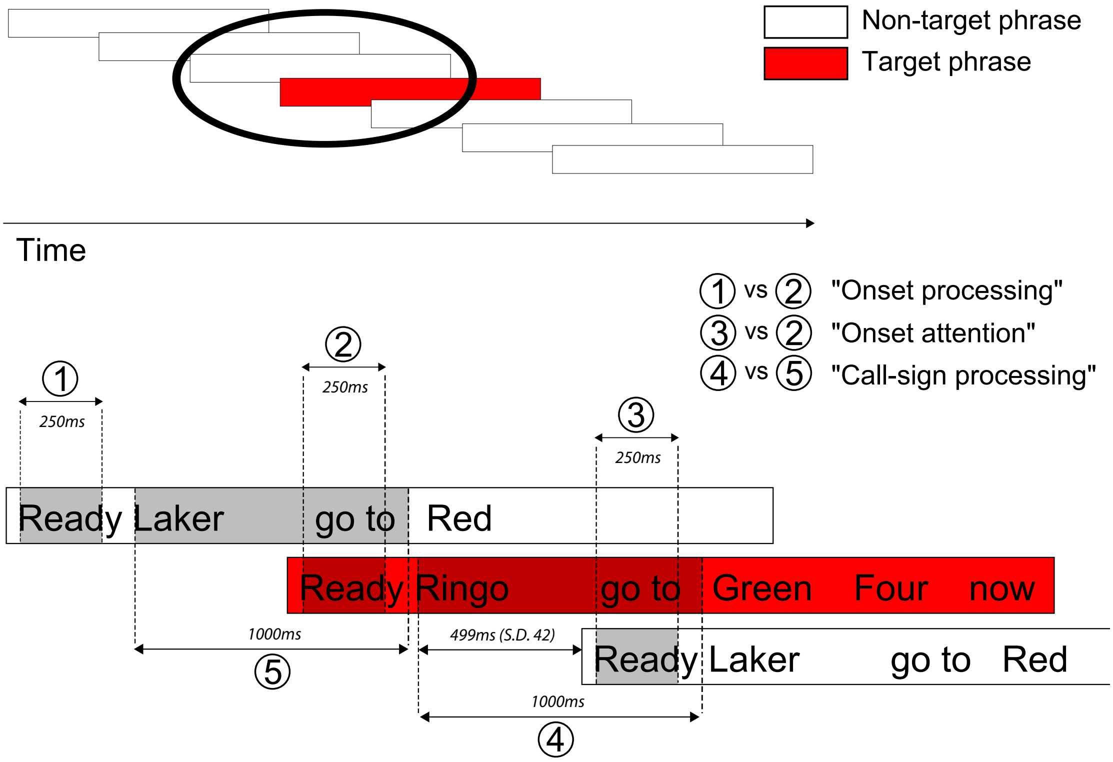

Figure 7.9. Top: The sequence of phrases within each stimulus. The target stimulus is shown in red. Bottom: magnified section of the sequence marked with the oval. Windows of interest synchronised to the onset of the 'Ready' and call-sign keywords were extracted from the MEG data for the target phrase and non-target phrases which preceded the target. The analyses of the MEG data compared pairs of epochs, indicated by the numerical labels (see Section 7.2.10.3).

ventricles. Unlike the process for identifying blinks, subsequent identification of cardiac artifacts based on the ECG data was done automatically due to the regularity of the cardiac cycle. Once all cardiac artifacts had been identified, an average topography was calculated based on the unfiltered MEG data. The unfiltered data were averaged to include the large number of frequency components in the cardiac artifacts.

Principle Component Analysis (PCA) was used to extract a sufficient number of topographic components which together explained over 90% of the variance in the average artifact data. For for blink artifacts, a single component was sufficient to reach this threshold. For cardiac artifacts, two components were required to exceed the threshold. The selected components were then used to estimate the contribution of the artifacts to the measured MEG data using the surrogate model correction function in the BESA software.

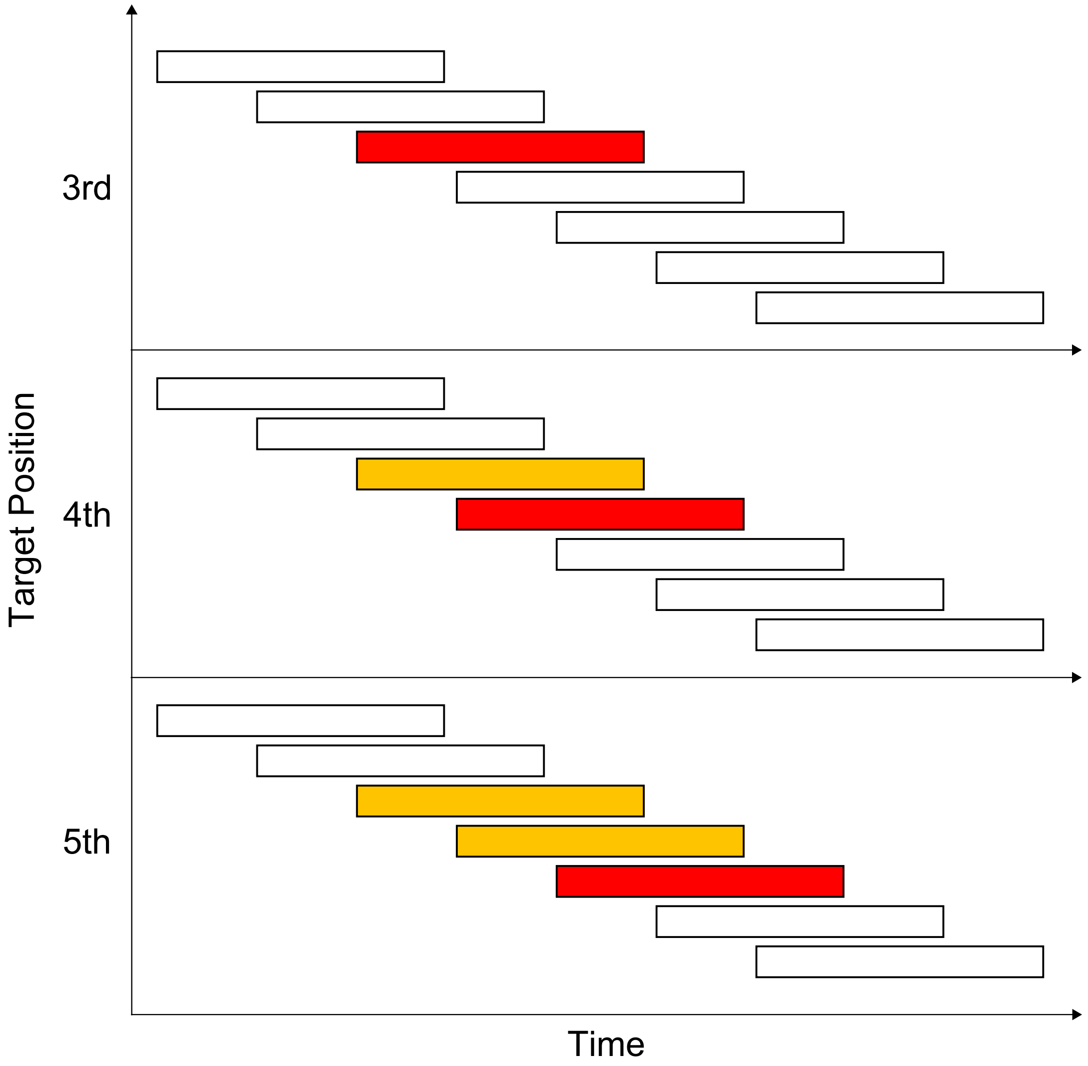

Figure 7.10. A schematic diagram of how data windows in the non-target phrases (orange) were chosen relative to the phrase containing the target call-sign (red). Windows were defined for phrases in slots 3–5 only which were all overlapped by the same number of phrases.

Rejection of Other Artifacts

Prior to averaging the MEG data within each of the epochs, the data were scanned for other perturbations which would obscure signals originating from the brain. These perturbations could have arisen from head or response-related movements, swallowing, etc. Artifact-contaminated epochs were identified using two criteria. The first criterion was peak-amplitude. Any epochs exhibiting signal levels above 3.5 pT were rejected. The second criterion was based on the rate of change of the MEG data. Any epoch which contained a difference between two adjacent samples greater than 2.5 pT was rejected.

7.2.9.4 Filtering

Before subjecting the averaged epochs or the individual epochs to further analysis, the data were digitally filtered into several standard frequency bands: 0.5–4 Hz (delta, δ), 4–8 Hz (theta, θ), 8–13 Hz (alpha, α), 13–30 Hz (beta, β), and 40–80 Hz (gamma, γ). The choice of frequency bands was determined by the requirements of each analysis and is detailed in the subsequent MEG analysis sections. The BESA software package was used to filter the data. Zero-phase Butterworth filters were used with a slope of 12 dB/octave for their low and high cut-offs. For the γ band, an additional band-stop filter centred on 50 Hz with a width of 2 Hz was used to remove electrical noise.

Steps were taken to avoid contaminating the beginnings and ends of epochs with artifacts created by the filtering process. For the evoked analyses, all epochs were averaged with several seconds of additional data before and after the required data window. This padding was removed after the data had been filtered. For the time-frequency and spatial filtering analyses, the entire MEG data set was filtered at the desired frequencies prior to the extraction of individual epochs.

7.2.9.5 Noise Estimation & Bad Channels

Estimates of the level of noise across the MEG sensor array were incorporated in the analyses of the data at the source level. A continuous 2-minute block of empty room data was acquired on the same day as each MEG scan to estimate noise levels across the sensor array. Each 2-minute block was divided into sub-blocks of 744 temporally-adjacent samples (three times the number of sensors, 248). The covariance matrix of each sub-block was calculated as:

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(X_i-\bar{X}\right)^T \qquad (7.2.1)$$

where $n = 744$ or the number of samples in the sub-block, ^{T} is the transpose operation which converts a row vector to a column vector, X_i is a row vector with 248 values representing the i^{th} sample of empty room data within the sub-block, and

$\bar{X}$ is sub-block mean defined as:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7.2.2}$$

The covariance matrix for the entire 2-minute block of empty room data was then calculated as the average covariance matrix across the sub-blocks. This technique prevented low frequency drifts from contaminating the noise estimates.

Bad channels in the MEG sensor array were identified for each MEG session using the noise estimates. The lead diagonal of the covariance matrix for the empty room data represented an estimate of the noise variance at each sensor. Any channel which exhibited a noise variance value which was more than 3 standard deviations away from the mean variance across all sensors was designated as bad. Bad channels were identified on a per-acquisition basis, and were removed from all subsequent analyses.

7.2.10 Analysis

7.2.10.1 Behavioural Performance

Performance on the spatial listening task was analysed as the percentage of stimuli for which the colour and number keywords in the target phrase were correctly identified. Differences between the two age groups were examined using an independent-samples t-test. The Kolmogorov-Smirnov and Levene's tests were used to check that the assumptions of normally distributed data and equality of variances between the two age groups had not been violated. Significance was assessed at the two-tailed level. Visual inspection of boxplots was used to identify possible outliers in the performance data. Any score with a corresponding z-score greater than 3 was removed from further analysis.

7.2.10.2 MRI Analysis

To perform an analysis on the MEG data at the cortical level, a model of the cortical surface was required for each participant. The Freesurfer image analysis suite (Freesurfer Development Team, 2007) was used to reconstruct a 3-dimensional mesh of the boundary between the grey matter and white matter. The reconstruction process involves stripping away all the material around the brain (Ségonne, Dale, Busa, Glessner, Salat, Hahn, & Fischl, 2004), segmenting the white matter and deep structures such as the hippocampus (Fischl, Salat, Busa, Albert, Dieterich, Haselgrove, van der Kouwe, Killiany, Kennedy, Klaveness, Montillo, Makris, Rosen, & Dale, 2002; Fischl, Salat, van der Kouwe, Makris, Ségonne, Quinn, & Dale, 2004), and creating tessellated surfaces at the boundaries between the grey & white matters and between the grey matter and cerebral spinal fluid (Dale & Sereno, 1993; Dale, Fischl, & Sereno, 1999; Fischl & Dale, 2000). The grey-white matter boundary surface of each participant was inflated to a sphere and registered to a standard anatomical atlas

(Fischl, Sereno, Tootell, & Dale, 1999b). This transformation enabled solutions to be averaged across participants and also provided an MNI Talairach co-ordinate transformation for each participant.

To perform the source analyses, a discrete number of source locations was required. By default, the cortical mesh models created by the Freesurfer software are highly detailed, comprising over 200,000 vertices. To compute source solutions efficiently, a smaller number of source locations was required. The MNE software suite (Hämäläinen, 2007) was used to decimate the grey-white matter model. This was achieved by inflating each hemisphere of the individual's cortical model to a sphere and registering it to a five-times recursively divided icosahedron. The resulting icosahedron always contained 10,242 vertices. This created a decimated cortical model with an approximate source spacing of 3.1 mm.

Cross-participant averaging and statistics were performed by morphing the individual source analysis solutions, created using the decimated individual cortical models, to the cortical mesh of an average subject based on a 305-subject average MRI volume (Evans, Collins, Mills, Brown, Kelly, & Peters, 1993). Figure 7.11 shows the processing stream for the cortical models and source solutions composed using those models. Individual source-space solutions on the decimated cortical mesh were smoothed onto the original detailed cortical mesh of the same participant using 3

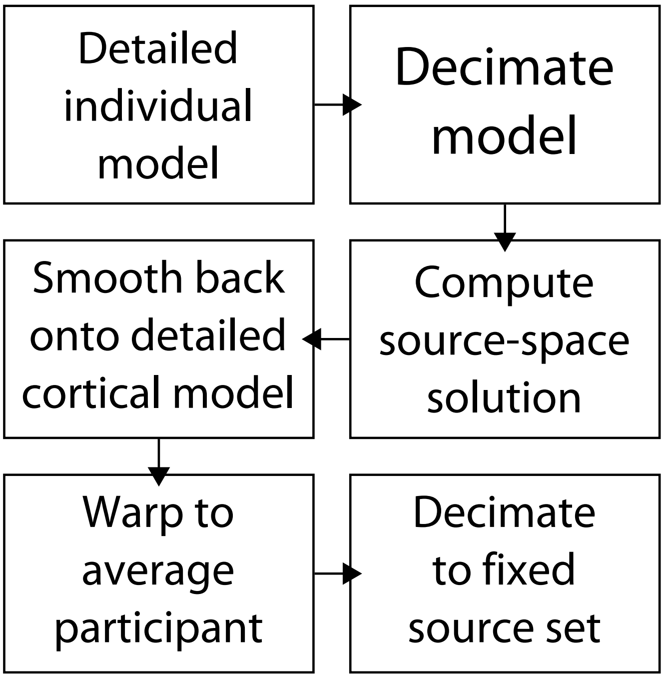

Figure 7.11. The processing stream for the 3-dimensional cortical models extracted from each participant's MRI. Decimated models were used to compute source solutions for computational efficiency. A warping process allowed for the averaging of source-space solutions across participants.

iterations of a linear smoothing method, and the resulting smoothed solution was morphed to the average subject's cortical mesh (Hämäläinen, 2007). A decimation procedure, identical to that used to decimate the individual cortical surfaces, was used to reduce the number of source points on the smoothed average solution to 10,242 vertices. The source solution was decimated by selecting those solution values on the spherical surface which were nearest to the icosahedral vertices.

In summary, the source solution for each individual was morphed to the cortical surface of an average subject and reduced to a fixed number of data points with identical spatial locations across participants. This facilitated the computation of averages and statistical maps. For display purposes, source solutions and statistical maps were overlaid on the partially inflated cortical surface of the average subject (Fischl, Sereno, & Dale, 1999a).

7.2.10.3 MEG Analysis: Overview

The MEG analyses were designed to identify differences in cortical activity at key moments during the spatial listening task. Analyses were performed in sensor and source spaces. The sensor-space analyses were carried out directly on the artifact-corrected sensor data. This approach did not take into account the position of the participant's head within the scanner. The source-space analyses reconstructed the activity measured with MEG as activation at the cortical level. The position of the participant's head within the scanner was factored into the source-space analysis along with the cortical anatomy of each individual.

The MEG analyses examined the aspects of the signal which were tightly phase-locked to the onset of the keywords (*evoked* data) and which were not strictly phase-locked to those onsets (*induced* data). All analyses compared pairs of epochs which were aligned either with the onset of a phrase (the word 'Ready') (the "Onset attention" and "Onset processing" comparisons) or with the onset of a call-sign keyword (the "Call-sign processing" comparison) (see Section 7.2.9.2).

The "Onset processing" analysis compared the activity related to the onset of the target phrase with the activity related to the onset of a non-target phrase. This analysis tested whether there were differences in cortical activity between target and non-target phrases associated with attending to a new phrase onset prior to the call-sign. The analysis was performed to check that no strong systematic differences in activity arose until the target call-sign was presented. This analysis also examined whether the data contained order effects which may have arisen due to the inclusion of target phrases that occurred in later positions in the phrase sequences compared to the non-target phrases (Section 7.2.9.2).

The "Call-sign processing" analysis compared the activity related to a call-sign keyword when it matched the target and when it did not match the target. This analysis identified the processes related to the recognition of the target call-sign, and the activity which resulted as a consequence of that recognition. The analysis was performed to test three hypotheses: 1) target and non-target call-signs are associated with different patterns of cortical activity, 2) two temporally-distinct 'phases' of processing follow the onset of target call-signs, 3a) the first phase involves cortical regions associated with the recognition and focussing of attention on the target call-sign, and 3b) the second phase involves cortical regions associated with resisting distraction from the onset of the following phrase.

The call-sign analysis encompassed the onset of the following phrase, which started at approximately 500 ms after the onset of the call-signs (Figure 7.9). This inclusion allowed for the analysis of the processes related to resisting distraction from the phrase onset following the target call-sign relative to the onset of the call-sign keyword; i.e. the hypothesised second 'phase' of processing (hypothesis 3b above). Due to the natural variation in the timing of the phrases across the 8 talkers, the latency between the onset of a call-sign and the onset of the phrase which followed it varied within the stimuli (Figure 7.5). Therefore, a third comparison was made.

This "Onset attention" analysis compared the activity related to the phrase onset which followed a target call-sign with activity related to the phrase onset which immediately preceded a target call-sign. This analysis identified differences in cortical activation in response to a phrase onset to which participants were attending and an onset to which they were not attending. The analysis was performed to establish whether the activity related to the phrase onset which followed the target call-sign was associated with processes linked to resisting distraction.

Table 7.1 provides a summary of the three comparisons.

7.2.10.4 MEG Analysis: Sensor Space

The analysis of the sensor data determined whether there were significant differences in the power of the magnetic fields recorded with MEG between the three pairs of epochs. As the sensor analyses did not take into account the position of the participant's head inside the scanner, power data were averaged across groups of sensors to create basic spatial divisions of the sensor data. The sensor-space analyses were performed for a group of channels in each hemisphere ('Left' & 'Right' channel groups in Figure 7.12) and all channels ('Global'). The total power p for a group of n channels was calculated as:

$$p_i = \sqrt{\frac{1}{n}\sum^{n}\left(B_{n,i}\right)^2} \tag{7.2.3}$$

where $B_{n,i}$ is the magnetic field value at the n^{th} channel and i^{th} sample, and p_i is the root mean square (RMS) power for that sample.

Separate analyses were performed to examine the *evoked* and *induced* aspects of the MEG data. Evoked differences between the two epochs in each of the three epoch pairs were examined using the average data for each data epoch (see Figure

7.7). Significant differences were determined using a method of permutation testing. Computing *t*-tests between the epochs at each data sample would have inflated the family-wise error rate (FWER) as the values of adjacent samples are likely to be correlated. Cluster-based permutation testing controlled for the inflation of the FWER due to the computation of multiple statistics across related samples (Maris & Oostenveld, 2007).

The mean power across the chosen group of sensors was calculated for each participant and for each sample within the two epochs to be compared. The data were permuted by relabelling the data at an individual level; i.e. randomly altering which epoch the data were associated with. For each permutation of the data, a maximum cluster statistic was determined as follows:

1. Paired *t*-tests were performed on the power data between the epochs at each sample.
2. The *t*-values were subjected to threshold which was specified a priori.

Epoch Pair	Summary
"Onset processing"	1. Compared phrase onsets occurring before target and non-target call-signs. 2. Examined whether attending to a new phrase onset produced a similar pattern of cortical activity when it was followed by a target call-sign or a non-target call-sign.
"Call-sign processing"	1. Compared target and non-target call-signs. 2. Examined whether there was a difference between the activity related to target and non-target call-signs and whether that activity was organised into two temporally-distinct 'phases'. 3. Tested whether differences between the activation associated with target and non-target call-signs involved processes of recognition and focussing of attention (first phase) and resisting distraction (second phase).
"Onset attention"	1. Compared the phrase onset which follows the target call-sign to the onset which directly preceded the target call-sign. 2. Examined differences in cortical activation between attended and non-attended phase onsets. 3. Tested whether the phrase onset following the target call-sign was associated with activity in cortical regions linked to resisting distraction.

Table 7.1. Summary of the three pairs of epochs which were compared using a variety of analysis methods in sensor- and source-space.

3. Samples whose absolute t-value exceeded the threshold were grouped into clusters based on temporal adjacency.

4. For each cluster, the sum of the t-values within the cluster was calculated. This value was denoted the “cluster statistic”.

5. The absolute value of the largest cluster statistic for the current permutation was recorded.

The threshold value was chosen as the t-value which corresponded to a two-tailed α-level of 95% for the appropriate degrees of freedom. The sequence of 5 steps was repeated for each permutation of the data, excluding the original ordering of the data. This process yielded a distribution of maximum cluster statistics. The clustering process was then performed on the original ordering of the data. Clusters from the original data ordering whose absolute statistic was larger than 95% of the observed maximum cluster statistics were selected as being statistically significant. Figure 7.13 shows an example of the output from the permutation analysis.

The *induced* (non-phaselocked) activity was examined using a time-frequency analysis. Prior to averaging, the data for each occurrence of an epoch in the MEG data

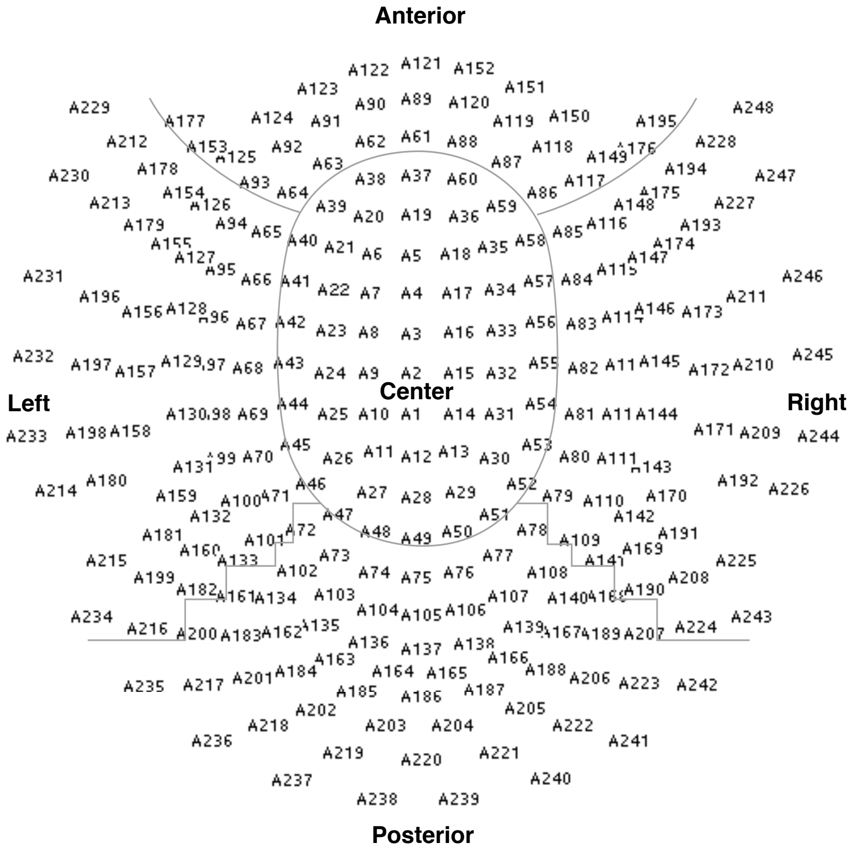

Figure 7.12. A 2-dimensional map of the 248 magnetometers in the MEG sensor array. Average magnetic field power was calculated for the ‘Left’ and ‘Right’ sub-groups of the sensors using the spatial divisions shown (after 4-D Neuroimaging, 2005).

were transformed into a time-frequency representation using the Stockwell transform (Stockwell, Mansinha, & Lowe, 1996). The frequency and temporal resolution of the transform vary as a function of frequency—the temporal resolution increases and frequency resolution decreases at higher frequencies. Each epoch was padded with extra data by extending the data window around the epoch prior to performing the transform. The padding minimised 'edge' artifacts resulting from the transformation.

To identify significant differences in induced power between the epochs in each pair, a method of permutation testing was used similar to that which was applied to the evoked analysis. A single-threshold test (Nichols & Holmes, 2002) was used for the induced data comparisons. A cluster-based test was not appropriate due to the non-uniform resolution of the Stockwell transform over time and frequency (Stockwell et al., 1996). The statistical test involved the computation of *t*-tests between the epochs at each time-frequency sample in the Stockwell transforms. This created a

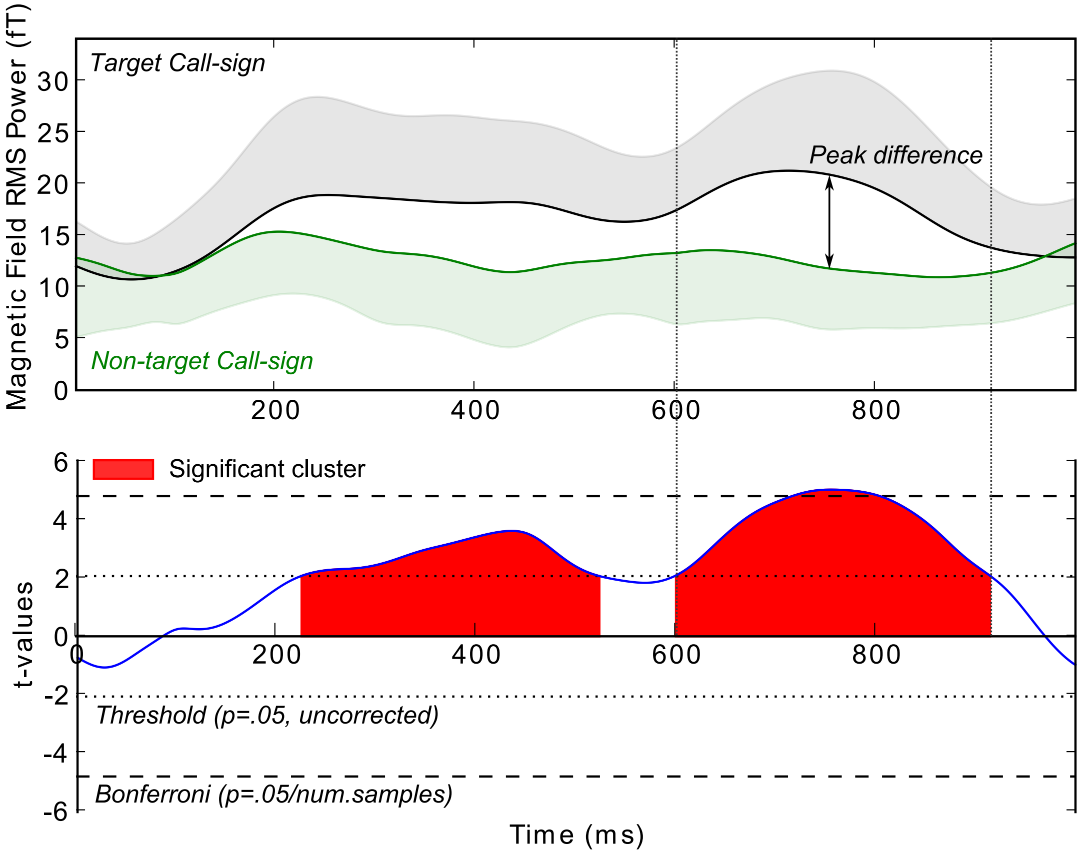

Figure 7.13. Example results from the permutation analysis of the "Call-sign processing" comparison is shown. Top: the means (bold lines) and standard deviations (shaded area) for the average power evoked by target (black) and non-target (green) call-signs in the 0.5–4 Hz frequency band. Bottom: The time windows (red) identified by the permutation analysis. The dotted line indicates the threshold used for the cluster analysis. For reference, the critical t-value after Bonferroni correction based on the number of samples in the epoch is shown as a dashed line. Peak differences in power were identified within each window for the analysis of individual differences (Section 7.2.10.7).

two-dimensional map of *t* values. For each permutation of the data, the maximum statistic in the map was recorded, yielding a distribution of maximum statistics. The critical threshold was selected as that *t*-value which was larger than 95% of the observed maximum statistics. Any time-frequency samples exceeding this threshold in the *t*-value map for the original ordering of the data was selected as statistically significant.

Selection of Frequency Bands

To select suitable frequency bands for each analysis, the average frequency spectrum of evoked and induced power across the MEG sensor array was examined using the Fast Fourier Transform (FFT). Figure 7.14 shows the frequency data for an example epoch synchronised to the onset of the target call-sign. The FFT analysis revealed that the majority of the energy in the evoked signal was below 30 Hz. The evoked analyses were therefore performed within four discrete frequency bands (0.5-4 Hz, 4-8 Hz, 8-13 Hz, and 13-30 Hz) and a broad band (0.5-30 Hz).

The induced data contained energy across a broad range of frequencies. The analysis of induced data examined a wider range of frequencies compared to the analysis of the evoked data. The lowest frequency included in the induced analysis was selected as the frequency with a period of half the length of the epoch. This

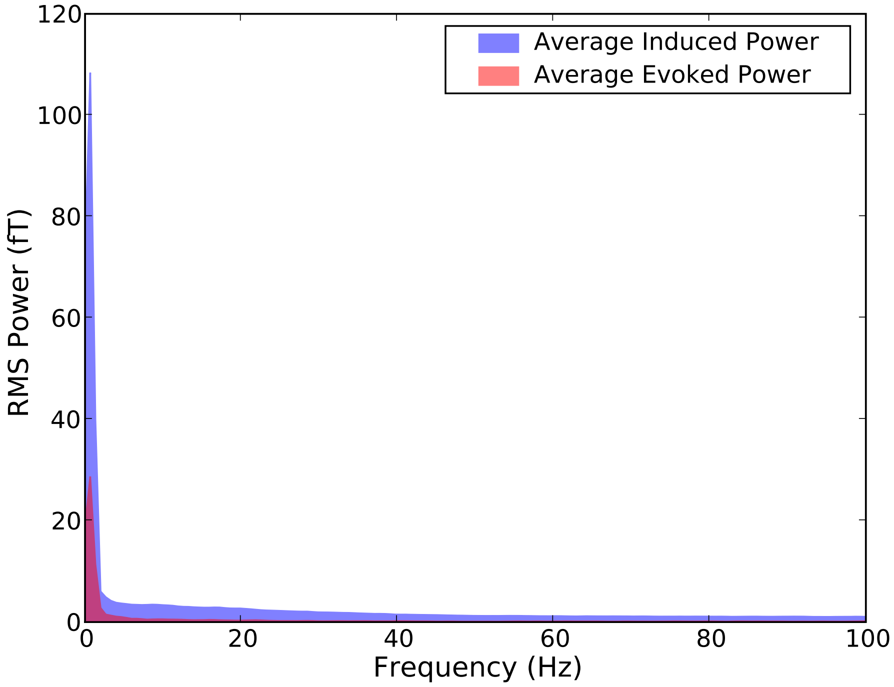

Figure 7.14. Average frequency spectrum (FFT) for the evoked (red) and induced (blue) power across the MEG sensor array. The FFTs were calculated from the response to the target call-sign keyword in the target phrase. The averaging procedure which produced the evoked data retained only a small proportion of the original signals which were tightly phase-locked to the onset of the keyword.

ensured that sufficient energy would be available in the signals at the lowest frequencies. For the 1 sec epochs synchronised to the call-sign keyword, the data were analysed from 2–100 Hz. For the epochs synchronised to the 'Ready' keyword, the 250 ms window was analysed from 8–100 Hz.

7.2.10.5 MEG Analysis: Source Space

The source-space analyses were computed using a model of the participant's individual cortical anatomy. Two contrasting methods of source analysis were used. The minimum-norm analysis examined the cortical activity which was evoked by (i.e. *phase-locked to*) keywords of interest. This method was used to identify changes in low-frequency or 'slow-wave' activity up to 8 Hz. The spatial filtering analysis examined changes in activity which were evoked and induced by (i.e. *not strictly phase-locked to*) the same keywords. This method was used to identify differences in activity at frequencies up to 80 Hz.

Minimum-norm

The aspects of the MEG data which were evoked by keywords of interest were examined at a cortical level using weighted ℓ_2 minimum-norm estimates of cortical activity (Chapter 4, Section 4.4, p. 77). As minimum-norm analysis produces a separate solution for each time sample of MEG data, it is suitable for observing changes in evoked activity over time. Solutions for the MEG data epochs were calculated for two low frequency bands: 0.5–4 Hz and 4–8 Hz.

The vertices from the decimated model of each individual's cortex were used as the source locations for the solution. To control for a bias towards more superficial source locations, the lead fields for each source location were normalised (Fuchs et al., 1999, see Chapter 4, Section 4.4.8, p. 86). A noise covariance matrix was incorporated into the computation of the solutions and was a diagonal matrix of sensor weights representing the reliability of the sensors. The sensor weights were calculated as the normalised noise variances across the sensors which had been estimated from the empty room data (Section 7.2.9.5).

The solutions were regularised to ensure that they would not be sensitive to small perturbations in the data arising from environmental or sensor noise. A method of Tikhonov regularisation was used and the regularisation parameter was estimated from a small percentage of the eigenvalues of the matrix to be inverted (Chapter 4, Section 4.4.5, p. 80). Several levels of regularisation were examined for each participant to find a parameter which did not over- or under-regularise the solution. A value of 1% was chosen which resulted in consistently stable solutions.

The individual minimum-norm solutions were warped to the cortical model of the average participant and decimated to a fixed set of source locations. This transformation yielded power values at a common set of source locations across participants. Paired *t*-tests were computed between the two conditions in each of the three epoch pairs at each source location and for each time-point in the epochs. This process created time-varying statistical parametric maps (SPMs) for each pair of epochs. This analysis determined whether there were significant differences in the mean level of activity within the epoch pairs.

The calculation of the SPMs did not correct for the inflation in the FWER due to multiple comparisons involving spatially and temporally adjacent data which were likely to be correlated. The calculation of cluster-based permutation statistics similar to those used for the sensor-space analysis would have imposed an excessive computational overhead. Therefore, the SPMs were subject to a stringent *t*-value threshold of 4.7, corresponding to an uncorrected significance level of $p < .0001$ for the appropriate degrees of freedom. These uncorrected SPMs were used to identify the location and latencies of peak differences.

Spatial Filtering

Estimates of power at a cortical level which included *evoked* and *induced* activation were estimated using a spatial filtering technique (Chapter 4, Section 4.5, p. 89). To facilitate comparisons between the different source analysis methods, the estimates of cortical power were calculated at the same set of source locations as the minimum-norm solutions. Cortical power was calculated using the form of the Neural Activity Index (NAI) proposed by Huang et al. (2004). To avoid errors in the reconstructed source activity due to high levels of correlation between bilateral auditory activity (Hillebrand et al., 2005), the NAI maps were computed separately for each hemisphere of the cortex using specific groups of sensors (Figure 7.15). The noise covariance matrix used in the calculations was estimated from the empty room data measured on the same day as each acquisition (Section 7.2.9.5).

