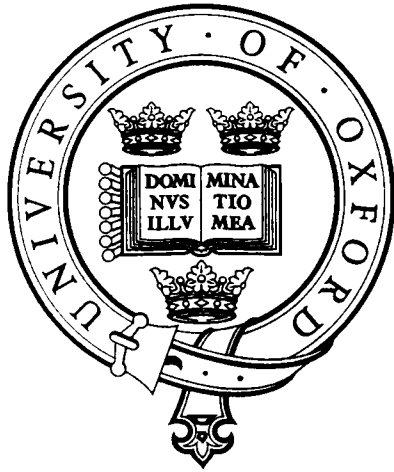




University of Oxford

Medical Sciences Division

University Laboratory of Physiology

Thomas Thesen

Merton College

Multisensory Processing in the Human Brain

**A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy at the University of Oxford**

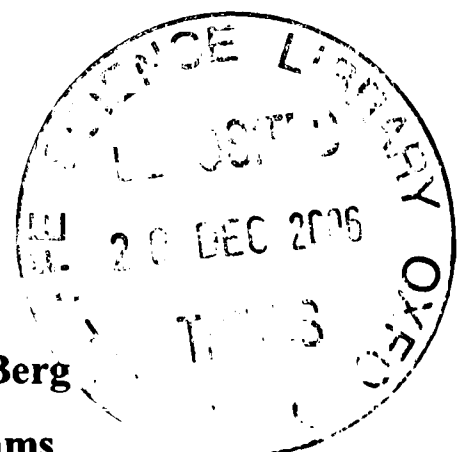
2005

Supervisor:

Dr. Gemma A. Calvert

Internal Examiner: Dr. Heidi Johansen-Berg

External Examiner: Professor Mikko Sams



Abstract

Multisensory Processing in the Human Brain

Thomas Thesen

Merton College

Doctor of Philosophy in Medical Sciences

Michaelmas 2005

Perception has traditionally been studied as a modular function where different sensory systems operate as separate and independent modules. However, multisensory integration is essential for the perception of a coherent and unified representation of the external world that we experience phenomenologically. Mounting evidence suggests that the senses do not operate in isolation but that the brain processes and integrates information across modalities. A standing debate is at what level in the processing hierarchy the sensory streams converge, for example, if multisensory speech information converges first in higher-order polysensory areas such as STS and is then fed back to sensory areas, or if information is already integrated in primary and secondary sensory areas at the early stages of sensory processing. The studies in this thesis aim to investigate this question by focussing on the spatio-temporal aspects of multisensory processing, as well as investigating phonetic and non-phonetic integration in the human brain during auditory-visual speech perception.

Section 1 reviews the literature on multisensory processing, with a special emphasis on bimodal speech perception, and discusses relevant findings from

psychophysics, anatomy, single-cell electrophysiology in animals and functional neuroimaging in humans.

Section 2 discusses the neuroimaging methodologies that have been used in this thesis. A particular strategy of the present investigation includes the use of multiple imaging techniques, namely functional Magnetic Resonance Imaging (fMRI), Magnetoencephalography (MEG), and intra-cranial Electroencephalography (iEEG) to study the spatio-temporal properties of the underlying neurophysiology. This section also reviews the analytic approaches for defining multisensory integration in the brain.

Section 3 introduces the auditory-visual stimuli and experimental conditions that were used for all studies in this thesis. This section also reports an fMRI study that discusses different strategies for defining multisensory integration. The section

Section 5 describes the cortical oscillations during auditory-visual speech processing in specific brain areas. Findings suggest that auditory areas are primed by the preceding visual speech stimulus.

Section 7 describes the cortical oscillations during multisensory speech processing on the superior temporal lobe, which confirm our findings in section 4 that visual speech influences auditory cortex before auditory stimulus onset.

Section 8 compares the results from the different imaging and analysis techniques in a common reference space. Results show a high degree of spatio-temporal overlap between imaging techniques and validate the use of certain hemodynamic criteria for defining multisensory integration.

Finally, section 9 summarizes the findings of this research and suggests a spatio-temporal model of auditory-visual speech processing based on the findings from the preceding sections and which emphasizes the precedence of the visual stimulus in natural speech. A novel categorization of multisensory effects is introduced.

Acknowledgements

I am grateful to my supervisor Dr. Gemma A. Calvert for her support and guidance, and for enabling me to conduct such an exciting research project.

My gratitude goes to my many collaborators whose expertise has helped to shape this thesis.

Special thanks go to Dr. Peter C. Hansen for his invaluable technical advice on fMRI and MEG design and analysis.

I am very grateful to Robert A. Osterbauer for help with fMRI analysis and invaluable scientific discussions.

I am indebted to the researchers at the Wellcome Centre for MEG Studies at the University of Aston in Birmingham, most notably Dr. Ian Holliday, Professor Krish Singh (special thanks for writing MRI3Dx) and Dr. Gareth Barnes for access to the MEG system and for their support and excellent technological expertise.

I am thankful to the researchers and neurosurgeons at the Department for Neurosurgery at the University of Iowa for providing me with the unusual luxury of working directly on the human brain. My special gratitude goes to Dr. Rick Reale, Professor John F. Brugge, Professor Matthew A. Howard III, and Professor Hiroto Kawasaki.

I would also like to thank Professor Ruth Campbell of the Department of Human Communication Sciences at University College London for her expertise on multisensory speech processing.

I am also thankful to the Centre for Functional MRI at the University of Oxford for access to the MRI scanner and for providing an excellent teaching and research environment from which I benefited very much.

I would also like to thank Dr. Robert Oostenveld and Dr. Eric Maris at the FC Donders Centre for Cognitive Neuroimaging at the University of Nijmegen for developing the FieldTrip analysis software toolbox and for fruitful scientific discussions.

My studies at Oxford and the research in this thesis were made possible by the generous financial support of The Wellcome Trust.

I am also thankful to my family for their support and to my friends for providing an important balance.

Thomas Thesen

Oxford, October 2005

Contents

ABSTRACT.....	3
ACKNOWLEDGEMENTS.....	6
CONTENTS.....	7
LIST OF FIGURES	11
LIST OF TABLES	13
GLOSSARY.....	14
1 INTRODUCTION TO MULTISENSORY PERCEPTION AND INTEGRATION..	16
1.1 GENERAL INTRODUCTION	16
1.2 THE AUDITORY SPEECH SYSTEM	18
1.2.1 <i>The peripheral auditory system</i>	18
1.2.2 <i>Anatomy of auditory cortex</i>	20
1.2.3 <i>Functional correlates of auditory cortex activity</i>	21
1.3 VISUAL SYSTEM	23
1.3.1 <i>The peripheral and central visual system</i>	23
1.4 MULTISENSORY INTEGRATION	26
1.4.1 <i>Behavioural studies of non-speech integration</i>	26
1.4.2 <i>Neuroanatomical findings</i>	27
1.4.3 <i>Multisensory integration at the cellular level</i>	28
1.4.4 <i>Electrophysiological evidence in humans</i>	32
1.4.5 <i>The case of auditory-visual speech</i>	33
1.5 CONCLUSIONS.....	37
2 TECHNIQUES FOR IMAGING THE PERFORMING HUMAN BRAIN	38
2.1 STRUCTURAL MAGNETIC RESONANCE IMAGING	40
2.2 FUNCTIONAL MAGNETIC RESONANCE IMAGING	41
2.2.1 <i>fMRI experimental designs</i>	42
2.3 ENCEPHALOGRAPHIC RECORDINGS (EEG/MEG)	43
2.3.1 <i>MEG</i>	44
2.4 INTRACRANIAL EEG.....	46
2.4.1 <i>Source reconstruction from iEEG</i>	47
2.5 ON THE SIMILARITY OF THE fMRI, MEG AND EEG SIGNALS	48
2.6 BRAIN OSCILLATIONS.....	51
2.7 ENCEPHALOGRAPHIC ANALYSIS STRATEGIES FOR MULTISENSORY RESEARCH	54
2.7.1 <i>Event-related averaging</i>	54
2.7.2 <i>Interaction effects in the frequency domain</i>	56
2.7.3 <i>Bimodal contrasts</i>	59
2.8 CONCLUSION	59
3 EVENT-RELATED FMRI OF AUDITORY-VISUAL SPEECH PERCEPTION.61	61
3.1 INTRODUCTION	61
3.2 METHODS	65
3.2.1 <i>Stimuli</i>	65
3.2.2 <i>Task</i>	69
3.2.3 <i>Subjects</i>	70
3.2.4 <i>Procedure</i>	70
3.2.5 <i>Scanning Protocol</i>	71
3.2.6 <i>Data Analysis</i>	71
3.2.6.1 <i>Statistical criteria for defining multisensory integration with fMRI</i>	73
3.3 RESULTS	74
3.3.1 <i>Behavioural</i>	74

3.3.2	<i>Auditory speech perception</i>	74
3.3.3	<i>Speechreading</i>	75
3.3.4	<i>Multisensory speech perception</i>	76
3.3.5	<i>Multisensory integration</i>	76
3.3.5.1	Non-phonetic interactions.....	76
3.3.5.2	$[A \cap V]$	77
3.3.5.2.1	$A > V$	80
3.3.5.2.2	$V > A$	80
3.3.5.3	$[AVcon > A \cap AVcon > V]$	81
3.3.5.4	$AVcon > (A + V)$	82
3.3.5.5	Sub-additivity.....	84
3.3.5.6	$AVcon > \text{mean}(A, V)$	86
3.3.6	<i>Phonetic interactions</i>	87
3.3.7	<i>Crossmodal activation of primary sensory cortex</i>	89
3.3.8	<i>Laterality</i>	92
3.3.8.1	ROI analyses.....	92
3.3.8.2	Laterality results.....	94
3.4	DISCUSSION.....	95
3.4.1	<i>Summary</i>	95
3.4.2	<i>Detailed discussion</i>	97
3.4.2.1	Auditory speech.....	97
3.4.2.2	Criteria for identifying multisensory areas.....	97
3.4.2.2.1	$A \cap V$	98
3.4.2.2.2	$AVcon > A$ AND $AVcon > V$	98
3.4.2.2.3	Super-additivity.....	99
3.4.2.2.4	Subadditivity.....	101
3.4.2.2.5	$AVcon > \text{mean}(A, V)$	101
3.4.2.3	Phonetic interactions.....	102
3.4.2.4	Laterality.....	102
3.4.3	<i>Crossmodal influence of primary sensory cortex</i>	104
3.4.3.1	Speechreading.....	104
3.4.3.2	Auditory speech.....	106
3.4.3.2.1.1	Different rules for cross-modal recruitment of primary cortices.....	107
3.5	CONCLUSION.....	108
4	THE TIME-COURSE OF MULTISENSORY SPEECH PERCEPTION.....	110
4.1	INTRODUCTION.....	110
4.2	METHODS.....	114
4.2.1	<i>Design</i>	114
4.2.2	<i>MEG Data Acquisition</i>	115
4.2.3	<i>Data analysis</i>	115
4.2.3.1	Source analysis.....	115
4.2.3.2	Independent MEG dipole source localization.....	116
4.2.3.3	FMRI-constraint dipole modelling.....	117
4.3	RESULTS.....	117
4.3.1	<i>Neuromagnetic responses to AV speech</i>	117
4.3.2	<i>Magnetic sources to AV speech</i>	119
4.3.3	<i>FMRI constrained modelling</i>	123
4.3.3.1.1	The timing of cortical events.....	124
4.3.3.1.1.1	Timing of auditory speech.....	126
4.3.3.1.2	Timing of crossmodal convergence.....	126
4.3.3.1.3	Timing of phonetic convergence.....	127
4.4	DISCUSSION.....	129
4.4.1	<i>General findings</i>	129
4.4.2	<i>Detailed discussion</i>	129
5	THE SUPERIOR TEMPORAL LOBE DURING BIMODAL SPEECH PERCEPTION.....	137
5.1	INTRODUCTION.....	137
5.2	METHODS.....	139
5.2.1	<i>Human subjects</i>	139
5.2.2	<i>Acoustic calibration and stimulus presentation</i>	141
5.2.3	<i>Electrophysiological recordings</i>	142
5.2.4	<i>Data analysis</i>	143

5.2.4.1	Spatio-temporal maps.....	143
5.2.4.2	Heschl's' gyrus	144
5.3	RESULTS	144
5.3.1	<i>Electrode grid locations</i>	144
5.3.1.1	S106	144
5.3.1.2	S100	144
5.3.2	<i>Event-related potentials</i>	145
5.3.2.1	Responses to auditory speech.....	147
5.3.2.1.1	S106	147
5.3.2.1.2	S100	149
5.3.2.2	Interactions between auditory and visual speech.....	150
5.3.2.2.1	S106	151
5.3.2.2.2	S100	152
5.3.3	<i>Phonetic interactions</i>	154
5.3.3.1	AVcon vs AVincon(V)	154
5.3.3.1.1	S106	154
5.3.3.1.2	S100	155
5.3.3.2	AVcon vs AVincon(A)	155
5.3.3.2.1	S106	155
5.3.3.2.2	S100	156
5.4	DISCUSSION	157
6	NEUROMAGNETIC FREQUENCY RESPONSES	162
6.1	INTRODUCTION	162
6.2	METHODS	165
6.2.1	<i>Subjects</i>	165
6.2.2	<i>SAM analysis</i>	165
6.2.3	<i>Virtual electrode locations</i>	167
6.3	STATISTICAL ANALYSIS	167
6.4	RESULTS	168
6.4.1	<i>Cortical power changes in AI</i>	168
6.4.2	<i>Cortical power changes in STS</i>	169
6.4.3	<i>Cortical power changes in Broca's area</i>	169
6.4.4	<i>Phonetic interactions</i>	172
6.4.5	<i>Cortical power changes in visual cortex</i>	174
6.5	DISCUSSION	175
7	BRAIN OSCILLATIONS ON THE SUPERIOR TEMPORAL LOBE DURING AV SPEECH PERCEPTION	179
7.1	INTRODUCTION	179
7.2	METHODS	181
7.2.1	<i>Subjects</i>	181
7.2.2	<i>Data analysis</i>	181
7.2.2.1	Time-frequency spectrograms.....	181
7.3	RESULTS	183
7.3.1	<i>Sensory responses and non-phonetic modulations</i>	183
7.3.1.1	Alpha band.....	183
7.3.1.1.1	S106	183
7.3.1.1.2	S100	185
7.3.1.2	Beta band	185
7.3.1.2.1	S106	185
7.3.1.2.2	S100	187
7.3.1.3	Gamma band	188
7.3.1.3.1	S106	188
7.3.1.3.2	S100	189
7.3.1.4	Omega band	190
7.3.1.5	S106	190
7.3.1.5.1	S100	190
7.3.2	<i>Phonetic modulations</i>	192
7.3.2.1	AVcon vs AVincon(V)	192
7.3.2.1.1	S106	192
7.3.2.1.2	S100	193
7.3.3	<i>Visual cuing in the Omega band</i>	194
7.3.4	<i>Coherence</i>	195
7.4	DISCUSSION	195
7.4.1	<i>Quantification of synchronization in the superior temporal lobe</i>	195
7.4.1.1	Alpha band.....	195

7.4.1.2	Gamma band	198
7.4.1.3	Omega band	198
7.4.1.4	Individual differences.....	199
7.4.1.5	Visual cuing	199
8	COMPARING RESULTS FROM DIFFERENT IMAGING MODALITIES.....	201
8.1	INTRODUCTION	201
8.2	SPATIAL COREGISTRATION.....	204
8.2.1.1	Functional MRI to structural MRI.....	204
8.2.1.2	MEG sensor space to structural MRI	205
8.2.1.3	IEEG to structural MRI	206
8.3	RESULTS	207
8.3.1	<i>Auditory (A)</i>	207
8.3.2	<i>Auditory-visual (Avcon)</i>	210
8.3.3	<i>Electrophysiological validation of hemodynamic integration criteria</i>	210
8.4	DISCUSSION	213
9	SUMMARY & CONCLUSIONS	214
9.1	SUMMARY OF MAIN RESULTS	215
9.1.1	<i>FMRI</i>	215
9.1.2	<i>MEG dipole analysis:</i>	216
9.1.3	<i>MEG frequency analysis:</i>	216
9.1.4	<i>IEEG ERP analysis:</i>	217
9.1.5	<i>IEEG frequency analysis:</i>	217
9.1.6	<i>Combining neuroimaging methods</i>	217
9.2	A SPATIO-TEMPORAL MODEL	218
9.2.1	<i>Spatio-temporal progression of neural activity</i>	218
9.2.2	<i>Visual precedence in AV speech processing</i>	220
9.2.3	<i>Auditory cortex during speechreading</i>	221
9.2.4	<i>Feedback or feedforward?</i>	223
9.2.5	<i>A proposal for the categorization of multisensory effects</i>	224
9.2.5.1	Different mechanisms serve crossmodal effects.....	224
9.2.5.2	Changes in neuronal network behaviour	228
9.3	SUGGESTIONS FOR FURTHER STUDIES	229
	BIBLIOGRAPHY	231
	APPENDIX.....	269

FIGURE 1-1: SUPERIOR TEMPORAL CORTEX OF THE HUMAN AND MACAQUE BRAIN.....21

FIGURE 1-2: PROJECTIONS OF SENSORY PATHWAYS25

FIGURE 1-3: SCHEMATIC DRAWINGS OF THE HUMAN BRAIN FROM THREE PERSPECTIVES.....31

FIGURE 2-1: COMPARISON OF TEMPORAL AND SPATIAL RESOLUTION OF FMRI, MEG & IEEG39

FIGURE 2-2: RELATIONSHIP BETWEEN CURRENT SOURCE AND THE ELECTRIC/MAGNETIC SIGNALS48

FIGURE 2-3: INTERACTION EFFECTS DURING PERFECT PHASE RELATIONSHIP.....57

FIGURE 2-4: INTERACTION EFFECTS DURING ANTI-PHASIC RELATIONSHIP58

FIGURE 3-1: SPECTROGRAM OF CV SOUND /DA/.67

FIGURE 3-3: TIME-COURSE OF STIMULUS PRESENTATION.....69

FIGURE 3-4: FMRI ACTIVATION TO A, V & AVCON75

FIGURE 3-5: FMRI ACTIVATION TO A, V & A∩V78

FIGURE 4-1: AVERAGED MEG DATA.....118

FIGURE 4-2: STRATEGY FOR DIPOLE SOURCE LOCALIZATION119

FIGURE 4-3:UNCONSTRAINED DIPOLE FIT (DIPOLE MOMENTS)122

FIGURE 4-4: FMRI CONSTRAINED DIPOLE FIT (DIPOLE MOMENTS).....124

FIGURE 4-5: TIMING OF DIPOLE ACTIVATION.....125

FIGURE 4-6: TIMING OF PHONETIC INTERACTIONS.....128

FIGURE 5-1: ELECTRODE GRID PLACEMENT DURING SURGERY.....141

FIGURE 5-2: LOCATION OF HESCHL’S GYRUS.....146

FIGURE 5-3: AVERAGED IERPs.....146

FIGURE 5-4: ACTIVATION TO AUDITORY SPEECH IN S106.....148

FIGURE 5-5: ACTIVATION TO AUDITORY SPEECH IN S100150

FIGURE 5-6: STATISTICAL COMPARISONS OF A VS AV IN S106.....152

FIGURE 5-7: STATISTICAL COMPARISONS OF A VS AVCON IN S106153

FIGURE 5-8: STATISTICAL COMPARISONS OF AVCON VS AVINCON(V) IN S106.....154

FIGURE 5-9: STATISTICAL COMPARISONS OF AVCON VS AVINCON(V) IN S100.....155

FIGURE 5-10: STATISTICAL COMPARISONS OF AVCON VS AVINCON(A) IN S106.....156

FIGURE 5-11: STATISTICAL COMPARISONS OF AVCON VS AVINCON(A) IN S100.....157

FIGURE 6-1: LOCATIONS OF VIRTUAL SAM ELECTRODES168

FIGURE 6-2: ERD/ERS GROUP DATA FOR A, AVCON AND V	170
FIGURE 6-3: TIME-POWER PLOTS OF LEFT A1 & STS	171
FIGURE 6-4: VISUAL CUEING IN LEFT A1.....	172
FIGURE 6-5: ERD/ERS GROUP DATA FOR AVCON AND AVINCON(V)	173
FIGURE 6-6: 20 HZ OSCILLATIONS IN RIGHT A1	174
FIGURE 6-7: ERD/ERS IN VISUAL CORTEX	175
FIGURE 7-1: ALPHA ERD/ERS IN S106	184
FIGURE 7-2: ALPHA ERD/ERS IN S100	185
FIGURE 7-3: BETA ERD/ERS IN S106.....	186
FIGURE 7-4: BETA ERD/ERS IN S106.....	187
FIGURE 7-5: GAMMA ERD/ERS IN S100	188
FIGURE 7-6: GAMMA ERD/ERS IN S100	189
FIGURE 7-7 OMEGA ERD/ERS IN S106.....	190
FIGURE 7-8: OMEGA ERD/ERS IN S106	191
FIGURE 7-9: AVCON VS AVINCON CONTRAST FOR S106	192
FIGURE 7-10: AVCON VS AVINCON CONTRAST FOR S106	193
FIGURE 7-11: VISUAL CUEING IN THE OMEGA BAND	194
FIGURE 8-1: COMPARING FMRI, MEG & IEEG (S106) RESULTS (A).....	209
FIGURE 8-2: COMPARING FMRI, MEG & IEEG RESULTS (AVCON).....	210
FIGURE 8-3: SPATIAL CONCORDANCE OF FMRI & IEEG INTEGRATION EFFECTS.....	212
FIGURE 9-1: A SPATIO-TEMPORAL MODEL OF AV PROCESSING	220

List of Tables

TABLE 3-1: AREAS OF ACTIVATION FOR $[AVCON > A \cap AVCON > V]$ 82

TABLE 3-2: AREAS OF ACTIVATION FOR $[AVCON < (A+V)]$ 85

TABLE 3-3: AREAS OF ACTIVATION FOR $[AVCON > MEAN(A, V)]$ 87

TABLE 3-4: AREAS OF ACTIVATION TO FOR $[AVINCON(A) > AVCON]$ 89

TABLE 4-1: UNCONSTRAINED DIPOLE FIT (DIPOLE LOCATIONS)..... 121

TABLE 4-2: FMRI CONSTRAINED DIPOLE FIT (DIPOLE LOCATIONS) 123

TABLE 5-1: IEEG SUBJECT DEMOGRAPHICS 140

Glossary

A	auditory only condition
AV	auditory-visual
AVcon	congruent auditory-visual condition
BOLD	blood oxygen level dependent
CNV	contingent negative variation
CV	consonant-vowel
dB	decibels
ECD	equivalent current dipoles
EEG	electroencephalography
EPI	echo-planar imaging
ER	event-related
ERD	event-related desynchronization
ERP	event-related potential
ERS	event-related synchronization
fMRI	functional magnetic resonance imaging
iEEG	intracranial EEG
IFG	inferior frontal gyrus
IP	intraparietal area
IPS	intra-parietal sulcus
ISI	interstimulus intervals
ITG	inferior temporal gyrus
LFP	local field potential
LGN	lateral geniculate nucleus
LI	laterality index
M1	primary motor cortex
MEG	magnetoencephalography
MMN	mismatch-negativity
MRI	magnetic resonance imaging
MT	motion-sensitive visual area (V5)
MTG	middle temporal gyrus
PET	positron emission tomography
PLST	posterior lateral superior temporal area
RF	radio frequency
RMS	root square mean
ROI	regions of interest
RT	reaction time
RTE	redundant target effect
S1	somatosensory cortex
SAM	Synthetic Aperture Magnetometry
SC	superior colliculus
SEM	standard error of the mean
SQUID	superconducting quantum interference device
STG	superior temporal gyrus
STP	supratemporal cortex
STS	superior temporal sulcus
thal	thalamus

TMS	transcranial magnetic stimulation
Tpt	temporo-parietal area
TR	repetition time
V	visual only condition
V1	primary visual cortex
V2	secondary visual cortex
V5	motion-sensitive visual area (MT)
VE	virtual electrode

1 Introduction to multisensory perception and integration

A major aim of this thesis is to investigate the processing of auditory and visual speech and how the two sensory streams interact to integrate information into a single coherent percept. This section provides an overview of the structural and functional anatomy of the auditory and visual systems, and discusses previous findings on multisensory integration of hearing and vision. We describe results from behavioural experiments in humans, single-unit recordings in animals, and more recent findings in humans from non-invasive neuroimaging techniques, such as EEG, MEG, PET and fMRI. We also discuss experiments specifically targeted towards the understanding of auditory-visual speech processing, which is complimented by more in-depth discussions in the relevant sections of this thesis.

The sum of the parts is not the whole!

Lao Tse (B.C. 600), 39th saying in Tao-te-King

1.1 General Introduction

The human brain samples its surroundings through a number of distinct peripheral sensors, each sensitive to very different forms of energy and chemicals in the environment. These signals are transduced at the peripheral level into neural impulses that are further relayed to the central nervous system, each sense separately. In the brain, signals from the individual sensory streams are processed simultaneously in a modality-specific fashion to create a sensory representation of the external world.

However, the brain also possesses synergistic properties to detect changes in the environment optimally and resolve ambiguity in a most adaptable fashion. In humans, the burgeoning psychological literature on this topic demonstrates the

existence of such multisensory effects across many sensory modalities. Investigations to date have mainly focused on interactions between our undoubtedly dominant and most studied sense, vision, with other modalities including audition, touch and olfaction, leading to the discovery of many perceptual alterations and behavioural performance changes associated with multisensory stimulus processing (Thesen et al., 2004).

Descriptions of the underlying physiological mechanisms by which the human brain accomplishes these tasks has largely been restricted to analogies to animal studies derived from invasive electrophysiological recordings of individual neurons (Stein & Meredith, 1993; King & Palmer, 1985). However, recent developments in non-invasive neuroimaging, such as functional magnetic resonance imaging (fMRI) (Jezzard, Matthews, & Smith, 2001), positron emission tomography (PET) (Haxby, Grady, Ungerleider, & Horwitz, 1991), and electro- and magnetoencephalography (EEG, MEG) (Regan, 1989) have contributed to our understanding of multisensory processing in the human brain. However, the cognitive neuroscience of auditory-visual (AV) integration is still in its beginning and many of the basic mechanisms are poorly understood. For example, the way by which the brain integrates information from different sensory streams remains unsolved and even the brain areas implicated in this task remain controversial (Calvert and Thesen, 2004). Further evidence for a differential involvement of brain areas and neuronal mechanisms of crossmodal integration for different stimuli types and tasks, such as speech and spatial localization (Calvert, 2001; Driver and Spence, 2000) strongly suggests that any theory of multisensory integration depends on such a context.

We have chosen, in this thesis, to focus on multisensory speech perception, which provides a special case of auditory-visual (AV) integration, as the speech sound and the movements of visible articulatory structures, such as jaw and lips, which create the sound, are inherently correlated and these associations have been over-

learned from early childhood (Bernstein et al., 2004; Massaro, 2004). We are specifically interested in characterizing which brain areas are recruited during AV speech processing and which cortical areas mediate the integration of both sensory streams. Of special interest in this respect is the role of auditory cortex in silent speech perception (speechreading), a question that remains controversial (Calvert et al., 1997; MacSweeney et al., 2000; Callan et al., 2001; Bernstein et al., 2002, Pekkola et al.,).

Another contentious issue in the literature is the exact timing of both crossmodal recruitment of primary sensory cortex and multisensory integration (Calvert, 2001; Calvert and Thesen, 2004). This thesis hopes to describe the exact time course of multisensory processing using electrophysiological methods with millisecond temporal resolution. We will also attempt to describe brain oscillations involved in auditory and visual speech processing, and their role in sensory integration, a particular area of neuroscience that has received only moderate attention (Pfurtscheller and Lopes da Silva, 2005).

In order to understand how the brain processes and integrates information from different sensory modalities it is useful to first give some background knowledge. The aim of the present section is therefore to discuss the structural and functional anatomy of the individual sensory systems, and to elaborate what is known about multisensory integration so far. Where relevant, this information is complemented in subsequent sections by additional references.