NAI solutions were calculated for the MEG data across multiple frequency bands to examine the differences in oscillatory activity within different frequency ranges. Separate analyses were performed on the 1-second call-sign epochs at 4–8 Hz, 8–13 Hz, 13–30 Hz, and 40–80 Hz. The 250-ms ‘Ready’ epochs were analysed at 8–13 Hz, 13–30 Hz, and 40–80 Hz.

The calculation of the NAI maps required an estimate of the data covariance. The covariance matrix of the MEG data was estimated using discrete sub-windows of data within each ‘Ready’ or call-sign window shown in Figure 7.9. The choice of window length was based on the range of frequencies that were to be analysed. To ensure sufficient energy in the reconstructed signal, a window length was chosen which provided at least two complete cycles of the lowest frequency to be analysed. For epochs synchronised to a call-sign keyword, a 500-ms window was adopted to allow for the reconstruction of source activity down to 4 Hz (period of 250 ms). To examine changes in the NAI maps over time, a series of 500-ms windows was analysed within each call-sign epoch. The onset of each window was separated from the onset of the previous window by 100 ms. Therefore, in the 1000-ms epoch, the windows started at 0 ms, 100 ms, 200 ms, 300 ms, 400 ms, and 500 ms. For the ‘Ready’ windows, a window size of 250 ms was chosen due to the shorter epoch length.

SPMs were created for the three epoch pairs by calculating paired t-tests across participants at each source location. This analysis examined differences in the mean level of induced cortical activity. As was done with the minimum-norm estimates, the calculation of the SPMs did not correct for the multiple comparisons involving spatially and temporally adjacent data points. A t-value of 4.7 was used to threshold the data, corresponding to an uncorrected significance level of $p < .0001$. The locations of peak differences were extracted from the SPMs.

7.2.10.6 MEG Analysis: Summary

A summary of the MEG analysis methods is shown in Table 7.2. The analyses examined the data evoked by (i.e. tightly phase-locked to) the ‘Ready’ and call-sign keywords and the data which was induced by (i.e. *not* phase-locked to) those keywords. In sensor space, the evoked analysis was carried out on the averaged power data for each participant and the induced analysis was conducted on the averaged time-frequency data for each participant. In source space, the evoked analysis applied the minimum-norm technique to the average MEG data for each participant and the induced analysis applied the spatial filter technique to estimates of the the data covariance, estimated from individual trials. Frequency bands for each analysis were chosen based on the requirements of the analysis methods.

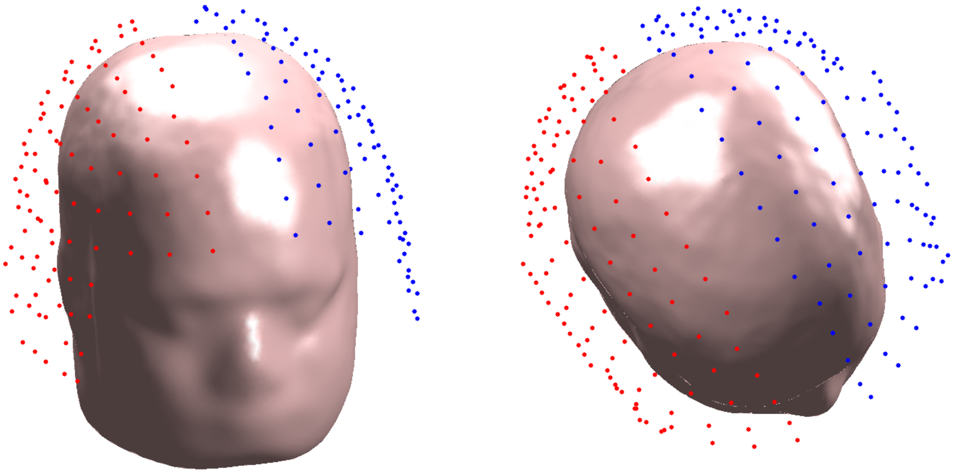

Figure 7.15. The groups of sensors over the right (red) and left (blue) hemispheres which were used to reconstruct the neural activity separately for the right and left cortical hemispheres, respectively. The use of localised sensor groups for each hemisphere reduced errors in the reconstruction of source activity due to correlated activity in bilateral auditory cortices.

7.2.10.7 Individual Differences

Relationships between performance on the spatial listening task in the MEG scanner, neuro-magnetic activity as measured by MEG, and the measures presented in Chapter 6 were assessed using rank-order correlations in the form of Kendall's τ_b. Correlations were performed across all participants, and within both age groups. Significance was assessed at the two-tailed level by converting τ_b values to z-scores using SPSS (SPSS Inc., 2006).

As outliers in a data set can influence measures of correlation (Field, 2005), including non-parametric rank-order correlations (Gideon & Hollister, 1987), visual inspection of boxplots was used to identify possible outliers which could bias estimates of correlation. The method described in Chapter 6 (Section 6.2.7.5, p. 148) was used to adjust for the calculation of multiple correlations involving the same variable. Significance values which have not been corrected are denoted by p_{uncorr}, and corrected values by p_{group}.

MEG Power Differences

The individual differences analysis included data extracted from the MEG power analyses in sensor- and source-space. For each participant, the peak difference in evoked power was identified within each significant time window from the permutation analysis of the evoked sensor data (Figure 7.13). The magnitude and latency of the peak difference was then extracted. This processes yielded an amplitude and latency measure for each individual, for each time window.

Analysis	**Data**	**Frequency band (Hz)**
Sensor space		
Evoked Power	Evoked	0.5–4
		4–8
		8–13
		13–30
		0.5–30
Time-frequency ('Call-sign' window)	Induced	2–100
Time-frequency ('Ready' windows)		8–100
Source space		
Minimum-norm	Evoked	0.5–4
		4–8
Spatial filtering	Induced	4–8
		8–13
		13–30
		40–80

Table 7.2. Summary of the different MEG data analysis methods in sensor- and source-space.

In source-space, the locations of the largest peak values in the SPMs from the minimum-norm analysis were recorded. As the power from a single source was smeared across adjacent areas by the linear solution, small patches of cortex were examined by extracting the nearest four source locations to each peak location. The time-varying minimum-norm solution was then extracted for the group of 5 locations. The average time-course of evoked activity across the patch of cortex was then calculated. These average region of interest (ROI) data were subjected to the same non-parametric permutation tests as the sensor-space data. This analysis identified time window(s) in which there was a significant difference in power between the epochs within the specified region of cortex. For each time window, the sizes of the peak differences in power were extracted as for the sensor-space data. Latencies of the peak differences were also extracted for each participant.

In addition to the source-space ROIs based on peaks in the SPMs, ROIs were also identified in the SPMs based on the significant time windows from the sensor-space analysis. The same procedure was used to extract the time-varying data from these ‘sensor ROIs’ as for the other source ROIs except that only a single source location was used. The choice of a single location allowed for the selection of more focal clusters of significant differences based on the a priori identification of the significant time window.

Significance values for the MEG power correlations across all participants were adjusted to correct for multiple comparisons using the technique described in Chapter 6 (Section 6.2.7.5, p. 148), and are denoted by p_{group}. The data from ROIs which were significantly correlated with performance across all participants after correcting for multiple comparisons were subjected to within-group correlations. These within-group correlations were not corrected for multiple comparisons.

7.3 Results

7.3.1 Behavioural Performance

Visual inspection of boxplots of the data indicated that the performance of one participant (45.8%) was a possible outlier. After standardising the scores, the data point was found to exceed the z-score threshold value of 3. To reduce the effect of the outlier, the MEG performance data were transformed from percent correct to percent incorrect and the square root of the values was taken, and are denoted by the subscript *trans*. Standardisation of the transformed scores confirmed that the data point in question no longer exceeded the outlier threshold. Homogeneity of variance and normality tests confirmed that the transformed data did not violate the assumptions of the independent-samples t-test.

On average, participants in both the young and older adult groups performed the task accurately (Overall Mean = 85.8%, Mean_{trans} = 3.5, σ_{trans} = 1.4). The performance of the young adults (Mean = 90.7%, Mean_{trans} = 2.9, σ_{trans} = 1.1) was significantly better than that of the older adults (Mean = 80.9%, Mean_{trans} = 4.1, σ_{trans} = 1.5) [$t(22) = -2.37$, $p < .05$, $r = .45$].

7.3.2 Response Latencies

Figure 7.16 shows the latency information for the behavioural responses. The latencies of the colour (Mean = 2477 ms, σ = 355) and number (Mean = 3062 ms, σ = 376) responses indicated that neither overlapped with the analysis window aligned to the onset of the target call-sign (Figure 7.16, shaded region). Contamination of the analyses by response-related motor artifacts was therefore unlikely, and not considered further.

7.3.3 Sensor-space

7.3.3.1 Evoked Analysis

Figure 7.17 shows the frequency content of the grand-average magnetic field power across the MEG sensor array for each of the three pairs of epochs. The evoked signal

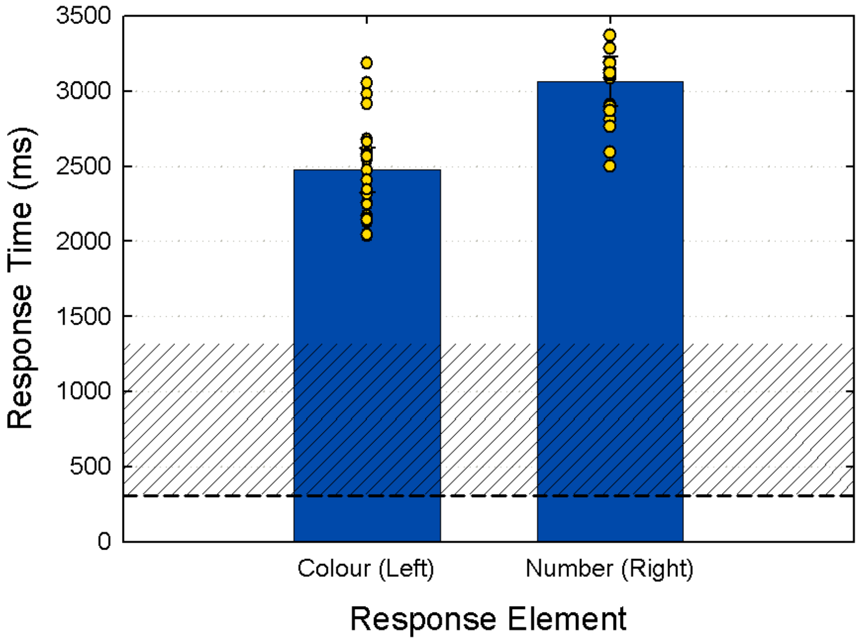

Figure 7.16. The mean (bars) and individual (symbols) latencies of the behavioural responses for the colour and number keywords, made using the left and right hands respectively. Response times are relative to the onset of the target phrase. The mean onset time of the call-sign keyword (dashed line) and the extent of the analysis window aligned to the onset of the target call-sign (shaded region) did not overlap with the responses. Error bars show 95% confidence intervals.

in all three epochs contained the most power at low frequencies (< 10 Hz). Visual inspection of the frequency spectra suggested that the largest differences for the “Onset attention” and “Call-sign processing” epoch pairs occurred at low frequencies (< 8 Hz). The spectra of the “Onset processing” epochs were found to be similar, only exhibiting small differences at frequencies above 10 Hz.

Figure 7.18 shows the results of the cluster-based permutation test analysis for the evoked data in the three sensor groups and in the widest frequency band (0.5–30 Hz). No significant differences were found between the “Onset processing” pair of epochs—there was no difference in the power evoked by the onset of a new phrase

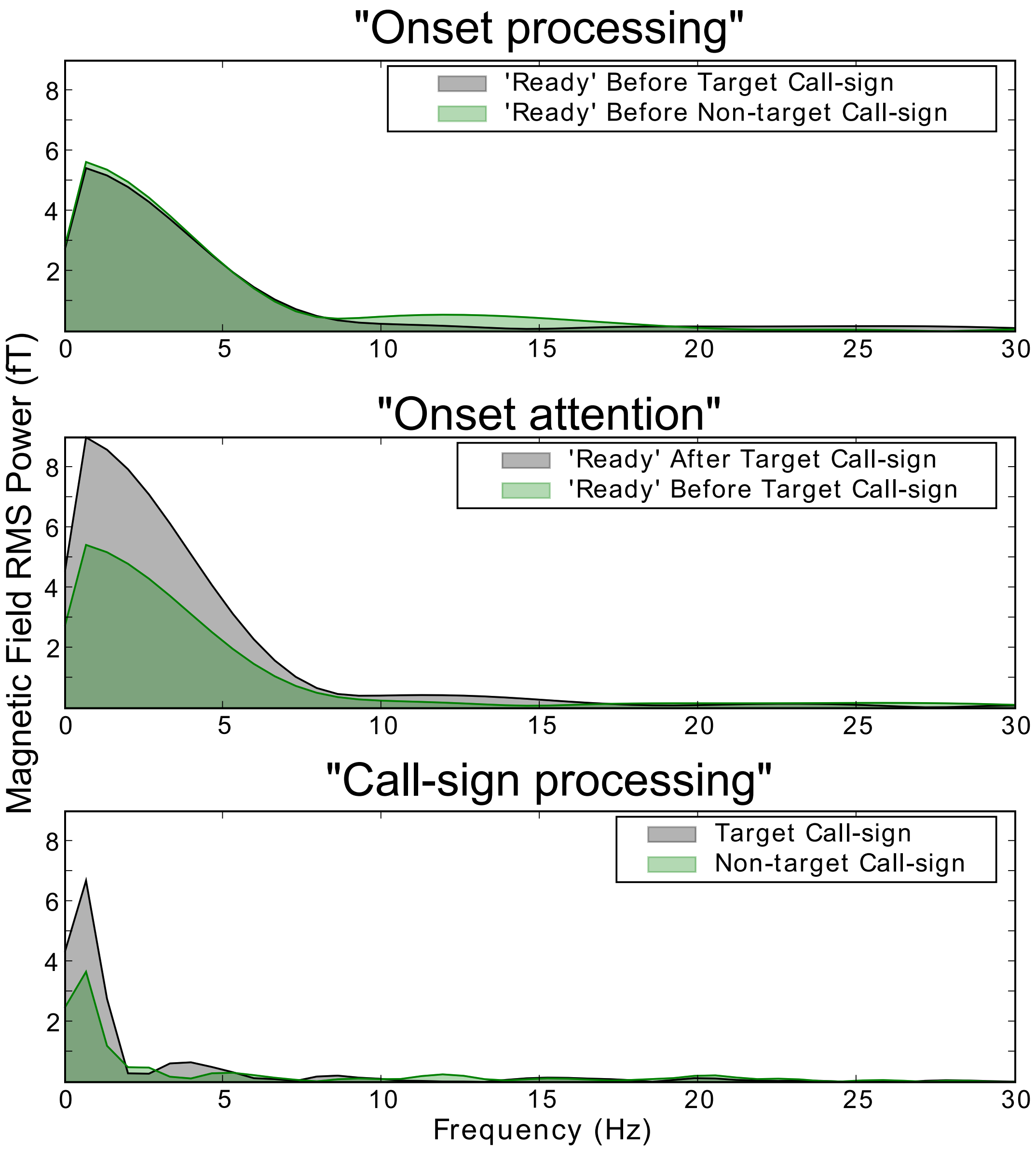

Figure 7.17. Frequency spectra of the mean evoked power across the sensor array for the “Onset processing” pair (top), “Onset attention” pair (middle), and “Call-sign processing” (bottom). The frequency resolution of the first two epoch pairs is lower compared to the bottom pair due to the shorter epoch length (250 ms and 1000 ms respectively).

when it preceded a target call-sign compared to when it preceded a non-target call-sign. For the “Onset attention” epoch pair, significantly more power was found in response to the ‘Ready’ keyword which occurred after the target call-sign compared to the ‘Ready’ which occurred directly before the target call-sign, equivalent to the onset of the target phrase. Analysis of the “Call-sign processing” pair of epochs revealed that significantly more power was evoked by the target call-sign compared to the non-target call-sign.

Significant differences in the data from 0.5–30 Hz were found in both hemispheres and across the entire sensor array (‘Global’). The differences in “Onset attention” occurred earlier in the group of sensors over the right hemisphere (26 ms after onset of ‘Ready’ keyword) compared to the sensors over the left hemisphere (100 ms after onset of ‘Ready’ keyword). A later difference starting at 190 ms was significant only for the left sensor group. Differences in “Call-sign processing” occurred in two time windows: an *early* window from approximately 350–500 ms and a *late* window around 625–825 ms after the onset of the call-sign keyword. The start of the earlier differences occurred at similar times in both hemispheres: 376 ms and 365 ms for the left and right sensor groups respectively. Later differences were identified in the left sensor group only, from 659–824 ms.

A summary of the results of the evoked permutation analysis for the different frequency bands is shown in Figure 7.19. No differences were found for the “Onset processing” pair in any frequency band. The results of the permutation tests for the “Onset attention” and “Call-sign processing” comparisons will be discussed in detail in the following sections.

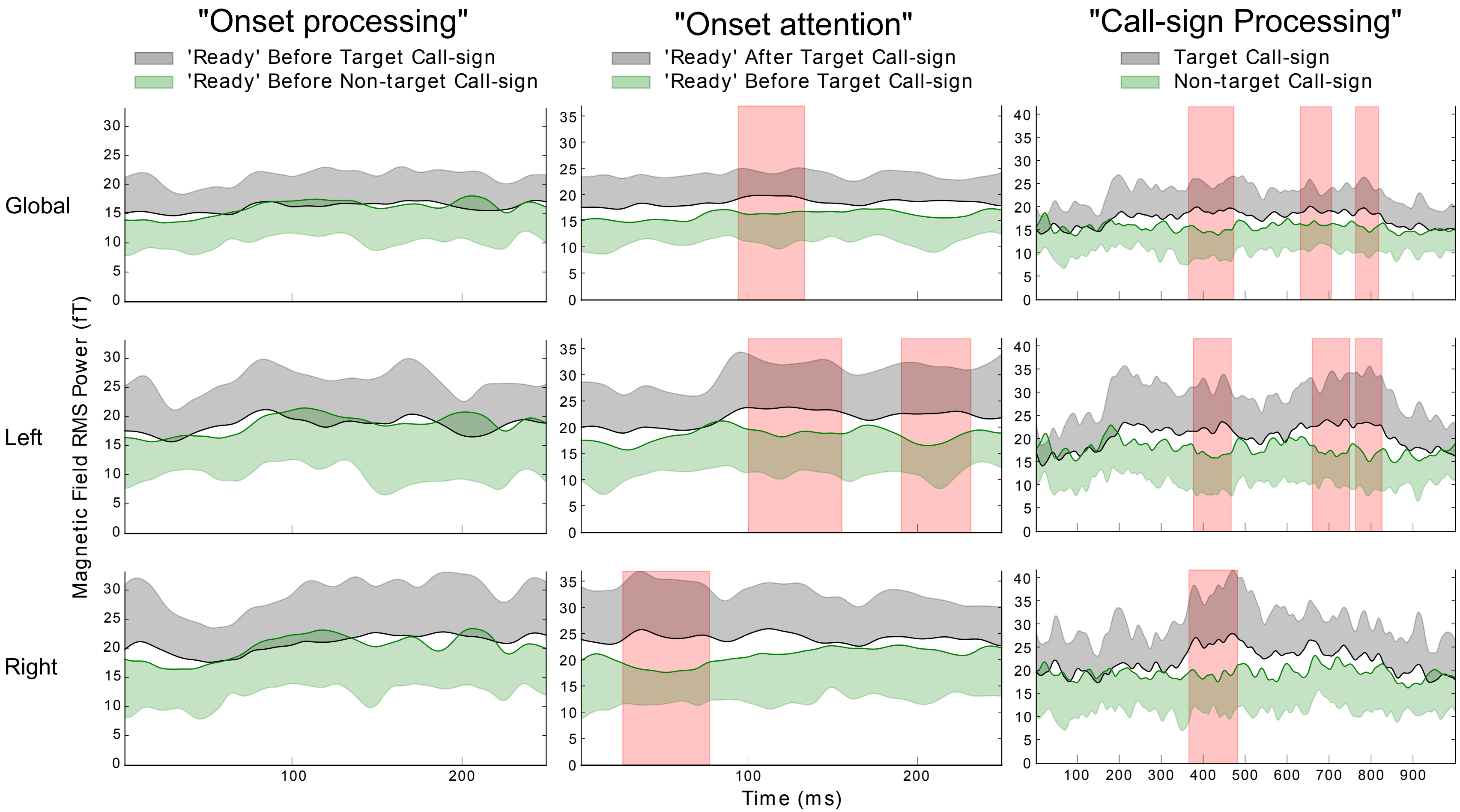

Figure 7.18. Mean evoked power (thick lines) and standard deviations (shaded areas) across different sensor groups for the three epoch pairs. The data were filtered from 0.5–30 Hz. The abscissa for each pair of epochs shows time relative to the onset of the keywords; i.e. 'Ready' for "Onset processing"/"Onset attention" and target/non-target call-signs for "Call-sign processing". Time-windows in which significant differences were found using a cluster-based permutation test are shown in red.

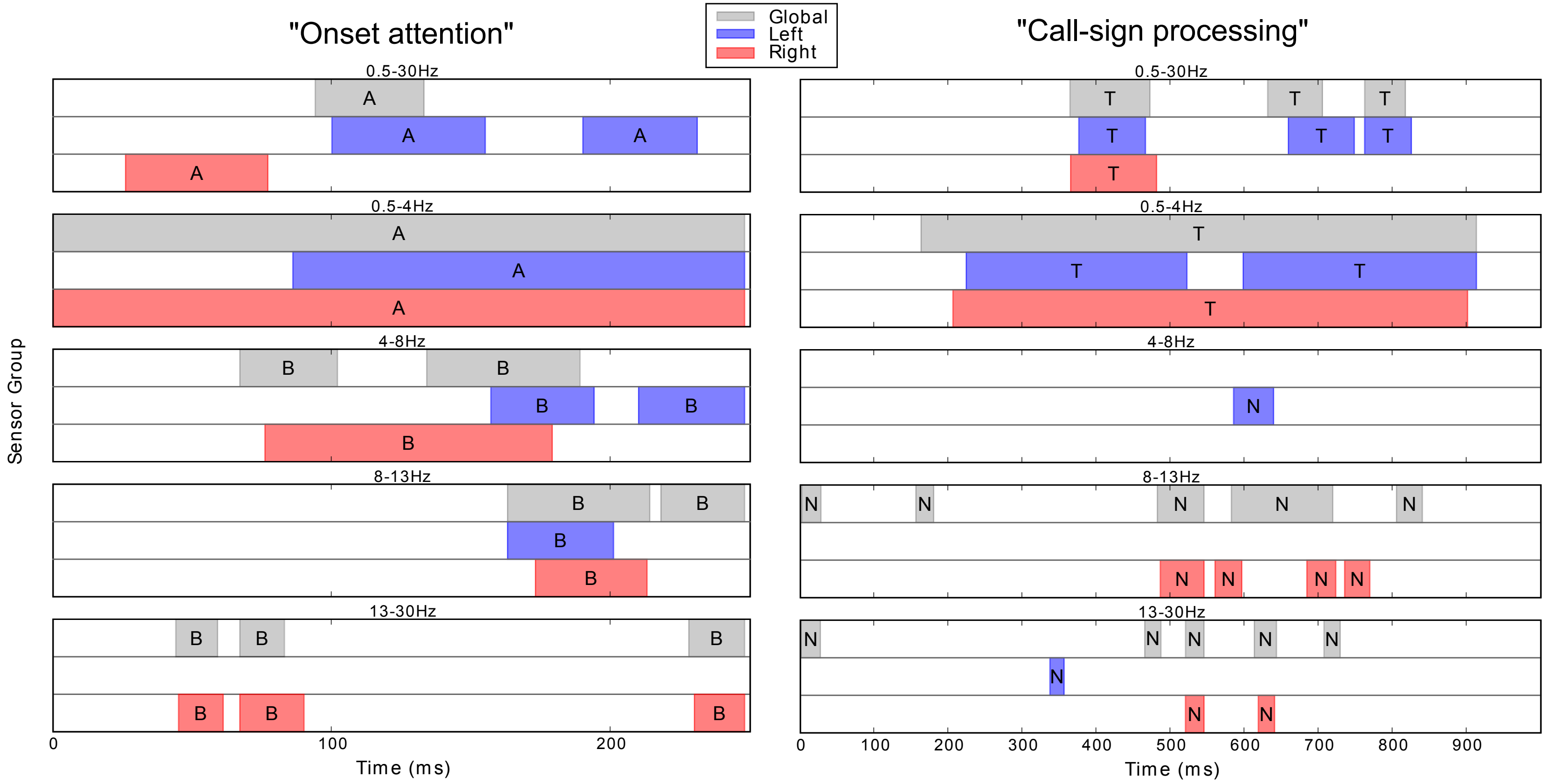

Figure 7.19. Time windows in which significant differences were found between epoch pairs. The different sensor groups are denoted by colour (Global: grey, Left: blue, Right: red). For the "Onset attention" pair (left), *B* denotes a greater response to the 'Ready' keyword before the target call-sign and *A* a greater response to the 'Ready' after the target call-sign. For the "Call-sign processing" pair (right), *T* denotes a greater response to the target call-sign and *N* a greater response to the non-target call-sign.

“Onset attention”

The comparison of the “Onset attention” epoch pair exhibited a different pattern of results at low and high frequencies (Figure 7.19). In the lowest frequency band (0.5–4 Hz), there was significantly more power evoked by the ‘Ready’ keyword which directly followed the target call-sign compared to the ‘Ready’ which preceded it. This difference spanned the entire epoch in the right hemisphere sensor group and started at 86 ms after the onset of the epoch in the left sensor group. At frequencies above 4 Hz, the direction of the effect was reversed. Significantly more power was evoked by the preceding ‘Ready’ in several time windows. The earliest differences above 4 Hz occurred in the right hemisphere sensor group at 76 ms after the keyword onset in the 4–8 Hz band and at 45 ms in the 13–30 Hz band. Later differences, starting after 150 ms, were found in both left and right hemisphere groups. For the left sensor group, differences were found starting at 157 ms and 163 ms in the 4–8 Hz band and the 8–13 Hz band, respectively. Late differences in the right sensor group started at 173 ms in the 8–13 Hz band and 230 ms in the 13–30 Hz band.

“Call-sign processing”

A similar contrast between significant differences in evoked power at frequencies above and below 4 Hz was observed for the “Call-sign processing” comparison (Figure 7.19). Within the low frequency band (0.5–4 Hz), significantly more power was evoked by the target call-sign compared to the non-target call-sign. As with the “Onset attention” comparison, the differences observed in the broad-band range (0.5–30 Hz) were similar in temporal position and direction to the low-frequency differences. The division of low-frequency differences into early and late time windows, as identified in the broad frequency band, was found for the left hemisphere sensor group. The early window started at 224 ms and the late window at 598 ms after the onset of the call-sign. In contrast, the right sensor group differences occurred throughout the range from 206–900 ms.

The majority of the higher frequency differences were grouped into a window from 450–850 ms with the earliest difference occurring in the left sensor group at 337 ms. At higher frequencies (> 4 Hz), all of the significant time windows showed greater evoked power in response to the non-target call-sign compared to the target call-sign.

7.3.3.2 Induced Analysis

The induced analysis included data which was phase-locked and data which was not strictly phase-locked to the start of the data epochs. Figure 7.20 shows the mean induced data transformed into time-frequency representations for the three pairs of epochs. Across all three, the maximum power in the induced activity was found in the frequency band between 15–25 Hz; i.e. within the β (beta) frequency band. For the “Onset processing” pair, power in this frequency band was observed throughout both epochs. The response to the 'Ready' keyword before a target call-sign (Figure 7.20 (a)) exhibited a more sustained response in the β-band than the 'Ready' before a non-target call-sign (Figure 7.20 (b)). This result will be discussed later in this chapter. Both epochs within the "Onset processing" pair contained an initial increase in power between 50–100 ms after the onset of the 'Ready' keyword.

The induced response to the 'Ready' which followed a target call-sign (Figure 7.20 (c)) contained the lowest absolute mean power across all epochs. The response exhibited peak power levels throughout the epoch within the β-band. A decrease in power across a range of frequencies between 40–80Hz, the γ (gamma) frequency band, was observed for the 'Ready' after the target call-sign compared to the 'Ready' which preceded it (Figure 7.20 (c) vs. (d)).

The "Call-sign processing" pair also exhibited a difference in power within the β and γ frequency bands. The response to the target call-sign (Figure 7.20 (e)) contained a decrease in power around 20 Hz extending from approximately 200–800 ms with the lowest power level occurring just after 600 ms. In contrast, the response to the non-target call-sign (Figure 7.20 (f)) exhibited a peak in power in the β-band at approximately the same time. A decrease in power between 40–80 Hz, centred on 55 Hz, was also observed in response to the target call-sign compared to the non-target call-sign. This decrease started at approximately 400 ms after the onset of the epoch.

A closer inspection of differences in the β and γ frequency bands is shown in Figure 7.21. The induced power in the 20 Hz and 55 Hz frequency bins is shown for each of the three epoch pairs. The left-hand panels show the similarity between the "Onset processing" epochs at both frequencies (Figure 7.21 (a) & (b)). In contrast, less induced power was evident in response to the phrase onset which followed the target call-sign compared to that which preceded it ("Onset attention") (Figure 7.21 (c) vs. (d)) and also in response to the target versus non-target call-sign ("Call-sign processing") (Figure 7.21 (e) vs. (f)). The difference spans the entire epoch for the "Onset attention" pair. The difference was found to emerge at approximately 300 ms in both the 20-Hz and the 55-Hz bins for the "Call-sign processing" pair.

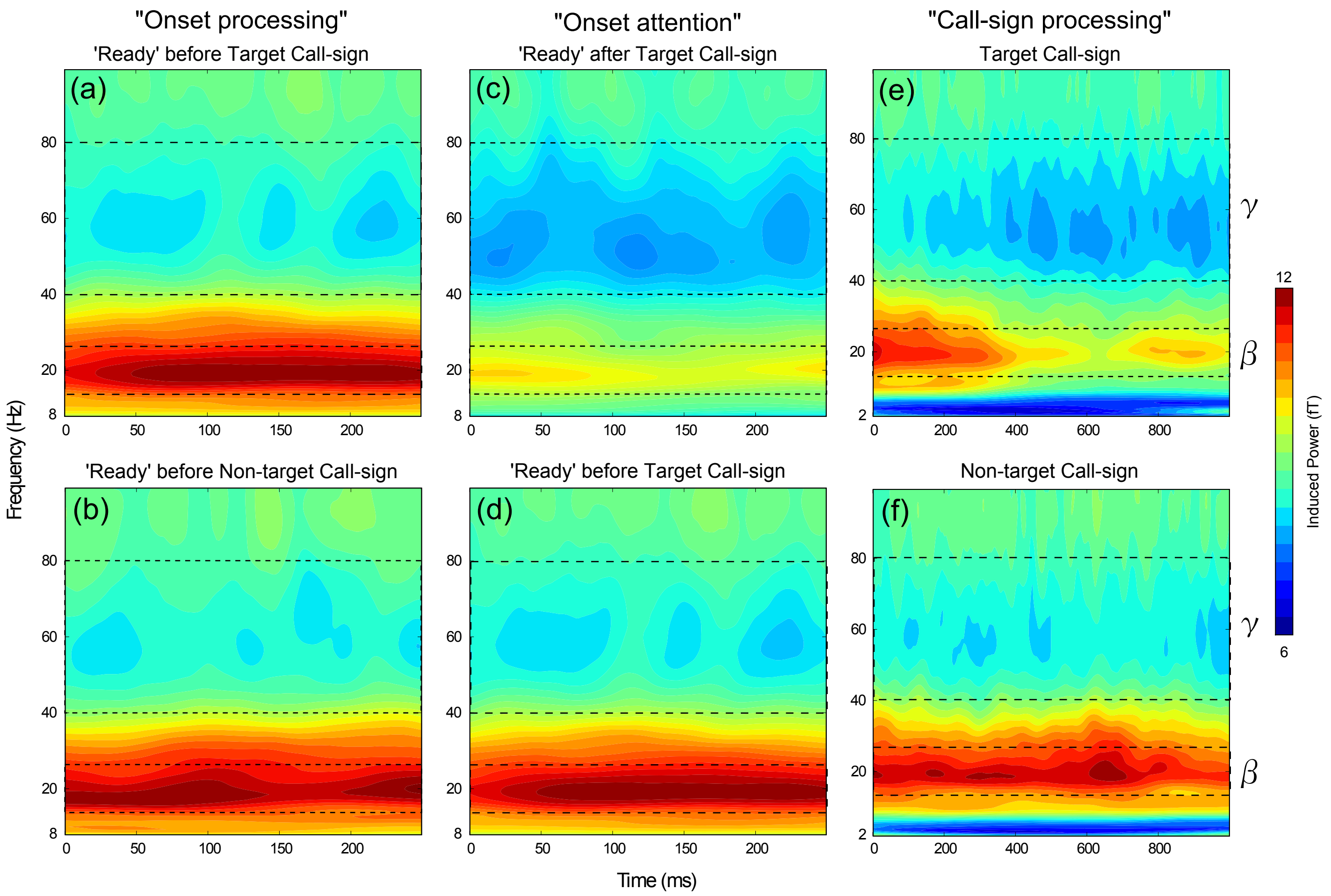

Figure 7.20. Time-frequency representations of the mean induced data for the three epoch pairs. Dashed boxes outline regions of the beta (β, 13–30 Hz) and gamma (γ, 40–80 Hz) frequency bands.

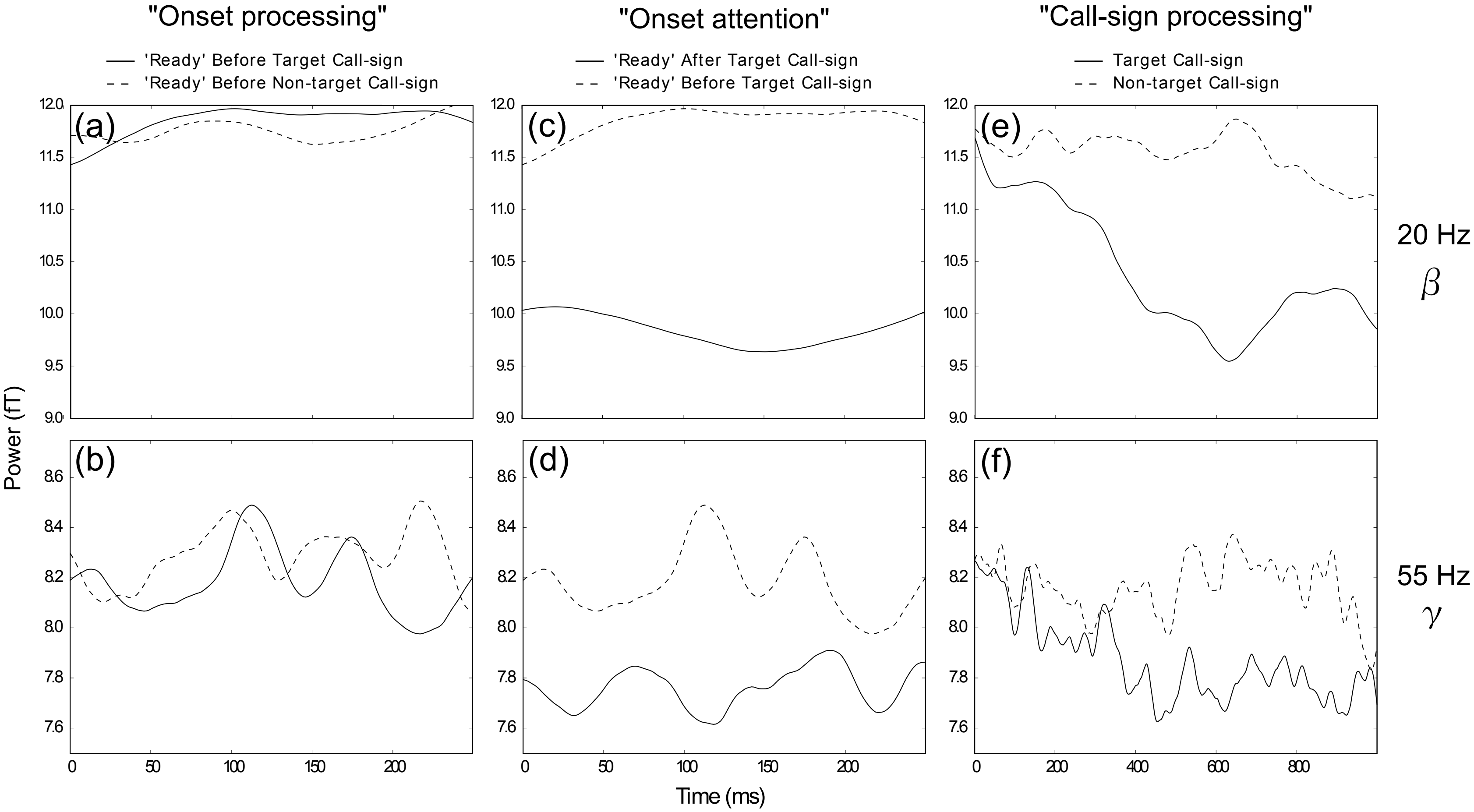

Figure 7.21. Mean induced power across all sensors for two frequency bins of interest extracted from the average time-frequency representations (Figure 7.20). The top row shows global power at 20 Hz (β band) and the bottom row shows global power at 55 Hz (γ band).