1.2 The auditory speech system

1.2.1 The peripheral auditory system

The human auditory system analyzes sound according to changes in frequency and intensity over time. To accomplish this task, the auditory input undergoes several transduction and computational processing steps. The pressure waves of sounds first

reach the outer ear and are funnelled by the pinna into the ear canal, where they are converted to mechanical vibrations by the tympanic membrane (ear drum) and a series of small bones, the ossicular chain, leading to the cochlea of the inner ear. In the cochlea, a snail-shaped structure within the temporal bone of the skull, information is further converted from mechanical vibrations to vibrations in the perilymphic fluid with which the cochlea is filled. Within the cochlea, inside the cochlear duct and placed in the organ of Corti, are the sensory receptors for hearing. Travelling waves within the cochlear duct cause a shearing of hair cells which results in an electrochemical excitation of the nerve fibres attached to the hair cells (Ashmore, 1987). Already at the level of the cochlea an analysis of complex sounds into component frequencies is accomplished and stimulus intensity is preserved. The auditory nerve, or cranial nerve VII, consists of a bundle of about 30,000 nerve fibres connected to the hair cells in the cochlea (Young and Sachs, 1979). The frequency analysis performed by the organ of Corti is further refined by lateral inhibition, a process which inhibits the response of surrounding cells and nerve fibres of a stimulated place and thereby sharpens the effect. Axons from the cochlea terminate in the cochlear nuclear complex at the junction of pons and medulla (Shofner and Young, 1985). Post-synaptic neurons in this structure project to other centres in the brain via three pathways: the trapezoid body, dorsal acoustic stria, and intermediate acoustic stria. Amongst other projections, postsynaptic axons from the superior olivary nucleus project to the central nucleus of the inferior colliculus in the midbrain via the lateral lemniscus (Oertel, 1991). From there projections are sent to the deep layers of the adjacent superior colliculus (SC) and the medial geniculate nucleus of the thalamus, which in turn projects directly to the primary auditory cortex (A1) of the temporal lobe (Hudspeth, 2000).

1.2.2 Anatomy of auditory cortex

Initial processing of auditory information in the cerebral cortex occurs in the primary or core auditory cortex. In humans, this area lays approximately on the caudo-medial half of the transverse temporal gyrus, also called Heschl's Gyrus (HG) in the superior temporal lobe (Rademacher et al., 1993; Rademacher et al., 2001), and corresponds to area 41 of Brodmann (Brodmann, 1908). Its comparative counterparts are A1 (Hackett et al., 1998); (Kaas and Hackett, 1998) and KA (auditory koniocortex, (Pandya and Sanides, 1973; Pandya and Rosene, 1993) in various species of monkeys. The primary auditory cortex is the only cortical structure to receive dense input from the main thalamic relay, the ventral medial geniculate nucleus (Rauschecker et al., 1997). Surrounding A1 rostrally, posteriorly and laterally is the auditory belt region comprising the secondary auditory cortex (A2), which is equivalent with BA 42 (see Figure 1-1). A2 receives input both transcortically from A1 and thalamo-cortically from the ancillary nuclei of the medial geniculate and the medial pulvinar (Rauschecker et al., 1995; Hackett et al., 1998). [REMOVED REPITION]

Extending over the lateral surface of the superior temporal gyrus (STG) is the parabelt area which surrounds A1 and A2 and corresponds to BA 22. This portion of the STG receives extensive projections from secondary auditory cortex (A2) (Hackett et al., 1998); (Luethke et al., 1989). Another distinct auditory responsive area has been identified by projections from parabelt regions to the superior temporal sulcus (STS) (Seltzer and Pandya, 1994; Kaas and Hackett, 2000). In humans, the planum temporale (PT) is situated immediately posterior to HG (Shapleske et al., 1999) and coincides with part of Wernicke's area. PT is believed to consist cytoarchitectonically of secondary auditory cortex. Posteriorly, area Tpt differs in cytoarchitecture and is located in the medial posterior supratemporal cortical plane, at its junction with the inferior parietal lobe. This area represents a transition between auditory association

cortex and the inferior parietal lobule, and compromises the posterior portion of BA 22.

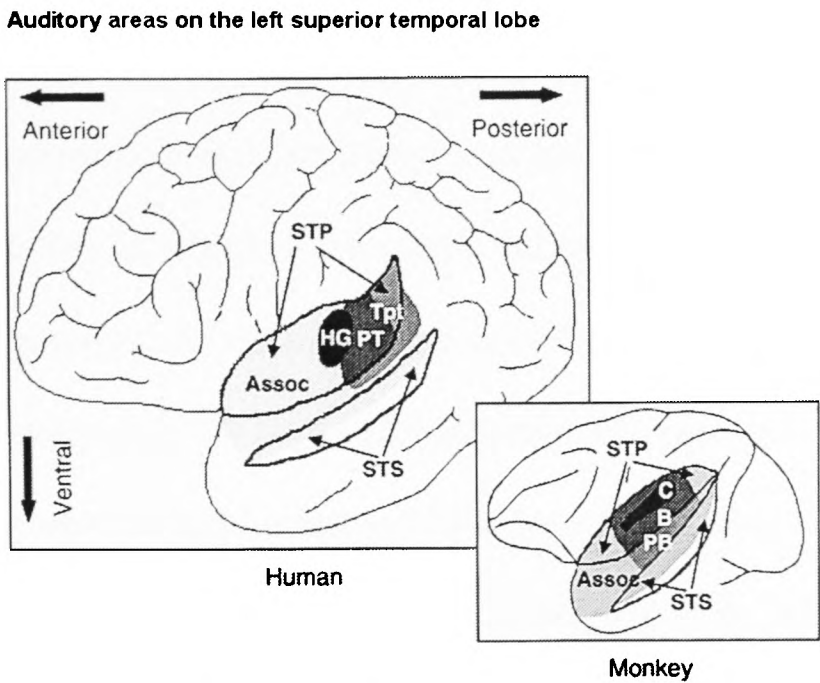


Figure 1-1: Superior temporal cortex of the human and macaque brain. The plane of the supratemporal cortex (STP) and the inside of the superior temporal sulcus STS are exposed. Human brain: HG = Heschl's gyrus (including primary auditory cortex); Tpt = supratemporal cortex posterior to HG; PT = planum temporale, part of the supratemporal cortical plane immediately posterior to HG (Shapleske et al., 1999); Assoc = auditory association cortex lateral and anterior to the previous three regions. Monkey brain: C = core (primary auditory cortex); B = belt; PB = parabelt; Assoc = auditory association cortex surrounding the previous three regions. Adapted from (Wise et al., 2001).

1.2.3 Functional correlates of auditory cortex activity

Electrophysiological studies in monkeys have shown a tonotopic organization of A1, where best frequencies are represented in strict spatial order (Merzenich and Brugge, 1973). Similar evidence for such processing of pitch has also been found by MEG (Pantev et al., 1988), PET (Lauter et al., 1985) and fMRI (Bilecen et al., 1998). However, other features such as intensity tuning, sensitivity, response latency and binaural interactions have been shown in monkeys (Recanzone et al., 1999). Evidence from human imaging studies suggest that A1 does not show any specificity for speech

stimuli and responds to speech and non-speech stimuli in an equal fashion (Binder et al., 2000; Scott et al., 2000). Regions of the belt, or A2, do not show a strict tonotopic organization (Merzenich and Brugge, 1973) but are more selective to complex sounds, such as species-specific calls. for example (Rauschecker et al., 1995). Processing of acoustic features becomes increasingly specialized during the flow of auditory information from lower to higher areas (Rauschecker and Tian, 2000). The STS seems, at least partially, involved in the acoustic analysis of speech sounds (Binder et al., 2000). The PT in humans is larger on the left than on the right (Geschwind and Levitsky, 1968), for a recent review, see (Shapleske et al., 1999), which has been attributed to the dominance of the left hemisphere to speech and language processing (Geschwind and Galaburda, 1985). However, not all studies have found such structural asymmetries (Westbury et al., 1999) nor is the speech-specificity of the left PT unambiguously established (Binder et al., 1996). Recent functional imaging studies have found increasing evidence for a strong role of the lateral auditory projections of STG and STS in speech processing.

Whereas neurons in the primary auditory cortex respond best to pure tones of a specific frequency (Kosaki et al., 1997), it is only in areas of the lateral projections that neurons respond to more complex acoustic features inherent in vocalizations (Rauschecker et al., 1995; Kosaki et al., 1997). Besides from the medial-lateral distinction of supratemporal cortex (STP), the anterior-posterior axis has also distinct functional significance (Rauschecker, 1998; Kaas and Hackett, 1999). The posterior STP, for example, responds to speech per se, whereas the anterior portion only responds to intelligible speech (Scott et al., 2000).

Comparing speech stimuli with non-speech sounds of similar acoustic characteristics, neuroimaging studies have consistently found an enhanced recruitment of left STG and STS regions. These results suggest that sub-lexical speech analysis is a higher-order auditory process which occurs relatively late in the

processing hierarchy and involves auditory areas of the belt and the multisensorially responsive posterior STS. However, the issue whether this reflects speech-specific processing is still unresolved as recent evidence suggests that the left STS is specialized in general processing of rapidly changing acoustic stimuli, a category of which consonant-vowels sounds are a part, and not only in speech-specific processing (Zatorre and Belin, 2001). Furthermore, recent evidence by Belin et al. (2002), who found that speech vocal sounds induced greater response than nonspeech vocalizations suggests that primary auditory cortex has a far greater involvement in speech processing than was previously thought.

1.3 Visual system

1.3.1 The peripheral and central visual system

As light hits the retina, visual information is transduced via two types of photoreceptors (cones & rods) into neural impulses. Cones have a very high density in the center of the retina (fovea) and contain three different photo pigments, whereas rods only contain one photo pigment and are more numerous in the segments of the retina corresponding to the peripheral visual field. Rods are also more sensitive to light. Thus, cones are largely responsible for day-time high acuity and color vision in the center of the visual field, while rods are used under night-time scotopic lightening conditions and for peripheral vision. Before information exists the retina through the optic nerve, the first processing of visual information has already occurred at the level of the retina, which is still considered to be part of the peripheral visual system. The retina sends axons via the retinofugal projection to the optic chiasm, where visual information from the nasal retina crosses to the other side, and form the optic tract, which runs just under the pia along the lateral surface of the diencephalon. A small number of neurons form synaptic connections with cells in the hypothalamus, whereas ~ 10% bypass the thalamus to innervate the superior colliculus in the midbrain. Most

of the optic tract neurons, however, innervate the lateral geniculate nucleus (LGN) in the dorsal thalamus, from where they project to the primary visual cortex (V1), which is the sole recipient of magno -and parvocellular layers of the LGN (Felleman and Van Essen, 1991). Like the auditory core, V1 codes for basic stimulus features and has a detailed retinotopic mapping of visual space, with neurons sensitive to orientation, movement, spatial frequency, wavelength and luminance (Hubel and Wiesel, 1968; Hubel and Wiesel, 1977; Cavanaugh et al., 2002; Lee, 2003).

Visual association areas, such as V2, V4, and V5 have direct connections with V1, and are functionally more specialized. For example, V5 or area MT is a region in extrastriate visual cortex which is heavily myelinated and specialized for visual motion processing (Allman and Kaas, 1971; Newsome et al., 1990), and thus relevant for visual speechreading. In monkeys, MT is located in the posterior bank of the STS (Zeki, 1983) and its analogue in humans is situated near the junction of the inferior temporal sulcus and the lateral occipital sulcus (Watson et al., 1993). MT comprises of several motion-sensitive areas which are often grouped together under the collective term MT+ (Beauchamp et al., 1997). Even though there is evidence for attention-related enhancement of sign-language processing in deaf individuals in MT, speech-specific neural coding is likely to occur at a higher visual processing level involving more complex visual object coding, such as faces. This occurs in the fusiform, inferior temporal and middle temporal gyri, the so-called ventral visual stream related to object ("What") recognition (Ungerleider and Haxby, 1994). Face-selective activations to complex visual objects are observed in the lateral occipital complex on the lateral bank of the fusiform gyrus (Puce et al., 1996; Halgren et al., 1999). Face processing in the fusiform gyrus seems to be general to faces as a category and does not differentiate between the identities of different faces, whereas face recognition might rely on the additional retrieval of semantic knowledge (Halgren et al., 1999).

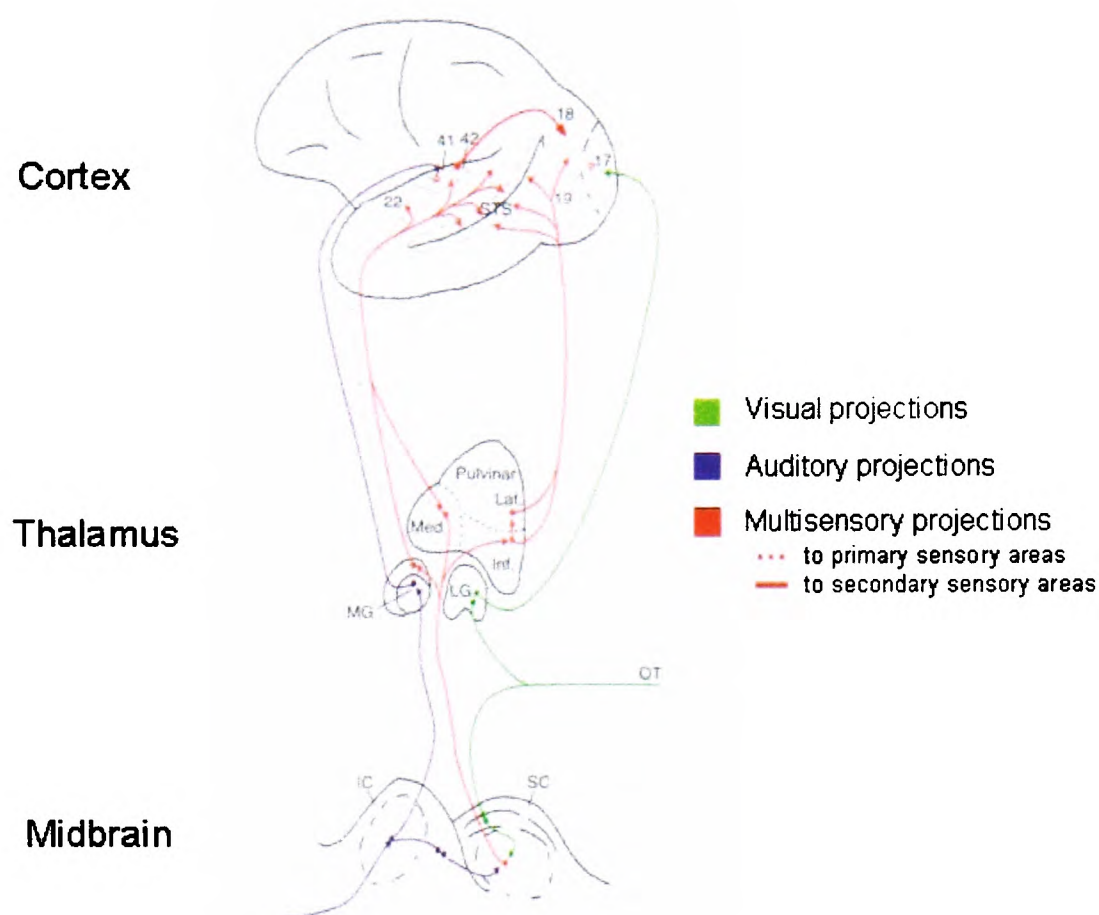


Figure 1-2: Projections of sensory pathways

Based on tracing studies in monkeys.

Green arrows indicate visual projections from the retina via the optic tract (OT) into the primary visual nucleus of the thalamus, the lateral geniculate nucleus (LGN). The LGN sends major projections into the primary visual cortex (area 17) in the occipital lobe. The OT also sends sparse projections into the superficial layers of the superior colliculus (SC) in the midbrain and from there to deeper layers of SC, where auditory and visual signals are integrated (Stein and Meredith, 1993).

Blue arrows show the ascending auditory pathway through the brain stem into the central nucleus of the inferior colliculus (IC). From there projections reach the primary auditory nucleus of the LGN, the medial geniculate (MG). From there dense projections reach both core and belt regions of the auditory cortex (BA 41, 42). Sparser projections are sent from the MG into the external regions of the SC from where they innervate deeper layers of SC.

Red arrows show projections from multisensory convergence zones in the deep layers of SC into primary and secondary sensory areas. Projections enter the non-primary nuclei of the thalamus. The lateral and inferior pulvinar

nuclei project to layer 4 (input layer) of higher visual areas (areas 18 and 19), while the medial pulvinar and medial geniculate nuclei project to higher level auditory areas (areas 42 & 22). Also seen are projections from core auditory areas (all straight red lines) to visual cortex serving the visual periphery (area 18). Shown in dotted red lines are projections into primary visual cortex (area 17) and primary (core) auditory cortex (area 41). Adapted and modified from (Bernstein et al., 2004).

1.4 Multisensory integration

1.4.1 Behavioural studies of non-speech integration

In an early study of crossmodal phenomena, Todd (1912) demonstrated that reaction time (RT) in a target detection task can be speeded by the presence of a non-specific accessory stimulus in another modality, i.e. a stimulus that bears no meaningful relationship other than temporal proximity. Subsequent investigations into this crossmodal 'redundant target effect' (RTE) have replicated and extended these findings (Hershenson, 1962; Bernstein et al., 1969; Gielen et al., 198; Doyle and Snowden, 2001; Schroger and Widmann, 1998) and provided further evidence that the observed crossmodal facilitation is not simply due to a statistical probability summation effect alone (Gondan et al., 2002; Molholm et al., 2002). Consequently, "race models" of the RTE that sought to explain the phenomenon on the basis of a probabilistic interpretation, have been largely superseded by "co-activation models" (Miller, 1991) in which signals from the different sensory channels are integrated prior to initiation of the motor response.

Behavioural studies have also explored the conditions under which crossmodal interactions occur. Two key determinants of intersensory binding are synchronicity and spatial correspondence (Radeau, 1994). Thus, when two or more sensory stimuli occur at the same time and place, they are typically bound into a single percept and detected more rapidly than either input alone. In contrast, slight discrepancies in the

onset and location of two crossmodal cues can be significantly less effective in eliciting responses than isolated unimodal stimuli (Stein et al., 1989; Sekuler et al., 1997). Similar instances of crossmodal facilitation have also been shown to effect detection thresholds. For example, Frassinetti et al. (2002) found that subject's sensitivity to visual stimuli presented below luminance threshold was increased by a simultaneous accessory sound burst presented at the same spatial location. This effect was eliminated when the two sensory inputs were separated in space or offset by more than 500ms. Similar crossmodal influences have also been reported in the case of auditory and tactile detection thresholds (for reviews see Loveless et al., 1970; Welch and Warren, 1986).

1.4.2 Neuroanatomical findings

A widely held conception of the human cortex is the distinction of areas into primary and secondary sensory unimodal cortices and heteromodal "association" cortices. These heteromodal zones were defined on the basis that they were found to receive converging afferents from multiple sensory modalities and contained neurons responsive to stimulation in more than one modality. A large number of such areas have been identified (see Figure 1-3). These include anterior portions of the STS (Benevento et al., 1977; Desimone and Gross, 1979; Bruce et al., 1981; Baylis et al., 1987; Watanabe and Iwai, 1991) posterior portions of the STS, including the temporo-parietal association cortex (Leinonen et al., 1980; Desimone and Ungerleider, 1986) parietal cortex, including the ventral and lateral intraparietal areas (IP) (Linden et al., 1999; Lewis and Van Essen, 2000; Bremmer F. et al., 2001), and premotor and prefrontal cortex (Watanabe, 1992; Graziano et al., 1999). Multisensory convergence zones have also been identified in sub-cortical structures, including the superior colliculus (SC) (Fries, 1984), the claustrum (Pearson et al., 1982), the pulvinar

(Mesulam and EJ, 1982) (Mufson and Mesulam, 1984), and amygdala (Turner et al., 1980).

Recent anatomical evidence supporting a direct and early influence of audition on visual processing has been reported in a recent tracer study in non-human primates by Falchier and colleagues (2002) (see also (Rockland & Ojima, 2001)). Retrograde labelling of core and parabelt areas of the auditory cortex revealed previously unknown direct neuronal connections to areas of primary visual cortex. Interestingly, these projections terminated in an area of primary visual cortex subserving the peripheral visual field.

[REMOVED DISCUSSION OF SHAMS et al.]

1.4.3 Multisensory integration at the cellular level

The most detailed studies of cross-modal interactions at the neuronal level have been conducted in the mammalian SC (see (Stein and Meredith, 1993)). As Figure 1-3 shows, the SC receives convergent input from both the optic tract and the ascending auditory pathway onto single neurons (Benevento et al., 1977; Stein, 1978; Stein, 1981; Stein et al., 2004). Single-unit recordings from this subcortical structure, which is thought to be involved in orientation and attention behaviours, suggest certain neuronal mechanisms and rules by which multisensory convergence is achieved. For example, multisensory neurons in the SC display overlapping auditory and visual receptive fields. When two or more sensory cues occur in close temporal and spatial proximity, the response of these neurons can be substantially enhanced, sometimes exceeding 12-fold enhancements in firing rate beyond that expected by summing the impulses exhibited by each unimodal input in isolation, a phenomenon referred to as 'super-additivity' (Stein and Meredith, 1993). Because the output no longer resembles the response obtained to either input, there is a de facto assumption that the information obtained from two sources has been combined to form a single

(new) output signal (Meredith et al., 1992; Stein et al., 1993). This process is referred to as multisensory integration. The observed facilitation of the neuronal response is often maximal when the responses to the individual inputs are weakest, a principle known as inverse effectiveness. In contrast, cross-modal stimuli that show spatial or temporal disparity can induce profound response depression. This means that the response to an unimodal stimulus can be severely lessened, even eliminated, by the presence of an incongruent stimulus from another modality (Kadunce et al., 1997). These principles of multisensory integration have also been shown to apply to superior colliculus-mediated functions such as orientation and attentive behaviours (Stein et al., 1988).

Ascending projections from SC have been demonstrated with tracer injections into the deep layers of SC in monkeys (Benevento and Fallon, 1975). These ascending SC outputs provide input into the ancillary and multisensory thalamic nuclei, the lateral, inferior and medial divisions of the pulvinar and the satellite nuclei of the medial geniculate complex. From these nuclei projections reach the auditory and visual cortices. The lateral and inferior pulvinar projects to the visual cortex and STS and the medial pulvinar and medial geniculate axons intervene auditory cortex including A2, STG and STS (Hashikawa et al., 1995) (see solid red lines, Figure 1-2). In both instances, ascending nerve fibres end in layer 4, the input layer of the cerebral cortex. Ascending fibres into primary sensory areas (A1, V1) innervate cortical layer 1. Thus, brainstem mediated multisensory interactions can affect cortical processing whenever auditory and visual signals are present and share features that are important for SC integration, such as spatial proximity and temporal coincidence.

However, there are also descending projections from unisensory cortical neurons to deep multisensory layers of the SC (Wallace et al., 1993) (not shown in Figure 1-2). That these connections are important for certain types of multisensory integration has been shown by cortical deactivation studies, where cooling and

application of local anaesthetics in cortex causes an inhibition of both behavioural and neuronal firing gains in response to multisensory orienting behaviours (Wilkinson et al., 1996; Stein et al., 2002). The role and significance of these cortico-tectal projections in governing higher-order multisensory integration in humans has not been established yet, but they have been proposed to be shaped by experience and to "ensure that the multisensory integrative functions of the SC and the behaviours that depend on them are appropriate for the specific context in which these cues are encountered" (Stein et al., 2002).

Apart from the SC, neurons exhibiting multisensory receptive fields have also been shown to be present in cortical structures of the monkey (Graziano and Gross, 1998; Mistlin and Perrett, 1990; Duhamel et al., 1991), cat (Wallace et al., 1992; Wilkinson et al., 1996) and rat (Barth et al., 1995). However, detailed observations of multisensory response properties in cortex are comparably sparse and sometimes vary from those of the SC. For example, in the cat multisensory integration responses in neurons of the anterior ectosylvian fissure and the lateral sulcus were less restrained by the precise temporal and spatial congruency of the multisensory stimuli (Wallace et al., 1992). This suggests that multisensory processing in the cerebral cortex may serve different and a wider range of functions, most of which remain to be explored. Single-units in macaque STS have been shown to code both visual and auditory information in the upper bank and fundus of the STS (Hikosaka et al., 1988) (~12 % of neurons) (Bruce et al., 1981) (~38 % of neurons) and in the lower bank of the STS (Benevento et al., 1977) (~36% of neurons). These neurons showed responses to a large variety of simple and complex stimuli, and although AV interactions were not systematically studied, there was a tendency for sound to attenuate visual responses and evidence that some neurons only responded to combined AV stimulation. Recent evidence using more complex stimuli shows that a subset of neurons in monkey STS exhibit auditory-visual integrative responses for biologically meaningful actions

(Barraclough et al., 2005). This study found that about 23% of visually responsive STS (both in upper and lower bank) neurons are modulated by a corresponding auditory stimulus, and show in equal numbers response augmentation (86% mean increase) or attenuation (46% mean decrease). Interestingly, response augmentation only occurred if the auditory stimulus matched the visually presented action, suggesting that STS integrates based on higher-level stimulus analysis compared to SC, which integrates largely based on spatial and temporal coincidence (Stein and Meredith, 1993).

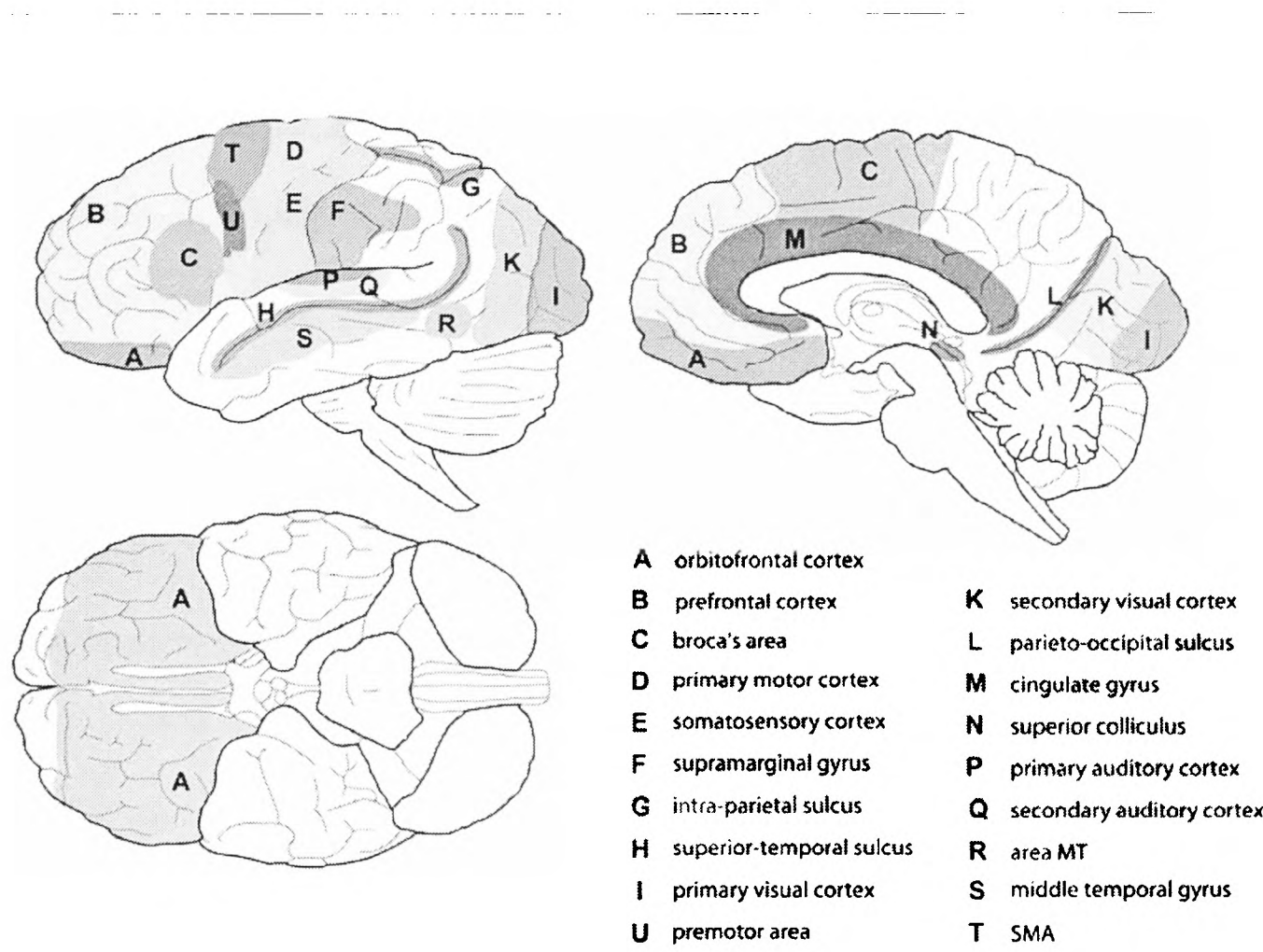


Figure 1-3: Schematic drawings of the human brain from three perspectives. Shaded areas correspond to brain structures discussed in the text. Upper left: Lateral view on the outer surface of the left hemisphere. Upper right: Sagittal view at the brain midline. Lower left: Ventral view showing the brain from below
[ADDED H and T to Figure]

1.4.4 Electrophysiological evidence in humans

Many multisensory experiments employ a redundant target paradigm, which requires subjects to make speeded detection responses to spatially concordant audio-visual targets and to the corresponding unimodal auditory and visual signals (Miller, 1982). Measured RTs are shorter to the bimodal event as compared to the unimodal counterparts in isolation, a robust effect that has been consistently replicated (Welch & Warren, 1986).