The results of the single-threshold permutation tests are shown in Figure 7.22 and Figure 7.23. The time-frequency statistical maps show *t*-values above the critical threshold which corresponded to $p < .05$ after correcting for multiple comparisons.

"Onset processing"

The differences observed for the "Onset processing" comparison were mainly localised in the γ-band and were transient (Figure 7.22). The peak difference, indicating significantly less power for the 'Ready' which preceded the target call-sign compared to that which preceded a non-target call-sign, occurred 220 ms after the onset of the epoch at a frequency close to 60 Hz.

"Onset attention"

The "Onset attention" comparison (Figure 7.23 (a–c)) revealed a difference in induced power spanning the length of the epoch. The differences indicated a broadband decrease in activity in response to the phrase onset which followed the target call-sign compared to that which preceded it. The peak of this decrease was identified across all sensors (Figure 7.23 (a)) in the upper β and lower γ frequency bands at 125 ms, and at 10–12 Hz between 150–200 ms. An analysis of the hemispheric sensor groups

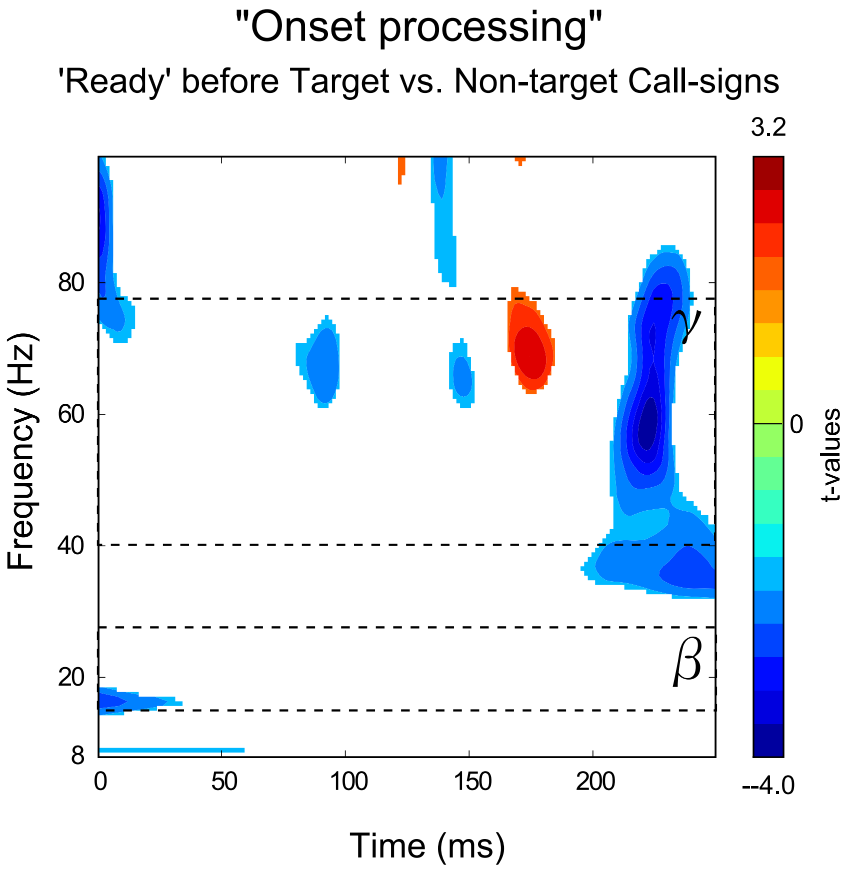

Figure 7.22. Significant differences in the power of induced activity across all sensors (Global) for the "Onset processing" pair of epochs.

revealed peak decreases in power in the β-band which peaked earlier in the right hemisphere (60 ms) (Figure 7.23 (c)) than in the left hemisphere (100 ms) (Figure 7.23 (b)) .

"Call-sign processing"

Comparison of the "Call-sign processing" pair of epochs (Figure 7.23 (d–f)) revealed a decrease in power from 200–1000 ms after the onset of the call-sign keywords. Less

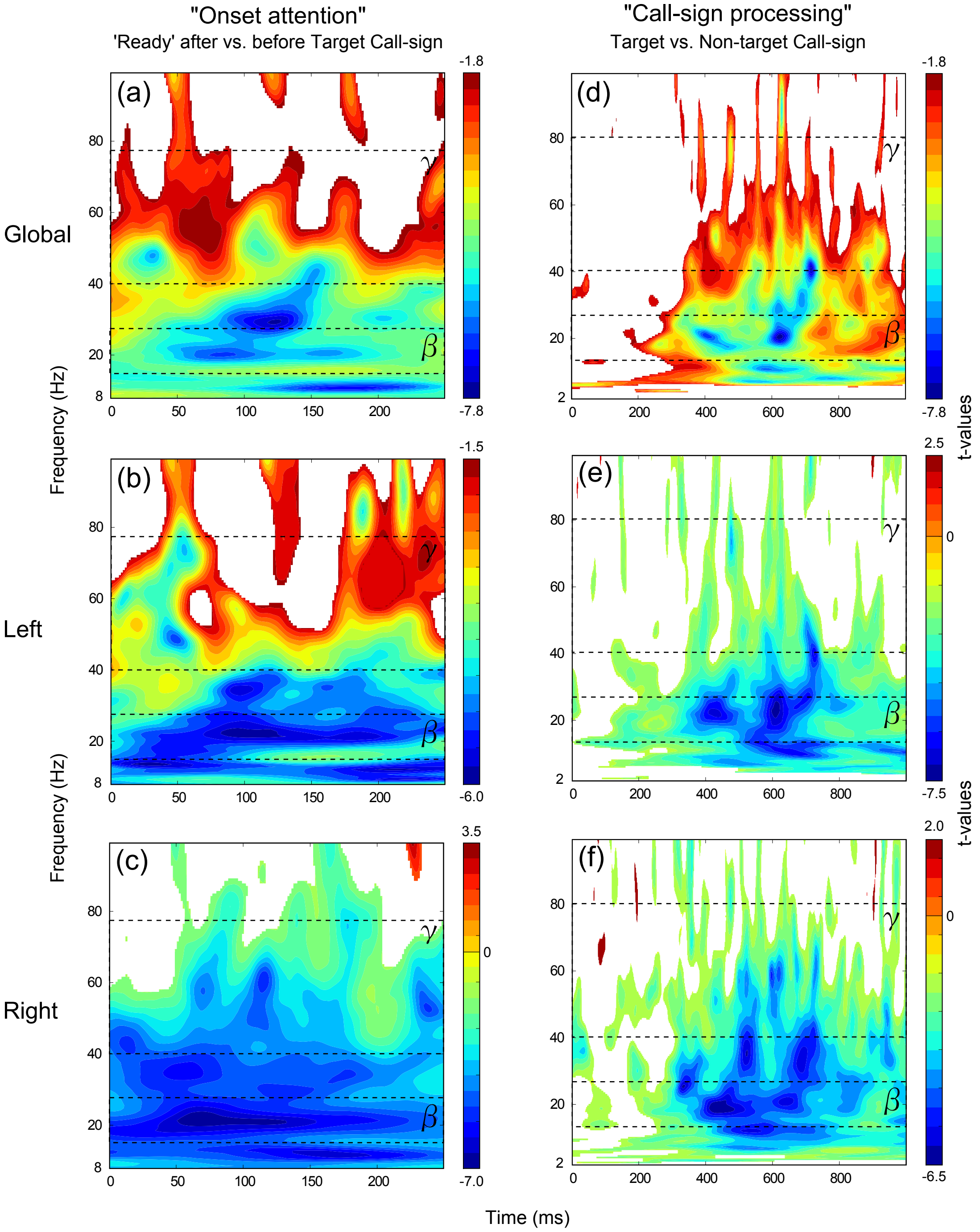

Figure 7.23. Significant differences in the power of induced activity for all sensors (Global) and the left and right hemisphere sensor groups.

induced power was found in response to the target call-sign compared to the non-target call-sign. As identified in the mean induced data (Figure 7.20 (e) vs. (f)), peak differences in power were found globally within the β-band around 20 Hz. The differences included early (400 ms) and late (600 ms) peaks (Figure 7.23 (d)). A broad significant decrease in power was also identified in the γ-band between 300–950 ms. Peak differences in the β-band included early (400 ms) and late (600 ms) peaks bilaterally.

7.3.4 Source-space

The locations of peak differences and the associated statistical values from the minimum-norm and spatial filtering analyses are listed in Appendix G.

7.3.4.1 Minimum-norm

"Onset processing"
No significant differences in the minimum norm source solutions were identified for the "Onset processing" comparison within any frequency band.

"Onset attention"
Figure 7.24 shows the SPMs for the "Onset attention" comparison at moments containing significant differences. As in the evoked sensor data, more evoked power was found at low frequencies in response to the phrase onset which followed the target call-sign compared to the onset which preceded it. The opposite pattern was identified at higher frequencies. At low frequencies (0.5–4 Hz), the earliest differences were identified in the temporal lobes bilaterally and were strongly right lateralised (Figure 7.24 (a)). This early activity was more anterior in the right hemisphere, located medially on the transverse temporal sulcus and more laterally on the superior temporal gyrus. The early left hemisphere activity was in the posterior portion of the superior temporal sulcus, close to the temporal-parietal junction (Figure 7.24 (c)). Later differences in the right hemisphere were found in the inferior frontal gyrus and inferior parietal gyrus (Figure 7.24 (b)). The later left hemisphere differences were localised in the intra-parietal sulcus, and in the middle frontal and pre-central gyrii of the frontal lobe (Figure 7.24 (d)).

Differences in the higher frequency band (4–8 Hz) were all identified within 100 ms of the start of the epoch. Focal differences were found bilaterally in the parietal lobe: in the intra-parietal sulcus of the right hemisphere and the left post-central gyrus (Figure 7.24 (f) & (e)). A difference was also found on the left superior temporal gyrus (Figure 7.24 (g)). Peak differences, their locations, and associated t-values for the "Onset attention" comparison are listed in Table G.1 (Appendix G, p. 275).

"Call-sign processing"

The results of the minimum-norm analysis for the "Call-sign processing" comparison at low frequencies (0.5–4 Hz) are displayed in an early (< 500 ms, Figure 7.25) and a late (> 500 ms, Figure 7.26) window. Similar to the evoked sensor data at low frequencies, all of the differences indicated greater levels of evoked power in response to the target compared to the non-target call-sign. Significant differences in the

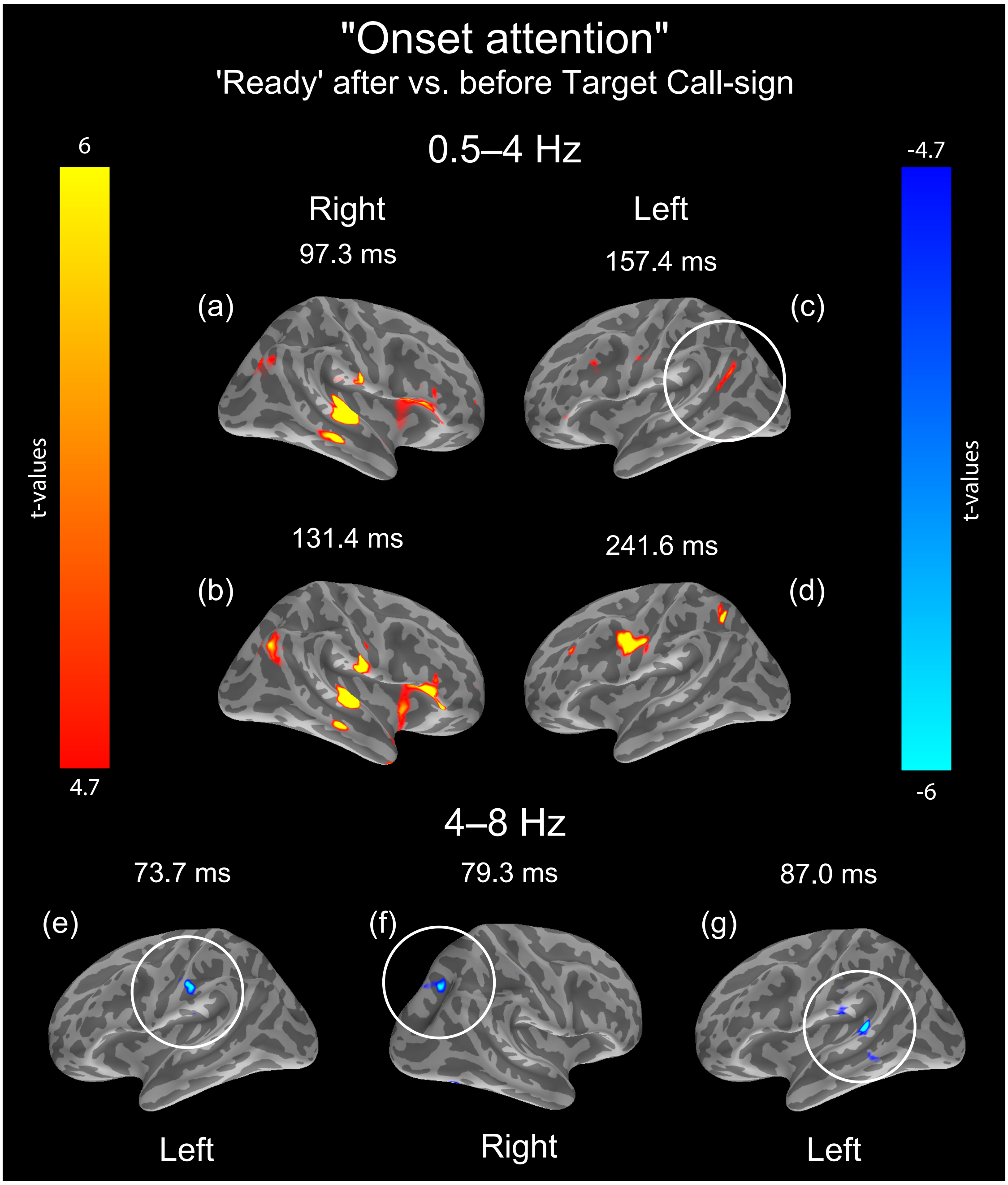

Figure 7.24. The results of the minimum-norm analysis for the "Onset attention" comparison which compared the cortical activity associated with the phrase onset following the target call-sign to the activity associated with the phrase onset which directly preceded the target call-sign. SPMs are shown for those moments at which peak differences were identified in the 0.5–4 Hz (a–d) and 4–8 Hz (e–g) frequency bands. The times are relative to the phrase onsets and more focal differences have been circled.

early window were found bilaterally in the temporal lobe. The left temporal activity (Figure 7.25 (b)) occurred more anterior to that in the right hemisphere, with peaks in posterior STS and adjacent in the planum temporale, with the activity extending into the Sylvian fissure along the circular sulcus of the insula. Right temporal differences (Figure 7.25 (e)) were found in the posterior STS and middle-temporal gyrus.

Differences were also found in the parietal lobe of the left hemisphere and were localised to the inferior parietal gyrus and intra-parietal sulcus (Figure 7.25 (a)). Right hemisphere differences were identified in the superior occipital gyrus, and in the middle-frontal gyrus and inferior pre-central sulcus of the frontal lobe (Figure 7.25 (c) & (d)). Table G.2 lists the peak differences for the early window (Appendix G, p. 276).

Significant differences were found in a more distributed range of cortical regions in the late window (Figure 7.26). Differences were found bilaterally in temporal, parietal, and frontal lobes. Temporal differences were found in bilateral posterior STS (Figure 7.26 (b) & (h)), with more anterior differences in STG, MTG, and planum temporale in the right hemisphere (Figure 7.26 (e)). Parietal differences included bilateral inferior parietal gyrus and post-central sulcus (Figure 7.26 (a), (d) & (h)), the anterior aspect of the right IPG (Figure 7.26 (h) & (i)) and left intra-parietal sulcus (Figure 7.26 (d)).

Differences in the frontal lobe were generally left-lateralised, with peaks in inferior and middle frontal sulcii, middle frontal gyrus, central sulcus, and inferior pre-central sulcus of the left hemisphere (Figure 7.26 (a) & (b)). Right hemisphere frontal differences were localised close to the border between frontal and parietal lobes in the pre-central sulcus, and in the inferior frontal gyrus (Figure 7.26 (h)).

Medial differences were identified in bilateral anterior cingulate cortex and left posterior cingulate sulcus (Figure 7.26 (c) & (f)). Other medial differences included the left precuneus gyrus (Figure 7.26 (c)), and right calcarine sulcus (Figure 7.26 (g)). The details of the peak differences in the late window at low frequencies are listed in Table G.3 (Appendix G, p. 277).

Figure 7.27 shows the results of the minimum-norm analysis for the “Call-sign processing” comparison for the data in the 4–8 Hz frequency band. Focal differences were found in early (< 500 ms) and late (> 500 ms) time windows. The early differences were increases in power in response to the target call-sign. The differences were localised to the left superior temporal gyrus and right middle frontal gyrus (Figure 7.27 (a) & (b)). The later window included a decrease in power at a similar latency to differences identified in the evoked sensor-space results. The difference, at 644 ms, was located in left post-central sulcus (Figure 7.27 (c)). A late focal differences was also identified in the anterior portion of the right superior temporal sulcus (Figure 7.27 (d)). Table G.4 lists the details of the peak differences between 4–8 Hz in the “Call-sign processing” comparison (Appendix G, p. 278).

7.3.4.2 Spatial Filtering

"Onset processing"

No significant differences were found in the SPMs from the spatial filtering analysis for the "Onset processing" comparison between 8–13 Hz, 13–30 Hz, or 40–80 Hz.

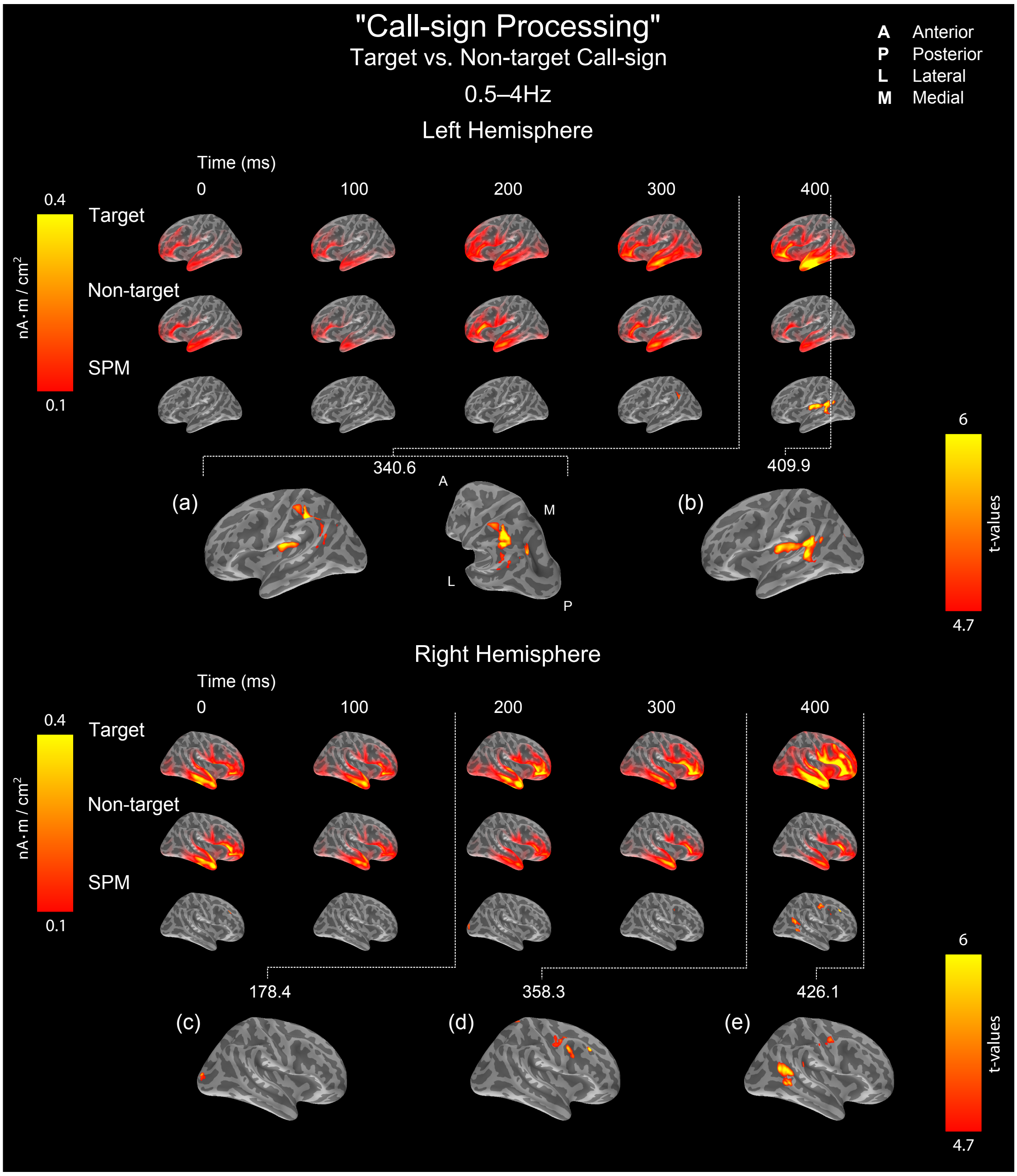

Figure 7.25. Results of the minimum-norm analysis for the first half of the "Call-sign processing" comparison in the 0.5–4 Hz band. The analysis compared the cortical response to target and non-target call-signs. Average minimum-norm solutions ('Target' and 'Non-target') and statistical parametric maps (SPMs) are shown for the left (top half) and right (bottom half) hemispheres. Enlarged SPMs (a–e) are shown for those moments at which peak differences were observed. Times are relative to the onset of the call-sign keywords.

"Onset attention"

Significant decreases in power were identified in the results of the spatial filtering analysis in response to the phrase onset following the target call-sign compared to the onset which directly preceded it. Figure 7.28 shows the SPMs for the three frequency bands which were analysed. In the right hemisphere, significant activity was found in the parietal and frontal lobes, specifically within the 40–80 Hz band (Figure 7.28

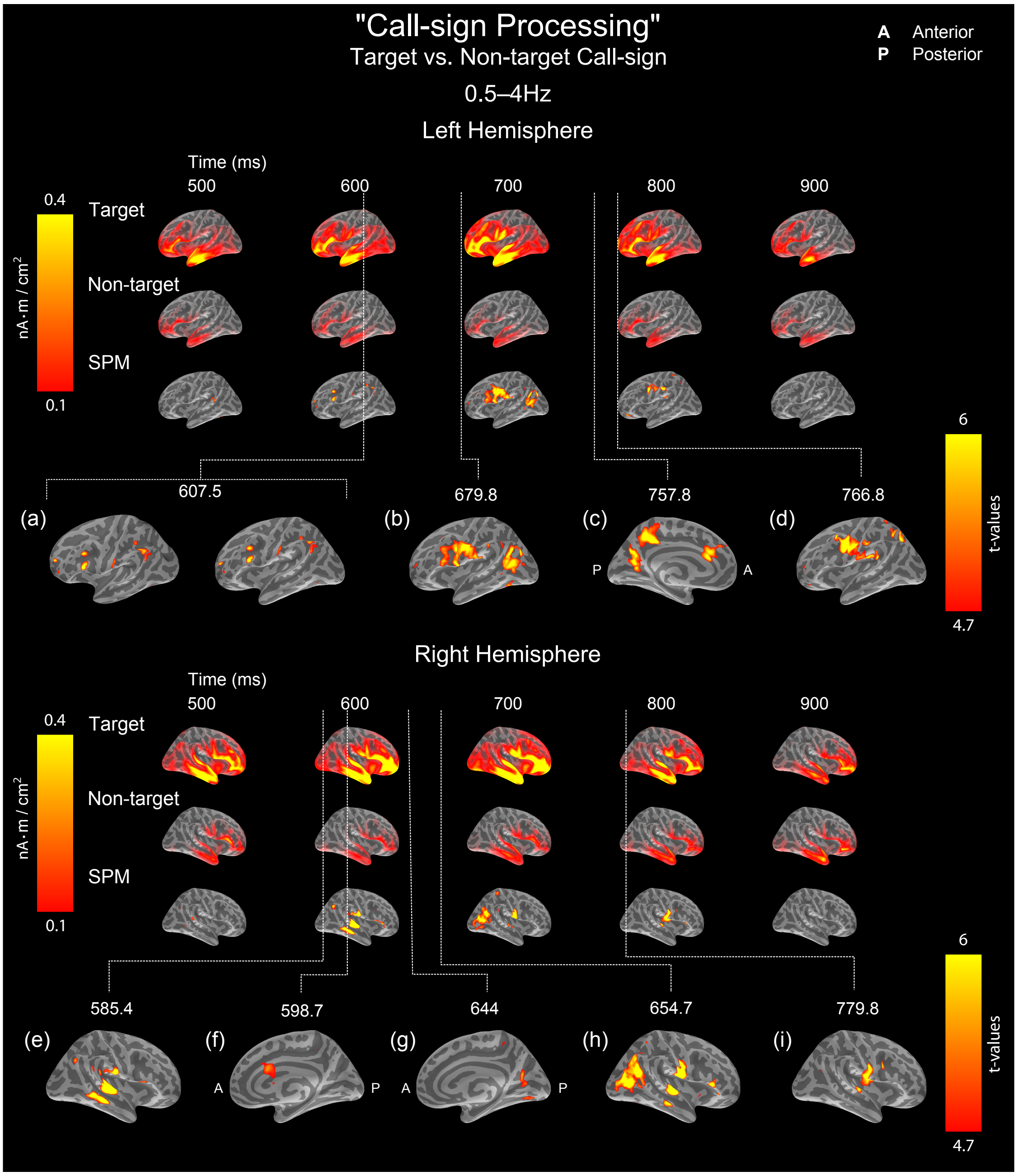

Figure 7.26. Results of the minimum-norm analysis for the second half of the "Call-sign processing" comparison in the 0.5–4 Hz band. The analysis compared the cortical response to target and non-target call-signs. Average minimum-norm solutions ('Target' and 'Non-target') and statistical parametric maps (SPMs) are shown for the left (top half) and right (bottom half) hemispheres. Enlarged SPMs (a–i) are shown for those moments at which peak differences were observed. Times are relative to the onset of the call-sign keywords.

(c)), with peaks in the intra-parietal sulcus, inferior parietal gyrus, the inferior pre-central sulcus, and the inferior frontal sulcus and gyrus. Left hemisphere activity was localised along the superior and middle temporal gyrii, with more widespread and stronger differences occurring in the 8–13 Hz band (Figure 7.28 (a)). All three bands featured activity in the superior occipital sulcus. The 8–13 Hz band included activity on and inferior to the inferior frontal gyrus. Tables G.5 and G.6 list the locations and t-values of the peak differences (Appendix G, p. 278).

“Call-sign processing”

SPMs for the comparison of target and non-target call-signs using the spatial filtering technique are shown in Figure 7.29 for three frequency bands. The pattern of significant differences varied between the left and right hemispheres and across the frequency bands. All differences indicated a decrease in power in response to the target call-sign. The left hemisphere exhibited differences mainly in posterior temporal, inferior parietal, and frontal lobes. The frontal activity, specifically in inferior pre-frontal cortex, is strongest in the 4–8 Hz band, with more anterior frontal activity in the fronto-marginal gyrus occurring only at higher frequencies (8–13 Hz and 13–30 Hz). The largest peaks in parietal and posterior temporal regions were found between 8–13 Hz, in the posterior STS, inferior parietal gyrus, and the superior occipital sulcus. Temporal activity was also found lateral to Heschl’s gyrus on the STG and MTG. The pre-central and middle-temporal differences were observed in the earlier time windows, between 0–600 ms, with the frontal and parietal differences occurring between 300–1000 ms.

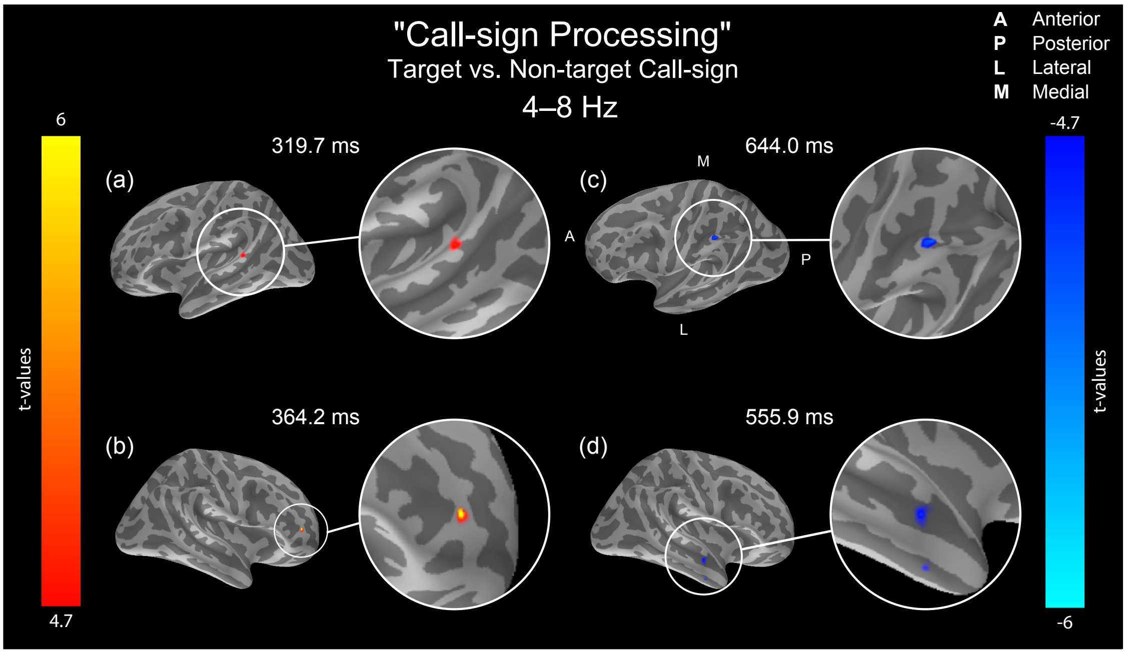

Figure 7.27. The results of the minimum-norm analysis for the “Call-sign processing” comparison between 4–8 Hz. SPMs are shown (a–d) at those moments at which significant differences in power were observed.

Significant differences in the right hemisphere were found mainly within the parietal lobe within all three frequency bands. The activity was more anterior and superior to the parietal activity observed in the left hemisphere, located in the anterior portion of the inferior parietal lobe close to the planum temporale, in the posterior IPG adjacent to the tempro-parietal junction, and in the superior parietal gyrus. Differences in the temporal lobe were found for the data between 4–8 Hz, including the transverse, superior, and middle temporal gyrii. Frontal activity was found mainly between 4–8 Hz, in the IFG and the fronto-marginal gyrus.

Differences in γ-band activity (40–80 Hz, Figure 7.30) were characterised by strong temporal differences in the right hemisphere between 100–400 ms, including the superior and inferior temporal sulcii, and in the left frontal lobe from 0–400 ms. The left frontal differences included the orbital, inferior frontal, and pre-central gyrii. Late differences, after 400 ms, were strongly right lateralised and localised to the right intra-parietal sulcus. Tables G.7, G.8, and G.9 list the peak differences for the "Call-sign processing" comparison for each of the frequency bands (Appendix G, p. 280).

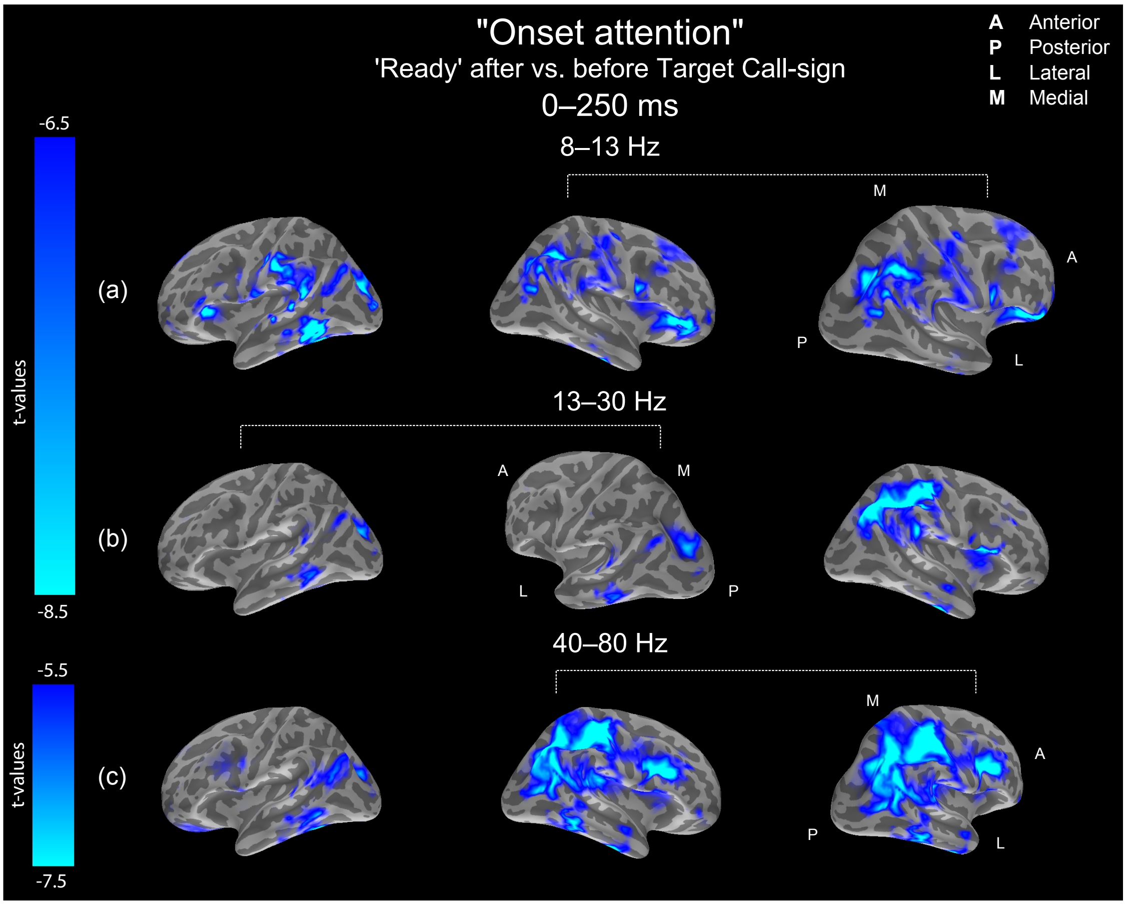

Figure 7.28. SPMs from the spatial filtering analysis for the "Onset attention" comparison in both hemispheres. SPMs calculated over a 250 ms window (0–250 ms) are displayed for the α (alpha, 8–13 Hz), β (beta, 13–30 Hz), and γ (gamma, 40–80 Hz) frequency bands. Bracketed SPMs are identical except for the viewing angle.