The neural correlates of the redundant target effect during object recognition and detection were investigated by Giard & colleagues (Giard & Peronnet, 1999; Fort et al., 2002). For example, subjects discriminated two objects based either on their visual, auditory or auditory-visual features while scalp EEG activity was recorded simultaneously (Fort et al., 2002). As expected, RTs to the bimodal presentation (247 ms) were significantly shorter than to either auditory (276 ms) or visual (310 ms) cues alone. To determine superadditive integration effects, the sum of the event-related potential (ERP) waveforms to both unimodal stimuli was subtracted from the activity to bimodal stimulation ($[AV - A+V]$) (see (Calvert & Thesen, 2004) for a discussion of different analysis strategies). The high temporal resolution of EEG allowed the investigators to detect significant interaction effects in the ERP waveform at around 50 ms post-stimulus over occipital cortex with scalp distributions typical of activities in visual cortex. The topography and timing of this effect over primary visual cortex suggest a mechanism for modulation of visual cortex by auditory input at early stages in visual processing before a complete sensory analysis of the stimulus has been achieved.

Similarly, Molholm et al. (2002) recorded EEG activity during a typical redundant target paradigm. Cortical interactions were detected at -46 ms at posterior electrodes over visual cortex corresponding to the initial visual responses in visual cortex occurring at around 40-55 ms post-stimulus in form of the C1 ERP wave over

primary visual cortex (Clark, 1995). These effects show that AV multisensory interaction can occur as early in the processing hierarchy as the sensory analysis stage in primary visual cortex, an area previously thought to process visual information only.

Under specific stimulus conditions, bisensory stimulation does not only result in detection advantages but also in a perceptual modulation of visual experiences by auditory co-stimulation. One such example is the "illusory flash effect", an auditory-visual illusion first reported by Shams et al. (2000) where a phenomenological change in the perception of a visual stimulus is induced by sound. More specifically, when a single flash is presented concurrently with multiple short beeps, the single flash is perceived as multiple flashes. The illusion is so compelling and automatic that even observers informed about the physical nature of the stimulus report a visual experience of seeing multiple flashes. In a follow-up study, Shams et al. (2001) recorded flash-evoked ERPs and introduced a sound to elucidate the temporal and spatial occurrence of the auditory-visual interactions. In this paradigm, sound modulation affected activity at around 170 ms poststimulus again at the level of visual cortex. However, this effect occurred at a considerably later time-window than those observed by Fort et al. (2002) and Molholm et al. (2002) and is therefore suggestive of a feedback modulation from higher-order multisensory areas.

1.4.5 The case of auditory-visual speech

Visual speech is a multifaceted stimulus involving high-level visual object and complex motion coding, and probably also mapping of visual to phonological feature representation. So far, only a few neuroimaging studies have investigated the neural correlates of speechreading and have found a large network of brain areas mediating this task. Cortical areas include regions of the left A1, STG, STS, middle temporal gyrus (MTG), right MTG and bilateral inferior temporal gyrus (ITG), premotor

cortex, inferior frontal gyrus (IFG), and MT (Sams et al., 1991; Calvert et al., 1997; Campbell et al., 2001; Pekkola et al., 2005). However, the involvement of A1 in speechreading remains controversial (Bernstein et al., 2002; Paulesu et al., 2003). The question whether these effects are speech-specific or represent general AV integration phenomena is rarely tested.

Simulated biological speech motion with point-light stimuli activated motor-related areas, suggesting an involvement of the human mirror neuron system in visual speech perception (Rizzolatti et al., 1996). Further support for an involvement of the motor system during both auditory and visual speech perception is by a transcranial magnetic stimulation (TMS) study by (Watkins et al., 2003), who stimulated the lip area of the motor cortex. Presenting auditory and visual speech stimuli together augmented the motor-evoked potential recorded from the lips compared to a control condition. This effect was specific to the left hemisphere and suggests that the left primary motor cortex (M1) is part of the network for processing and, potentially, integrating auditory-visual speech signals.

Areas of the left superior temporal lobe also activate to stationary images of visual speech (Calvert and Campbell, 2003). Facial gurning (non-speech) movements do not activate the left STG and STS as extensively as articulatory mouth movements, further suggesting that this region of the superior temporal lobe is specialized for the extraction of speech-specific features from visual signals (Calvert et al., 1997; Campbell et al., 2001). Evidence that this effect is experience-dependent comes from the finding that congenitally deaf individuals activate the left STS and STG less robustly to visual speech as normal hearing subjects (MacSweeney et al., 2001).

Crossmodal interactions and facilitations have been reported with more complex and socially relevant stimulation. Human speech perception is an inherently multisensory phenomenon, as a speaker's articulatory mouth movement is intrinsically correlated with the time-frequency sound waveform. Having information available

from both auditory and visual sources can have profound effects on the resulting speech percept. For example, under noisy acoustic conditions, being able to see the talkers mouth movements can significantly enhance speech comprehension to the equivalent of changing the acoustic signal-to-noise ratio by 11 decibels (dB) (i.e. increasing the physical intensity by a factor of 4 (Sumby & Pollack, 1954; MacLeod & Summerfield, 1987). A classic example of an auditory-visual spatial localization illusion is the ventriloquism effect, where the presentation of auditory speech, whose source is located away from the articulating face, is mislocalized and captured by the spatial position of the visual signal (Bertelson and Radeau, 1981; Driver, 1996).

Concurrent visual speech information can also influence the resulting auditory speech percept, as is shown in another classical illusion by McGurk & MacDonald (1976). Here, the incongruous pairing of an auditory /ba/ and a visual /ga/ results in the perception of /da/, producing a very robust effect of auditory-visual integration in speech perception. The McGurk effect is very robust and is perceived even if the subject is aware that the AV signals are incongruent or if auditory and visual stimuli are presented with spatial disparity (Jones and Munhall, 1997).

In a first physiological study into the McGurk effect, Sams and colleagues (1991) showed that the auditory neuromagnetic response, usually localised within the auditory cortex, was modulated by mouth movements. Similarly, functional MRI studies investigating the neural basis of cross-modal speech processing have shown that the auditory cortex can be activated by silent lipreading (Calvert, et al., 1997; MacSweeney et al., 2000), providing an avenue for visual speech to influence the perception of auditory speech at a pre-lexical stage in primary sensory cortex. However, several groups report conflicting results. Some investigators have reported modulation of auditory cortex during auditory processing by simultaneously presented visual speech using MEG and EEG techniques (Sams, 1991; Callan et al., 2001; Mottonen et al., 2002; Klucharev et al., 2003, Pekkola et al., 2005). Similarly, a

previous fMRI study has found activation to silent speechreading in auditory cortex (belt and parabelt regions) and STS (Calvert et al., 1997). Other studies, however, have failed to report involvement of primary auditory cortex in speechreading (MacSweeney et al., 2000; Campbell et al., 2001). As a consequence, the full extent by which visual speech activates superior temporal areas is not yet fully understood and remains controversial (Bernstein et al., 2002).

In an fMRI study using interaction effects to determine audio-visual integration sites in the human brain, Calvert et al. (2001) presented subjects with both unimodal as well as congruent and incongruent audio-visual speech in a multiplexed block design. In the congruent condition, subjects heard a human voice reading a story while simultaneously seeing the corresponding lip movements through a back-projection screen. In the incongruent condition, the audio and visual streams were shifted in time so that the visual information did not correspond to what the subjects heard. Using the superadditive model ($AV - (A + V)$) for determining cross-modal interactions, Calvert et al. (2001) identified a network of brain areas showing super-additive response enhancements and response depressions for synchronous and asynchronous audio-visual language inputs. These included area IP, insula, SC, STS and regions of the medial ventral and dorsal prefrontal cortex. A similar role of the STS in the integration of arbitrary audio-visual information was found in a recent MEG study. Raij et al. (2000) investigated the human brain's audio-visual integration mechanisms for phonetic and graphemic representations of letters, which, in literate people, have been extensively paired through associative learning. Subjects were presented with auditory, visual, and auditory-visual letters in a target identification task while neuromagnetic activity was recorded from the skull surface. Reaction time findings showed a clear behavioural advantage for audio-visual stimuli (425 ms) compared to auditory (505 ms) and visual (520 ms) stimuli alone. Time windows of audio-visual interaction were determined by calculating $[AV - (A + V)]$ from the

averaged evoked responses at around 380 ms after stimulus presentation revealed the left posterior STS as a main area of AV convergence. Further results from previous studies on brain areas involved in AV speech processing and their time-course of activation are discussed in more detail in the relevant sections of this thesis.

1.5 Conclusions

In summary, the field of multisensory research is advancing rapidly. Neuroimaging experiments have revealed audio-visual interaction patterns at various cortical locations and latencies, extending into primary sensory cortices. Stimulus feature and maybe task requirements (Calvert, 2001) play an important role in explaining the variability between findings and underline the dynamic nature of AV bisensory processing. It is also evident that speech distinguishes itself as a form of multisensory processing not only behaviourally but also neurologically, as it recruits specialized cortical areas for AV stimulus integration. Together with simple AV stimuli, speech shows also bisensory interactions in sensory-specific auditory and visual cortices. The present thesis will attempt to further describe the neuronal correlates of auditory and visual speech processing and test for areas where the two sensory streams interact. We will use an experimental paradigm by which we can test both speech-unspecific and speech-specific interaction areas. Using neuroimaging methods with both high spatial and high temporal resolution we will be able to describe the spatio-temporal interactions of auditory and visual speech in great detail. It is expected that our research will help to disambiguate some of the findings in the multisensory literature and at the same time generate novel findings which will help us to ask new and increasingly specific and sophisticated questions about the nature of multisensory integration.

2 Techniques for imaging the performing human brain

Recent developments in neuroimaging technology allow for a much better description of brain function than was previously possible based on neuropsychological approaches and lesion studies. These new methods have enabled us to directly and repeatedly observe the human brain while it is active. Each functional imaging technique offers individual advantages, but also has its distinct disadvantages, which mostly pertain to the spatial or temporal resolution of the measurement, or its relationship with neural activity. In this section we describe the basics behind each of the imaging techniques that are used in this thesis, namely functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and intracranial electroencephalography (iEEG), and discuss their strength and weaknesses. By combining these techniques in the investigation of one neuroscience question, AV speech processing, we are hoping to circumvent the limitations dictated by any particular methods and to create a more complete and integrated picture of the underlying functional neurophysiology of multisensory processing.

"Though this be madness, yet there is method in't"

-From *Hamlet* (II, ii, 206)

The various neuroimaging methods that have been used in the investigation of human multisensory brain mechanisms fall into the two categories of hemodynamic/metabolic methods, of which the most prominent techniques are fMRI and PET and electrical/magnetic methods, which includes EEG and MEG.

These imaging techniques differ, not only in terms of their temporal and spatial resolution (see Figure 2-1), but also in the source of their respective signals, which can have profound consequences on the interpretation of the data obtained by these methods. In the following section we provide a description of the techniques and methodologies of fMRI, MEG and iEEG.