7.3.5 Individual Differences

7.3.5.1 Spatial Listening Task and Hearing Sensitivity

Table 7.3 lists rank-order correlation coefficients between the transformed performance scores for the spatial listening task during MEG imaging, the principle factor extracted from spatial listening task performance in the laboratory, and participants' better-ear average hearing level. Across all 24 participants, significant relationships were observed between performance during MEG imaging and performance on the related tasks in the laboratory, such that better laboratory performance was associated with fewer errors during MEG imaging. When the groups were analysed separately, this relationship was significant only within the young adult group. Poorer hearing sensitivity was significantly associated with poorer performance during MEG imaging when the groups were combined. When the groups were correlated separately, this relationship was significant only within the older adult group.

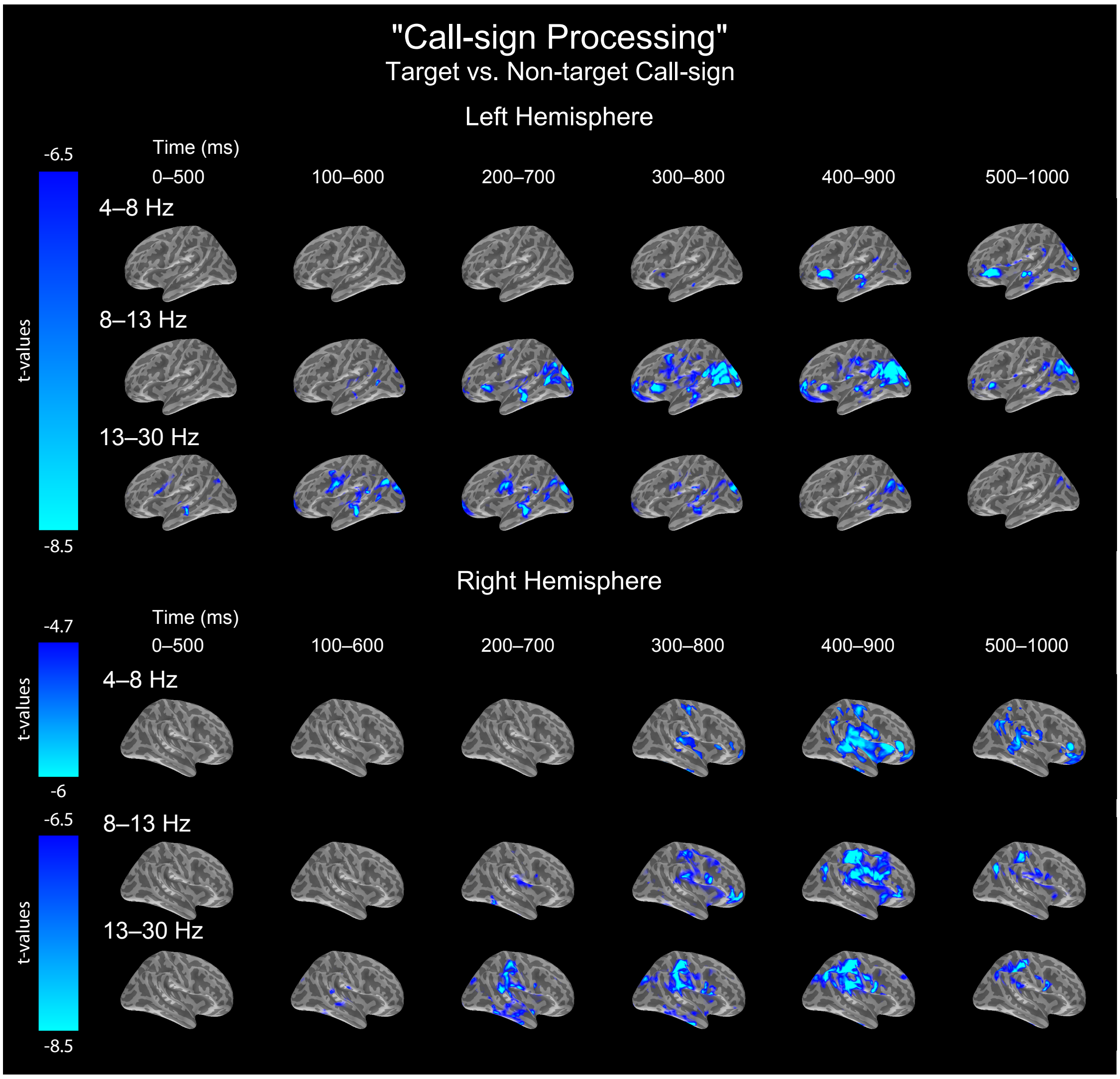

Figure 7.29. SPMs from the spatial filtering analysis for the "Call-sign processing" comparison. SPMs for each 500 ms time-window are displayed for the theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) frequency bands.

7.3.5.2 Attentional Abilities

Performance during MEG imaging was related to several of the tasks from the TEA (Table 7.4). Higher accuracy in the MEG listening task was associated with a higher number of identified items in the visual *Map search - 2 min* task (1B), a faster visual search speed in the *Telephone search* task (6), and a higher score in the auditory *Elevator counting with reversal* task (5). No significant correlations were observed within the young adult group. A relationship with the *Lottery* task (8) was observed within the older adult group. This relationship was not found to be significant after correcting for multiple comparisons.

7.3.5.3 Self-reported Difficulties

The analysis of self-reported difficulties examined relationships between performance during MEG imaging with individual questions in the SSQ, the 10 sub-scales of the SSQ, and three questions specifically related to the spatial listening task. A significant relationship was found between MEG performance and Question 2 from the Speech section of the SSQ ("You are talking with one other person in a quiet,

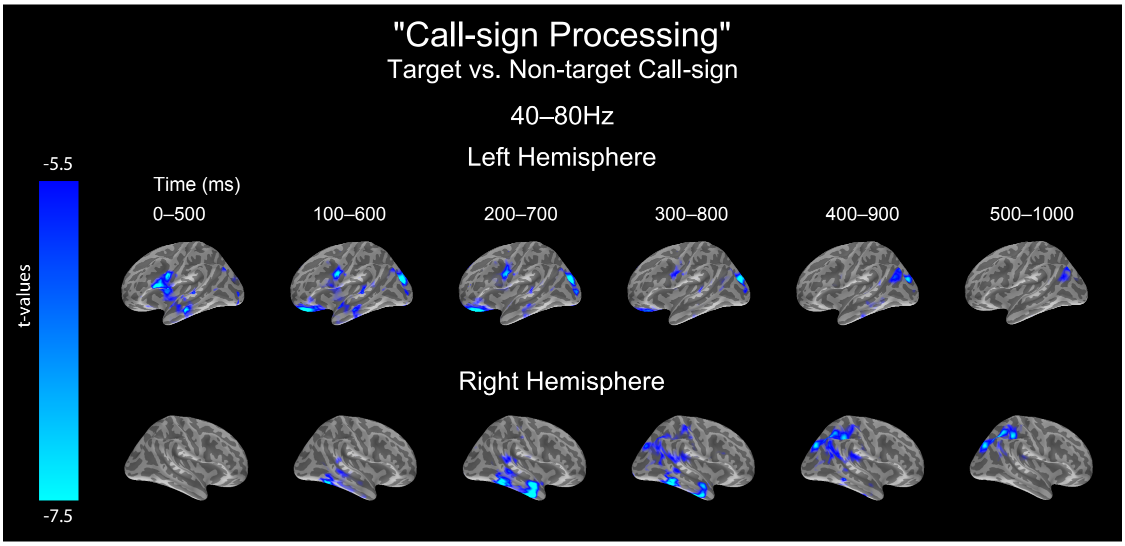

Figure 7.30. SPMs from the spatial filtering analysis for the "Call-sign processing" comparison. SPMs for each 500 ms time-window are displayed for the gamma band (40–80 Hz).

	MEG Performance		
	All	Young	Older
SLT Factor	$-.56^{***}$	$-.53^{*}$	$-.39$
BEA	$.52^{***}$	$.38$	$.48^{*}$

Table 7.3. Rank-order correlations between performance during MEG imaging (transformed scores), performance in the laboratory spatial listening tasks (SLT Factor), and the better-ear averages (BEA) (* $p < .05$, *** $p < .001$, two-tailed).

carpeted lounge-room. Can you follow what the other person says?") such that poorer MEG performance was associated with greater levels of self-reported difficulty [$\tau_b = -.42$, $p_{uncorr} = .01$, $p_{group} > .05$ ns]. This relationship did not remain significant after correcting for multiple comparisons. No significant correlations were observed within the young or older adult age groups.

A significant correlation was found between MEG performance and self-reported difficulties in the *Speech in Quiet* sub-scale [$\tau_b = -.41$, $p_{uncorr} = .01$, $p_{group} > .05$ ns] but did not meet the multiple comparison criteria. An examination of the component questions in this sub-scale revealed that this relationship was also being driven by Speech Question 2. No correlations were observed within either age group. In addition, no significant relationships were found between performance during MEG imaging and the three questions of particular interest (Speech Questions 3,4, & 12).

7.3.5.4 Cortical Activation ROIs

Relationships between MEG power differences in the three epoch pairs, spatial listening performance during MEG imaging, and performance in the laboratory were assessed at the sensor and source level.

Sensor-space

Peak difference measures extracted from nine of the time windows identified from the evoked analysis of the sensor data were correlated significantly with spatial listening performance during MEG imaging. Three of the windows were from the "Onset attention" comparison (O1–3) and six were from the "Call-sign processing"

	MEG Performance		
TEA Subtest	All	Young	Older
1A (V)	−.26	−.19	.02
1B (V)	$-.41^{**}$	−.23	−.31
2 (A)	−.16	−.34	†
3 (A)	−.09	−.37	.29
4A V	−.05	−.19	−.05
4B (V)	.27	.17	−.02
5 (A)	$-.35^{*}$	−.21	−.05
6 (V)	$.33^{*}$	.29	−.02
7 (A/V)‡	.23	.27	−.15
8 (A)	−.31	−.33	$-.53^{*+}$

Table 7.4. Rank order correlations between subtests of the TEA and performance during MEG imaging for all participants and individual age groups (* $p_{uncorr} < .05$, ** $p_{uncorr} < .01$, two-tailed; $p_{group} < .05$ for all *uncorrected* correlations after correcting for multiple comparisons apart from those marked with $^{+}$ $p_{group} > .05$). †Zero variance. ‡An outlier from the older adult group was removed for the correlations with subtest 7 of the TEA.

comparison (C1–6) (Table 7.5, see also Figure 7.19).

The results of the sensor-space correlation analysis are shown in Table 7.6. Measures of the magnitude and latency of the peak differences in power were significantly correlated with the transformed scores on the spatial listening task performed during MEG imaging. Three of the windows were significantly related to MEG performance within the young adult group. None of these relationships was significant within the older adult group. Performance in the laboratory spatial listening task (SLT Factor) was also significantly related to the size of the MEG power differences within two of the windows. The power differences in a window spanning 230–247 ms after the onset of the "Onset attention" epoch pair (window O1) was significantly related to laboratory performance across all participants, but not within either age group. The power difference in a late window from 585–638 ms after the onset of the "Call-sign processing" epoch pair (window C4), was significantly related to laboratory performance both within and across the two age groups.

Source-space

Table 7.7 lists the details of the measures from the source-space region of interest analysis for which a significant relationship was found with performance during MEG imaging. The ROIs labelled ROI O1 and C1–9 were identified from peaks in the the minimum-norm SPMs. The ROIs labelled S1–2 were identified based on the significant time windows from the sensor-space analysis. The measure extracted at each ROI was either the magnitude ('Peak') or latency ('Latency') of the peak difference within the specified time window. Figure 7.31 shows the locations of each of the ROIs which correlated with performance during MEG imaging. The coordinates of the ROIs are listed in Table G.10 (Appendix G, p. 283).

Table 7.8 lists the rank order correlations between each of the source-space

	Freq. band	Sensor group	Measure	Window (ms)
"Onset attention"				
Window O1	13–30 Hz	Right	Peak	230–247
Window O2	13–30 Hz	Right	Peak	67–89
Window O3	13–30 Hz	Global	Latency	228–247
"Call-sign processing"				
Window C1	13–30 Hz	Global	Peak	1–26
Window C2	8–13 Hz	Global	Peak	1–26
Window C3	13–30 Hz	Left	Peak	337–355
Window C4	4–8 Hz	Left	Peak	585–638
Window C5	13–30 Hz	Global	Latency	520–544
Window C6	8–13 Hz	Right	Latency	486–544

Table 7.5. Details of the windows from the sensor-space analysis in which the magnitude ('Peak') or latency ('Latency') of the peak differences correlated with performance during MEG imaging.

measures, the transformed performance measure during MEG imaging, and the factor extracted from the spatial listening task performance in the laboratory from Experiment 5. Within the “Onset attention” comparison, the latency of the peak difference in left posterior superior temporal sulcus (ROI O1) (Mean latency 118.5 ms, $\sigma = 79.1$) was related to performance.

A single relationship with MEG performance was identified as significant within either of the age groups—the magnitude of the peak power difference in left post-

	MEG Performance (transf)			**SLT Factor**		
	All	Young	Older	All	Young	Older
“Onset attention”						
Window O1	.33*	.47*	.09	−.30*	−.36	−.33
Window O2	.40**	.60**	.33	−.21	−.55*	−.03
Window O3	−.31*	−.24	−.45	.06	−.13	.14
“Call-sign processing”						
Window C1	.42**	.14	.33	−.28	.03	−.27
Window C2	.36*	.26	.36	−.15	−.15	.00
Window C3	.34*	.50*	.18	−.24	−.21	−.12
Window C4	−.36*	−.35	−.18	.53***	.46*	.55*
Window C5	−.32*	−.16	−.36	.21	−.11	.27
Window C6	−.30*	−.34	−.31	.12	.05	−.03

Table 7.6. Rank-order correlations between the sensor-space measures and the transformed performance scores of participants during MEG imaging (* $p_{uncorr} < .05$, ** $p_{uncorr} < .01$, *** $p_{uncorr} < .001$, two-tailed; $p_{group} < .05$ for all significant (uncorrected) correlations across all participants after correcting for multiple comparisons; within-group correlations were not corrected for multiple comparisons).

	Region	**Freq.**	**Hemi**	**Measure**	**Window (ms)**
“Onset attention”					
ROI O1	pSTS	0.5–4 Hz	LH	Latency	0–240
“Call-sign processing”					
ROI C1	STS	4–8 Hz	RH	Latency	541–588
ROI C2	AG	0.5–4 Hz	RH	Latency	318–825
ROI C3	MFG	0.5–4 Hz	RH	Peak	238–878
ROI C4	SOG	0.5–4 Hz	RH	Peak	420–799
ROI C5	POS	0.5–4 Hz	LH	Peak	324–878
ROI C6	POS	0.5–4 Hz	RH	Peak	491–836
ROI C7	CiS	0.5–4 Hz	LH	Peak	172–859
ROI C8	AG	0.5–4 Hz	RH	Peak	308–813
ROI C9	MTG	0.5–4 Hz	LH	Peak	224–430
ROI S1	PoCS	4–8 Hz	LH	Peak	579–612
ROI S2	PoCS	4–8 Hz	LH	Peak	637–668

Table 7.7. Details of the regions of interest from the minimum-norm analysis, identified from peaks in the SPMs, in which the peak differences or peak latencies correlated with task performance during MEG imaging.

central sulcus (ROI S1) within the older adult group. An analysis of the individual peak latencies revealed that the peak difference occurred at 601.0 ms ($\sigma = 16.5$) on average.

Two of the 11 measures from the “Call-sign processing” comparison were found to be significantly correlated with laboratory performance. The first measure was the latency of the peak difference observed in the angular gyrus (inferior parietal lobe) of the right hemisphere (ROI C2). The average latency of the peak difference was 609.5 ms ($\sigma = 136.9$). The second measure was the size of the peak difference in left middle temporal gyrus (ROI C9). The average latency of the peak difference was found to be 329.5 ms ($\sigma = 78.8$). This relationship was also found to be significant within the young adult group (Mean latency 323.3 ms, $\sigma = 72.7$).

Figure 7.32 shows four examples of the neural correlations of performance, two from the sensor-space analysis and two from the minimum-norm ROI analysis.

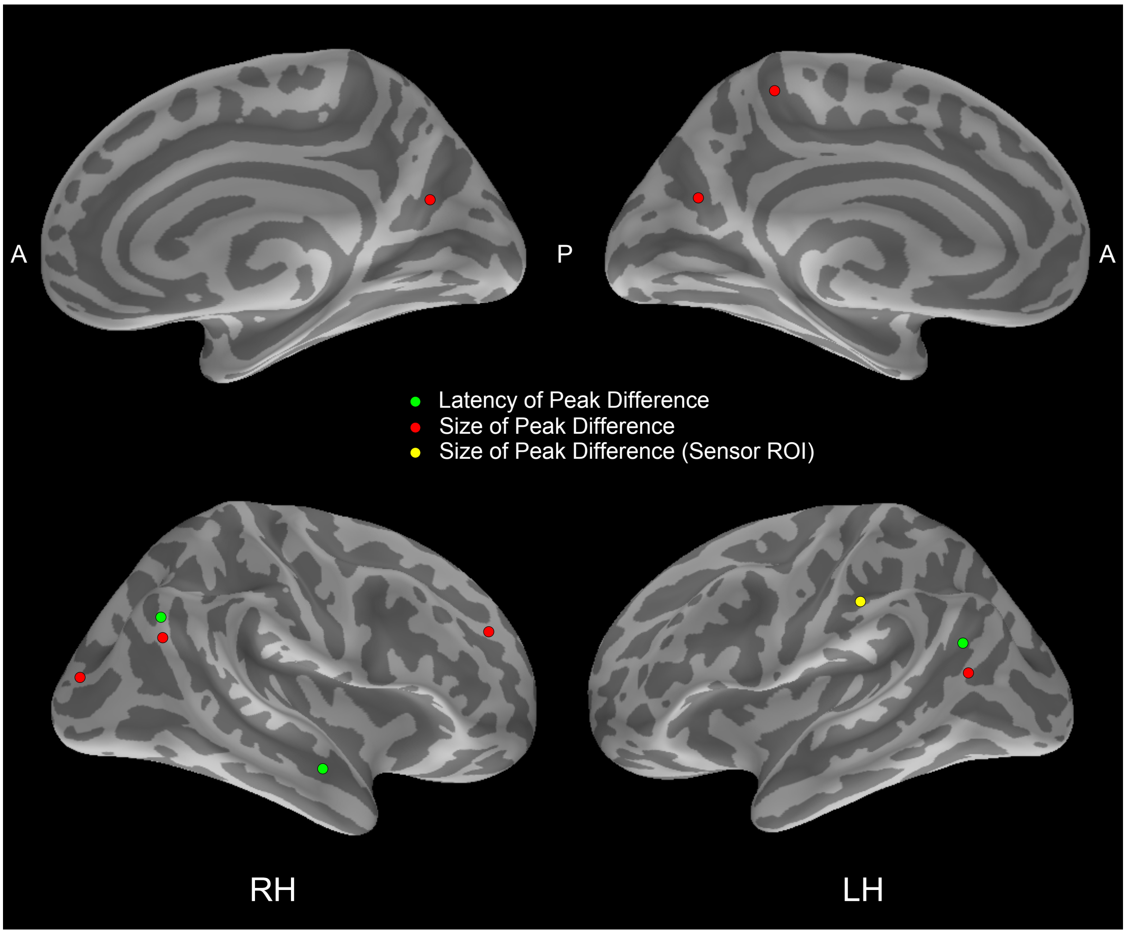

Figure 7.31. The location of minimum-norm ROIs which correlated with performance during MEG imaging. The colour code denotes whether the latency (green) or magnitude (red) of the peak difference exhibited the significant relationship. ROIs which were chosen based on the timing of significant relationships at the sensor level are shown in yellow.

7.4 Discussion

This chapter has described comparisons of measures of brain activity between three pairs of epochs at key moments during the spatial listening task. The key moments included times when attention had to be focussed on the onset of a new phrase ("Onset processing"), when participants had to distinguish their target call-sign from other non-target call-signs ("Call-sign processing"), and when a phrase onset had to be ignored so that attention could be sustained on the target phrase ("Onset attention").

The strong relationship observed between performance on the spatial listening task during MEG imaging and the similar tasks performed in the laboratory confirmed that the adapted version of the task was successful in translating the attention-demanding task from the laboratory to MEG. The use of a single target-to-non-target ratio of -3 dB was successful in producing an average performance level (86%) which was below ceiling but well above chance. In addition, the difference in performance between the young (91%) and older (81%) adult groups observed in Experiment 5 was preserved. As in Experiment 5, an association was observed between performance and hearing level, although this was significant only within the older adult group. It is possible that the older adults were particularly affected by difficulties adapting to the tube-phones, the limited frequency range of the stimuli, the impoverished frequency cues from the use of generic pinnae when recording the stimuli, or a combination of

	MEG Perf. (transf)			**SLT Factor**		
	All	Young	Older	All	Young	Older
"Onset attention"						
ROI O1	−.29*	.02	−.39	.24	−.19	.30
"Call-sign processing"						
ROI C1	−.33*	−.26	−.30	.23	.21	.03
ROI C2	−.35*	−.26	−.42	.36*	.39	.36
ROI C3	.30*	.23	.09	−.17	−.10	.03
ROI C4	.30*	.29	.00	−.24	−.33	.06
ROI C5	.29*	.32	.24	−.07	−.12	.12
ROI C6	.29*	.41	.00	−.09	−.03	.12
ROI C7	.31*	.26	.12	−.18	−.06	.12
ROI C8	.29*	.08	.09	−.09	.27	.09
ROI C9	.37*	.35	.18	−.41**	−.55*	−.18
ROI S1	−.30*	−.17	−.61**	.09	.03	.36
ROI S2	−.36*	−.41	−.18	.18	.00	.12

Table 7.8. Correlations between peak differences or peak latencies from the minimum-norm ROIs and task performance during MEG imaging (* $p_{uncorr} < .05$, ** $p_{uncorr} < .01$, two-tailed; $p_{group} < .05$ for all significant (uncorrected) correlations across all participants after correcting for multiple comparisons; within-group correlations were not corrected for multiple comparisons).

those factors. Greater variability in BEA thresholds within the older adult group may also have contributed to the observed relationship.

A brief summary of each of the three pairs of epochs will be presented followed by a detailed discussion of the results.

7.4.1 "Onset processing"

Differences in the response to phrase onsets which preceded target and non-target call-signs were identified only in the induced sensor-space analysis and were absent at the source level. The induced differences were more focal and transient that the induced differences found for the other two comparisons. The target phrase onsets included phrases in the 3^{rd} to 5^{th} slots in the sequence of phrases, whereas the non-target onsets only included phrases in the 3^{rd} and 4^{th} slots. Therefore, these

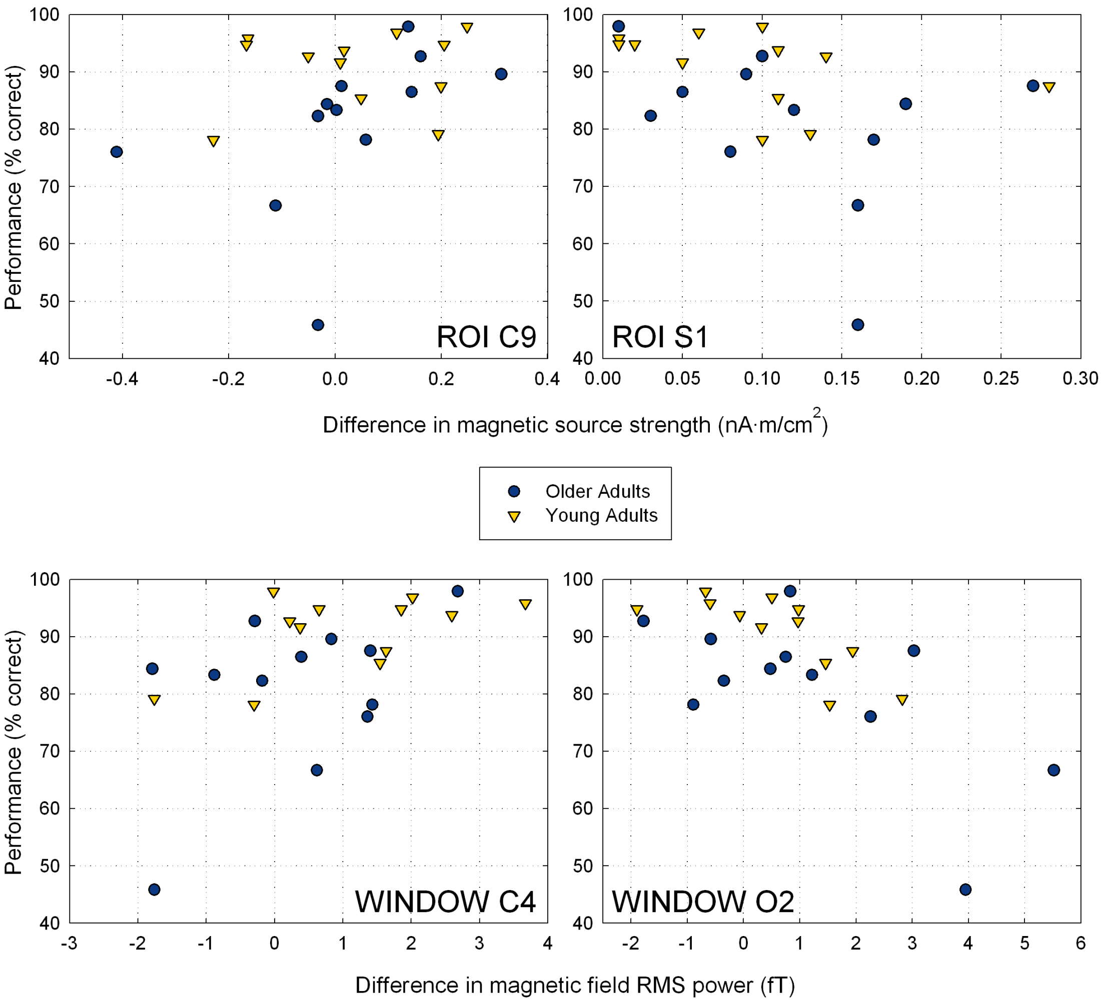

Figure 7.32. Scatter plots of performance during MEG imaging and differences in source power at two source locations (top) and field power at the sensors in two time-windows (bottom). Significant relationships (τ_b) were observed for the four pairs of variables. Data are shown for the young (yellow triangles) and older (blue circles) adults.

differences may have arisen due to slight differences in signal-to-noise ratios of the target onsets compared to the non-target onsets. Alternatively, the differences may be due to a greater level of adaptation of the cortical response to the phrase onsets which occurred later in the sequence of phrases.

The inconsistency of differences in the “Onset processing” comparison across the different analyses, when compared to the “Onset attention” and “Call-sign processing” comparisons, suggests that the underlying pattern of cortical activity was similar in response to a phrase onset regardless of the following call-sign. Furthermore, no measures extracted from the comparison correlated with task performance during MEG imaging. Together, these finding provide evidence that a new phrase onset induces a pattern of cortical activity that is independent of the call-sign which follows the onset. However, it is unclear whether the observed differences occurred by chance or as a result of the method used to analyse the data incorporated in the comparison. The differences in the “Onset processing” comparison will not be discussed further.

7.4.2 “Call-sign processing”

In contrast to the “Onset processing” comparison, clear differences were found between the responses to target and non-target call-sign keywords at the sensor and source levels. Significant increases in evoked power at low frequencies ($< 4\ Hz$) and decreases in power at higher frequencies ($> 4\ Hz$) were localised to a wide array of cortical regions, including frontal, parietal, and temporal cortices. These results confirmed the hypothesis that different patterns of cortical activity are observed for target and non-target call-signs, and that those differences include regions of the cortex previously related to processes of attention. This neuroimaging evidence for the recruitment of attentional processes was supported by the significant relationships between MEG performance and measures of attentional ability in the auditory and visual domains.

A distinction was observed between activation in the early and late portions of the “Call-sign processing” window. The latencies of differences in the later window were congruent with activity related to the onset of the phrase following the target call-sign. On average, the phrase onset occurred 499 ms after the onset of the call-sign—late differences were observed in sensor-space at 450 ms and in source space at 550 ms. This early vs. late distinction supported the hypothesis that there would be two distinct phases of processing in response to the target call-sign: 1) recognise and focus attention on the target phrases, and 2) resist distraction from the phrase onset following the target call-sign.

7.4.3 “Onset attention”

The “Onset attention” epoch pair compared this distracting onset which followed the target call-sign to the onset, to which participants attended, which immediately preceded the target call-sign. Compared to the late window of the “Call-sign processing” comparison, a similar group of cortical areas exhibited differences in the onset comparison. Attention-related activity was strongly right-lateralised, with low-frequency increases in right temporal regions and high-frequency decreases in right parietal and frontal regions. These findings confirmed the hypothesis that the phrase onset following the target phrase is associated with activation in attention-related regions of the cortex, possibly reflecting the resistance of distraction from the target phrase.

In the following section, I will discuss these differences at key moments within the spatial listening task in detail. I will present evidence from the sensor- and source-space results supporting the suggestion that attention plays an important role in demanding spatial listening tasks, that a wide network of cortical structures is recruited by the task, and that both focussed attention and resistance from distraction are required for successful performance. Secondly, I will discuss the observed relationships between task performance and levels of activation at both sensor and source levels.

7.4.4 Neural Bases of Focussing Attention and Resisting Distraction

The “Call-sign processing” epoch compared the MEG responses to target and non-target keywords and the processing which resulted from that discrimination process. To successfully perform the task, participants had to complete three key objectives: 1) recognise their target call-sign, 2) focus their attention on the location and talker who spoke the target call-sign to prepare for the colour and number information, and 3) resist distraction from any phrase onsets which followed their target call-sign. Failure to achieve any of these goals would be detrimental to performance; i.e. the ability to hear out the colour and number keywords in the target phrase. The first two processes are likely to occur concurrently according to the model of attention proposed by Knudsen (2007). The results of Experiments 1–4 showed that the onset of a new phrase at an unattended spatial location was successful at ‘grabbing’ attention automatically. In the terms of Knudsen’s model, the sensory information related to the new onset is assigned a strong weighting in the competitive processes which select from the large amount of incoming sensory information. This strong signal strength could arise from priming for phrase onsets facilitated by task training or prior experience of successfully coping with difficult listening situations. Both could lead to the voluntary adoption of a *dispersed* or non-focussed attentional state. Once information relating to the phrase onset gains access to working memory, top-down bias signals 'tune' the selection processes to favour information associated with the new phrase. In the current task, this tuning may include the reorienting of attention in the spatial or frequency domains. Thus, the acoustical and linguistic analysis of the new phrase leads to the continued focussing or sustaining of attention on that phrase, and the communication between these processes forms a "recurrent loop" of information within the attentional system (Knudsen, 2007).

7.4.4.1 Focussing Attention and Call-sign Analysis

Several significant time windows in the results of the evoked sensor-space data were identified as starting within 200 ms after the onset of the call-sign keyword. Such early differences may have arisen due to the use of anticipatory coarticulatory information available prior to the onset of the call-signs (at negative latencies relative to the "Call-sign processing" epochs). An alternative explanation is that the unique initial phonemes across the 8 call-signs could have aided discrimination based on small amounts of acoustical information. A supplementary experiment, described in Appendix F, examined whether it was possible to discriminate between target and non-target call-signs using only coarticulatory information. This experiment also determined the minimum amount of acoustical information necessary for discrimination performance to reach an asymptote. The results indicated that discrimination performance was above chance when only coarticulatory information was available, and that performance did not improve significantly when more than 80 ms of acoustical information was provided. Therefore, the results support the hypotheses that the observed early differences were possibly related to the call-sign discrimination process aided by coarticulatory information or the distinctiveness of the call-signs.

The results of the source-space analysis for the early part of the "Call-sign processing" epoch support the simultaneous analysis of and focusing of attention on the call-sign keyword. At low frequencies, the minimum-norm analysis indicated enhanced levels of activation in the left temporal lobe, including superior temporal gyrus and sulcus (STG and STS), and in the planum temporale (PT). These regions encompass areas of the cortex which have been associated with the processing of speech, including Wernicke's area (Hickok & Poeppel, 2000; Scott & Wise, 2003; Vigneau et al., 2006). Attention-related increases in activation in these regions have been found during acoustically-difficult tasks (Hashimoto et al., 2000; Nakai et al., 2005; Salvi et al., 2002) suggesting that the increased activation may be indicative of enhanced processing of the acoustical information. The importance of correctly identifying the target call-sign may have invoked a more detailed processing of the target call-sign compared to non-target call-signs. Alternatively, the increase in power may be related to the extraction of features from the target call-sign which would aid the tuning of attention, such as the fundamental frequency of the talker who spoke the target call-sign keyword.

Additional increases in power in response to the target call-sign were also observed in the left parietal and right frontal lobes. The left parietal differences in power were localised to the inferior parietal gyrus (IPG) close to the temporal parietal junction and the intra-parietal sulcus (IPS). The IPG and IPS are preferentially activated when orienting attention in either spatial or temporal dimensions, along with right lateral premotor cortex (Coull & Nobre, 1998). All of these areas exhibited target-related increases in activation (Figure 7.25), providing support for the orienting or continued focussing of attention shortly after the onset of the target call-sign. While both left and right parietal lobes have been implicated in the orienting of attention, the left parietal lobe in particular has been suggested to be involved in the orienting of attention at the 'local' rather than 'global' level (Posner & Petersen, 1990). In this case, it may imply a fine tuning of attention on the spatial location or individual features of the target talker, a process which may not be performed in the case of a non-target call-sign, resulting in the observed target-related increase in power. This assertion is further supported by the activation of left PT in the early window which has been found to be involved in the representation of space in the auditory domain (Deouell, Heller, Malach, D'Esposito, & Knight, 2007).

The notion of a "recurrent loop" within the attentional system implies the exertion of top-down bias signals, originating in processes of working memory, which serve to tune the competitive selection processes towards the attended object or information stream (Knudsen, 2007). Increased activation for target call-signs was identified in the middle frontal gyrus (MFG) of the right cerebrum. The dorso-lateral prefrontal cortex (DLPFC), encompassing the MFG, has been implicated as contributing to the central executive component of working memory (Baddeley, 2003). The activation of this region provides further support for the selective focussing of attention on, and analysis of, the target call-sign in the first 500 ms following the onset of the keyword.

Oscillatory Activity

Oscillatory activity at common frequencies across different regions of the brain has been proposed as reflecting communication between those regions (Salinas & Sejnowski, 2001) or as a method of associating different categories of information relating to the same stimulus or perceptual object (Tallon-Baudry & Bertrand, 1999). The spatial filtering analysis of the early part of the "Call-sign processing" comparison revealed two networks of activity showing oscillatory activity in common frequency bands. In the β band (13–30 Hz), decreases in oscillatory activity were found mainly in the left hemisphere, in frontal and temporal regions. Bilateral activity was also found in PT. The left lateralisation of β-band activity in the frontal lobe has been identified previously in response to target detection, with an increase in phase synchrony at β frequencies between left frontal and right parietal regions associated with an increased state of attentional alertness (Gross et al., 2004). While no strong right parietal differences were observed in the early part of the "Call-sign processing" window, the left fronto-temporal network did encompass many regions implicated in spatial attention such as the transverse frontal sulcus (Coull & Nobre, 1998) and PT (Deouell et al., 2007). This network may therefore indicate the communication between a group of target-related processes which were involved in the spatial focussing of attention on the target call-sign. In the γ band (40–80 Hz), decreases were found between 0–500 ms in left inferior frontal regions such as the IFG and bilaterally in middle-temporal regions. These are well-established as key regions for speech perception (Belin et al., 2000; Hickok & Poeppel, 2000). More anterior temporal regions which have been associated with the 'what' pathway and are modulated by intelligibility (Scott & Wise, 2003). Such a target-related network at γ frequencies may therefore be involved in the analysis of information related to the target call-sign keyword.