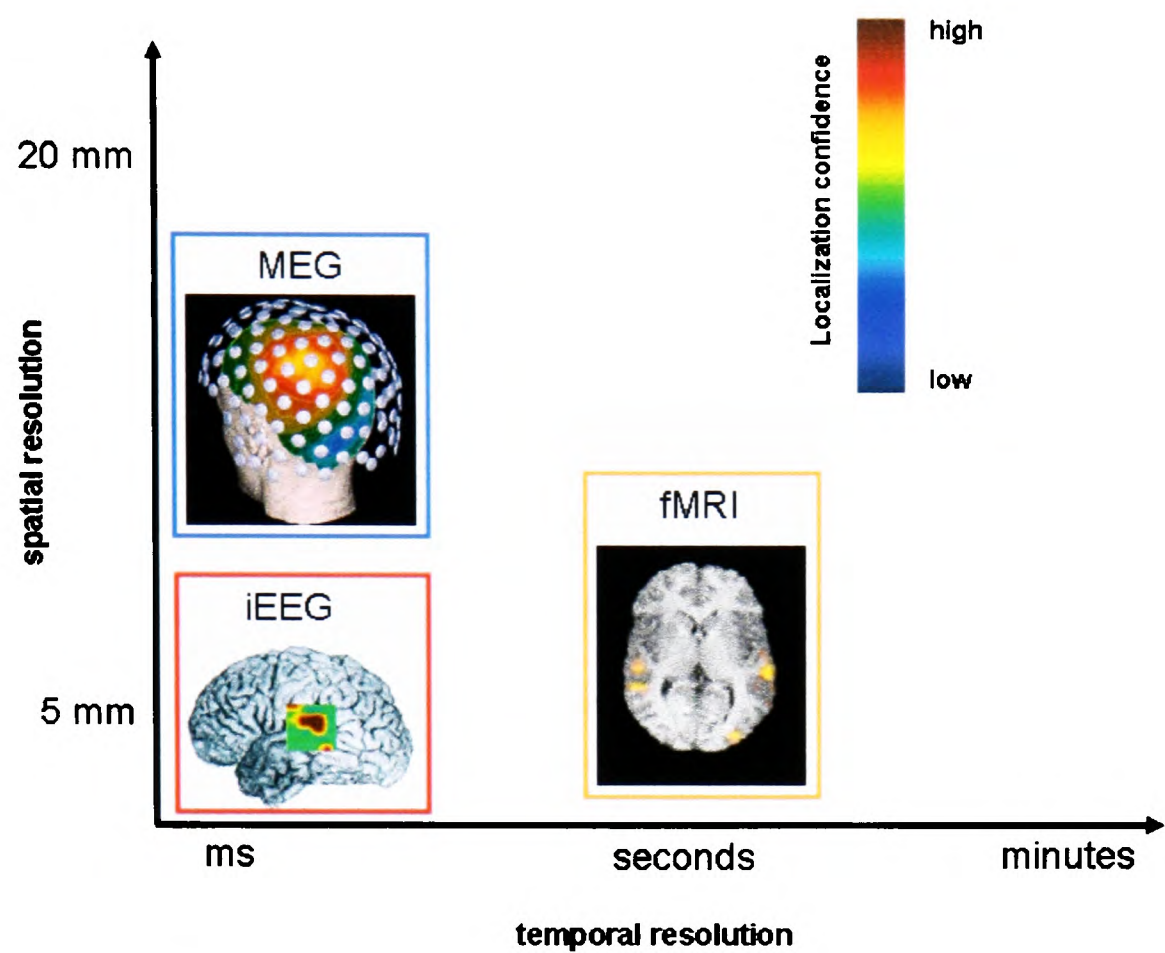


Figure 2-1: Comparison of temporal and spatial resolution of fMRI, MEG & iEEG

Color-coded frames represent the degree of uncertainty which is associated with the localization of the actual source of neural activity.

FMRI offers high spatial resolution on the millimetre scale across the whole brain, but due to the sluggishness of the hemodynamic coupling, its absolute temporal resolution is on the order of seconds. Partial volume effects and the spatial specificity of the vascular system together with large-vessel effects compromise the localization confidence to some degree. MEG records neuronal activity on a millisecond timescale, but suffers from inferior and ambiguous spatial resolution due to the inverse problem. Recordings directly from the human scalp with iEEG offer both high temporal and spatial resolution, whereby only the spatial extend of the sampled area is restricted but localization confidence is high.

2.1 Structural Magnetic Resonance Imaging

Structural MRI has become a popular tool for clinical diagnosis and research (see Jezzard and Clare, 2001) for a detailed technical review). It exploits the fact that hydrogen atoms are abundant in human tissue, especially in water and fat. The nuclei of H atoms consist of single protons, which have magnetic characteristics. When placed in a high magnetic field, most of a population of protons will align either parallel to the magnetic field in a low energy state or orthogonal to the magnetic field in a high-energy state. This ensures that the net magnetization vector of the protons and the magnetic field will be aligned with the magnetic field at equilibrium along the Z-axis. Protons in this state will absorb a photon of energy if the energy has the same frequency as the resonating frequency of the proton's procession about the magnetic field. Absorbing a photon with the appropriate frequency causes a proton to shift from a low-energy to a higher-energy state. MRI utilizes this property of protons and introduces photons into the magnetic field at the known resonating frequency of protons, which is in the radio frequency (RF) range, emitted by the RF coil.

What MRI then measures is how long it takes for the protons to return to their baseline (low energy) state after having absorbed a photon which caused a transition to the high-energy state. The time that it takes for the proton to return into the low-energy alignment is referred to as T1 relaxation time (or "spin-lattice relaxation" time). After RF pulse administration nearly all protons align and begin to process together in the same direction. Immediately, the spins begin to dephase and this relaxation is called the T2 relaxation time or the "spin-spin" relaxation. T1 and T2 relaxations are part of the same process and can be measured separately. Different MRI contrast can be used to produce different images, such as T1-weighted, T2-weighted or proton density, by weighting each voxel by their T1 or T2 relaxation times. In different media in the brain (white matter, grey matter, cerebrospinal fluid,

bone, tumour etc.) the protons will have different T1 and T2 relaxation times, thus allowing the distinction between different tissue types.

By introducing magnetic field gradients (which are variations in the magnetic field based on the spatial position of gradient coils) in specific orientations and of specific widths (which determines slice thicknesses), it is possible to determine where the T1 and T2 relaxation times differ and to reconstruct a 2-D or 3-D image of these magnetic field changes (thus producing a 2D or 3D image of the underlying structure of the human brain and yielding maps with high spatial resolution of around 1mm^3 or even higher anatomical detail). High-resolution structural images are most often T1-weighted scans where a whole volume is recorded once for a subject during an experiment. Their spatial resolution is on a mm^3 scale and can discriminate between cortical structures, such as gyri and sulci. Structural MRI scans are often realigned with functional MRI scans to match functional activation with detailed structural brain anatomy.

2.2 Functional Magnetic Resonance Imaging

Over the last decade, fMRI has overtaken PET as the method of choice for studying brain function in vivo (for a review see (Matthews, 2001)). This is largely due to the fact that FMRI does not require injection of a radioactive substance, but relies instead on a natural contrast agent - the blood oxygen level dependent (BOLD) effect (Kwong et al., 1992; Ogawa et al., 1993). BOLD contrast exploits the different magnetic properties of oxygenated and deoxygenated blood. When a subject is placed in a high magnetic field (the MR scanner), task-induced changes in brain metabolism alter the ratio of oxy- and deoxy-hemoglobin locally, causing measurable changes in MR signal intensity (more precisely in the $T2^*$ weighting, which is T2 relaxation adjusted for local field inhomogeneities (Kwong et al., 1992; Ogawa et al., 1993; Turner, 1995)). Deoxygenated hemoglobin concentration is reduced as blood flow

increases, causing an increase in the MR signal on a T2* weighted image (Ogawa et al, 1993, Turner, 1995). Detection of these MR signals changes allows the source of the underlying neuronal activity to be localized to within a few millimetres. The fact that the signal can be recorded with a conventional MR scanner of a hospital (some additional hardware is required, e.g. gradient coil) makes fMRI an accessible means of studying brain processes.

But while this provides fMRI with very high spatial resolution, its temporal resolution is restricted by the sluggish nature of the BOLD response (peaks at approximately 4-8 s post stimulus onset) (Menon & Goodyear, 2001). Recent data have shown that the BOLD signal indeed is a marker of neuronal activity and that it correlates well with local field potentials and neural spike activity (Logothetis et al., 2001; Mukamel et al., 2005). Due to the extreme stability of the hemodynamic response within one region and given the right experimental design, relative timing differences of less than 1 s can theoretically be detected within and between regions. However, considering that there are variable timing differences in the neuro-vascular coupling between brain regions (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000) and that the dynamics of brain responses occur generally on the order of tens of milliseconds (Enns & Di Lollo, 2000), it becomes clear that fMRI at its present state of development is not ideal for answering questions regarding the exact timing of cortical events. However, the crucial advantage of using fMRI in multisensory experiments remains the high degree of certainty with which spatially exact determinations of neuronal activity can be made

2.2.1 FMRI experimental designs

Early fMRI studies employed statistically powerful block designs, similar to the way PET studies are performed, in which a subject is stimulated repeatedly during the so-called ON period in a block of trials. Interleaved into the stimulation blocks are

typically (but not necessarily) OFF periods where the stimulation of interest is absent (all other things being equal), often called a rest condition. This design allows for high-signal-to-noise recording and for powerful cross-correlation or statistical-parametric mapping statistics to be performed for analysis. Recent advances in fMRI experimental design allowed for determining the responses to a single type of stimulus by averaging the responses to multiple presentations time-locked to a specific event, similar to the ERP in EEG/MEG studies (for review see D'Esposito et al., 1999). Such event-related designs are particularly useful for certain cognitive experiments in which an unpredictable and random order of events are important and where grouping of event types can be computed post-hoc (i.e. according to the subject's response or reaction time).

2.3 Encephalographic recordings (EEG/MEG)

Both MEG and EEG can be used for recording ongoing as well as event-related brain activity. Both measures of the MEG/EEG signal correlate well with the subject's state and are modulated by stimulus characteristics and task (Regan, 1989; Pfurtscheller, 2001). The activity related to a specific sensory event of interest can be extracted from the background activity by averaging over multiple event trials, time-locked to the stimulus onset. The resulting waveform pattern in EEG, the ERP, is comprised of a deflection of positive and negative peaks over time (Lopes da Silva, 2005). A large body of research has focused on identifying stimulus and cognitive correlates that modulate the latency and amplitude of these ERP components to stimulation in all sensory modalities (audio, visual, olfactory, gustatory, tactile and vestibular).

The basis of the MEG/EEG signal is the macroscopic current flow in neural assemblies. If these neurons are aligned in parallel and fire synchronously the signal summates and becomes detectable over the ambient noise at the level of the sensory

or electrode (Hämäläinen et al., 1993). Because of these considerations, the main contributor to the magneto- and electroencephalographic signal are post-synaptic currents of pyramidal neurons of the cortex. Post-synaptic potentials last about 10 ms, a time frame which allows magnetic fields and electrical current to summate. It usually requires the combined activity of thousands of spatially aligned and synchronously firing neurons to create a detectable signal.

The distinct advantage of the neuromagnetic/electrical methods is their high temporal resolution. The signal is directly related to neuronal activity and the transmission of neuronal currents through the brain and to the sensors is virtually instantaneous, and is only limited by the sampling frequency of the recording equipment. This renders MEG and EEG as ideally suited for testing hypotheses concerning the exact time course of multisensory events in the human brain. For example, neuromagnetic/electrical methods are able to answer questions about the temporal occurrence of first interactions of the two sensory streams (Giard and Peronnet, 1999; Molholm et al., 2002).

2.3.1 MEG

The ionic currents originating from biochemical sources at the neural level of the central nervous system generate not only electric potentials but also magnetic fields. The magnetic field generated by the brain is very weak ($\sim 10^{-8}$ gauss compared to the earth magnetic field of ~ 0.5 gauss (Reite et al., 1976) and requires superconducting sensor devices to measure it. A first device consisted of an induction coil magnetometer to detect magnetic alpha rhythms. However, since then the introduction of SQUID (superconducting quantum interference device) and other developments, commercially available MEG systems are now highly sophisticated pieces of machinery with a high number of sensors that cover the whole scalp. The sources of MEG are usually described as primary currents which are a result of ionic

currents in neural cells. It has been found that both intrinsic and synaptic current can contribute to the magnetic field (Murakami et al., 2002; Murakami et al., 2003). The fissural structure of cortex, however, decreases the significance of this problem and allows MEG to detect a larger number of sources across the outer sheet of cortex.

The spatial resolution of MEG is somewhat restricted and can provide ambiguous answers. In dipole analysis, the local neuronal foci are usually modelled as equivalent current dipoles (ECD) whose number, strength and locations are estimated based on the externally measured magnetic field distribution. The dipole parameters (a current dipole is completely described by six parameters, three to establish its position within the head, two to define its orientation and one to define its strength) can be calculated at any given time-point by a non-linear least-squares search (Hämäläinen et al., 1993). The model can include several spatially and temporally distinct dipolar sources within the same data by using a spatio-temporal dipole model (Scherg and von Cramon, 1985).

This procedure poses a non-trivial challenge because of the fact that there is no mathematically unique solution to the problem of inferring the numbers and locations of dipoles that could, theoretically, produce the observed pattern of activity on the surface of the skull, (i.e. there is an infinite number of source configurations that could produce exactly the same measured field). This is generally referred to as the inverse problem (Helmholtz, 1853; Nunez, 1990). Solving the inverse problem for one local source of activation is straightforward since the laws governing the propagation of electrical/magnetic fields through skull and tissue, the so-called forward problem, are well understood (Hämäläinen et al., 1993). However, if several sources are contributing to the observed field map, as can be expected in multisensory speech experiments, solving the inverse problem becomes increasingly more difficult and complex. In practice, however, the experimenter uses a-priori knowledge of physiology and functional anatomy, often derived from other neuroimaging

modalities, to incorporate feasible constraints into the model (Hämäläinen et al., 1993; Hari, 1996; Dale and Halgren, 2001).

2.4 Intracranial EEG

Intracranial EEG (iEEG) is recorded from electrodes implanted inside the cranial cavity of neurosurgical patients. These patients often suffer from medially intractable epilepsy and iEEG recordings are used to help in the diagnosis of the disorder and in defining appropriate treatment. Whereas the selection of electrode sites and the duration of the implantation are decided solely on clinical grounds, patients often give informed consent to participate in research studies. An advantage of iEEG over MEG is that intracranial recordings are not so susceptible to artifactual contamination from muscle movements and eye blinks, which regularly impair the quality of MEG recordings.

One of the prerequisites of recording directly from the human brain is that the subject is a neurological patient who most often suffers from severe epilepsy. This raises the issue that the brain responses recorded from these patients contain abnormal neuronal patterns and may not generalize to the general, healthy population. This possibility can never be ruled out, but steps can be taken to minimize its likelihood, such as using patients with different pathologies and anticonvulsant medication. Data contamination by interictal spiking is a prevalent problem and can be minimized by selecting testing periods where the subject is under anticonvulsive medication and by increasing the number of stimulus trials in case data have to be discarded.

The spatial resolution of iEEG is higher than any of the other methods discussed but varies depending on recording equipment. Investigators have used single and multi-unit recordings from extra-cellular microelectrodes (Bechtereva et al., 1992; Ulbert et al., 2001), but the multitude of iEEG recordings uses surface strips

or grids of electrodes. The latter is used in the present investigation and all further discusses focuses on this approach.

Human iEEG grid recordings measure local field potentials (LFP), which reflect the summated activity from coherent neuronal assemblies. As such, their origin is closely related to that of signals from neuronal spikes (Shadlen and Movshon, 1999), scalp EEG (Regan, 1989), MEG (Hämäläinen et al., 1993) and fMRI (Logothetis et al., 2001). Whereas spikes are thought to reflect the output of neurons, LFP are considered to the summation of post-synaptic potentials and thus represent the input of a neuronal population to an area (Shadlen and Newsome, 1998). It has to be noted that changes in LFP amplitude can result from both changes in excitation of a neuronal population and in the synchronization of the firing patterns within that population.