Temporal Sequence of Activation

In addition to the localisation of the target-related increases in activation, an insight into the temporal dynamics of the attentional system can be obtained from the time-course of the differential activations. The earliest peaks in the SPMs showing increases in power for the target were found in right frontal and left parietal lobes; i.e. those areas which have been implicated in attentional (Duncan, 2006; Posner & Petersen, 1990) and working memory (Baddeley, 2003) processes. Later peaks of activity were observed in the temporal lobe bilaterally 60–80 ms after the earlier peaks. The later temporal activation may indicate changes in processing based on top-down directives, either due to spatial or feature tuning (left parietal) or a bias signal affecting information selection processes (right DLPFC).

7.4.4.2 Resisting Distraction

A key part of successfully responding with the colour and number in the phrase containing the target call-sign was the ability to resist distraction from the onsets of new phrases after the target call-sign had been spoken. Figure 7.33 illustrates the timing of the phrase onset following the target call-sign and its relationship to the data windows defined for the analyses. Relative to the onset of the target call-sign, the next phrase onset starts at approximately 500 ms, with the variation in onset times arising from the variable onset times of the call-signs within each phrase. Thus, within the "Call-sign processing" pair of epochs, a difference in processing should occur after the 500 ms point, reflecting attention to the new phrase onset in the case of a preceding non-target call-sign and the resistance from distraction in the case of a preceding target call-sign.

This processing difference was identified as the *late* window in the "Call-sign processing" comparison, as reflected in the sensor- and source-space results. In the minimum-norm analysis, a group of late significant increases in power were identified starting at approximately 550 ms after the onset of the call-sign keywords. The pattern of these differences was largely distinct from the pattern of differences in the earlier part of the "Call-sign processing" comparison. Differences in activation in the late window arose in the posterior temporal and prefrontal regions in the left hemisphere, posterior temporal, middle temporal, and inferior prefrontal regions in the right hemisphere, and inferior parietal activity bilaterally. Several of these areas have been previously implicated in difficult tasks of speech perception. Activation of bilateral pre-central regions has been observed when listeners segregate two speech streams spoken by the same talker (Nakai et al., 2005). The bilateral posterior activity spanned the inferior parietal lobe, including the supramarginal and angular gyrii, and posterior temporal lobe. It has been suggested that both gyrii contribute to auditory selective attention (Nakai et al., 1999; Sabri et al., 2008), and have been associated more generally with selective attention as the "posterior parietal network" (Posner & Petersen, 1990).

Shifts of Attention

The foci of activity within the *late* window in the right temporal lobe encompassed

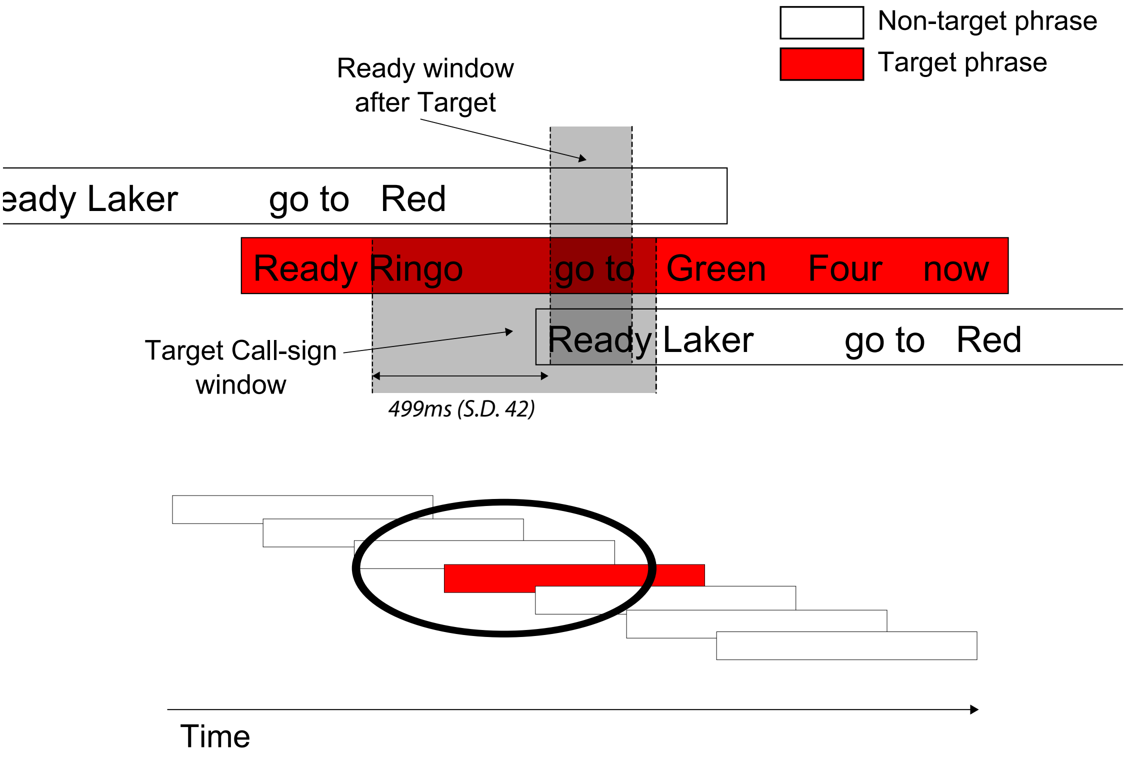

Figure 7.33. Top: A schematic representation of the phrase onset following the target call-sign and the overlap between the chosen data windows. Bottom: A diagram of the sequence of phrases. The outlined section of the sequence corresponds to the magnified section shown above.

the superior temporal gyrus lateral to Heschl's gyrus, middle temporal gyrus, and PT. Attention-related increases in activity in similar regions of the STG have been identified in a range of studies in which attention is directed towards an auditory stimulus (Hashimoto et al., 2000; Nakai et al., 2005; Salvi et al., 2002; Woodruff et al., 1996). In the current study, it was hypothesised that attention would have to be sustained on the target phrase to perform the task successfully, and that distraction by the onset of the phrase following the target phrase would have to be resisted. The activation of areas directly lateral to the primary auditory regions on Heschl's gyrus may be an index of a top-down influence on the processing of and attention to incoming auditory events. Sabri, Liebenthal, Waldron, Medler, & Binder (2006) identified the right superior and middle temporal gyrii as being involved in shifts of auditory attention to infrequent deviant stimuli. This suggestion was compatible with a peak observed from 60–110 ms in the scalp-recorded EEG, the P3a, which is an ERP associated with involuntary shifts of attention (Näätänen, 1990; Soltani & Knight, 2000). In addition, Sabri et al. (2006) found a later enhanced response (210-340 ms) in the EEG difference wave (standard vs. deviant) when the difficulty of the deviance detection task was increased. It was suggested that the later modulation of activity in STG was indicative of a top-down influence on the level of processing resources allocated to the task. In the current experiment, increases in activity in the late time window were observed on the STG at comparable latencies (585 ms and 783 ms) relative to the onset of the phrase onset relative to the call-sign onset (499 ms, $\sigma = 42$). Therefore, the strong right-temporal activation may represent processes related to the involuntary shift of attention induced by the phrase onset and/or the effects of top-down attentional control on the inhibition or correction of such shifts. The middle temporal gyrus activation provides additional support for the distracting effect of the phrase onset, having been associated with changes in the width of spatial attention (Chen, Marshall, Weidner, & Fink, 2008).

ERP studies which have examined the effects of top-down control on attentional shifts have found that attention to one stream within multiple auditory streams reduces the depth at which the non-attended streams are processed (Sussman, Bregman, Wang, & Khan, 2005) and that high predictability of distracting events can eliminate involuntary attentional shifts (Sussman, Winkler, & Schröger, 2003). The results of Experiments 1–4 suggested that in the spatial listening task, despite the temporally-regular and therefore predictable sequence of phrases, the onset of a new phrase was successful in distracting the listener and impeded their ability to sustain their attention on the phrase containing the target call-sign. It is therefore not surprising that the predictability of the phrase onset in the current task did not result in the absence of any related cortical response, as was identified for ERP responses (Sussman et al., 2003). An additional consequence of the strong attentional effect of phrase onsets is that selective attention to one auditory stream (the target phrase) would not be likely to completely inhibit a response to a distracting auditory event in another stream (a new phrase onset), also suggested from ERP results (Sussman et al., 2005). However, the focus of activity in STG around the moment of distraction suggests that the extent to which the new phrase onset was processed might have been limited to a superficial level based on top-down control.

Medial Activation

Medial differences were also found in the latter part of the “Call-sign processing” comparison in bilateral anterior cingulate, left posterior cingulate, and left precuneus. The anterior cingulate cortex (AAC) in particular has been identified in a range of studies of auditory attention which impose high attentional demands on the listener (Nakai et al., 2005; Sabri et al., 2008; Salvi et al., 2002). It has also been found to be modulated by increasing the number of target items (Posner et al., 1988) or when attention must be focussed on a single source when there are multiple streams of information (Salvi et al., 2002; Posner & Petersen, 1990). The simultaneous activation of prefrontal, posterior parietal, and ACC regions, which are known to share connections (Goldman-Rakic, 1988), suggests the involvement of the wider fronto-parietal attentional network (Duncan, 2006; Posner & Petersen, 1990) at a time when attention must be strongly focussed on the target phrase despite multiple competing sources of information.

Maintaining Attentional Focus

The “Onset attention” comparison provided an alternative examination of the cortical response to the phrase onset following the target call-sign, as it was synchronised to that phrase onset (Figure 7.33). This synchronisation provided a more accurate representation of the data phase-locked to the onset. The pattern of differences was very similar to that observed in the later portion of the “Call-sign processing” comparison. In addition to superior and middle temporal gyrii, the right temporal activity was localised immediately adjacent to Heschl’s gyrus, on the transverse temporal sulcus. Other similarities with the later window from the “Call-sign processing” comparison included increases in power in bilateral posterior parietal regions, anterior cingulate cortex, and decreases in power in response to the target call-sign relative to non-target call-signs at 4–8 Hz. These similarities support the suggestion that the later differences in the “Call-sign processing” comparison were related to the onset of the following phrase.

Further evidence supporting the role of attention in the resistance of distraction from the phrase onset following the target call-sign was found in the spatio-temporal analysis of the induced data at the sensor level and in the spatial filtering analysis at the source level. Significant decreases in power in β and γ frequency bands were observed for the target call-sign in the “Call-sign processing” comparison and for the phrase onset following the target call-sign in the “Onset attention” comparison. An examination of the evolution of the decreases in power after the onset of the call-sign keywords, specifically around 20 Hz and 55 Hz (Figure 7.21), showed the largest target-related decrease in oscillatory activity at 600 ms, or approximately 100 ms after the phrase onset which followed the call-sign. Oscillatory activity as measured with MEG requires the synchronisation of large populations of neurons (Jensen et al., 2007). Likewise, decreases in oscillatory power might reflect the de-synchronisation of such populations or a reduction in the number of neurons firing. Such de-synchronisation might be used to control the gain of a signal within a network of cortical regions (Salinas & Sejnowski, 2001). As synchronisation possibly enhances the processing of a stimulus by coupling together cortical regions involved in representing different aspects of stimulus (Tallon-Baudry & Bertrand, 1999), de-synchronisation may serve to inhibit such a network. Changes in oscillatory activity have therefore been implicated in the adoption of, or considered as a reflection of, different attentional states (Tiitinen et al., 1993; Gross et al., 2004). The frequency of oscillation may reflect communication over different distances, with neural models suggesting that γ is suitable for short range or local communication and β frequencies for more distant communication (Bibbig et al., 2002).

Using spatial filtering, decreases in oscillatory activity were localised to the right parietal lobes in both the latter part of the “Call-sign processing” comparison and relative to the following phrase onset in the “Onset attention” comparison. In the β band, the differences were found in the posterior parietal lobe, inferior frontal gyrus, and cingulate cortex. In the γ band, the majority of the activity was also localised to the posterior parietal lobe and cingulate cortex, with more dorsal frontal activity in the inferior frontal sulcus. The strong differences in posterior parietal areas imply processes of attention, specifically the switching of the attentional focus (Posner & Petersen, 1990). The observed target-related decrease in oscillatory activation may indicate the inhibition of such a network. The broader network of regions in both frequency bands has been found to be activated by a wide range of tasks which require attention and are almost identical to the “multiple demand” network (Duncan, 2006). In the θ band (8–13 Hz), linked to the suppression of cortical processes (Ward, 2003), decreases in activity were found in the intra-parietal sulcus, adjacent to the superior parietal gyrus, and the inferior frontal gyrus of the right hemisphere. Chen et al. (2008) associated both regions with the ‘zooming out’ of attentional focus, or the adoption of a wider attentional focus. The differential θ activity may therefore imply the inhibition of this function at the key moment when an involuntary shift of attention may be induced by the onset of a new phrase. These results, based on the analysis of the induced data, further support the notion that resisting distraction from a new phrase onset involves a complex and widespread network of cortical regions, with the de-synchronisation of neural activity possibly acting to inhibit shifts of attention to, or the processing of, the distracting phrase onset.

7.4.4.3 Interim Summary

In summary, the results of the source-space analysis suggest that the strong attention-grabbing effect of a new phrase onset, as identified by the results of Experiments 1–4, requires voluntary intervention either to inhibit an involuntary, stimulus-driven, shifts of attention or to refocus attention on the target phrase in the event of a shift.

7.4.5 Neural Correlates of Performance

The correlational analysis of the latencies and magnitudes of significant differences in the MEG field power at the sensor level and reconstructed neural activation power at the source level examined the links between cortical activity and successful performance on the spatial listening task. Despite the gross nature of the analyses at the sensor level, measures from 9 time windows identified from the evoked sensor-space analysis and from 12 regions of interest in the source-space analysis correlated with performance. All but 4 of the measures were either extracted from the "Onset attention" comparison or from the latter part of the "Call-sign processing" comparison. Visual inspection of scatter plots of the data (Figure 7.32) suggested that the correlations did not arise solely due to differences in the mean values of the two age groups on the variables being correlated. Furthermore, the majority of the source ROIs were localised to posterior and inferior parietal, frontal, and cingulate cortices. These regions of the cortex have been consistently associated with processes involved in the selective focussing and/or shifting of attention (Duncan, 2006; Posner & Petersen, 1990). These results support the hypothesis that the central processes which are elicited at key moments within the spatial listening task, specifically at the time which the listener must resist distraction from the onset of a new phrase, are task-critical.

Of the 15 magnitude measures which correlated with performance, all but three showed positive correlation values such that an increase in the number of errors on the task was associated with larger differences in cortical or sensor power. The majority of the measures which showed a positive correlation, 8 of the 12, were extracted from latencies which were compatible with activity related to the onset of the phrase following the target call-sign; i.e. in the later part of the "Call-sign processing" comparison or within the "Onset attention" comparison. It is therefore possible that the association between a smaller difference in power and performance reflects the cost of recruiting additional processes at key moments. The timing of the peak differences may suggest that greater differences in power reflect the distracting effects of the phrase onset following the target call-sign. The localisation of the differences to regions including the posterior parietal lobe and cingulate cortex provides some support for this assertion, as such structures have been related to selective attention (Corbetta et al., 1993; Coull & Nobre, 1998; Posner & Petersen, 1990). However, this ‘irrelevant processing’ hypothesis is not directly testable using the current data and further study is required.

All of the latency measures in both sensor- and source-space which were significantly related to performance exhibited negative correlations with the transformed performance scores. In other words, an increase in the number of errors in performance was associated with a shorter latency of the peak difference. This finding does not support the hypothesis that poorer performance in the older adult group, suggested to arise from a decline in speed of processing from the results of Experiment 5, would be associated with longer latencies in evoked power. The latency of cortical activity has traditionally been used to assess decreases in the speed of cognitive processing with advancing age. Many previous studies have used the P300, an event-related potential commonly evoked using odd-ball tasks and interpreted as an index of attentional and memory processing (Soltani & Knight, 2000), to study the relationship between evoked cortical activity, cognition, and ageing. It is a general finding that the amplitude of the P300 decreases and the latency of its peak increases with advancing age (Picton, 1992). The finding that ERP components which occur earlier than the P300 in response to a stimulus, reflecting sensory processing, do not exhibit age-related increases in latency is taken as evidence that the slowing of the P300 is indicative of a decline in cognitive function (Dujardin et al., 1993). Likewise, shorter latencies have been associated with high levels of cognitive performance (Polich, 1996). It is therefore unlikely that the significant relationships between the latency and performance measures observed in the current experiment are indicative of a general effect of ageing on the speed of cognitive performance. The location of the latency measures to left MTG and right anterior STS suggest that the differences in activity are arising from the processing of the auditory input rather than higher cognitive functions. The extraction of measures from the MEG data which better reflect reductions in the speed at which information can be processed may require a more detailed analysis at the level of individual participants.

Finally, it is important to acknowledge that measures related to peak differences in activation are not the exclusive candidates for neural correlates of attention. Further analyses examining the absolute level of cortical activation in either condition, relationships between the activation levels in different regions of the cortex, and the phase synchrony of oscillatory activity amongst related regions are desirable.

To summarise, neural correlates of successful performance were identified at key moments during the spatial listening task. The correlates included difference measures of magnetic field power and the magnitude of cortical activation. The results supported the hypothesis that the activity of brain processes at key moments in the task are critical to successful performance.

7.5 Direction of Source-space Effects

Systematic differences in the direction of effects between the analyses of evoked and induced data were observed in source space. Generally, evoked power as reconstructed by the minimum norm technique was found to be significantly greater for target compared to non-target call-signs (“Call-sign processing”) and for phrase onsets following a target call-sign compared to phrase onsets immediately prior to a target call-sign (“Onset attention”). The direction of the effects was reversed for the spatial filtering analysis of induced data. This difference may have arisen for two reasons. Firstly, the techniques are measuring different aspects of the MEG data. Minimum norm only measures that portion of the MEG signal which arises due to the synchronisation of large populations of neurons in response to a specific event, and the activity of those neurons is therefore tightly phase-locked with respect to the event onset. However, aspects of the MEG data which are not tightly phase-locked to the stimulus event will only be revealed by the spatial filtering analysis. Secondly, the choice of analysis technique was confounded with frequency—spatial filtering was mainly used to look at activity above 8 Hz and minimum norm was applied to the data below 8 Hz. Therefore the differences in the direction of the results may not be related to the specific techniques used, but rather due to the differences in the direction of the effects at low and high frequencies. The observation that the majority of sensor-space effects over 4 Hz had a different direction than those below 4 Hz for both evoked and induced analyses provides some support for this hypothesis.

7.6 Conclusions

The ability of listeners to hear out information spoken by a target talker while many other people are speaking at the same time requires a wide range of central processes including those related to the focussing and shifting of attention and resisting distraction. Indices of the activation of these processes can be used to predict performance in a demanding task of spatial listening.

7.7 Summary

- Brain activity was recorded using MEG while participants performed a spatial listening task.

- The task was an adapted version of the listening task which was also performed by the same participants in a laboratory setting.

- Key moments were identified in the spatial listening task at which appropriate brain activity was hypothesised as being crucial to successful task performance.

- As performance on the laboratory version of the task had been found to be related to auditory and visual measures of attention, it was expected that a wide range of cortical regions associated with the selective focussing of attention, shifting attentional focus, and resisting distraction would be activated by the task.

- The results confirmed that there were significant differences between the cortical response at key moments, including in response to target and non-target call-sign keywords, and that these differences were localised to regions of the brain associated with attentional processes.

- The differences were categorised as occurring in early (< 500 ms) and late (> 500 ms) time windows relative to the onset of the call-sign keyword. These analyses supported the hypothesis that two phases of processing were required after the onset of the target call-sign—focussing on the phrase containing the target call-sign and resisting distraction from the phrase onset which followed the target call-sign.

- The early differences involved regions of the brain associated with auditory selective attention and language processing. The later differences involved regions previously implicated in shifts of attention, and exhibited oscillatory activation which suggested that these regions may have been suppressed as a result of top-down attentional control. Similar results were observed from the analysis of the data synchronised to the distracting phrase onset.

- Taken together, these results supported the hypothesis that listeners had to exert attentional control to avoid distraction from the phrase which followed the target call-sign keyword to maintain attentional focus on the phrase containing the target call-sign.

- Measures of the magnitude and latency of cortical activation were extracted from the MEG data at sensor and source levels. Several of the measures were found to be significantly correlated with performance. The source-space measures were localised to cortical regions encompassing attention-related processes. These relationships were interpreted as a confirmation of the hypothesis that central processes, including cognitive functions, played a critical role in achieving successful performance on the spatial listening task.

Chapter 8

Summary and General Discussion

This chapter summarises the main findings from the six experiments in this thesis. Issues arising from the research are discussed, and directions for future research are proposed.

8.1 Recap of Research Aims

The overall goal of the experiments presented in this thesis was to examine the role of cognition, specifically attentional ability, in the perception of speech in situations where many people are speaking at the same time. The first aim was to develop an attention-demanding task of spatial listening for speech which recreated the cognitive demands that are placed on listeners in demanding situations that arise in everyday life. Experiments 1–4 assessed the importance of voluntary and involuntary attention to talkers within a multi-talker environment among young normally-hearing listeners using the spatial listening task. The ability of a new talker to induce shifts in attentional focus automatically, and the extent of the time window during which a new talker could distract a listener, were investigated. The experiments also examined the ability of listeners to overcome such stimulus-driven shifts in attentional focus through the voluntary control of attention by providing information about *who* to listen for, *where* they would speak from, and *when* they would speak. The second aim was to investigate the relationships between performance on the spatial listening task, attentional ability, hearing sensitivity, and difficulties that listeners report in everyday situations. The fifth experiment examined these relationships among young and older normally-hearing adults using an attentional test battery comprising both visual and auditory tasks, and a questionnaire designed to elicit self-report measures of listening difficulties. The third aim was to record the brain activity of listeners using magneto-encephalography (MEG) while they performed the spatial listening task. The sixth experiment investigated the pattern and time-course of this cortical activity among the young and older adult listeners from Experiment 5.

8.2 Summary of Findings

8.2.1 Main Findings of Experiments 1–4

1. When talkers started speaking one at a time in a regular sequence, listeners could hear out the target phrase at negative SNRs. A small benefit arose from knowledge about *who* would speak the target phrase. No benefit was received from knowing *when* it would appear or *where* it would be spoken from.

2. When pairs of talkers started speaking simultaneously, performance was poorer compared to when phrases started one at a time. Listeners were only able to report information from the target phrase at negative SNRs when *a priori* information about the target phrase was provided.

3. The distracting effect of a new person starting to speak shortly before or after a target talker was apparent over a broad time window, and was related to uncertainty about when the target phrase would be spoken. When there was a high degree of uncertainty about when the target phrase would be spoken, the introduction of an asynchrony of up to 320 ms between target and masker phrases did not reduce the distracting effect of the masker phrase. When uncertainty about when the target phrase would be spoken was reduced, listeners were able to focus attention on the target phrase at lower SNRs when the target and masker phrases were separated by 160 ms or more. The window in which the masker phrase had a distracting effect, the limits of which were defined by the asynchronies which were associated with a 3 dB decrease in SRTs, was −640 ms to +320 ms relative the onset of the target phrase.

8.2.2 Main Findings of Experiment 5

1. Older adults performed poorer on the spatial listening task compared to young adults, but derived as much benefit as the younger adults from *a priori* information about a target phrase both when phrases started one at a time and in pairs.

2. Young and older adults were equally susceptible to distraction from a masker phrase which started simultaneously with the target phrase and was more intense than the target phrase.

3. Poorer hearing sensitivity was related to poorer performance on the spatial listening task, both between and within the young and older adult groups.

4. Older adults performed less well than younger adults on 5 of the 8 tasks of attention, including purely visual tasks. An analysis of the requirements of the

tasks indicated that good performance on four of the tasks required fast speed of processing.

5. Performance on the spatial listening task was related to four of the attention tasks for which age-related differences were found. Three of those tasks were influenced by the speed of processing.

6. Two models were proposed to account for the contributions of hearing sensitivity and attentional abilities on performance on the spatial listening task. The observed pattern of relationships in the data could not distinguish between the models.

7. Poorer performance on the spatial listening task was also related to difficulties reported by participants with following a conversation in a group of people in everyday life.

8.2.3 Main Findings of Experiment 6

1. Different patterns of cortical activity were identified at key moments in the spatial listening task. The key moments included discriminating between target and non-target keywords and resisting distraction from a new talker starting to speak. The differences in the power of cortical activation were localised to parietal and frontal regions of the brain that have previously been associated with attention.

2. Differences at latencies congruent with processes related to the discrimination of target and non-target keywords were identified in regions of the brain associated with auditory selective attention and language processing.

3. Differences within the time window in which distraction from a new talker had to be resisted involved regions previously implicated in shifts of attention, and exhibited oscillatory activation suggesting that activity in these regions may have been suppressed as a result of top-down attentional control.

4. The amplitude or latency of the differences in cortical activity at 12 locations in the brain were found to be significantly correlated with performance. The majority of the correlates were localised to the parietal lobe, in regions associated with attentional processes rather than auditory or language processing.

8.3 General Discussion

8.3.1 Effects of Age: Assessing the Impact of Cognitive Ability

The results of Experiments 5 and 6 provide insights into the contribution of cognitive deficits to speech perception difficulties among older adults. Poorer performance was observed among the older adults compared to the young adults on visual, auditory, and multi-modal tasks of attention from the Test of Everyday Attention (TEA) (Robertson et al., 1996). All but one of those tasks were associated with speed of processing—the ability to perform the tasks at a high speed, or to process information rapidly, were likely to improve performance.

Although the results of previous studies which have examined relationships between speech perception and cognitive factors have suggested that a general cognitive decline is not responsible for age-related deficits in speech perception (Pichora-Fuller et al., 1995; Helfer & Freyman, 2008), several general effects have been associated with the age-related decline in cognitive performance. These effects include a decrease in the speed of processing (Salthouse, 1996) and the "complexity effect" (Kok, 2000) which associates increased task complexity with increased age-related deficits in performance. These effects are related to general physiological changes that accompany the natural ageing process which are not specific to any single functional area; e.g. a reduction in synaptic connections (Willott, 1996), demyelination (Albert, 1993; Bartzokis, 2004), and degradation of white matter fibers (O'Sullivan et al., 2001). As a result of these general changes, it has been suggested that tasks which involve a large array of processes, distributed throughout the cortex, will exhibit greater age-related differences compared to tasks which involve processes which are more local to each other; e.g. low-level processes within a single sensory modality (Kok, 2000).

While it is unlikely that a decline in speed of processing was solely responsible for all of the group differences observed on the tests of the TEA, the results of Experiment 5 suggest that speed of processing had a significant impact on the performance of the older adults on the attentional tests and the complex task of spatial listening, as reflected by the correlations between the attentional measures and SRTs. Evidence for an effect of task complexity was also evident, with poorer performance among the older adults on the more demanding tasks from the TEA (subtests 1, & 4–7) and on the spatial listening task, and no age-related effects for those TEA tasks which did not place high cognitive demands on participants (subtests 2, 3, and 8).

Experiment 5 also provided evidence to support the suggestion that not all cognitive functions are impaired with age. For example, the ability of older adults to benefit from *a priori* information about the target phrase in the spatial listening task (Mean benefit 6.5 dB) was not significantly different from the group of young adults

(Mean benefit 6.7 dB). Likewise, both groups showed an equal level of susceptibility to stimulus-driven attentional capture by a more intense phase when two phrases started simultaneously (*Paired-Unpaired* SRTs; Young +7.0 dB, Older +6.0 dB). Thus, the results agree with previous findings that certain aspects of cognitive function, specifically volitional control over attention and involuntary attentional shifts, are well preserved with age (Folk & Hoyer, 1992; Madden, 1990).

While the results of Experiment 5 suggested that the older adults were equally susceptible to stimulus-driven attentional capture, the timing and location of the measures extracted from cortical activity may indicate that the older adults were poorer at resisting distraction while their attention was focussed on the target phrase. Differences in cortical power, in a time window surrounding the onset of a new phrase which they had to ignore, were located largely in the parietal lobes, and correlated with performance on the task during MEG imaging. These neural correlates suggest that deficits in the ability to maintain attentional focus on the target phrase contributed to poorer task performance.

In summary, the results of the behavioural and neuroimaging experiments presented in this thesis have shown that listening to what one person is saying when many people are speaking at the same time involves a broad array of attentional processes, including those required to sustain attention and resist distraction. The results confirm the general finding that complex cognitive tasks reveal age-related differences in performance, regardless of whether they are in the visual or auditory domain. The results also suggest that the ability to cope with difficult listening tasks requires the rapid processing of sensory information. These conclusions imply that age-related deficits in coping with these listening situations are likely to be due, at least in part, to slower processing speeds and a decline in specific cognitive functions—those related to resisting distraction and maintaining attentional focus.

8.3.2 Accounting for Individual Differences in Performance

Higher performance levels on the listening task, both in the laboratory and during MEG imaging, were associated with better attentional ability as measured by the TEA, and to better hearing sensitivity, among a group of young and older normally-hearing adults. The network of relationships that was identified in Experiments 5 and 6 is shown in Figure 8.1.

The proportion of variability in performance on the spatial listening task that could be explained by the chosen predictors was similar to previous studies which have examined individual differences in performance on speech-in-noise tasks together with predictors of performance, including hearing sensitivity, age, and cognitive measures. In a review of such studies, Houtgast & Festen (2008) concluded that no single factor completely explains speech perception performance. Apart from a single study which included individuals with noise-induced hearing loss, predictor variables explained only 70% of the variance in speech perception performance on average. Houtgast & Festen (2008) offered several suggestions for the 'missing 30%' of variance, including the choice of linear models which may not always be appropriate, and a failure to incorporate estimates of measurement error. However, it was also possible that additional variables which contributed to performance deficits may not have been considered.

Based on these conclusions, it may be possible to explain a higher proportion of variance in the spatial listening task performance data from Experiment 5 if a more complex model is constructed to include additional predictors, such as age, and multiple measures from the TEA to simultaneously account for deficits in a variety of cognitive functions. However, the current data did not support the construction of a more complex model due to the small sample size, which could be addressed in future studies.

8.4 Future Research

8.4.1 Understanding the Nature of Difficulties in Multi-talker Tasks

Experiments 1–5 were successful in identifying those individuals who experienced difficulties with listening to what one person was saying when many people were speaking at the same time. To gain a thorough understanding of why those difficulties

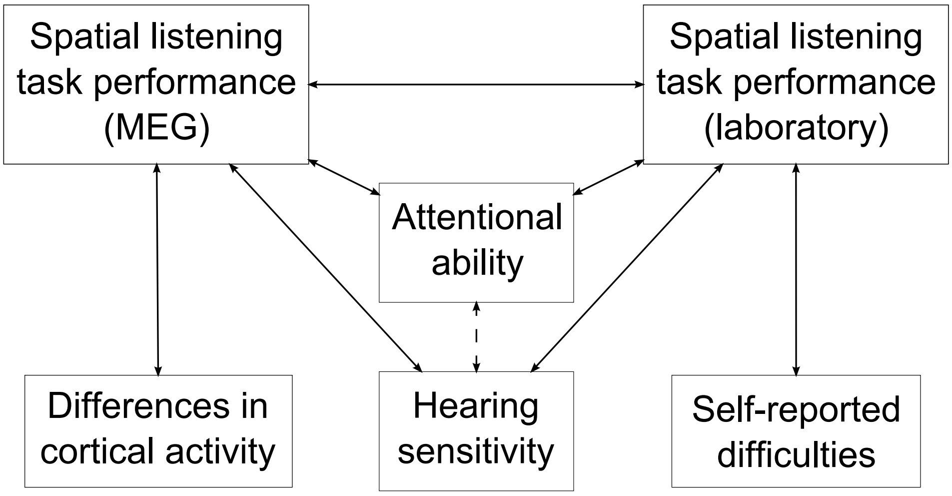

Figure 8.1. Network of significant rank-order correlations (solid lines) between measures of speech perception, attentional ability, hearing sensitivity, self-reported difficulties with listening in everyday environments, and measures extracted from the cortical activity of participants. The dotted line denotes a significant relationship between purely-visual tasks of attention and hearing sensitivity—a relationship which was attributed to differences in the means of the young and older groups for both variables.

arose, it would be desirable to examine the nature of the errors that participants made. An analysis of errors might help identify the aspects of the task which each listener found most challenging. This type of analysis has been applied successfully to data from multi-talker listening tasks. For example, Kidd et al. (2005) performed an error analysis on the results of a multi-talker listening task to identify whether errors were random or due to confusion between target and masker phrases, and whether the pattern of errors changed as the predictability of the target phrase location was varied.

The design of the current experiments did not permit a similar analysis of errors for two reasons. First, the number of incorrect trials was insufficient to perform the analysis. The use of adaptive routines in the laboratory resulted in variable numbers of incorrect trials across participants at a range of SNRs, and the deliberate selection of an SNR which resulted in a high proportion of correct responses during MEG imaging resulted in too few incorrect trials for an analysis of errors. Second, unlike the task used by Kidd et al. (2005), the spatial listening task used in Experiments 1, 5, & 6 comprised phrases which started in sequence rather than at the same time. This reduced the possibility of classifying errors unambiguously as intrusions from words in masker phrases, because the call-sign, colour, and number keywords in different phrases did not overlap in time.

Nonetheless, an analysis of errors would help provide a better understanding of bases of individual differences in performance observed in Experiments 1–5. Those differences could have arisen as a result of 1) a failure to identify the target call-sign and thus the target phrase, or 2) a failure to maintain the focus of attention on the target phrase once identified and to resist the stimulus-driven capture of attention by talkers who started to speak after the target phrase. Experiment 6 showed that individual differences in performance were related to differences in cortical power in regions associated with the resistance of distraction and shifts of attention. This result is compatible with the idea that individual differences in performance on the spatial listening task did not arise due to a failure to identify the target phrase, but rather from a failure to maintain attention on it. A further experiment could examine this hypothesis by distinguishing between the two possible performance deficits. Participants would be required to make an additional response to indicate when they heard the target call-sign. Participants would be required to respond before the next call-sign was spoken. The inclusion of catch-trials in which no target phrase was presented would confirm that participants' responses were specifically related to the presence of the target phrase.

Individuals for whom difficulties with performing the task were primarily related to the identification of the target call-sign would be expected to produce a large number of call-sign identification errors. For those trials in which the call-sign was correctly identified, those individuals should produce few errors related to the identification of the colour-number co-ordinate, indicating the ability to maintain attention on the target phrase and resist distraction. Individuals who were poor at resisting distraction from the onset of new phrases once the target phrase has been identified would be expected to only make a small number of call-sign identification errors. However, for those trials in which the target call-sign is correctly identified, those listeners would be expected to make a large number of colour-number identification errors.

8.4.2 Disassociating Hearing Loss and Cognitive Difficulties

The comorbidity of hearing loss (Davis, 1989; Hannaford et al., 2005), cognitive deficits (Kok, 2000; Verhaeghen & Cerella, 2002), and speech perception difficulties (Wiley et al., 1998) among older adults makes the process of disassociating the contributions of age-related peripheral and central changes to speech perception difficulties challenging. To address this challenge, a future experiment could examine all three variables on groups of young and older adults, both containing individuals with and without sensori-neural hearing loss. This design has been applied successfully to investigate the role of age and hearing loss on speech perception (Dubno et al., 1984), and would facilitate the assessment of the independent contributions of hearing loss and cognitive decline to the ability to cope with difficult listening tasks.

An alternative approach to disassociating the impact of hearing loss and cognitive decline on performance in difficult listening situations is to compare performance on the spatial listening task with performance on a purely visual task of attention designed to recreate the cognitive demands of the spatial listening task. The visual task could comprise multiple sequences of symbols with each sequence corresponding to a phrase in the listening task. Each visual 'phrase' would comprise several symbols designed to mimic the 'Ready', call-sign, 'go to', colour, and number aspects of the CRM phrases (Figure 8.2). As in the listening task, the 'phrases' would start at regular intervals and the target phrase would be identified as the visual sequence which contained a unique target symbol. The response requirements would be identical to the spatial listening task—identify the target 'phrase' and report the colour and number co-ordinate. The intervals at which new phrases start, the speed of presentation, and the spatial distribution of the phrases could be manipulated to increase the demands of the task to equate performance levels to that of the spatial listening task among young normally-hearing individuals. The performance of older adults on the visual and auditory tasks could then be compared to examine difficulties on attention-demanding tasks in general, while differences in performance between the tasks would assess the impact of hearing loss on complex tasks when they are in the auditory domain.