Assessing the exact spatial resolution of iEEG is not straight-forward and is subject to similar uncertainties as scalp EEG. Each electrode, in effect, records a weighted sum of the electric field sources of the entire brain. The weight of each source decreases with the square of the distance between source and electrode (Morris and Luders, 1985). This can mean that sources far away from the recording point can have a significant influence on the recorded signal if these sources are very strong relative to those created in closer proximity to the electrode.

2.4.1 Source reconstruction from iEEG

The visualization of the spatial distribution of the electric potential sometimes allows an estimation of the localization of the generator currents. For example, if a polarity inversion between neighbouring electrodes is present, the source of that activity is likely to be in a plane orthogonal to the line joining the two electrodes. If the source can be modelled as a dipole, the orientation of the dipole, i.e. the modelled current flow, is parallel to that line pointing into the direction of the positive potential

(Halgren et al., 1995). An indicator that one is recording from a local structure is a large voltage gradient within the recordings from one electrode that is not visible in surrounding electrodes. A rule of thumb is, the larger the spatial distribution of voltage gradients, the deeper the source. Figure 2-2 illustrates the relationship between current flow and magnetic field and electrical potential.

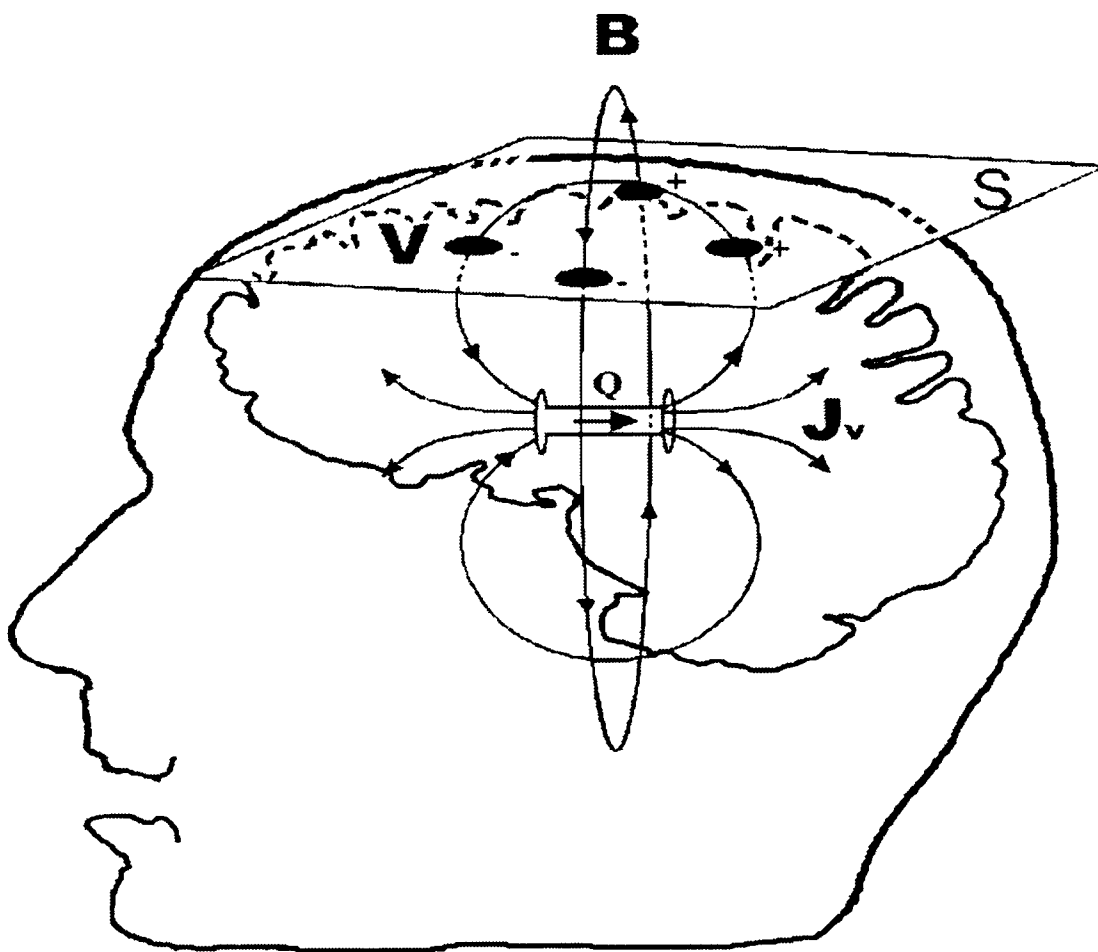


Figure 2-2: Relationship between current source and the electric/magnetic signals

This graphic depicts how the magnetic fields and potentials outside the head are related to electrical currents in the brain. A current source with strength Q causes a current flow J_v within the brain. This current flow causes a potential difference V on the electrode grid S (measured with EEG) and a magnetic field B outside the head (measured with MEG). (Image modified from (Aston, 2005).

2.5 On the similarity of the fMRI, MEG and EEG signals

MEG and EEG signals are closely related. The primary current generators of both signals are the same (the resulting signals is only rotated 90 degrees with respect

to each other, see Figure 2-2), but there are also systematic differences between the recorded signals from both imaging modalities and their relation to neural anatomy. MEG has a higher spatial resolution because the magnetic signal only decays with the square root of the distance and is not influenced by brain tissue, cerebro-spinal fluid or the skull. The EEG distribution on the scalp, however, is subject to considerable distortion and spatial smearing due to volume conduction (Nunez, 1990). This makes MEG more apt to study activation with multiple sources and overlapping temporal activity across different brain regions. Furthermore, the electric measurements are reference-dependent and therefore the electric potential distribution is dependent on the placement of the reference electrode, whereas magnetic signals are recorded reference-free. Both methodologies also show differential sensitivity to sources of different orientation and location. MEG is most sensitive to tangential currents whereas EEG is sensitive to both tangential and radial currents (Hämäläinen et al., 1993). A radial component, detected by EEG, does not generate a magnetic field outside a sphere-shaped volume conductor, i.e. the head, and thus does not contribute to the MEG signal. Thus, it is clear that there are similarities between the signals, but also that both imaging complement each other to reflect a wider spectrum of neural activity. Further elaboration of the differences and similarities between MEG and EEG signals and their functional interpretation is presented in section 8, which compares and discusses the results from the different imaging modalities used in this thesis.

One of the implicit assumptions of coregistering and combining fMRI with MEG and EEG is that the signals originate from the same cortical generator. However, even though this assumption is supported by various studies linking metabolic and hemodynamic activity with local neuronal synaptic and electrical/magnetical events (Rosen et al., 1998; Logothetis et al., 2001), the coupling is not 100 percent. EEG and MEG are selective measures of current source activity

and often only sample a subset of the total synaptic activity of a cortical structure (Halgren, 2004). MEG sources are usually described as “primary currents” resulting from ionic currents in neurons. Both synaptic and intrinsic cellular currents can contribute to the observed magnetic field (Murakami et al., 2002), but currents can cancel each other out on the local and larger scale. For example, the currents within dendrites of symmetrically distributed stellate cells produce local electromagnetic fields that cancel each other out (‘closed field structure’) so that no current will be recorded outside the dendritic field (Tenke et al., 1993). Stellate cells contribute about 15% to the total neuronal population in cortex (Wilson et al., 1994) and possibly have a higher metabolic demand compared to other neuronal types because of higher net firing frequencies (Connors and Gutnick, 1990). Current cancellation on the larger scale depends on the particular arrangements of cortical folding, orientation and depth. MEG is most sensitive to currents closest to the sensors and sensitivity declines rapidly with distance, however, dipoles in gyral crowns are radially oriented and thus together can cancel each other out when summated over a larger area. Cortical folding also has a large effect on current cancellation due to the fact that the majority of the cortical surface is located in sulci. Two simultaneously active dipoles at opposite sides of the sulcus would often cancel each other out (Halgren, 2004). In fact, a simulation study by (Ahlfors et al., 2002) estimated that ~10 % of random activation in the cortical surface results in more than 90 % of the MEG response at the sensors being cancelled. This means a lot of cortical synaptic activity remains invisible to extracranial MEG sensors, and this all because of reasons that do not apply to fMRI. Furthermore, the amplitude of the MEG and EEG signal is largely dependent on the phase relationships within and between oscillating neuronal networks (Varela et al., 2001; Makeig et al., 2002), a factor that is likely to be largely independent of local metabolism and hemodynamics (Nunez and Silberstein, 2000). One example of this phenomenon is the occipital alpha rhythm, which is reduced with eyes open, showing

that EEG amplitudes and metabolic demand can change directions because of cortical synchrony and the specific oscillatory nature of local and global neuronal networks (Goldman et al., 2002). However, direct and quantitative comparisons between the fMRI BOLD response and electrophysiological recordings have been conducted during simultaneous recordings in monkeys. (Logothetis et al., 2001) found that LFP nevertheless correlated well with the BOLD response, much better than spiking activity, for example. Since LFP are the main contributors to the observed MEG and EEG signal this suggests that these measures should be similarly well correlated with the fMRI BOLD signal. The data we expect to collect as part of this thesis will give us the opportunity to assess this hypothesis by comparing the spatial overlap between MEG/iEEG and fMRI results.

2.6 Brain oscillations

Other important information that can be extracted from the MEG and EEG signals is related to the frequency content of the measured response. ERP averaging is useful for detecting phase-locked activity related to stimulus processing, but is insensitive to stimulus induced activity which is phase variant (Varela et al., 2001). Calculating the signal power in the frequency domain over time detects non-phase-locked oscillatory activity. The spontaneous EEG activity contains distinct rhythmic components that peak in specific frequency bands (e.g. Alpha = 8 Hz - 13 Hz, Beta = 14 Hz - 30 Hz; Gamma = 31 Hz - 60 Hz, Omega = 60-120 Hz) correlate well with the subject's state and are modulated by stimulus characteristics and task (Niedermeyer, 2005). It has to be noted that the temporal resolution in the frequency domain depends on the particular frequency band. For example, a 10 Hz alpha rhythm is defined as an oscillation at 10 cycles per minute and thus needs 100 ms to complete one full cycle, whereas 40 Hz gamma needs 25 ms for a complete cycle. This, of course, is an inert

restriction of the temporal resolution of the frequency method, especially in the lower frequency ranges.

These stimulus-related changes in the power spectrum are often termed event-related desynchronization (ERD) if the power decreases and event-related synchronization (ERS) if power increases compared to a passive baseline (for a review, see (Pfurtscheller and Lopes da Silva, 2005)). There is ample evidence that neuronal populations have synchronous or oscillatory patterns of activity, and that their frequencies and temporal dynamics are associated with distinct behavioural states. Changes in cortical oscillatory power have been found to correlate with a wide range of sensory, perceptual and cognitive events (Berger, 1930; Pfurtscheller et al., 1977; Pfurtscheller and Neuper, 1992; Makeig, 1993; Pfurtscheller and Lopes da Silva, 1999; Pfurtscheller, 2001; Krause et al., 1996; Hari et al., 1997; Salenius et al., 1997). Occipital rhythmic activity in the alpha band (8-13 Hz) for example, is dependent on visual input, whereby closing the eyes augments the power of the oscillatory signal and opening the eyes dampens it (Berger, 1930). In comparison, a 10 Hz mu rhythm over rolandic regions is not affected by opening/closing of the eyes but diminishes during hand movements (Pfurtscheller and Neuper, 1994), suggesting that these oscillations are distributed across functional brain areas and may serve different functions. These cortical rhythms represent both sensory and cognitive processing and there is evidence for both a bottom-up thalamo-cortical modulation of these rhythms (Galambos et al., 1952; Cotillon et al., 2000) and for trans-cortical top-down processing, such as motor imagery (Pfurtscheller and Neuper, 1997). Several studies have shown that averaged ERPs are not separate from ongoing cortical processes, but are rather generated by phase synchronization and partial resetting of ongoing activity (Makeig et al., 2002). The power in the 8-13 Hz alpha band, for example, correlates well with ERP amplitude (Brandt et al., 1991; Rahn and Basar, 1993; Rahn and Basar, 1993).

More recently, high frequency oscillations in the gamma band (>30 Hz) have been become widely discussed in the literature and are thought to reflect higher cognitive functions and to provide an underlying mechanism for the integration of stimulus-related features across brain regions ('binding problem') (Singer, 1993; Singer and Gray, 1995; Treisman, 1996; Treisman, 1998). As such, it is possible that the brain could also use these mechanisms to bind information across the sensory modalities. However, the topic of feature binding via high-frequency synchrony still remains actively debated (Shadlen and Movshon, 1999; Gray, 1999) and there remain large gaps in our understanding of the relationship between cortical oscillations, neuronal firing and information processing.

The relationship between synchronization/desynchronization and neuronal activity is not straightforward either. The ERD of the posterior alpha rhythm for example is thought to be resulting from thalamo-cortical stimulation and thus be a correlate of excited neural networks or activated cortical area (Niedermeyer, 2005). This view is supported by simultaneous EEG and fMRI recordings, where increased alpha power over occipital regions was correlated with decreased MRI signal in multiple cortical regions, including in the occipital and superior temporal cortices. Increased alpha power, however, correlated with activity in thalamus and insula (Goldman et al., 2002). These findings suggest that the cortical, posterior/visual alpha rhythm is an index of cortical 'idling' of a neuronal network (Pfurtscheller and Neuper, 1994) generated in part by the thalamus.

Using these oscillatory properties of the EEG/MEG signals, Von Stein et al. (1999) presented subjects with representations of the same objects through different sensory modalities (audio & visual), and were able to elucidate some of the mechanisms by which supramodal feature representation is achieved in the human brain. Coherence analysis (Rappelsberger et al., 1986) revealed an enhanced coherence between temporal and parietal electrodes in the 13 - 18 Hz frequency

range, which was common to both modalities of presentation and absent in the control condition. This study is an example that neuroelectrical/magnetical methods can provide unique information about large-scale oscillatory brain mechanisms involved in multisensory processing and object representation that cannot be tapped by hemodynamic methods. More relevant studies are discussed in the appropriate sections in the remainder of this thesis.