8.4.3 Improving Performance on Demanding Listening Tasks

Self-reported difficulties in attention-demanding listening situations have been found to be related to a self-reported measure of handicap, independent of hearing loss (Gatehouse & Noble, 2004). This relationship suggests that deficiencies in attentional abilities may contribute to difficulties with listening in everyday situations, which in turn could lead to avoidance and isolation (Noble et al., 1995). Thus, improving the cognitive abilities of adults may improve their ability to cope with such demanding listening situations, which in turn may have a positive effect on their quality of life.

In light of the results discussed in Section 8.3.1, it is possible that increasing the speed at which older adults can process information may improve their ability to cope with complex listening tasks. Previous research has demonstrated that speed of processing can be improved among older adults through training regimes—87% of 712 adults between the ages of 65–94 showed improvements in speed of processing after 10 sessions of training on visual search tasks over a period of 18 months (Ball, Berch, Helmers, Jobe, Leveck, Marsiske, Morris, Rebok, Smith, Tennstedt, Unverzagt, Willis, & Advanced Cognitive Training for Independent and Vital Elderly Study Group, 2002). Future research could examine the effects of cognitive training designed to

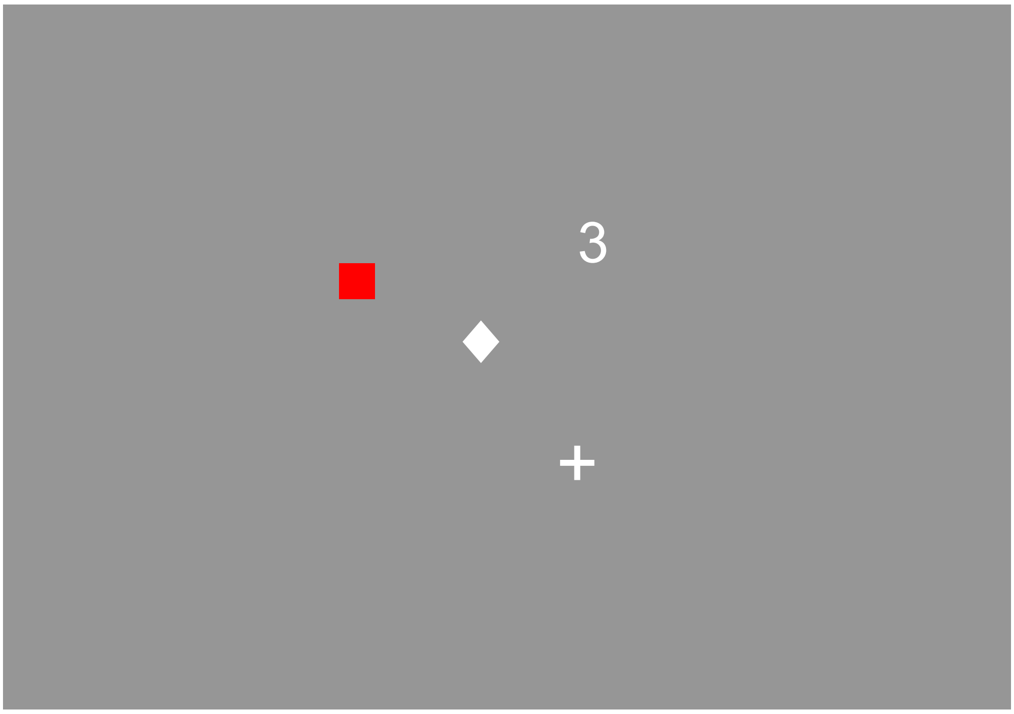

Figure 8.2. An example moment from the proposed visual task designed to create attentional demands comparable to that of the spatial listening task for speech. Each 'phrase' occupies a unique spatial location, and consists of a sequence of symbols corresponding to the keywords in the CRM phrases. The target 'phrase' is identified by the presence of a unique symbol within the sequence. Four 'phrases' are shown, each at a different position in the sequence of symbols: (left to right) colour, 'call-sign', 'ready', and number.

increase speed of processing on complex listening tasks for speech. Based on the results of Ball et al. (2002), the training regimes could include visual search tasks such as the *Map search* task from the TEA. Possibly, visual tasks like the one sketched in Figure 8.2 with similar cognitive demands as the complex listening task used in the current research might also lead to improved performance on attention-demanding listening tasks.

A further experiment could therefore assess the relative benefits to performance on the spatial listening task from training regimes using different tasks designed to recreate the attentional demands of difficult listening situations. The tasks could included the non-verbal visual task outlined in Section 8.4.2, a verbal version of the visual task which used words instead of symbols, and the spatial listening task itself. Performance on the spatial listening task would be assessed prior to and after training. In addition, the brain activity of participants would be recorded while they performed the spatial listening task using MEG, before and after training. The experiment would test the hypothesis that performance deficits on the spatial listening task among older adults are largely due to a slower speed of processing and difficulties with attentional control, rather than deficits in central auditory and language processing. Based on the conclusions in Section 8.3.1, it would be expected that training on any of the tasks would have similar benefits on performance—all three tasks would be as likely to improve speed of processing and attentional control. It would also be expected that differences in brain activity recorded before and after training would be localised to regions associated with attentional processing, e.g. parietal and frontal regions, and would possibly be related to improvements in performance. However, if deficits in central auditory and/or language processing contribute significantly to poorer performance on the spatial listening task, then training on the listening task would be expected to result in a larger increase in performance relative to those individuals who received training on the visual tasks of attention.

Dietary supplements have also been found to improve cognitive performance among older adults. Daily consumption of essential fatty acids, specifically n-3 long-chain polyunsaturated fatty acids (LCPs), has been associated with improved cognitive function (Uauy & Dangour, 2006). The ability of LCPs to delay age-related cognitive decline is currently being investigated in a randomised controlled trial (Dangour, Clemens, Elbourne, Fasey, Fletcher, Hardy, Holder, Huppert, Knight, Letley, Richards, Truesdale, Vickers, & Uauy, 2006). A further experiment could assess the possible benefits of dietary supplements such as LCPs on the performance of older adults on the spatial listening task, on measures from the TEA, and on self-reported difficulties with listening in everyday situations.

8.5 Conclusion

Following what one person is saying when many people are speaking at the same time requires a listener to focus, divide, shift, and sustain attention efficiently and rapidly. Age-related difficulties in performing a multi-talker spatial listening task are associated with reductions in peripheral sensitivity and poorer performance on a range of cognitive tasks. Differences in brain activity at key moments in the task arise in regions of the brain associated with attention, and a sub-set of those differences are related to performance in the task. The results reported in this thesis predict that interventions which increase the speed at which information can be processed, and improve cognitive function more generally, would reduce the level of difficulty with listening that older adults report in everyday life.

Appendix A

Vectors and Matrices: Notation, Terminology and Properties

A.1 Vectors

A.1.1 Definition

A vector is a set of one or more values. Vectors of n-dimensions, or containing n values, are denoted by the accent $\vec{\cdot}$, e.g. $\vec{u}$.

A.1.2 Vector Operations

The cross-product of two vectors in 3-dimensional space is denoted as $\vec{u} \times \vec{v}$, and is that operation which produces a vector perpendicular to $\vec{u}$ and $\vec{v}$. The dot product of two equal-length vectors, $\vec{u} \cdot \vec{v}$, is defined as:

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i \tag{A.1.1}$$

A.1.3 Norm of a Vector

The delimiters $|\cdot|$ refer to the Euclidean or ℓ_2 norm of a vector which represents its *size* or *length*. It is defined as the square root of the sum of its elements squared:

$$|\vec{u}| = \sqrt{\sum_{i=1}^{n} u_i^2} \tag{A.1.2}$$

A.2 Matrices

A.2.1 Definition

A matrix is a rectangular table of values. The letters m and n are used to describe the size of the first and second dimensions, or number of rows and columns, of a matrix, respectively. Single-dimension matrices which comprise a single column ($m \times 1$), also known as a *column vector*, or a single row ($1 \times n$), or a *row vector*, are represented by lowercase letters such as u. Capital letters are used to denote matrices whose dimensions are larger than 1; e.g. U. Individual rows and/or columns of a matrix are referred to using subscript letters: $u_{i,j}$ refers to the value at the i^{th} row and j^{th} column of the matrix U. The *rank* of a matrix is defined as the number of linearly independent columns in a matrix (Horn & Johnson, 1990). The transpose of the matrix U, when its rows are replaced by its columns and vice versa, is U^T.

A.2.2 Norm of a Matrix

The *Euclidean* or ℓ_2 norm of a matrix, referred to as the *Frobenius* norm, is denoted by the delimiters $\|\cdot\|$, e.g. $\|U\|$, and like the vector equivalent is a measure of the size of the matrix. It is defined as the square root of the sum of its elements squared:

$$\|U\| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(u_{i,j}\right)^2} \tag{A.2.1}$$

A.2.3 Eigenvalues and Eigenvectors

If there exists a non-zero vector x and a value λ which satisfy the following equation:

$$Ax = \lambda x \tag{A.2.2}$$

then x is said to be an *eigenvector* and λ an *eigenvalue* of the matrix A. The eigenvalues of a matrix play an important role when that matrix is part of a system of linear equations, and are discussed further in Section A.3.

A.2.4 Singular Values and Vectors

If there exists two non-zero vectors, u and v, and a positive value σ such that:

$$\begin{aligned} Av &= \sigma u \\ A^T u &= \sigma v \end{aligned} \tag{A.2.3}$$

where A is a real matrix, i.e. comprises real numbers only, and A^T is the transpose of A, then σ is a *singular value* and u & v are *singular vectors* of A. When Eq. A.2.3 is expanded to include all of the singular values and vectors of A, the vectors u and v are replaced by the matrices U & V, and the singular value σ is replaced by the diagonal matrix Σ, a matrix with the singular values on its main diagonal and zeros for all off-diagonal values. By including all singular values and vectors, Eq. A.2.3 becomes:

$$\begin{aligned} AV &= U\Sigma \\ A^T U &= V\Sigma^T \end{aligned} \tag{A.2.4}$$

From these equations we can derive an expression for the real matrix A in terms of its singular values and vectors:

$$A = U\Sigma V^T \tag{A.2.5}$$

which is the *singular value decomposition* (SVD) of A. Singular values are important factors of matrices which translate between two vector spaces.

A.2.5 Square Matrices

A square matrix is defined as a matrix in which the numbers of rows and columns are equal; i.e. $m = n$. The trace of a square matrix, trace(U), is the sum of the values on its main diagonal:

$$\text{trace}(U) = tr\ U = \sum_{i=1}^{n} u_{i,i} \tag{A.2.6}$$

The identity matrix, a square matrix with the values on its main diagonal being 1 and all other values being 0, is denoted by I:

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{A.2.7}$$

Square matrices have some unique properties. The *eigenvalues* of a square matrix are equivalent to its *singular values*. The sum of the *eigenvalues* of a square matrix is equal to the sum of the elements on its main diagonal; i.e. the trace of the square matrix. This property allows for the estimation of the smallest and largest eigenvalues without requiring the decomposition of the matrix, which can be computationally expensive for large matrices.

A.2.6 Covariance Matrices

A *covariance matrix* is a matrix containing the variances and covariances for a set of two or more variables (Stevens, 1986). If there is a matrix, X, for which each column represents a variable and each row an observation, then the covariance matrix $cov(X)$ is a square matrix whose values on the main diagonal are the variance of the individual variables and the off-diagonal values are the covariances between the variables. Thus, the value at the i^{th} row and j^{th} column of $cov(X)$ is the covariance between the i^{th} and j^{th} variables. The covariance matrix is a *symmetric matrix* as $cov(X)$ is equal to its transpose; i.e. $cov(X) = cov(X)^T$. This property is a result of that fact that $cov(var_1, var_2) = cov(var_2, var_1)$ and therefore values on either side of the main diagonal are equal or $cov(X)_{i,j} = cov(X)_{j,i}$.

A.2.7 The Matrix Inverse

The square matrix U is said to be *invertible* if another matrix, U^{-1}, exists such that $U^{-1}U = UU^{-1} = I$. If U is not square, such that $m \neq n$, the true matrix inverse, U^{-1}, does not exist. An inverse of a non-square matrix can be derived by computing its *generalised inverse, L^+*, also termed the *pseudoinverse* or *Moore-Penrose Pseudoinverse* (Moore, 1920; Penrose, 1955) (see Appendix B, Section B.1, p. 254).

A.2.8 Singularity

A two-dimensional matrix is *non-singular* if it only produces an output of 0 if the input is also 0 (Horn & Johnson, 1990); i.e.:

$$Ux = 0$$

only when

$$x = 0$$

The concept of non-singularity is related to the matrix inverse. If the true inverse of a square matrix exists, then the matrix is *non-singular*. Non-singular matrices, as well as being invertible, have other important properties. The rank of a non-singular square matrix is equal to the number of columns. Thus, for a matrix to be *non-singular* and *invertible*, its columns must be linearly independent of one other. In addition, a *non-singular* matrix does not have any *eigenvalues* (or equivalently *singular values*) equal to 0 (Horn & Johnson, 1990). If a matrix is not square, i.e. $m \neq n$, then it is *singular*.

A.2.9 Condition Number

Another concept linked to matrix inversion is the *condition number* of a matrix. The condition number of a square matrix is defined as the ratio between its largest and smallest *eigenvalues* (Press et al., 2002). Thus, matrices with eigenvalues which approach zero have very large condition numbers. If a matrix is *singular*, and therefore *non-invertible*, its condition number is infinite.

If the condition number is large, such that its reciprocal is close to the floating-point precision of the machine used to calculate the inverse (approximately 10^{-12}

when using double-precision), the matrix is *invertible* but is said to be *ill-conditioned.* If the condition number is small (near 1), such that the sizes of the largest and smallest eigenvalues of the matrix are similar, then the matrix is said to be *well-conditioned* (Press et al., 2002).

From the above definition, it follows that if a square matrix contains singular values which are very small (near zero), then its condition number is large, and the matrix is *ill-conditioned.* The condition number of a matrix is an important property to consider in the context of a system of linear equations, a concept which is described in the next section.

A.3 System of Linear Equations

A system of linear equations having m observations or known solutions ($b_{1-\mathbf{m}}$), and n unknown values ($x_{1-\mathbf{n}}$) with coefficients ($a_{1-\mathbf{m},1-\mathbf{n}}$) is expressed as:

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \cdots & + & a_{1,\mathbf{n}}x_\mathbf{n} & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \cdots & + & a_{2,\mathbf{n}}x_\mathbf{n} & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a_{\mathbf{m},1}x_1 & + & a_{\mathbf{m},2}x_2 & + & \cdots & + & a_{\mathbf{m},\mathbf{n}}x_\mathbf{n} & = & b_\mathbf{m} \end{array} \tag{A.3.1}$$

The same system can be described using matrices:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & \cdots & a_{1,\mathbf{n}} \\ a_{2,1} & a_{2,2} & \cdots & \cdots & \cdots & a_{2,\mathbf{n}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{\mathbf{m},1} & a_{\mathbf{m},2} & \cdots & \cdots & \cdots & a_{\mathbf{m},\mathbf{n}} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ \vdots \\ x_\mathbf{n} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_\mathbf{m} \end{bmatrix} \tag{A.3.2}$$

or in matrix notation as:

$$Ax = b \tag{A.3.3}$$

where x is the column vector of unknown values to be estimated or the unknown equation *variables*, A is a matrix containing the known (or estimated) variable *coefficients* for each equation, and b is a column vector of the known *solutions* to the equations.

The matrix A can be described as a *transposition* matrix which acts as a translation operator between the variable space defined by x and the solution space defined by b. If the number of knowns, m, is less than the number of unknowns, n, then the system is said to be *underdetermined.* If $m > n$ then the system is *overdetermined.*

The condition number of the matrix A can be interpreted as a measure of how stable the estimate of x, derived by solving the system of equations and denoted as $\hat{x}$, will be (Golub & Van Loan, 1996; Press et al., 2002). If the system in Eq. A.3.3 is stable, then a small numerical change in b should result in a small change in $\hat{x}$. This is only the case if A is *well-conditioned*. If A is *ill-conditioned*, the system will be unstable and overly sensitive to small perturbations of b. In that case, even if the derived estimate, $\hat{x}$, produces a modelled solution which is numerically close to the known solution (b), the estimate could still be different from the true values of x (Horn & Johnson, 1990).

This situation can be overcome by modifying the problem in Eq. A.3.3 so that it is changed to a *well-conditioned* problem that is numerically "close" to the original problem (Hansen, 1992). The selection of a suitable *well-conditioned* problem is dependent on some form of a priori information about x such as its degree of smoothness, a method known as *regularisation*. Regularisation methods for inverse problems associated with MEG are discussed in Chapter 4 (Section 4.4.5, p. 80).

Appendix B

Calculations

B.1 Moore-Penrose Pseudoinverse

We start by restating the forward problem which determines the relationship between the source strengths which we want to estimate, s, and our measured data at a single time point, d, using the lead fields, L:

$$d = Ls \tag{B.1.1}$$

Based on the linear relationship of the forward problem in Eq. B.1.1, we want to solve for s so that it explains fully the measured data such that we do not distinguish signal and noise, and that the solution does not contain any elements which would not create a measured signal at any of the sensors, i.e. radial components for MEG. We also make the presumption that our measurements are independent of each other. We can then express s as a linear combination of the source *lead fields* (Mosher et al., 2003); i.e. the rows of the forward solution, and some arbitrary factor yet to be determined, z (Hauk, 2004):

$$s = L^T z \tag{B.1.2}$$

We can then derive an expression for z by combining Eqs. B.1.2 and B.1.1:

$$d = LL^T z \tag{B.1.3}$$

which gives us:

$$z = (LL^T)^{-1} d \tag{B.1.4}$$

Finally, we can combine Eqs. B.1.4 with B.1.2 to arrive at an expression for the source strengths, s, in terms of quantities which we either measure or can estimate:

$$s = L^T (LL^T)^{-1} d = Wd \tag{B.1.5}$$

From this equation, we can extract our inverse operator, W, as:

$$W = L^T(LL^T)^{-1} = L^+ \tag{B.1.6}$$

which is the *Moore-Penrose Pseudoinverse,* or *minimum norm* inverse (Hauk, 2004; Wang & Kaufman, 2003). It is only possible to calculate the inversion $(LL^T)^{-1}$ providing the rows of L are linearly independent, which implies that our measurements are independent of one another. If Eq. B.1.2 was altered to include a source weighting matrix, C_s, i.e.

$$s = C_s L^T z \tag{B.1.7}$$

which could be used to incorporate priors from fMRI (Dale & Sereno, 1993) or the estimated covariance between the sources, then our source solution in Eq. B.1.5 would become:

$$s = C_s L^T (LC_s L^T)^{-1} d \tag{B.1.8}$$

If $C_s = I$, i.e. all sources are given equal weight and are assumed to be independent with equal variance, then the solution reduces to the Moore-Penrose Pseudoinverse (Eq. B.1.5), or the classical minimum norm solution. If $C_s \neq I$, then Eq. B.1.8 is the *weighted* minimum norm solution.

B.2 LCMV Optimised Spatial Filter

Recall that we want to derive a set of weights, w, which when applied to the measured sensor data, D, will yield the reconstructed activity, S, of a source at a particular location and of a certain orientation:

$$S = w^T D \tag{B.2.1}$$

We can then derive an expression for the total power of the activity related to that source as the sum of this activity squared:

$$Q = (S)^2 = (w^T D)^2 = w^T D D^T w = w^T C w \tag{B.2.2}$$

where C is the covariance of the sensor data over the time window of interest. From Eq. B.2.2 we can see that this power estimate is constructed across the whole time window for which the data covariance is calculated. We then wish to construct a spatial filter which satisfies two constraints: that the reconstructed power, Q, is minimised and the output of the filter at the target source location is a pass-band:

$$\min_{Q} \{Q\} \qquad \text{subject to:} \qquad w^T \ell = 1 \tag{B.2.3}$$

where ℓ is the lead field of the target source. To find the weights we use a mathematical approach termed a *Lagrange Multiplier*, which is a constrained optimisation technique for differentiable functions. In other words, this technique is used to find an optimal solution to a function with one or more parameters, i.e. find the maximum or minimum of the function, subject to a constraint which may take the form of an equality or inequality. This constraint limits the range of points on the function which we are interested in, termed the *feasible region*.

The extremes of a function, either maxima or minima, occur at stationary points and thus can be found where the gradient is zero, i.e. for some function f the stationary points are found by satisfying:

$$\nabla_x f(x) = 0 \tag{B.2.4}$$

Additionally, we want to constrain the solution by specifying a fixed value of another related function, $G(x)$, i.e.:

$$G(x) = c \tag{B.2.5}$$

where c is the constraint value that is specific to the problem we are solving. If we put both of these equations in the context of the LCMV spatial filtering method, we want to find a certain set of weights, w, which produce a minimum on our output power function:

$$F(w^T) = w^T C w \tag{B.2.6}$$

i.e. find w^T so that $\nabla_{w^T} F(w^T) = 0$, and which also satisfy the pass-band constraint in Eq. B.2.3, which we can express as a function:

$$G(w^T) = w^T \ell = 1 \qquad \text{or alternatively} \qquad G(w^T) - 1 = w^T \ell - 1 = 0 \tag{B.2.7}$$

In summary, we want to search our output power function to find a set of weights which lie on a minimum of the power function, but only for that subset of weights which also occur where our constraint function is equal to 1. In other words, if we move along the function $G(w^T)$ where it equals 1 for different values of the weights (this is our pass-band constraint for the source location of interest) we want to find the point where we are also on the power function, $F(w^T)$, but where it is at a minimum. The theory behind the *Lagrangian* method is that the optimal solution occurs at a point where both functions touch but do not overlap. If at a certain point both functions overlapped, i.e. crossed each other, then moving in either direction on $G(w^T)$ would change the value of $F(w^T)$, and consequently $F(w^T)$ is not at a stationary point, i.e. a minimum or maximum. We are helped by the fact that it is only when both functions are *tangential* to each other that they touch but do not overlap. This has the consequence that the gradients of both functions at that point are parallel, which can be re-stated as one being a multiple of the other. We can express this mathematically using an arbitrary multiplication factor, $-\lambda$, which is termed the *Lagrangian Multiplier*. Our entire problem of finding the spatial filter weights can now be stated in terms of solving for two equations:

$$\nabla_{w^T} F(w^T) = -\lambda \nabla_{w^T} G(w^T) \qquad \text{(Functions are tangential)}$$

$$G(w^T) = 1 \qquad \text{(We satisfy a pass-band constraint)}$$

We can conveniently rewrite this as a single function L, referred to as a *Lagrangian*, on which we then look for a stationary point, i.e. a minimum:

$$L(w^T, \lambda) = F(w^T) + \lambda(G(w^T) - 1) \tag{B.2.8}$$

or in expanded form:

$$L(w^T, \lambda) = w^T C w + \lambda(w^T \ell - 1) \tag{B.2.9}$$

As stated previously, we obtain a minimum by finding where the gradient of the *Lagrangian* is zero:

$$\nabla_{w^T} L(w^T, \lambda) = \frac{\partial}{\partial w^T}(w^T C w) + \lambda \frac{\partial}{\partial w^T}(w^T \ell - 1) = 0 \tag{B.2.10}$$

where

$$\frac{\partial}{\partial w^T}(w^T C w) = 2Cw$$

$$\frac{\partial}{\partial w^T}(w^T \ell - 1) = \ell$$

$$\therefore \nabla_{w^T} L(w^T, \lambda) = 2Cw + \lambda \ell = 0 \tag{B.2.11}$$

We can now rearrange Eq. B.2.11 to obtain an expression for the spatial filter weights, w:

$$w = \frac{-\lambda}{2} C^{-1} \ell \tag{B.2.12}$$

and substitute this expression into our constraint equation $w^T l = 1$ to solve for the multiplier, $-\lambda$, which is the final unknown term:

$$\frac{-\lambda}{2} \left(C^{-1}\ell\right)^T \ell = 1$$

$$\therefore -\lambda = \frac{2}{\ell^T C^{-1} \ell} \tag{B.2.13}$$

Finally, we can combine Eqs. B.2.12 and B.2.13 to obtain an expression for the weights in terms of the source lead fields and the data covariance matrix:

$$w = \frac{C^{-1}\ell}{\ell^T C^{-1} \ell} \tag{B.2.14}$$

This is the fundamental equation of the LCMV spatial filtering ('beamforming') method.

B.3 Equivalence

It is possible to show that many approaches to solving the inverse problem are fundamentally very similar, and can be distinguished based on the form of the source covariance matrix. In fact, the LCMV spatial filtering method has be shown to be a particular case of the minimum norm approach in which the reconstructed source power is constrained by an a priori assumption of uncorrelated sources (Mosher et al., 2003). To illustrate this point, an expression for the output of the spatial filter weights will now be derived based on the linear inverse equations which underpin the minimum norm method. From the generalised linear inverse approach that we have been using to derive the minimum norm methods of source analysis, we know that the source amplitudes can be expressed as:

$$s = C_s L^T (L C_s L^T + C_n)^{-1} d \tag{B.3.1}$$

where s is the source amplitudes, L is the forward solution matrix (the lead fields), d is the measured data, and C_s and C_n are the source and noise covariance matrices, respectively. More generally, if we know the covariance of the measured data, then Mosher et al. (2003) showed that the expression reduces to:

$$s = C_s L^T C_d^{-1} d \tag{B.3.2}$$

where C_d is the data covariance matrix. In applying the LCMV spatial filtering approach to MEG data, all sources are assumed to be uncorrelated—C_s is diagonal. The source covariance matrix for n sources can therefore be written as:

$$C_s = \left\{ \begin{array}{ccc} \sigma_i^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{array} \right\} \tag{B.3.3}$$

where σ_i^2 is the variance of the i^{th} source. Taking the individual variance value from this matrix, the expression for the amplitude of the i^{th} source becomes:

$$s_i = \sigma_i^2 \ell_i^T C_d^{-1} d \tag{B.3.4}$$

where ℓ_i^T is the i^{th} column (or *forward field*) of the forward solution, L, related to the i^{th} source amplitude. In the case of LCMV spatial filtering, the source variance, σ_i^2, is estimated from the measured data. When the spatial filter has perfect resolution, Mosher et al. (2003) defined the variance estimate as:

$$\sigma_i^2 = (\ell_i^T C_d^{-1} \ell_i)^{-1} \tag{B.3.5}$$

Thus we arrive at our estimate of source amplitude based on the data covariance by combining Eqs. B.3.5 and B.3.4:

$$s_i = (\ell_i^T C_d^{-1} \ell_i)^{-1} \ell_i^T C_d^{-1} d = w_i^T d \tag{B.3.6}$$

where w_i^T is the row vector of filter weights for the i^{th} source. This equation is identical to the expression for the spatial filtering weights when derived from LCMV filtering principles (Van Veen et al., 1997) (Section B.2, p. 255).

As is evident from Eq. B.3.5, the variance of the sources is based on an estimate of the data covariance which is derived from the measured data. This is distinct from the minimum norm approach, in which the data covariance is estimated from assumptions about the source covariance and other source priors, and can be expressed as (Lin et al., 2006b):

$$C_d = \frac{LC_s L^T}{\lambda} + C_n \tag{B.3.7}$$

where λ is the regularisation parameter.

Appendix C

Effects of Voice Characteristics and Location on Speech Perception

C.1 Talker Intelligibility

C.1.1 Introduction

By constraining *who* would say the target phrase within a block of trials in Experiments 1 and 2 (Chapter 5), the speech recognition performance of the listener would be partially dependent on the intelligibility of the particular target talker used for that block. Therefore, it was necessary to employ multiple target talkers and average the performance levels across the talkers to obtain a measure of speech recognition performance which would not be dependent on any individual target talker. A pilot study was conducted to choose four talkers, two male and two female, from the eight available talkers.

C.1.2 Methods

C.1.2.1 Participants

Three male normally-hearing volunteers participated in the pilot study. Two had previous experience with the CRM stimuli. The loudspeaker array and stimuli were the same as those used in Experiments 1 and 2 (Chapter 5).

C.1.2.2 Stimuli & Procedure

The trial layout was identical to that of Experiment 1, with thirteen phrases being presented in an overlapping sequence. Thresholds were measured using the same procedure as Experiment 1, except that the second phase of the adaptive procedure lasted for ten reversals. Within each block of trials, the voice of the target talker, *who*, was fixed to one of the eight possible talkers. The location and position of the target talker in the sequence of stimuli were selected randomly for each trial.

C.1.2.3 Design

Participants completed eight blocks of trials, each with a different target talker. The order of target talkers was partially counterbalanced across participants by ensuring that a different order was used for each of the three participants.

C.1.2.4 Analysis

Mean speech-reception thresholds (SRTs) and standard deviations were calculated for each target talker across the group of participants.

C.1.3 Results

Table C.1 shows the mean SRTs for the eight target talkers. Average speech recognition performance varied by a maximum of 2.7 dB across the eight talkers. The mean performance level across all participants and talkers was -13.1 dB.

C.1.4 Discussion & Conclusions

Four target talkers, two male and two female, were identified as giving performance levels closest to the average performance level of -13.1 dB: male 1, male 2, female 3, and female 4. These four talkers were selected for use as target talkers in Experiments 1 and 2.

C.2 Effect of Location on Speech Recognition

C.2.1 Introduction

In Experiments 1 and 2, providing the listener with knowledge about the location of the target phrase would have been accompanied by the location of the target phrase

	M1	M2	M3	M4
SRT (dB)	−12.8	−14.1	−12.2	−14.5
σ	2.0	3.9	1.3	3.5

	F1	F2	F3	F4
SRT (dB)	−11.9	−12.6	−14.1	−12.8
σ	2.6	3.2	2.1	1.9

Table C.1. Mean speech reception thresholds (SRTs) and standard deviations (σ) for the four male (M) and four female (F) target talkers.

being fixed within a block of trials. However, constraining the target phase to appear at certain locations could have provided an additional performance advantage over other fixed locations, by improving signal-to-noise levels at the ears. Pilot testing was used to determine whether constraining the target to the loudspeaker directly in front of the listener improved performance relative to when the target was positioned at other locations.

C.2.2 Methods

C.2.2.1 Participants

Two male normally-hearing listeners participated in the pilot study (Group 1). Both had previous experience with the CRM stimuli. Additionally, each of the listeners who participated in Experiment 1 also took part in the study (Group 2), although data from those participants were collected after participation in Experiment 1.

C.2.2.2 Stimuli & Procedure

The loudspeaker array and stimuli were the same as those used in Experiments 1 and 2. The trial layout was identical to that of Experiment 1, with thirteen phrases being presented in an overlapping sequence. Speech-reception thresholds (SRTs) were measured using the same adaptive procedure as Experiment 1. Within each block of trials, the location of the target phase was constrained to one of three positions: −90°, 0°, and +90°, where positive azimuths are to the listener's right, and 0° is directly in front of the listener. The talker who spoke the target phrase and position of the target phrase in the sequence of phrases were selected randomly for each trial.

C.2.2.3 Design

Participants completed six blocks of trials, with each of the three possible target locations being repeated once. The order of the target locations was partially counterbalanced across participants such that a different order was used for each participant.

C.2.2.4 Analyses

Mean SRTs and standard deviations were calculated for both groups of participants. To determine whether the location of the target phrase had a significant effect on performance, the data from Group 2 was submitted to a repeated-measures ANOVA with one within-subjects factor, *location*. Omega squared (ω^2) was used to estimate the effect size for the *location* factor (Chapter 5, Section 5.4.2.6, p. 120). Planed contrasts were performed to compare performance when the target phase was presented in front of the listener with performance when it was presented at

±90°. This analysis would determine whether presenting the target phrase directly in front of the listener improved performance compared to when it was located at other locations. Effect sizes for the contrasts were calculated by converting the F-values to correlation values (Chapter 5, Section 5.2.2.10, p. 110).

C.2.3 Results

The results for both groups are listed in Table C.2. For Group 1, there was no obvious difference in performance across the three target locations, when related to the degree of variability in the data. For Group 2, performance improved by 2 dB when the target was presented at −90°, and improved by 3.1 dB when presented at +90°. An ANOVA assessed the effect of target location on performance for Group 2. The main effect of location was significant [$F(2,14) = 10.487,\ p < .01,\ \omega^2 = .32$]. Planned contrasts revealed that performance improved significantly when the location of the target phrase was constrained to the left [$F(1,7) = 11.061,\ p < .05,\ r = .78$] or right [$F(1,7) = 13.105,\ p < .05,\ r = .81$] of the listener compared to when it was located directly in front of the listener.

C.2.4 Discussion & Conclusions

For Group 1, the small differences in thresholds between target locations suggested that constraining the target at the location directly in front of the listener did not afford the listener an extra advantage compared with other locations. For Group 2, constraining the target location directly in front of the listener led to significantly lower performance levels compared to when the target was positioned at other locations. Therefore, any improvements in performance that arose in Experiments 1 and 2 in conditions where the location of the target phrase was fixed in front of participants, as compared to conditions in which its location varied from trial to trial, were not likely to be due to the particular spatial location used for the target phrase.

		−90°	0°	+90°
Group 1	SRT (dB)	−15.7	−14.7	−15.1
	σ	3.6	3.8	4.4
Group 2	SRT (dB)	−14.8	−12.8	−15.9
	σ	1.9	1.5	1.9

Table C.2. Mean speech reception thresholds (SRTs) and standard deviations (σ) for each of the three target locations.

Appendix D

Hearing Health Questionnaire

Participant Questionnaire

1. Do you wear, or have you ever been advised to wear a hearing aid?
2. Have you ever had surgery to either ear?
3. Do you suffer from tinnitus (ringing, whistling or other noises in the ear(s))?
4. Do you have trouble with your balance and/or vertigo?
5. Are you experiencing or have you recently had any of the following:
    (a) Pain in either ear
    (b) Discharge from either ear
    (c) Inflammation in either ear
    (d) A blockage in either ear
    (e) An injury of any kind to either ear
    (f) A cold or flu
6. Are you currently on or have you recently taken medication related to your ears or hearing?
7. Have you ever had a head injury requiring hospitalisation?
8. Have you been exposed to loud noise in the past few days?
9. Do you have any previous experience with listening/hearing tests? If so, please give details.
10. Are you planning or likely to be exposed to any loud noise in the following month? If so, pleas give details.

Appendix E

Speech, Spatial, and Qualities of Hearing Questionnaire

Speech

1. You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says?
2. You are talking with one other person in a quiet, carpeted lounge-room. Can you follow what the other person says?
3. You are in a group of about five people, sitting round a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?
4. You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation?
5. You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?
6. You are in a group of about five people in a busy restaurant. You cannot see everyone else in the group. Can you follow the conversation?
7. You are talking to someone in a place where there are a lot of echoes, such as a church or railway terminus building. Can you follow what the other person says?
8. Can you have a conversation with someone when another person is speaking whose voice is the same pitch as the person you're talking to?
9. Can you have a conversation with someone when another person is speaking whose voice is different in pitch from the person you're talking to?

10. You are listening to someone talking to you, while at the same time trying to follow the news on TV. Can you follow what both people are saying?

11. You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying?

12. You are with a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?

13. Can you easily have a conversation on the telephone? [use one aid, which one, why?]

14. You are listening to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both speakers?

Spatial

1. You are outdoors in an unfamiliar place. You hear someone using a lawnmower. You can't see where they are. Can you tell right away where the sound is coming from?

2. You are sitting around a table or at a meeting with several people. You can't see everyone. Can you tell where any person is as soon as they start speaking?

3. You are sitting in between two people. One of them starts to speak. Can you tell right away whether it is the person on your left or your right, without having to look?

4. You are in an unfamiliar house. It is quiet. You hear a door slam. Can you tell right away where that sound came from?

5. You are in the stairwell of a building with floors above and below you. You can hear sounds from another floor. Can you readily tell where the sound is coming from?

6. You are outside. A dog barks loudly. Can you tell immediately where it is, without having to look?

7. You are standing on the footpath of a busy street. Can you hear right away which direction a bus or truck is coming from before you see it?

8. In the street, can you tell how far away someone is, from the sound of their voice or footsteps?

9. Can tell how far away a bus or a truck is, from the sound?

10. Can you tell from the sound which direction a bus or truck is moving, for example, from your left to your right or right to left?
11. Can you tell from the sound of their voice or footsteps which direction a person is moving, for example, from your left to your right or right to left?
12. Can you tell from their voice or footsteps whether the person is coming towards you or going away?
13. Can you tell from the sound whether a bus or truck is coming towards you or going away?
14. Do the sounds of things you are able to hear seem to be inside your head rather than out there in the world?
15. Do the sounds of people or things you hear, but cannot see at first, turn out to be closer than expected when you do see them?
16. Do the sounds of people or things you hear, but cannot see at first, turn out to be further away than expected when you do see them?
17. Do you have the impression of sounds being exactly where you would expect them to be?

Qualities of Hearing

1. Think of when you hear two things at once, for example, water running into a basin[a power-tool being used][a plane flying past] and, at the same time, a radio playing[the sound of hammering][a truck driving past]. Do you have the impression of these as sounding separate from each other?
2. When you hear more than one sound at a time, do you have the impression that it seems like a single jumbled sound? (If you have this experience, can you give examples of the sounds in question?)
3. You are in a room and there is music on the radio. Someone else in the room is talking. Can you hear the voice as something separate from the music?
4. Do you find it easy to recognise different people you know by the sound of each one's voice?
5. Do you find it easy to distinguish different pieces of music that you are familiar with?
6. Can you tell the difference between different sounds, for example, a car versus a bus; water boiling in a pot versus food cooking in a frying pan?

7. When you listen to music, can you make out which instruments are playing?
8. When you listen to music, does it sound clear and natural?
9. Do everyday sounds that you can hear easily seem clear to you (not blurred)?
10. Do other people's voices sound clear and natural?
11. Do everyday sounds that you hear seem to have an artificial or unnatural quality?
12. Does your own voice sound natural to you?
13. Can you easily judge another person's mood from the sound of their voice?
14. Do you have to concentrate very much when listening to someone or something?
15. If you turn one hearing aid/implant off, and do not adjust the other, does everything sound unnaturally quiet? [for long-term BL only]
16. When you are the driver in a car can you easily hear what someone is saying who is sitting alongside you? [use one aid, which one, why?]
17. When you are a passenger can you easily hear what the driver is saying sitting alongside you? [use one aid, which one, why?]
18. Do you have to put in a lot of effort to hear what is being said in conversation with others?
19. Can you easily ignore other sounds when trying to listen to something?

Appendix F

Assessing the Discrimination Point of Call-sign Keywords

F.1 Introduction

When the cortical responses to target and non-target call-signs were compared in Experiment 6 (Chapter 7), differences were identified as starting as early as 150 ms after the onset of the call-sign keywords. It was not known whether the short latencies of these differences excluded them from being related to the discrimination of target and non-target call-signs. A gating experiment was conducted to examine the duration of the acoustical information necessary to discriminate between target and non-target call-signs. The experiment was designed to be as similar to Experiment 6 as possible. Therefore, the gated stimuli were created from the same stimuli that were used in Experiment 6 and they were presented using tube-phones.

The ability to discriminate between target and non-target call-signs with minimal acoustic information may be possible based on the presence of coarticulatory information prior to the onset of the call-sign, or due to the unique initial phonemes of the 8 call-signs ("arrow," "baron," "charlie," "eagle," "hopper," "laker," "ringo," and "tiger"). It was hypothesised that discrimination performance would be above chance based on coarticulatory information prior to the onset of call-signs, and that performance would asymptote close to 100% with only a few tens of milliseconds of acoustical information.

F.2 Methods

F.2.1 Participants

Twelve paid listeners, four male and eight female, between the ages of 18–25 (Mean age 19.8 years, $\sigma = 1.9$) participated. Two of the participants had previously participated in Experiment 4 (Chapter 5, Section 5.5, p. 123). No participants reported difficulties with hearing and all had a normal hearing health history (Appendix D).

F.2.2 Stimuli

The stimuli were chosen from the sequences of phrases from Experiment 6 (Chapter 7). Two groups of stimuli, *target* and *non-target*, were created for each of the target call-signs; i.e. 'Baron', 'Tiger', and 'Ringo'. The *target* group contained stimuli in which the target call-sign was gated, and the *non-target* group contained stimuli in which any of the non-target call-signs were gated.

For each of the three target call-signs, the stimuli were divided into three groups based on the position of the phrase containing the target call-sign; i.e. in the 3^{rd}, 4^{th}, or 5^{th} slot. For each target position, 10 stimuli were chosen at random. These were the *target* stimuli and were gated relative to the onset of the target call-sign.

For the *non-target* stimuli, 20 sequences were chosen at random from those original stimuli in which the target was in the 4^{th} or 5^{th} slots—these stimuli always contained a non-target call-sign in the 3^{rd} slot. An additional 10 sequences were chosen from those stimuli in which the target was in the 5^{th} slot—these stimuli contained a non-target call-sign in the 4^{th} slot. The *non-target* stimuli were gated related to the onset of the non-target call-sign.

Each stimulus was gated at 0 ms, 20 ms, 40 ms, 80 ms, 160 ms, and 320 ms relative to the onset of the target call-sign in the case of target stimuli, or relative to the onset of the non-target call-sign for the non-target stimuli. This created a total of 180 target and 180 non-target stimuli. For both target and non-targets, a 50 ms raised cosine ramp was applied to the end of the stimuli after gating to avoid audible artifacts that could have arisen from the abrupt gating of the stimuli.

F.2.3 Training

To familiarise participants with the stimuli and the apparatus, a block of twelve trials was presented prior to the main experiment. The block include 6 trials with a gating of 160 ms and 6 with a gating of 320 ms. For each gate length, half of the trials comprised stimuli in which the target call-sign was gated. Participants were reminded of the task before and after the training block.

F.2.4 Procedure

Participants sat at a desk in a 1.0 × 1.2 m double-walled sound-attenuated IAC booth. The stimulus delivery system was constructed to be as similar as possible to the one used in MEG (Chapter 7). Tube-phones terminated with foam ear-tips were inserted into the ear canals of the participant and a short audio test confirmed that input was being received in both ears. Prior to the training block, participants were allocated a target call-sign for the duration of the experiment. On each trial, a gated sequence of overlapping phrases was presented. The message “Was the call-sign TARGET?” was displayed on a monitor positioned in front of the participant, where ‘TARGET’ was the participant’s target call-sign. A 2-alternative forced-choice task was used. Participants indicated their choice, ‘Yes’ or ‘No’, using a mouse. Feedback based on the accuracy of responses was provided on each trial. The stimuli in the experimental block of trials were presented in a random order.

F.2.5 Design

All 12 participants completed the training and experimental blocks of trials in a single session.

F.2.6 Analysis

A one-sample t-test was performed against chance performance level (50%) for the trials with a gate length of 0 ms. This test was used to determine whether performance was above chance when only coarticulatory information prior to the onset of the call-sign was available.

To determine the gate length required for discrimination performance to reach an asymptote, data from the condition with the longest gate length (320 ms) was compared to each successive shortening of the gate length using paired-samples t-tests. The gate length at which performance reached an asymptote was defined as the shortest gate length at which performance did not differ significantly from the condition with the longest gate length (320 ms). Significance values were corrected using Bonferroni correction to control for the inflation of the FWER due to the calculation of multiple comparisons involving the same data, denoted by p_{bf}. Effect sizes were calculated by converting the t-values to correlation values (Chapter 5, Section 5.4.2.6, p. 120).

F.3 Results

Figure F.1 shows the mean discrimination performance levels across all trials for each of the gate lengths. When only coarticulatory information was available, performance was significantly above chance (50%) [Mean = 62.78%, $\sigma = 11.22$, $t(11) = 3.944$, $p < .01$, $r = .77$].

Performance in the condition with a gate length of 320 ms did not differ significantly from the condition with a gate length of 160 ms [$t(11) = -.832$, $p_{bf} > .05$ ns, $r = .24$] or 80 ms [$t(11) = -2.294$, $p_{bf} > .05$ ns, $r = .57$]. Performance was significantly worse at the gate lengths of 40 ms [$t(11) = -5.8$, $p_{bf} < .001$, $r = .87$], 20 ms [$t(11) = -5.134$, $p_{bf} < .001$, $r = .84$], and 0 ms [$t(11) = -9.468$, $p_{bf} < .001$, $r = .94$] compared to performance at a gate length of 320 ms.

F.4 Discussion & Conclusions

The results of the experiment confirmed the two predictions: target and non-target call-signs were discriminable at a level above change when only coarticulatory information was available; performance reached an asymptote when only a brief duration of acoustical information (80 ms) was available.

When only prior coarticulatory information was available, participants were able to discrimination between their target call-sign and non-target call-signs with an accuracy that exceeded chance (63%). This high level of discrimination performance may also reflect the distinctiveness of the call-signs. Each call-sign (“arrow,” “baron,” “charlie,” “eagle,” “hopper,” “laker,” “ringo,” and “tiger”) has a unique initial phoneme. The results suggest that participants were able to discriminate between their target and the other call-signs with minimal acoustic information.

The comparisons between the most extreme condition (320 ms) and each of the shorter gate lengths indicated that increasing the gate beyond 80 ms did not lead to

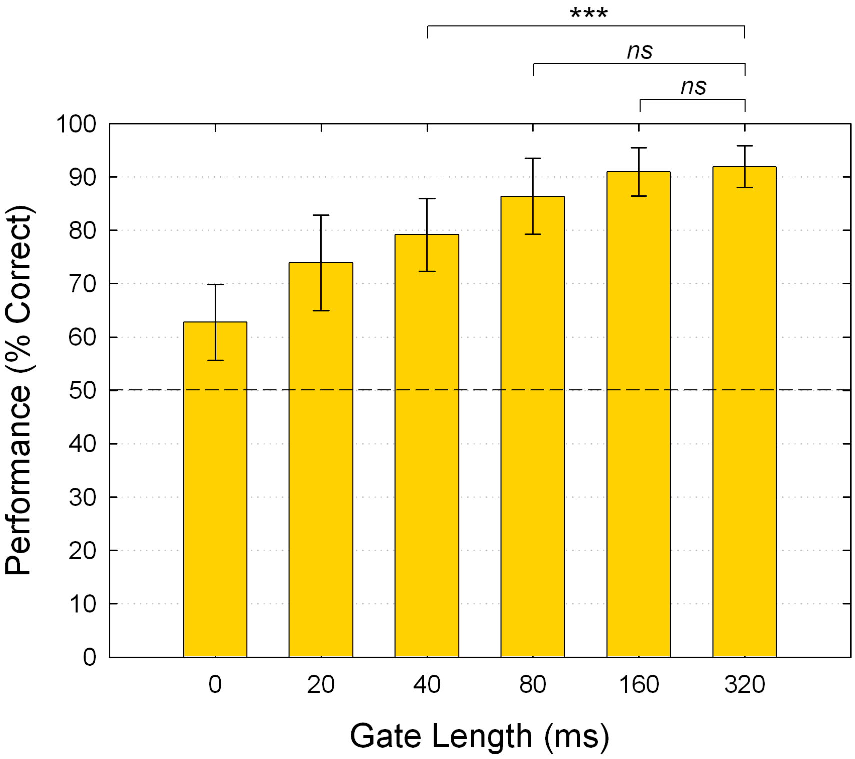

Figure F.1. Mean performance expressed as the percentage of correct target/non-target discriminations as a function of gate length. Performance at the longest gate length was compared to performance at each of the shorter gate length to determine the gate length at which performance reached asymptote (80 ms). The dotted line indicates performance at chance. Error bars show 95% confidence intervals (*** $p < .001$, *ns* not significant).

significant increases in target/non-target discrimination performance. These results suggest that participants only required 80 ms of acoustical information to reliably perform the call-sign discrimination. Petten, Coulson, Rubin, Plante, & Parks (1999) examined word identification performance using a similar gating technique. A corpus of word stimuli were gated at 50 ms steps, from 50 ms after word onset until the end of the word. Participants were presented with the phrase "The next word is", a 400 ms silence, and then the gated word. Identification performance for the initial phoneme was 78% when words were gated at 50 ms and 100 ms, and was 89% when gated at 150 ms. In the current study, slightly higher performance levels were found at equivalent gate lengths: 86% at 80 ms and 91% at 160 ms on average. This difference can be accounted for by the unique initial phonemes across all 8 call-signs and the use of a two-alternative discrimination task rather than a more challenging identification task.

For Experiment 6, it was important to estimate the earliest neural responses that were likely to be related to target/non-target call-sign discrimination. Petten et al. (1999) measured ERPs while participants listened to sentences which ended with a word which was either congruent or incongruent with the preceding context. The final sentence words were ungated versions of the stimuli used in the word identification experiment discussed above. The data from the word identification task were used to determine the isolation point (IP) for each final word, defined as the point at which 7 of the 10 participants correctly identified the word. The data were then analysed to identify the N400, an ERP elicited by incongruent sentence-final words, and to compare the time-course of the response to the previously identified IP for each word. Petten et al. (1999) found a difference between the response to congruent and non-congruent sentence-final words starting between 150–200 ms after word onset. The difference wave started 200 ms before the IP. Thus, a difference in the scalp-recorded ERP related to semantic processing was identified as starting after a time interval similar in duration to the minimum gate length required for reliable word identification, between 150–200 ms. This difference was found to begin before the word identification process, represented by the IP, had completely finished.

As the current study indicated that the target/non-target discrimination process could be performed at a high level of accuracy (86%) with only 80 ms of information, the results of Petten et al. (1999) suggest that evoked differences in cortical processing related to the discrimination process could start in the region of 100 ms after call-sign onset.

Appendix G

Cortical SPMs: Peak Activations

Abbreviations

Temporal

TTG, Transverse Temporal Gyrus; TTS, Transverse Temporal Sulcus; STG, Superior Temporal Gyrus; STS, Superior Temporal Sulcus; aSTS, Anterior STS; pSTS, Posterior STS; MTG, Middle Temporal Gyrus; ITG, Inferior Temporal Gyrus; ITS, Inferior Temporal Sulcus; PT, Planum Temporale; LF, Lateral Fissure; sInG, Short Insular Gyrus; CirS, Circular Sulcus (insula); SCirS, Superior Circular Sulcus (insula).

Parietal

SPG, Superior Parietal Gyrus; IPG, Inferior Parietal Gyrus; aIPG, Anterior IPG; pIPG, Posterior IPG; IPS, Intra-Parietal Sulcus; PoCG, Post-Central Gyrus; PoCS, Post-Central Sulcus; SubPS, Sub-parietal Sulcus; CuG, Cuneus Gyrus; PrCuG, Precuneus Gyrus.

Frontal

SFG, Superior Frontal Gyrus; SFS, Superior Frontal Sulcus; OG, Orbital Gyrus; IFG, Inferior Frontal Gyrus; IFS, Inferior Frontal Sulcus; MFG, Middle Frontal Gyrus; MFS, Middle Frontal Sulcus; FTG, Frontal Transverse Gyrus; TFS, Transverse Frontal Sulcus; FMG, Fronto-marginal Gyrus; CS, Central Sulcus; PrCG, Pre-Central Gyrus; SPrCG, Superior Pre-Central Gyrus; PrCS, Pre-Central Sulcus; SPrCS, Superior Pre-Central Sulcus; IPrCS, Inferior Pre-Central Sulcus; SCG, Sub-central Gyrus.

Occipital

SOG, Superior Occipital Gyrus; SOS, Superior Occipital Sulcus; MOG, Middle Occipital Gyrus; MOS, Middle Occipital Sulcus; POS, Parieto-Occipital sulcus; CalcS, Calcarine Sulcus/fissure.

Cingulate

CiG, Cingulate Gyrus; aCiG, Anterior CiG; CiS, Cingulate Sulcus; pCiS, Posterior CiS.

Peak Activations

			MNI Talairach			
Region	**Hemi**	**Peak (ms)**	*x*	*y*	*z*	**t-value**
0.5–4 Hz						
Temporal						
TTS	RH	97.3	54.82	−21.48	3.48	9.08
STG	RH	123.4	52.64	5.02	−12.52	5.95
pSTS	LH	76.7	−41.03	−67.39	27.22	5.71
	LH	157.4	−37.98	−59.17	25.73	5.47
MTG	RH	109.1	63.38	−25.74	−12.35	7.41
sInG	RH	131.2	34.91	12.34	−3.19	5.68
Parietal						
IPG	RH	131.4	42.47	−65.68	32.24	5.95
IPS	LH	235.6	−25.30	−56.78	48.26	6.43
Frontal						
IFG	RH	139.4	50.57	26.62	4.66	6.4
MFG	LH	235.6	−40.08	32.03	28.80	5.87
PrCG	LH	241.6	−55.42	−2.22	36.22	6.25
Cingulate						
aCiG	RH	117.4	5.83	12.81	29.30	5.95
	LH	187.5	−5.70	25.92	20.55	5.46
4–8 Hz						
Temporal						
STG	LH	87.0	−62.71	−38.67	11.78	−6.57
Parietal						
IPS	RH	79.3	20.99	−62.98	38.76	−6.09
PoCS	LH	73.7	−52.61	−21.58	31.88	−5.8

Table G.1. Minimum-norm: Peak differences for the “Onset attention” comparison.

			MNI Talairach			
Region	**Hemi**	**Peak (ms)**	*x*	*y*	*z*	**t-value**
0.5–4 Hz						
Temporal						
STS	LH	416.5	−51.98	−40.76	5.33	5.87
pSTS	RH	416.5	49.39	−55.72	15.65	5.81
MTG	RH	426.1	59.95	−50.19	2.83	5.52
PT	LH	405.5	−56.09	−43.00	16.16	6.22
CirS	LH	346.7	−34.05	−15.95	16.56	6.0
Parietal						
IPG	LH	340.6	−31.61	−70.69	39.08	5.63
IPS	LH	340.6	−44.80	−49.60	39.94	6.29
Frontal						
MFG	RH	358.3	34.69	28.27	38.58	6.64
IPrCS	RH	342.1	39.58	1.00	36.95	5.46
Occipital						
SOG	RH	178.4	24.90	−93.27	12.26	5.7

Table G.2. Minimum-norm: ROIs from 0.5–4 Hz in the early window of the “Call-sign processing” comparison.

			MNI Talairach			
Region	**Hemi**	**Peak (ms)**	*x*	*y*	*z*	**t-value**
0.5–4 Hz						
Temporal						
STG	RH	585.4	62.55	−25.48	0.95	7.56
	RH	783.5	63.37	−16.13	3.20	5.82
pSTS	RH	669.7	48.14	−60.96	22.09	6.78
	LH	679.8	−41.89	−64.75	12.18	5.82
MTG	RH	585.4	62.50	−41.32	−4.71	6.51
PT	RH	585.4	57.66	−21.59	4.55	8.39
Parietal						
IPG	RH	585.4	61.18	−17.12	20.78	6.18
	LH	607.3	−47.46	−56.30	41.04	5.49
	RH	634.1	42.58	−62.71	38.47	6.17
aIPG	RH	532.3	59.29	−41.55	16.22	5.24
pIPG	RH	532.3	45.98	−63.01	31.44	5.39
IPS	LH	766.8	−25.30	−56.78	48.26	5.85
PoCS	RH	684.2	24.73	−43.00	56.30	5.38
	LH	790.8	−49.21	−24.73	32.62	6.07
PrCuG	LH	757.8	−8.32	−43.04	42.56	6.48
Frontal						
IFG	LH	607.3	−49.04	17.11	21.00	6.2
	RH	640.3	50.11	25.58	6.64	6.76
IFS	LH	607.3	−38.23	18.16	31.05	5.77
MFG	LH	757.8	−40.46	11.99	47.96	6.17
MFS	LH	614.6	−22.82	53.46	11.08	5.93
FTG	RH	532.3	21.96	56.57	1.94	5.37
CS	LH	679.8	−45.16	−8.99	27.48	6.12
PrCS	RH	779.8	35.38	4.61	26.63	5.73
IPrCS	LH	766.8	−44.81	−2.52	33.75	7.52
SCG	LH	679.8	−57.14	2.92	9.76	5.8
Occipital						
MOG	RH	669.4	46.49	−71.38	10.22	7.78
	LH	699.1	−38.64	−75.52	27.24	5.99
POS	RH	644.0	23.64	−60.69	21.72	5.28
	LH	735.8	−17.38	−60.79	23.79	6.25
CalcS	RH	629.3	21.66	−71.00	7.30	5.25
Cingulate						
CiG	RH	610.5	6.06	15.36	27.92	5.18
	LH	756.4	−5.70	25.92	20.55	5.74
pCiS	LH	798.2	−16.38	−44.99	57.10	6.35

Table G.3. Minimum-norm: ROIs from 0.5–4 Hz in the late window of the “Call-sign processing” comparison.

Region	Hemi	Peak (ms)	MNI Talairach *x*	*y*	*z*	t-value
4–8 Hz						
Temporal						
STG	LH	319.7	−62.26	−42.69	11.36	5.48
STS	RH	555.9	47.98	−12.05	−18.63	−5.62
Parietal						
PoCS	LH	644.0	−44.21	−40.26	39.46	−5.08
Frontal						
MFG	RH	364.2	38.54	48.98	9.29	6.66

Table G.4. Minimum-norm: Peak differences for the “Call-sign processing” comparison from 4–8 Hz.

Region	Hemi	MNI Talairach *x*	*y*	*z*	t-value
8–13 Hz					
Temporal					
STG	LH	−64.02	−32.12	8.07	−8.30
STS	RH	37.99	−65.73	20.08	−8.14
PT	LH	−59.09	−42.05	14.81	−8.66
ITG	LH	−55.33	−51.92	−11.35	−8.81
	RH	54.42	−17.12	−33.32	−8.32
SCirS	LH	−28.88	29.28	5.80	−8.41
Parietal					
PoCS	LH	−52.61	−21.58	31.88	−8.95
IPG	LH	−40.12	−70.30	32.63	−7.74
	RH	45.98	−63.01	31.44	−8.17
IPS	RH	32.36	−52.25	37.84	−8.74
Frontal					
SFG	RH	8.87	31.95	34.30	−7.81
IFG	LH	−46.05	29.75	9.74	−7.74
	RH	47.70	32.30	−4.13	−8.87
IPrCS	RH	51.01	3.50	19.64	−8.29
SPrCS	RH	29.82	−13.41	53.30	−7.77
aCiG	RH	8.00	38.27	5.68	−8.05
CiG	RH	4.17	−9.15	35.59	−8.00
Occipital					
SOS	LH	−22.11	−81.62	16.46	−8.67
POS	LH	−20.81	−70.38	20.14	−8.29
	RH	22.86	−63.38	23.76	−8.17

Table G.5. Spatial filtering: Peak differences from the “Onset attention” comparison from 8–13 Hz.

			MNI Talairach		
Region	**Hemi**	*x*	*y*	*z*	**t-value**
13–30 Hz					
Temporal					
PT	LH	−61.43	−41.91	14.11	−7.65
STG	LH	−62.65	−19.66	−0.89	−7.81
MTG	LH	−58.01	−46.88	−8.37	−7.78
ITG	RH	54.42	−17.12	−33.32	−8.77
LF	RH	44.59	−33.11	26.64	−8.41
Parietal					
IPG	LH	−41.52	−69.81	30.42	−7.53
	RH	51.92	−29.77	40.22	−8.16
POS	LH	−20.77	−62.27	22.52	−7.83
IPS	RH	37.03	−57.12	39.84	−9.59
PoCG	RH	45.27	−25.58	49.86	−8.88
Frontal					
IFG	RH	51.99	12.26	7.06	−9.17
CiS	RH	16.73	−27.52	37.50	−8.19
CiG	RH	4.37	−12.04	35.88	−7.80
40–80 Hz					
Temporal					
aSTS	RH	49.49	−10.62	−20.89	−6.66
pSTS	RH	44.91	−57.01	28.11	−7.26
	LH	−41.15	−65.65	22.90	−6.81
ITG	LH	−54.78	−50.21	−9.84	−6.86
ITS	RH	56.58	−46.26	−13.40	−7.08
PT	RH	61.78	−35.07	15.06	−7.06
	LH	−61.11	−45.56	15.73	−6.60
Parietal					
SPG	RH	15.44	−55.87	59.70	−6.73
IPS	RH	37.63	−60.84	42.83	−8.28
PoCS	RH	41.46	−26.57	58.99	−8.10
PrCuG	RH	7.13	−48.14	49.93	−7.09
	LH	−7.11	−66.42	27.70	−6.62
Frontal					
IFS	RH	38.97	23.28	23.75	−7.93
Occipital					
SOG	LH	−21.25	−83.25	18.64	−7.04
POS	RH	19.42	−56.29	14.27	−6.93
	LH	−20.81	−70.38	20.14	−6.88
Cingluate					
CiS	RH	17.03	−30.46	37.51	−7.42

Table G.6. Spatial filtering: Peak differences from the “Onset attention” comparison from 13–30 Hz and 40–80 Hz

		MNI Talairach			
Region	**Hemi**	*x*	*y*	*z*	**t-value**
4–8 Hz					
Temporal					
TTG	RH	42.37	−26.49	9.26	−5.87
STG	RH	63.16	−32.56	11.82	−6.05
	LH	−62.70	−17.90	−0.11	−9.40
STS	RH	49.00	−34.05	3.59	−5.94
MTG	LH	−61.71	−30.61	−12.69	−7.78
CirS	RH	32.96	27.81	7.58	−6.69
	LH	−28.26	29.79	3.93	−9.44
Parietal					
IPG	LH	−56.72	−49.89	30.22	−7.78
	RH	50.10	−53.92	42.49	−5.85
SPG	RH	34.42	−46.17	60.41	−5.48
PrCuG	LH	−6.40	−68.31	33.47	−9.57
Frontal					
IFG	LH	−44.59	31.19	−1.45	−9.19
IFS	RH	44.15	35.56	3.56	−6.84
FMG	RH	30.83	55.85	−8.81	−5.52
PrCG	RH	58.16	2.42	26.98	−6.76
SPrCG	RH	37.09	−20.03	59.02	−5.78
Occipital					
MOS	LH	−23.16	−91.54	4.39	−8.69
SOS	LH	−20.13	−86.63	14.74	−8.44
POS	LH	−12.17	−75.33	18.47	−10.07
Cingulate					
CiG	RH	4.54	−21.04	35.29	−5.54

Table G.7. Spatial filtering: Peak differences from the “Call-sign processing” comparison from 4–8 Hz.

		MNI Talairach			
Region	**Hemi**	*x*	*y*	*z*	**t-value**
8–13 Hz					
Temporal					
STG	LH	−62.97	−19.18	0.58	−8.17
PT	RH	61.05	−36.23	16.60	−8.87
	LH	−55.09	−44.88	18.14	−8.98
pSTS	RH	38.12	−62.71	20.65	−8.15
	LH	−41.21	−62.94	14.86	−9.16
MTG	LH	−59.39	−31.10	−12.37	−8.18
CirS	RH	31.14	18.49	−3.38	−8.19
	LH	−30.43	31.63	3.23	−8.61
Parietal					
IPG	RH	45.98	−63.01	31.44	−8.57
	LH	−58.73	−50.97	24.35	−8.10
aIPG	RH	46.45	−26.59	21.24	−8.64
IPS	RH	33.42	−67.71	32.27	−8.30
PoCG	RH	39.44	−31.98	52.67	−9.55
SubPS	LH	−12.77	−48.88	29.41	−8.31
Frontal					
SFS	RH	25.71	6.21	46.54	−8.25
IFS	RH	40.63	21.38	28.37	−8.53
IFG	RH	47.78	34.40	−7.85	−8.34
FMG	RH	31.50	54.50	−7.21	−8.68
	LH	−28.91	47.93	2.29	−8.93
PrCG	LH	−40.94	3.29	46.49	−8.04
IPrCS	RH	45.85	3.41	24.25	−9.36
	LH	−37.90	7.03	42.39	−8.03
CS	LH	−38.60	−19.78	33.81	−7.87
Occipital					
MOG	LH	−45.82	−73.88	12.29	−8.62
MOS	LH	−23.95	−92.53	7.01	−8.93
SOS	LH	−25.94	−72.96	20.54	−8.66
POS	LH	−20.95	−60.90	19.19	−9.15
Cingulate					
CiG	RH	5.26	2.83	37.58	−8.67
	LH	−3.94	−14.00	35.28	−7.84
CiS	RH	16.61	−24.30	42.02	−8.48
	LH	−11.77	41.78	5.50	−7.81

Table G.8. Spatial filtering: Peak differences from the “Call-sign processing” comparison from 8–13 Hz.

		MNI Talairach			
Region	**Hemi**	*x*	*y*	*z*	**t-value**
13–30 Hz					
Temporal					
PT	LH	−60.35	−43.68	15.36	−8.37
	RH	50.03	−31.16	9.23	−7.82
MTG	LH	−60.95	−28.39	−12.66	−9.35
ITG	RH	52.74	−15.63	−32.69	−8.59
Parietal					
IPG	LH	−39.67	−72.21	33.05	−8.59
aIPG	RH	46.37	−27.33	23.31	−8.48
pIPG	RH	41.80	−63.62	43.02	−8.40
PoCG	RH	44.05	−28.99	54.14	−9.52
SubPS	LH	−15.23	−46.04	32.19	−8.21
Frontal					
SFG	RH	9.94	50.72	34.75	−8.44
TFS	LH	−15.68	62.61	−3.01	−7.84
MFG	LH	−38.37	6.39	44.48	−7.81
PrCG	LH	−55.35	3.18	30.88	−8.41
CS	RH	43.35	−9.92	29.11	−8.55
Occipital					
SOS	LH	−17.71	−86.87	20.76	−8.23
SOG	RH	23.67	−84.99	29.87	−8.15
CalcS	LH	−13.18	−76.88	12.75	−8.36
Cingulate					
CiG	RH	5.34	−0.12	38.62	−7.72
CiS	RH	15.35	−29.37	36.51	−8.27
40–80 Hz					
Temporal					
STG	LH	−63.38	−32.81	4.69	−6.20
STS	LH	−42.70	−67.18	23.65	−6.74
	RH	49.49	−10.62	−20.89	−7.83
MTG	LH	−62.82	−20.08	−16.55	−6.79
ITS	RH	49.52	−12.29	−29.25	−8.05
Parietal					
IPG	RH	47.57	−61.68	31.35	−6.70
IPS	RH	32.13	−42.45	37.68	−7.11
CuG	LH	−7.97	−69.94	10.52	−7.70
PrCuG	LH	−5.45	−58.96	15.45	−7.29
Frontal					
OG	LH	−15.51	41.33	−19.24	−7.52
IFG	LH	−52.34	14.31	12.36	−7.93
	RH	52.43	−51.10	−16.40	−7.22
PrCG	LH	−56.89	1.63	31.85	−7.31
Occipital					
SOS	LH	−20.06	−84.72	16.43	−7.86
SOG	RH	23.66	−81.59	30.27	−7.27

Table G.9. Spatial filtering: Peak differences from the “Call-sign processing” comparison from 13–30 Hz and 40–80 Hz.

	MNI Talairach		
	x	*y*	*z*
“Onset attention”			
ROI O1	−37.98	−59.17	25.73
“Call-sign processing”			
ROI C1	47.98	−12.05	−18.63
ROI C2	45.98	−63.01	31.44
ROI C3	34.69	28.27	38.58
ROI C4	24.90	−93.27	12.26
ROI C5	−17.38	−60.79	23.79
ROI C6	23.64	−60.69	21.72
ROI C7	−16.38	−44.99	57.10
ROI C8	48.14	−60.96	22.09
ROI C9	−41.89	−64.75	12.18
ROI S1	−50.66	−32.06	35.00
ROI S2	−50.66	−32.06	35.00

Table G.10. Co-ordinates of the regions of interest from the minimum-norm analysis in which the peak differences or peak latencies correlated with task performance during MEG imaging.

Glossary

ACC	Anterior Cingulate Cortex
ANOVA	Analysis of Variance
BEA	Better-ear Average
BOLD	Blood Oxygen Level-Dependent
CAS	Central Auditory System
dB	Decibels
dB HL	Decibels (Hearing level)
ECG	Electro-cardiogram
EEG	Electro-encephalography
EOG	Electro-oculogram
ERF	Event-Related Field
ERP	Event-Related Potential
FFA	Four-Frequency Average (Hearing level)
FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance Imaging
FWER	Family-Wise Error Rate
HATS	Head and Torso Simulator
Hz	Hertz
ICP	Iterative Closest Point
IFG	Inferior Frontal Gyrus
IID	Inter-Aural Intensity Difference
ILD	Inter-Aural Level Difference
IM	Informational Masking
ITD	Inter-Aural Time Difference
K-S	Kolmogorov-Smirnov
MD	Multiple-demand pattern
MMN	Mismatch Negativity
MSR	Magnetically-shielded room
Nd	Negative Displacement (ERP)
PCA	Principal Component Analysis
PET	Positron Emission Tomography
rCBF	Regional Cerebral Blood Flow

ROI	Region Of Interest
SLT	Spatial Listening Task
SNR	Signal-to-Noise Ratio
SPIN	Speech In Noise Test
SPM	Statistical Parametric Map
SRT	Speech-Reception Threshold
SSQ	Speech, Spatial, and Qualities of Hearing Scale
STG	Superior Temporal Gyrus
TEA	Test of Everyday Attention
TMR	Target-to-Masker Ratio
WM	Working Memory

References

4-D Neuroimaging (2005). *Magnes 2500/3600 WH - MSW Software Reference Manual*. San Diego, CA.

Akeroyd, M. A. (2006). The psychoacoustics of binaural hearing. *International Journal of Audiology, 45 Suppl*, 25–33.

Akeroyd, M. A., & Summerfield, A. Q. (1999). A binaural analog of gap detection. *The Journal of the Acoustical Society of America, 105*(5), 2807–20.

Albert, M. (1993). Neuropsychological and neurophysiological changes in healthy adult humans across the age range. *Neurobiology of Aging, 14*(6), 623–5.

Alho, K., Medvedev, S. V., Pakhomov, S. V., Roudas, M. S., Tervaniemi, M., Reinikainen, K., Zeffiro, T., & Näätänen, R. (1999). Selective tuning of the left and right auditory cortices during spatially directed attention. *Brain Research. Cognitive Brain Research, 7*(3), 335–341.

Allen, K., Carlile, S., & Alais, D. (2008). Contributions of talker characteristics and spatial location to auditory streaming. *The Journal of the Acoustical Society of America, 123*(3), 1562–1570.

ANSI (1960). American standard acoustical terminology.

Arbogast, T. L., & Kidd, J., G. (2000). Evidence for spatial tuning in informational masking using the probe-signal method. *The Journal of the Acoustical Society of America, 108*(4), 1803–10.

Arbogast, T. L., Mason, C. R., & Kidd, J., G. (2002). The effect of spatial separation on informational and energetic masking of speech. *The Journal of the Acoustical Society of America, 112*(5 Pt 1), 2086–98.

Arbogast, T. L., Mason, C. R., & Kidd, J., G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 117*(4), 2169–2180.

Bacon, S. P., & Moore, B. C. (1986). Temporal effects in masking and their influence on psychophysical tuning curves. *The Journal of the Acoustical Society of America, 80*(6), 1638–1645.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience, 4*(10), 829–839.

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine, 18*(6), 14–30.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Morris, J. N., Rebok, G. W., Smith, D. M., Tennstedt, S. L., Unverzagt, F. W., Willis, S. L., & Advanced Cognitive Training for Independent and Vital Elderly Study Group (2002). Effects of cognitive training interventions with older adults: a randomized controlled trial. *Journal of the American Medical Association, 288*(18), 2271–2281.

Barnes, G. R., & Hillebrand, A. (2003). Statistical flattening of MEG beamformer images. *Human Brain Mapping, 18*(1), 1–12.

Barnett, S. (1990). *Matrices: Methods and Applications*. New York: Oxford University Press.

Bartzokis, G. (2004). Age-related myelin breakdown: a developmental model of cognitive decline and alzheimer's disease. *Neurobiology of Aging, 25*(1), 5–18; author reply 49–62.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309–312.

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology, 90*(3), 229–241.

Bibbig, A., Traub, R. D., & Whittington, M. A. (2002). Long-range synchronization of gamma and beta oscillations and the plasticity of excitatory and inhibitory synapses: a network model. *Journal of Neurophysiology, 88*(4), 1634–1654.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27*(1), 32–48.

Boersma, P., & Weenink, D. (2008). *Praat: doing phonetics by computer v5.0.32*. Institute of Phonetic Sciences, University of Amsterdam.
URL `http://www.praat.org/`

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America, 107*(2), 1065–1066.

Bolia, R. S., Nelson, W. T., & Morley, R. M. (2001). Asymmetric performance in the cocktail party effect: implications for the design of spatial audio displays. *Human Factors, 43*(2), 208–16.

Box, G. E. P., & Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics, 29*(2), 610–611.

Broadbent, D. E. (1958). *Perception and communication.* Oxford: Oxford University Press.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica United with Acustica, 86*(1), 117–128.

Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America, 83*(4), 1508–16.

Bronkhorst, A. W., & Plomp, R. (1989). Binaural speech intelligibility in noise for hearing-impaired listeners. *The Journal of the Acoustical Society of America, 86*(4), 1374–83.

Bronkhorst, A. W., & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *The Journal of the Acoustical Society of America, 92*(6), 3132–9.

Brookes, M. J., Stevenson, C. M., Barnes, G. R., Hillebrand, A., Simpson, M. I., Francis, S. T., & Morris, P. G. (2007). Beamformer reconstruction of correlated sources using a modified source model. *Neuroimage, 34*(4), 1454–65.

Brookes, M. J., Vrba, J., Robinson, S. E., Stevenson, C. M., Peters, A. M., Barnes, G. R., Hillebrand, A., & Morris, P. G. (2008). Optimising experimental design for MEG beamformer imaging. *Neuroimage, 39*(4), 1788–1802.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 1101–9.

Brungart, D. S., & Simpson, B. D. (2002). Within-ear and across-ear interference in a cocktail-party listening task. *The Journal of the Acoustical Society of America, 112*(6), 2985–95.

Brungart, D. S., & Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: effects of masker uncertainty. *The Journal of the Acoustical Society of America, 115*(1), 301–10.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5 Pt 1), 2527–38.

BSA (2004). *Recommended procedure: Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels*. British Society of Audiology.

Chaba (1988). Speech understanding and aging. *The Journal of the Acoustical Society of America, 83*(3), 859–95.

Chapman, R. M., Ilmoniemi, R. J., Barbanera, S., & Romani, G. L. (1984). Selective localization of alpha brain activity with neuromagnetic measurements. *Electroencephalography and Clinical Neurophysiology, 58*(6), 569–72.

Chen, Q., Marshall, J. C., Weidner, R., & Fink, G. R. (2008). Zooming in and zooming out of the attentional focus: An fMRI study. *Cerebral Cortex*. Advance on-line access.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975–979.

Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 26*(4), 554–559.

Cocchiarella, L., Turk, M. A., & Andersson, G. (2000). Improving the evaluation of permanent impairment. *Journal of the American Medical Association, 283*(4), 532–533.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America, 119*(3), 1562–73.

Corbetta, M., Miezin, F. M., Shulman, G. L., & Petersen, S. E. (1993). A PET study of visuospatial attention. *The Journal of Neuroscience, 13*(3), 1202–1226.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience, 18*(18), 7426–7435.

Cuffin, B. N., & Cohen, D. (1977). Magnetic fields of a dipole in special volume conductor shapes. *IEEE Transactions on Bio-Medical Engineering, 24*(4), 372–81.

Culling, J. F., & Summerfield, Q. (1995). Perceptual separation of concurrent speech sounds: absence of across-frequency grouping by common interaural delay. *The Journal of the Acoustical Society of America*, *98*(2 Pt 1), 785–97.

Cusack, R., & Long, C. (2006). Simulator for tube phone auditory response. Retrieved December 12, 2006.
URL `http://www.mrc-cbu.cam.ac.uk/ rhodri/headphonesim.htm`

Czigler, I., Csibra, G., & Csontos, A. (1992). Age and inter-stimulus interval effects on event-related potentials to frequent and infrequent auditory stimuli. *Biological Psychology*, *33*(2-3), 195–206.

Dajani, H. R., Purcell, D., Wong, W., Kunov, H., & Picton, T. W. (2005). Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Bio-Medical Engineering*, *52*(9), 1614–1618.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179–194.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., & Halgren, E. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*(1), 55–67.

Dale, A. M., & Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *Journal of Cognitive Neuroscience*, *5*, 162–176.

Dangour, A. D., Clemens, F., Elbourne, D., Fasey, N., Fletcher, A. E., Hardy, P., Holder, G. E., Huppert, F. A., Knight, R., Letley, L., Richards, M., Truesdale, A., Vickers, M., & Uauy, R. (2006). A randomised controlled trial investigating the effect of n-3 long-chain polyunsaturated fatty acid supplementation on cognitive and retinal function in cognitively healthy older people: the older people and n-3 long-chain polyunsaturated fatty acids (OPAL) study protocol [isrctn72331636]. *Nutrition Journal*, *5*, 20.

Darwin, C. J. (2006). Contributions of binaural information to the separation of different sound sources. *International Journal of Audiology*, *45 Suppl*, 20–4.

Davis, A. C. (1989). The prevalence of hearing impairment and reported hearing disability among adults in Great Britain. *International Journal of Epidemiology*, *18*(4), 911–7.

Deouell, L. Y., Heller, A. S., Malach, R., D'Esposito, M., & Knight, R. T. (2007). Cerebral responses to change in spatial location of unattended sounds. *Neuron*, *55*(6), 985–996.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Deutsch, G., Papanicolaou, A. C., Bourbon, W. T., & Eisenberg, H. M. (1987). Cerebral blood flow evidence of right frontal activation in attention demanding tasks. *The International Journal of Neuroscience, 36*(1-2), 23–28.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology, 92*(Pt 1), 53–78.

Drullman, R., & Bronkhorst, A. W. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *The Journal of the Acoustical Society of America, 107*(4), 2224–35.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America, 76*(1), 87–96.

Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *The Journal of the Acoustical Society of America, 111*(6), 2897–907.

Dujardin, K., Derambure, P., Bourriez, J. L., Jacquesson, J. M., & Guieu, J. D. (1993). P300 component of the event-related potentials (ERP) during an attention task: effects of age, stimulus modality and event probability. *International Journal of Psychophysiology, 14*(3), 255–67.

Duncan, J. (2006). EPS Mid-Career Award 2004: Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology, 59*(1), 2–27.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23*(10), 475–483.

Duquesnoy, A. J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *The Journal of the Acoustical Society of America, 74*(3), 739–743.

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking. *The Journal of the Acoustical Society of America, 113*(6), 2984–2987.

Ebata, M. (2003). Spatial unmasking and attention related to the cocktail party problem. *Acoustical Science and Technology, 24*(5), 208–219.

Ermer, J. J., Mosher, J. C., Baillet, S., & Leahy, R. M. (2001). Rapidly recomputable EEG forward models for realistic head shapes. *Physics in Medicine and Biology*, *46*(4), 1265–81.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). 3d statistical neuroanatomical models from 305 MRI volumes. In *Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference*, (pp. 1813–1817).

Festen, J. M., & Plomp, R. (1983). Relations between auditory functions in impaired hearing. *The Journal of the Acoustical Society of America*, *73*(2), 652–662.

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications, 2nd ed.

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11050–11055.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355.

Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *Neuroimage*, *23 Suppl 1*, S69–S84.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999a). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, *9*(2), 195–207.

Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284.

Folk, C. L., & Hoyer, W. J. (1992). Aging and shifts of visual spatial attention. *Psychology and Aging*, *7*(3), 453–65.

Ford, J. M., & Pfefferbaum, A. (1991). Event-related potentials and eyeblink responses in automatic and controlled processing: effects of age. *Electroencephalography and Clinical Neurophysiology*, *78*(5), 361–77.

Freesurfer Development Team (2007). *Freesurfer image analysis suite v4.0.1*. Martinos Center for Biomedical Imaging. Retrieved January 2, 2008.
URL `http://surfer.nmr.mgh.harvard.edu/`

Frisina, D. R., & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hearing Research*, *106*(1-2), 95–104.

Frith, C. D., & Friston, K. J. (1996). The role of the thalamus in “top down” modulation of attention to sound. *Neuroimage*, *4*(3 Pt 1), 210–215.

Fuchs, M., Wagner, M., Kohler, T., & Wischmann, H. A. (1999). Linear and nonlinear current density reconstructions. *Journal of Clinical Neurophysiology*, *16*(3), 267–95.

Gatehouse, S., & Akeroyd, M. (2006). Two-eared listening in dynamic situations. *International Journal of Audiology*, *45 Suppl*, 120–4.

Gatehouse, S., Akeroyd, M., Deas, R., Glover, C., Howell, P., & Lawson, H. (2006). A large scale study of the drivers of auditory disability in adults aged 50 to 80: Attentional capacities. In *Annual Meeting of the British Society of Audiology*. Cambridge, UK.

Gatehouse, S., & Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology*, *43*(2), 85–99.

George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., & Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2362–2375.

Gideon, R., & Hollister, R. (1987). A rank correlation coefficient resistant to outliers. *Journal of the American Statistical Association*, *82*(398), 656–666.

Goldman-Rakic, P. S. (1988). Topography of cognition: parallel distributed networks in primate association cortex. *Annual Review of Neuroscience*, *11*, 137–156.

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, *21*, 215–223.

Golub, G. H., & Pereyra, V. (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on Numerical Analysis*, *10*, 413–432.

Golub, G. H., & Van Loan, C. (1996). *Matrix Computations*. Johns Hopkins University Press, 3rd ed.

Gordon-Salant, S. (2005). Hearing loss and aging: New research findings and clinical implications. *Journal of Rehabilitation Research and Development*, *42*(4 Suppl 2), 9–24.

Gorodnitsky, I. F., George, J. S., & Rao, B. D. (1995). Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. *Electroencephalography and Clinical Neurophysiology*, *95*(4), 231–51.

Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C. M., & Landis, T. (2001). Noninvasive localization of electromagnetic epileptic activity. I. method descriptions and simulations. *Brain Topography*, *14*(2), 131–7.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., & Schnitzler, A. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 13050–13055.

Hämäläinen, M. S. (2007). *MNE software suite v2.5.2*. Martinos Center for Biomedical Imaging. Retrieved November 21, 2007.
URL `http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/sofMNE.php`

Hämäläinen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*(2), 413.

Hannaford, P. C., Simpson, J. A., Bisset, A. F., Davis, A., McKerrow, W., & Mills, R. (2005). The prevalence of ear, nose and throat problems in the community: results from a national cross-sectional postal survey in Scotland. *Family Practice*, *22*(3), 227–33.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology*, *49*(3-4), 277–290.

Hansen, P. C. (1992). Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, *34*(4), 561–580.

Hashimoto, R., Homae, F., Nakajima, K., Miyashita, Y., & Sakai, K. L. (2000). Functional differentiation in the human auditory and language areas revealed by a dichotic listening task. *Neuroimage*, *12*(2), 147–158.

Hauk, O. (2004). Keep it simple: a case for using classical minimum norm estimation in the analysis of EEG and MEG data. *Neuroimage*, *21*(4), 1612–21.

Hawley, M. L., Litovsky, R. Y., & Colburn, H. S. (1999). Speech intelligibility and localization in a multi-source environment. *The Journal of the Acoustical Society of America*, *105*(6), 3436–48.

Hays, W. L. (1974). *Statistics for the Social Sciences*. New York: Holt, Rinehart and Winston, 2nd ed.

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and Hearing, 29*(1), 87–98.

Helfer, K. S., & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech and Hearing Research, 33*(1), 149–55.

Helmholtz, H. (1853). Ueber einige Gesetze der Vertheilung elektrischer Strome in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. *Annalen Der Physik Und Chemie, 165*(6), 211–233.

Hickok, & Poeppel (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4*(4), 131–138.

Hietanen, A., Era, P., Henrichsen, J., Rosenhall, U., Sorri, M., & Heikkinen, E. (2005). Hearing among 75-year-old people in three Nordic localities: a comparative study. *International Journal of Audiology, 44*(9), 500–8.

Hillebrand, A., & Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *Neuroimage, 16*(3 Pt 1), 638–50.

Hillebrand, A., & Barnes, G. R. (2005). Beamformer analysis of MEG data. *International Review of Neurobiology, 68*, 149–71.

Hillebrand, A., Singh, K. D., Holliday, I. E., Furlong, P. L., & Barnes, G. R. (2005). A new approach to neuroimaging with magnetoencephalography. *Human Brain Mapping, 25*(2), 199–211.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*(108), 177–180.

Hink, R., & Hillyard, S. (1976). Auditory evoked potentials during selective listening to dichotic speech messages. *Perception & Psychophysics, 20*, 236–242.

Hofman, M., & Van Opstal, J. (2003). Binaural weighting of pinna cues in human sound localization. *Experimental Brain Research, 148*(4), 458–70.

Horn, R. A., & Johnson, C. R. (1990). *Matrix analysis.* New York: Cambridge University Press.

Houtgast, T., & Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology, 47*(6), 287–295.

Huang, M. X., Mosher, J. C., & Leahy, R. M. (1999). A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. *Physics in Medicine and Biology, 44*(2), 423–40.

Huang, M. X., Shih, J. J., Lee, R. R., Harrington, D. L., Thoma, R. J., Weisend, M. P., Hanlon, F., Paulson, K. M., Li, T., Martin, K., Millers, G. A., & Canive, J. M. (2004). Commonalities and differences among vectorized beamformers in electromagnetic source imaging. *Brain Topography*, *16*(3), 139–58.

Humes, L. E. (1991). Understanding the speech-understanding problems of the hearing impaired. *Journal of the American Academy of Audiology*, *2*(2), 59–69.

Humes, L. E. (1996). Speech understanding in the elderly. *Journal of the American Academy of Audiology*, *7*(3), 161–7.

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America*, *120*(5), 2926–2937.

Ille, N., Berg, P., & Scherg, M. (1997). A spatial components method for continuous artifact correction in EEG and MEG. *Biomedical Techniques and Biomedical Engineering*, *42*(suppl. 1), 80–83.

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology*, *19*(2), 113–124.

Iwaki, S., & Ueno, S. (1998). Weighted minimum-norm source estimation of magnetoencephalography utilizing the temporal information of the measured data. *Journal of Applied Physics*, *83*(11), 6441–6443.

James, W. (1890). *The Principles of Psychology.*. New York: Henry Holt.

Jeffs, B., Leahy, R., & Singh, M. (1987). An evaluation of methods for neuromagnetic image reconstruction. *IEEE Transactions on Bio-Medical Engineering*, *34*(9), 713–23.

Jensen, O., Kaiser, J., & Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317–324.

Jerger, J., & Martin, J. (2004). Hemispheric asymmetry of the right ear advantage in dichotic listening. *Hearing Research*, *198*(1-2), 125–36.

Jerger, J., Stach, B., Pruitt, J., Harper, R., & Kirby, H. (1989). Comments on “Speech Understanding and Aging”. *The Journal of the Acoustical Society of America*, *85*(3), 1352–4.

Johnson, T. A., & Brown, C. J. (2005). Threshold prediction using the auditory steady-state response and the tone burst auditory brain stem response: a within-subject comparison. *Ear and Hearing*, *26*(6), 559–576.

Johnsrude, I. S., Mackey, A., Alexander, E. M., Macdonald, H., & Carlyon, R. P. (2008). Effects of voice familiarity and age on perceptual organization of sound from two competing talkers. *The Journal of the Acoustical Society of America*, *123*(5), 3566.

Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology*, *9*(4), 510–22.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351.

Kidd, J., G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *The Journal of the Acoustical Society of America*, *118*(6), 3804–15.

Kidd, J., G., Mason, C. R., Deliwala, P. S., Woods, W. S., & Colburn, H. S. (1994). Reducing informational masking by sound segregation. *The Journal of the Acoustical Society of America*, *95*(6), 3475–80.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex*, *3*, 163–178.

Kinsbourne, M. (1978). Biological determinants of functional bisymmetry and asymmetry. In M. Kinsbourne (Ed.) *Asymmetrical Function of the Brain*, (pp. 3–16). New York: Cambridge University Press.

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, *30*, 57–78.

Kok, A. (2000). Age-related changes in involuntary and voluntary attention as reflected in components of the event-related potential (ERP). *Biological Psychology*, *54*(1-3), 107–43.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): still no identification without attention. *Psychological Review*, *111*(4), 880–913.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology*, *133*(3), 339–54.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–97.

Lee, J. H., & Humes, L. E. (2005). Comparison of cues in auditory selective and divided attention. *The Journal of the Acoustical Society of America*, *117*(4), 2606.

Leek, M. R., Brown, M. E., & Dorman, M. F. (1991). Informational masking and auditory attention. *Perception & Psychophysics, 50*(3), 205–14.

Leibold, L., Neff, D., & Jesteadt, W. (2005). An examination of temporal signal uncertainty for sequences of noise or random-frequency maskers. *The Journal of the Acoustical Society of America, 117*(4), 2536–2536.

Leonowicz, Z., Karvanen, J., & Shishkin, S. L. (2005). Trimmed estimators for robust averaging of event-related potentials. *Journal of Neuroscience Methods, 142*(1), 17–26.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2), Suppl 2:467+.

Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology. Human Perception and Performance, 30*(6), 1077–1091.

Lin, F. H., Belliveau, J. W., Dale, A. M., & Hämäläinen, M. S. (2003). Minimum current estimates with a cortical orientation constraint. In *The 9th Annual Meeting of the Organization for Human Brain Mapping,* (p. s45).

Lin, F. H., Belliveau, J. W., Dale, A. M., & Hämäläinen, M. S. (2006a). Distributed current estimates using cortical orientation constraints. *Human Brain Mapping, 27*(1), 1–13.

Lin, F. H., Witzel, T., Ahlfors, S. P., Stufflebeam, S. M., Belliveau, J. W., & Hämäläinen, M. S. (2006b). Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage, 31*(1), 160–71.

Lin, F. H., Witzel, T., Hämäläinen, M. S., Dale, A. M., Belliveau, J. W., & Stufflebeam, S. M. (2004). Spectral spatiotemporal imaging of cortical oscillations and interactions in the human brain. *Neuroimage, 23*(2), 582–95.

Liu, H., Schimpf, P. H., Dong, G., Gao, X., Yang, F., & Gao, S. (2005). Standardized shrinking LORETA-FOCUSS (SSLOFO): a new algorithm for spatio-temporal EEG source reconstruction. *IEEE Transactions on Bio-Medical Engineering, 52*(10), 1681–91.

Lorenzi, C., Gatehouse, S., & Lever, C. (1999). Sound localization in noise in normal-hearing listeners. *The Journal of the Acoustical Society of America, 105*(3), 1810–20.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*(6601), 616–618.

Madden, D. J. (1990). Adult age differences in attentional selectivity and capacity. *European Journal of Cognitive Psychology, 2*, 229–252.

Madden, D. J., Spaniol, J., Whiting, W. L., Bucur, B., Provenzale, J. M., Cabeza, R., White, L. E., & Huettel, S. A. (2007). Adult age differences in the functional neuroanatomy of visual attention: A combined fMRI and DTI study. *Neurobiology of Aging, 28*(3), 459–76.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.

Martin, J. S., & Jerger, J. F. (2005). Some effects of aging on central auditory processing. *Journal of Rehabilitation Research and Development, 42*(4 Suppl 2), 25–44.

Mattout, J., Phillips, C., Penny, W. D., Rugg, M. D., & Friston, K. J. (2006). MEG source localization under multiple constraints: an extended bayesian framework. *Neuroimage, 30*(3), 753–67.

MEGIS Software GmbH (2008). *BESA ® v5.2*. Grafelfing, Germany.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology. Human Perception and Performance, 21*(2), 387–409.

Moore, B. C. J. (1986). *Frequency Selectivity in Hearing*. London: Academic Press.

Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing*. San Diego: Academic Press, 5th ed.

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society, 26*, 394–395.

Moore, T. J. (1981). Voice communications jamming research. In *AGARD Conference Proceedings 331: Aural Communication in Aviation*, (pp. 2:1–2:6). Neuilly-Sur-Seine, France.

Mosher, J. C., Baillet, S., & Leahy, R. M. (2003). Equivalence of linear approaches in bioelectromagnetic inverse solutions. In *2003 IEEE Workshop on Statistical Signal Processing*, (pp. 294–297).

Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999). EEG and MEG: forward solutions for inverse methods. *IEEE Transactions on Bio-Medical Engineering, 46*(3), 245–259.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *The Behavioral and Brain Sciences, 13*, 201–288.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329.

Näätänen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biological Psychology*, *8*(2), 81–136.

Näätänen, R., Paavilainen, P., Tiitinen, H., Jiang, D., & Alho, K. (1993). Attention and mismatch negativity. *Psychophysiology*, *30*(5), 436–50.

Nager, W., Dethlefsen, C., & Münte, T. F. (2008). Attention to human speakers in a virtual auditory environment: Brain potential evidence. *Brain Research*, *1220*, 164–170.

Nakai, T., Kato, C., & Matsuo, K. (2005). An fMRI study to investigate auditory attention: a model of the cocktail party phenomenon. *Magnetic Resonance in Medical Sciences*, *4*(2), 75–82.

Nakai, T., Matsuo, K., Kato, C., Matsuzawa, M., Okada, T., Glover, G. H., Moriya, T., & Inui, T. (1999). A functional magnetic resonance imaging study of listening comprehension of languages in human at 3 tesla-comprehension level and activation of the language areas. *Neuroscience Letters*, *263*(1), 33–36.

Neff, D. L., & Green, D. M. (1987). Masking produced by spectral uncertainty with multicomponent maskers. *Perception & Psychophysics*, *41*(5), 409–15.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, *15*(1), 1–25.

Noble, W., & Gatehouse, S. (2004). Interaural asymmetry of hearing loss, speech, spatial and qualities of hearing scale (SSQ) disabilities, and handicap. *International Journal of Audiology*, *43*(2), 100–14.

Noble, W., Ter-Horst, K., & Byrne, D. (1995). Disabilities and handicaps associated with impaired auditory localization. *Journal of the American Academy of Audiology*, *6*(2), 129–40.

Okada, Y. (2003). Toward understanding the physiological origins of neuromagnetic signals. In Z. L. Lu, & L. Kaufman (Eds.) *Magnetic Source Imaging of the Human Brain,* (pp. 43–76). New Jersey: Lawrence Erlbaum Associates.

Okada, Y. C., Lauritzen, M., & Nicholson, C. (1987). Magnetic field associated with neural activities in an isolated cerebellum. *Brain Research*, *412*(1), 151–155.

Okada, Y. C., Wu, J., & Kyuhou, S. (1997). Genesis of MEG signals in a mammalian CNS structure. *Electroencephalography and Clinical Neurophysiology*, *103*(4), 474–85.

O'Sullivan, M., Jones, D. K., Summers, P. E., Morris, R. G., Williams, S. C., & Markus, H. S. (2001). Evidence for cortical "disconnection" as a mechanism of age-related cognitive decline. *Neurology*, *57*(4), 632–8.

Pascual-Marqui, R. D. (1999). Review of methods for solving the EEG inverse problem. *International Journal of Bioelectromagnetism*, *1*(1), 75–86.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., & Milroy, R. (1982). The deterioration of hearing with age: frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, *72*(6), 1788–1803.

Pearson, J. D., Morrell, C. H., Gordon-Salant, S., Brant, L. J., Metter, E. J., Klein, L. L., & Fozard, J. L. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *The Journal of the Acoustical Society of America*, *97*(2), 1196–205.

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, *51*, 406–413.

Penrose, R. (1956). On best approximate solution of linear matrix equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, *52*(17-19), 17–19.

Peters, R. W., & Moore, B. C. (1992). Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects. *The Journal of the Acoustical Society of America*, *91*(1), 256–66.

Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, *103*(1), 577–587.

Petten, C. V., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *25*(2), 394–417.

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, *110*(11), 1842–57.

Phan, M. L., Schendel, K. L., Recanzone, G. H., & Robertson, L. C. (2000). Auditory and visual spatial localization deficits following bilateral parietal lobe lesions in a patient with balint's syndrome. *Journal of Cognitive Neuroscience*, *12*(4), 583–600.

Phillips, C., Rugg, M. D., & Friston, K. J. (2002). Systematic regularization of linear inverse solutions of the EEG source localization problem. *Neuroimage*, *17*(1), 287–301.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*(4), 456–79.

Platel, H., Price, C., Baron, J. C., Wise, R., Lambert, J., Frackowiak, R. S., Lechevalier, B., & Eustache, F. (1997). The structural components of music perception. A functional anatomical study. *Brain*, *120 (Pt 2)*, 229–243.

Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, *66*(5), 1333–1342.

Polich, J. (1996). Meta-analysis of P300 normative aging studies. *Psychophysiology*, *33*(4), 334–353.

Posner, M. I. (2008). Measuring alertness. *Annals of the New York Academy of Sciences*, *1129*, 193–199.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M. I., Petersen, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science*, *240*(4859), 1627–1631.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*(2), 160–174.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press, 3rd ed.

Pugh, K. R., Offywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., Shankweiler, D. P., Katz, L., Constable, R. T., Fletcher, J., Lacadie, C., Marchione, K., & Gore, J. C. (1996). Auditory selective attention: an fMRI investigation. *Neuroimage*, *4*(3 Pt 1), 159–73.

Python Software Foundation (2007). *Python v2.5*. Hampton, NH.
URL `http://www.python.org`

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL `http://www.R-project.org`

Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine, 13*, 214–232.

Ribary, U., Ioannides, A. A., Singh, K. D., Hasson, R., Bolton, J. P., Lado, F., Mogilner, A., & Llinás, R. (1991). Magnetic field tomography of coherent thalamocortical 40-hz oscillations in humans. *Proceedings of the National Academy of Sciences of the United States of America, 88*(24), 11037–11041.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The Test of Everyday Attention. *Journal of the International Neuropsychological Society, 2*(6), 525–34.

Robinson, S. E., & Vrba, J. (1999). Functional neuroimaging by synthetic aperture magnetometry (SAM). In T. Yoshimoto, M. Kotani, S. Kuriki, H. Karibe, & N. Nakasato (Eds.) *Recent Advances in Biomagnetism*, (pp. 302–305). Sendai: Tohoku Univ. Press.

Ropshkow, O. (2004). Cross-section of the human cochlea. Retrieved September 23, 2008.
URL `http://en.wikipedia.org/wiki/Image:Cochlea-crosssection.png`

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *Neuroimage, 39*(3), 1444–1456.

Sabri, M., Liebenthal, E., Waldron, E. J., Medler, D. A., & Binder, J. R. (2006). Attentional modulation in the detection of irrelevant deviance: a simultaneous ERP/fMRI study. *Journal of Cognitive Neuroscience, 18*(5), 689–700.

Sach, A. J., & Bailey, P. J. (2004). Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Perception & Psychophysics, 66*(8), 1379–87.

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews. Neuroscience, 2*(8), 539–550.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403–28.

Salvi, R. J., Lockwood, A. H., Frisina, R. D., Coad, M. L., Wack, D. S., & Frisina, D. R. (2002). PET imaging of the normal human auditory system: responses to speech in quiet and in background noise. *Hearing Research, 170*(1-2), 96–106.

Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine and Biology, 32*(1), 11–22.

Scholtz, A. W., Kammen-Jolly, K., Felder, E., Hussl, B., Rask-Andersen, H., & Schrott-Fischer, A. (2001). Selective aspects of human pathology in high-tone hearing loss of the aging inner ear. *Hearing Research*, *157*(1-2), 77–86.

Schuknecht, H. F., & Gacek, M. R. (1993). Cochlear pathology in presbycusis. *The Annals of Otology, Rhinology, and Laryngology*, *102*(1 Pt 2), 1–16.

Scott, S. K., & Wise, R. J. S. (2003). PET and fMRI studies of the neural basis of speech perception. *Speech Communication*, *41*(1), 23–34.

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, *22*(3), 1060–1075.

Sekihara, K., Sahani, M., & Nagarajan, S. S. (2005). Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. *Neuroimage*, *25*(4), 1056–67.

Shafiro, V., & Gygi, B. (2007). Perceiving the speech of multiple concurrent talkers in a combined divided and selective attention task. *The Journal of the Acoustical Society of America*, *122*(6), EL229–EL235.

Shaw, E. A. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *The Journal of the Acoustical Society of America*, *56*(6), 1848–61.

Sheskin, D. J. (2003). *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL: CRC Press, 3rd ed.

Snell, K. B., & Frisina, D. R. (2000). Relationships among age-related differences in gap detection and word recognition. *The Journal of the Acoustical Society of America*, *107*(3), 1615–26.

Soltani, M., & Knight, R. T. (2000). Neural origins of the P300. *Critical Reviews in Neurobiology*, *14*(3-4), 199–224.

Sommers, M. S. (1997). Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment. *The Journal of the Acoustical Society of America*, *101*(4), 2278–88.

Sommers, M. S., & Humes, L. E. (1993). Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners. *The Journal of the Acoustical Society of America*, *93*(5), 2903–14.

Souza, P. E., & Turner, C. W. (1994). Masking of speech in young and elderly listeners with hearing loss. *Journal of Speech and Hearing Research*, *37*(3), 655–61.

SPSS Inc. (2006). *SPSS (Release 15)*. Chicago, Illinois.

Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. New Jersey: Lawrence Erlbaum.

Stockwell, R., Mansinha, L., & Lowe, R. (1996). Localization of the complex spectrum: The S-Transform. *IEEE Transactions on Signal Processing*, *44*(4), 998–1001.

Stubbs, R. J., & Summerfield, Q. (1990). Algorithms for separating the speech of interfering talkers: evaluations with voiced sentences, and normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *87*(1), 359–72.

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630–637.

Sussman, E. S., Bregman, A. S., Wang, W. J., & Khan, F. J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cognitive, Affective & Behavioral Neuroscience*, *5*(1), 93–110.

Tallon-Baudry, & Bertrand (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162.

Tarantola, A. (2004). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: Society for Industrial and Applied Mathematics.

Teder-Sälejärvi, W., & Hillyard, S. (1998). The gradient of spatial auditory attention in free field: an event-related potential study. *Perception & Psychophysics*, *60*(7), 1228–42.

The MathWorks (2007). *MATLAB ® v7.1*. Cambridge, UK.

Tiitinen, H., Sinkkonen, J., Reinikainen, K., Alho, K., Lavikainen, J., & Näätänen, R. (1993). Selective attention enhances the auditory 40-hz transient response in humans. *Nature*, *364*(6432), 59–60.

Tikhonov, A. N., & Arsenin, V. Y. (1977). *Solutions of Ill-Posed Problems*. Washington D.C.: Winston & Sons.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*, 242–48.

Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, *76*, 282–299.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging, 17*(3), 453–467.

Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C., & Patterson, R. D. (1984). Auditory filter asymmetry in the hearing impaired. *The Journal of the Acoustical Society of America, 76*(5), 1363–8.

Tyler, R. S., Summerfield, Q., Wood, E. J., & Fernandes, M. A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 72*(3), 740–52.

Tzourio, N., Massioui, F. E., Crivello, F., Joliot, M., Renault, B., & Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *Neuroimage, 5*(1), 63–77.

Uauy, R., & Dangour, A. D. (2006). Nutrition in brain development and aging: role of essential fatty acids. *Nutrition Reviews, 64*(5 Pt 2), S24–33; discussion S72–91.

Uchida, Y., Nakashima, T., Ando, F., Niino, N., & Shimokata, H. (2003). Prevalence of self-perceived auditory problems and their relation to audiometric thresholds in a middle-aged to elderly population. *Acta Oto-Laryngologica, 123*(5), 618–26.

van Rooij, J. C., & Plomp, R. (1992). Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion. *The Journal of the Acoustical Society of America, 91*(2), 1028–33.

Van Veen, B. D., van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Bio-Medical Engineering, 44*(9), 867–80.

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: a review of meta-analyses. *Neuroscience and Biobehavioral Reviews, 26*(7), 849–857.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage, 30*(4), 1414–1432.

Vrba, J. (2002). Magnetoencephalography: the art of finding a needle in a haystack. *Physica C: Superconductivity, 368*(1-4), 1–9.

Vrba, J., & Robinson, S. E. (2001). Signal processing in magnetoencephalography. *Methods, 25*(2), 249–71.

Vrba, J., Robinson, S. E., & McCubbin, J. (2004). How many channels are needed for MEG? *Neurology & Clinical Neurophysiology, 99.*

Wang, J. Z., & Kaufman, L. (2003). Magnetic source imaging: Search for inverse solutions. In Z. L. Lu, & L. Kaufman (Eds.) *Magnetic Source Imaging of the Human Brain*, (pp. 101–134). New Jersey: Lawrence Erlbaum Associates.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559.

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *The Journal of the Acoustical Society of America*, *60*(5), 1176–86.

Webster, J. C., & Thompson, P. O. (1954). Responding to both of two overlapping messages. *The Journal of the Acoustical Society of America*, *26*(1), 396–402.

Wightman, F. L., & Kistler, D. J. (1989). Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, *85*(2), 858–67.

Wightman, F. L., & Kistler, D. J. (1992). The dominant role of low-frequency interaural time differences in sound localization. *The Journal of the Acoustical Society of America*, *91*(3), 1648–61.

Wiley, T. L., Cruickshanks, K. J., Nondahl, D. M., Tweed, T. S., Klein, R., & Klein, B. E. (1998). Aging and word recognition in competing message. *Journal of the American Academy of Audiology*, *9*(3), 191–198.

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research, Series A: Physical Sciences*, *2*, 149–168.

Willott, J. F. (1996). Anatomic and physiologic aging: a behavioral neuroscience perspective. *Journal of the American Academy of Audiology*, *7*(3), 141–51.

Willott, J. F., Aitkin, L. M., & McFadden, S. L. (1993). Plasticity of auditory cortex associated with sensorineural hearing loss in adult C57BL/6J mice. *The Journal of Comparative Neurology*, *329*(3), 402–11.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(18), 8722–8726.

Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature*, *406*(6797), 691.

Woodruff, P. W., Benson, R. R., Bandettini, P. A., Kwong, K. K., Howard, R. J., Talavage, T., Belliveau, J., & Rosen, B. R. (1996). Modulation of auditory and visual cortex by selective attention is modality-dependent. *Neuroreport*, *7*(12), 1909–1913.

Woods, D. L. (1992). Auditory selective attention in middle-aged and elderly subjects: an event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, *84*, 456–468.

Woods, D. L., Hillyard, S. A., & Hansen, J. C. (1984). Event-related brain potentials reveal similar attentional mechanisms during selective listening and shadowing. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(6), 761–777.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of Experimental Psychology*, *10*(5), 601–21.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology*, *16*(1), 121–34.

Yonan, C. A., & Sommers, M. S. (2000). The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychology and Aging*, *15*(1), 88–99.

Yost, W. A. (1997). The cocktail party problem: Forty years later. In R. A. Gilkey, & T. R. Anderson (Eds.) *Binaural and Spatial Hearing in Real and Virtual Environments*, (pp. 329–348). Hillsdale, NJ: Erlbaum.

Yost, W. A., Dye, J., R. H., & Sheft, S. (1996). A simulated “cocktail party” with up to three sound sources. *Perception & Psychophysics*, *58*(7), 1026–36.

Zekveld, A. A., Deijen, J. B., Goverts, S. T., & Kramer, S. E. (2007). The relationship between nonverbal cognitive functions and hearing loss. *Journal of Speech, Language, and Hearing Research*, *50*(1), 74–82.

Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, *13*(2), 119–152.