2.7 Encephalographic analysis strategies for multisensory research

2.7.1 Event-related averaging

In the forthcoming electro- and magnetoencephalographic studies we will use event-related averaging to infer about crossmodal processing. In most crossmodal MEG and EEG studies the activity related to a specific sensory event of interest is extracted from the background activity by averaging over multiple event trials, time-locked to the stimulus onset. The resulting waveform pattern, the evoked response, is comprised of a deflection of positive and negative peaks over time.

As with fMRI and PET, the detection of bimodally evoked responses that exceed the algebraic sum of that obtained to the two individual contributing components has also been used in the context of averaged ERP studies (e.g. (Giard and Peronnet, 1999)). In this case, amplitude values of ERP components are measured in response to stimulation in both modalities separately, as well as to concurrent bimodal stimulation. The crossmodal interaction effect is then defined as the difference waveform of [bimodal - (unimodal modality A + unimodal modality B)] at each electrode/sensor. This difference waveform can then be displayed as surface potential maps on the outer surface of the skull (Giard and Peronnet, 1999; Raji et al., 2000; Fort et al., 2002; Molholm et al., 2002) or subjected to a dipole analysis to obtain an estimate of the relative strength and location of the proposed crossmodal

interaction effect (Fort et al., 2002; Luetkenhoner et al., 2002; Teder-Salejarvi et al., 2002).

Calculation of interaction effects in MEG experiments may be subject to artifacts of interpretation that may not affect hemodynamic studies to the same extent. For example, (Teder-Salejarvi et al., 2002) observed that if there is a component X present in all 3 tasks (e.g. in A, V & AV), calculating a simple interaction effect $([A+V] - AV)$ results in the double addition of this component in the unimodal cases, but only once in the bimodal condition $([X+X] - X = X)$. This effectively leads to the presence of the component X in the difference waveform and makes delineation from true multisensory components impossible. Teder-Salejarvi et al. (2002) showed that an early component related to stimulus expectancy, namely the contingent negative variation (CNV, (Rosahl and Knight, 1995)), could precisely result in the appearance of a spurious early interaction effect. Similar confounding effects can be expected to result from late stimulus or task-related activity (e.g. P3 (Polich, 1986), motor response, etc). To avoid these confounds variable interstimulus intervals (ISI) or a high-pass filter (after visual inspection of pre-stimulus baseline) needs to be applied to minimize expectancy waveforms.

Unequal subtraction in simple interaction paradigms is, however, less of a problem in fMRI data analysis. Whereas a MEG/EEG sensor or electrode might receive signals that are generated by multiple cortical areas involved in different aspects of stimulus processing, the point of measurement in fMRI is an isolated three-dimensional unit (cubic voxels). Consequently, whereas a frontally generated CNV (Rosahl and Knight, 1995) in an MEG experiment might affect the estimation of an early auditory event-related fields (ERF) waveform at temporal sites, similar frontal activity induced in an fMRI experiment will not affect the assessment of the signal recorded from voxels in distal regions such as the auditory cortex.

2.7.2 Interaction effects in the frequency domain

Calculating a similar interaction for frequency power estimates is more problematic. When determining superadditive responses it is assumed that any non-linear summation denotes the interaction of the two sensory streams, and hence, is a measure of multisensory integration. For example, assume the following response profile:

$$A = 2 \text{ units}$$

$$V = 3 \text{ units}$$

$$AV = 6 \text{ units}$$

$$\text{Integration effect} = 6 - (2 + 3) = 1$$

Here, the response to the bimodal presentation AV cannot be predicted by the summation of the unimodal inputs alone, hence the difference (1) is thought to be related to multisensory integration. Frequency responses, however, are calculated by band pass filtering the EEG signal, squaring all samples to get a power estimate and then averaging over all trials (Niedermeyer, 2005). This process actually introduces a "non-linear" step of squaring into the process. Calculating the linear interaction as above could then result in erroneous estimates of the integration effect:

$$A = 3 \text{ units; squared} = 9$$

$$V = 3 \text{ units; squared} = 9$$

$$AV = 6 \text{ units; squared} = 36$$

$$\text{Integration effect} = 6^2 - (3^2 + 3^2) = 18$$

From the calculation it is quite evident that the EEG response to AV is a direct summation of A and V, showing that when converting the amplitude into power estimates the linearity assumption is violated by the squaring procedure. Simply taken the square root of the power after squaring might seem like an appropriate solution. In order to assess this possibility we carried out a simulation to evaluate the criterion of assumed linearity under different phase relationships (MatLab®, Mathworks, Inc.).

Figure 2-3 and 2-4 show simulated data with perfect and anti-phase phase relationships between A and V signals.

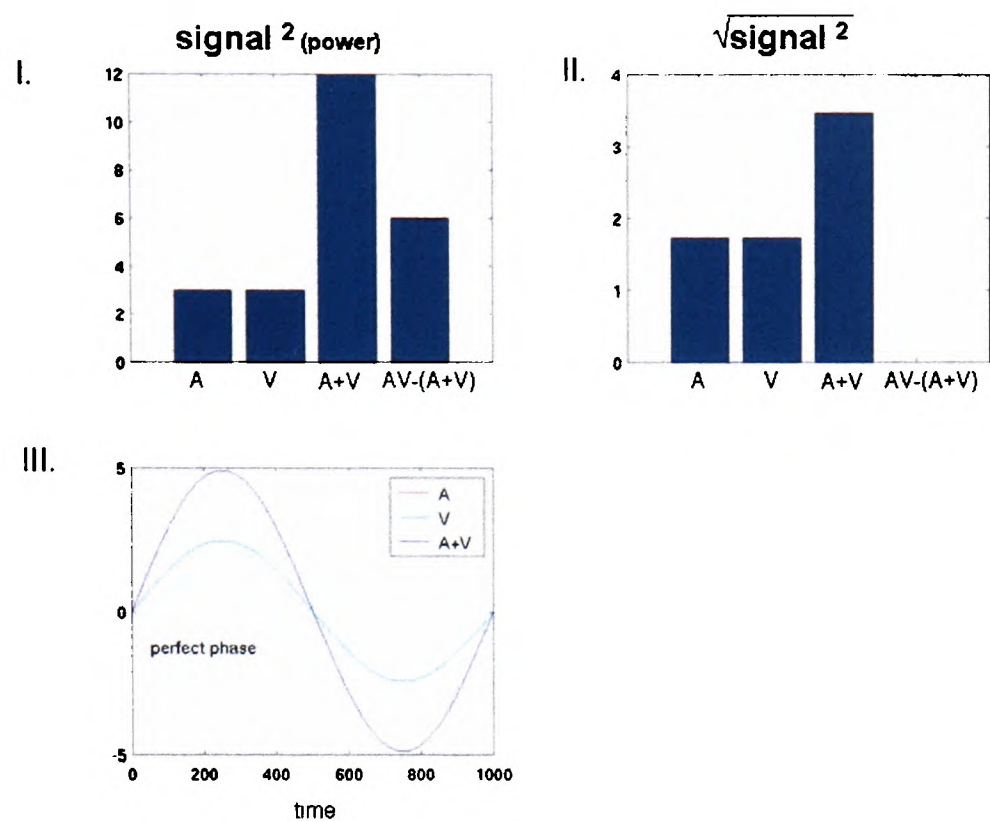


Figure 2-3: Interaction effects during perfect phase relationship

The bottom graph shows the phase relationship of the simulated sinusoidal signals for condition A and V, and their sum A+V. The top graphs (I.) show averages from signal processing to gain power estimates common in EEG frequency analysis (i.e. individual trials were squared, and (II.) the same analysis but here the square root was taken from the individual power estimates before calculating the 'interaction' effect. A+V shows the sum of signals for each approach, whereas AV - (A+V) shows the 'interaction effect'.

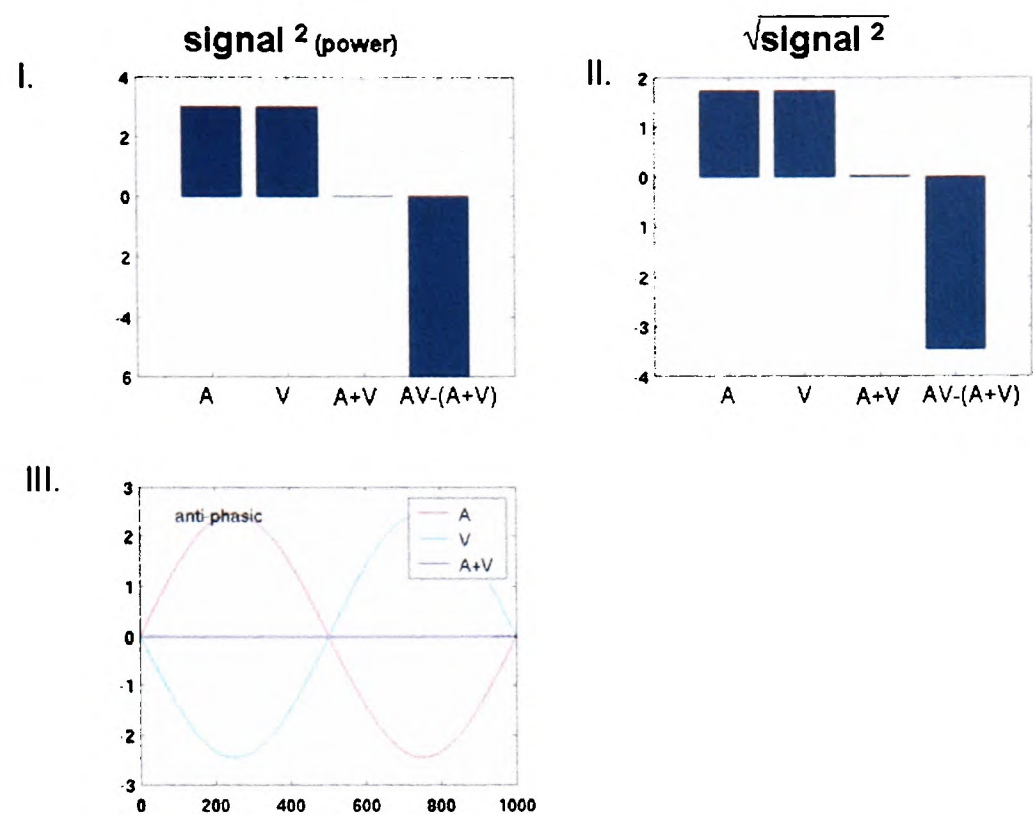


Figure 2-4: Interaction effects during anti-phasic relationship
(see Figure 2-3 for legend)

If $AV-(A+V) = 0$ (the right-most bar of the top right chart (II.) in both figures), the summation is linear and therefore the interaction effect can be estimated with linear methods. However, if $\text{diff} \neq 0$, this indicates that the linearity criterion is violated and the linear interaction effect cannot be calculated.

From the simulation we see that the interaction effect cannot be calculated based on power values (I.), irrespective of the phase relationship. The data also show that the square root of the estimated power of the sum of the two signals depends on their phase relationship. If the two signals are in perfect phase alignment in each trial (i.e. zero deg phase difference; Figure 2-3, III.), the amplitude (i.e. square-root of the power) depends linearly on the amplitude of the two signals. In this case the interaction effect can be correctly estimated. However, if they are in perfect anti-phase (180 degrees phase difference; Figure 2-4, III.), they will cancel out and violate the linearity assumption. A 'random' phase relationship will violate the assumption to varying degrees. Since there is at present no neural evidence for assuming a perfect

phase relationship between A and V conditions, we conclude that none of the above methods is suited for estimating the interaction effect in the frequency domain.

In a yet unpublished attempt, the approach of Senkowski et al. (2005) was to apply bootstrapping randomization techniques to frequency domain data in order to detect interactions. Even though this approach avoids any issues arising concerning phase relationships, it does not consider the noise in the measurement of A, V, and AV trials, and therefore their method systematically overestimates the A+V power in respect to the AV power. In conclusion, to date there is no theoretical basis for estimating interactions in the frequency domain when applied to neuroimaging data. Given the likely role of cortical oscillations in multisensory integration, future research should certainly consider this problem.

2.7.3 Bimodal contrasts

An alternative to the unequal subtraction problem in the neuroelectric/magnetic methods is a congruency interaction paradigm to determine AV integration dynamics. In this kind of experimental design, stimuli are always presented bimodally with congruent and incongruent features in one stimulus dimension (e.g. same/different spatial location, matching/non-matching AV speech) (Ojanen et al., 2005). With this kind of experimental design, task-related activity that is present in all conditions and is not related to sensory integration (e.g. CNV, P3) is effectively canceled out; and thereby the confounding effects of unequal subtraction discussed by Teder-Salejarvi et al. (2002) (see above) are avoided.

2.8 Conclusion

The field of cognitive neuroscience has benefited enormously from the development in non-invasive imaging techniques. Recent applications of neuroimaging to questions of multisensory processing in humans have employed different strategies to detect multisensory convergence. In the forthcoming

experimental sections we will use and evaluate different criteria in defining multisensory convergence during auditory-visual speech perception by using paradigms to detect integration on both the phonetic and non-phonetic level. Furthermore, we will employ a multimethod approach and tackle the same neural processes with different imaging techniques to gain a more complete understanding of the cortical processing of auditory-visual speech. Besides answering a multitude of questions regarding crossmodal processing, this approach will also tell us more about the relationship of signals from different neuroimaging techniques.

3 Event-related fMRI of auditory-visual speech perception

Previous fMRI studies have identified several cortical structures involved in auditory-visual speech integration. Most of these have used block design paradigms which may have introduced attention and context-related confounds. In this section we will use high magnetic field strength (3T) and an event-related paradigm to reduce the likelihood of such confounds while attempting to detect neuronal correlates of auditory-visual processing and integration. The results from this section will also be used to inform MEG source solutions in a subsequent section and will be compared and validated with electrophysiological recordings from iEEG.

3.1 Introduction

Recent functional imaging evidence indicates that multisensory integration affects higher-level association cortices, such as STS, and may also recruit ‘sensory-specific’ areas of the brain. Calvert et al. (2000) used fMRI to measure brain activity when subjects were presented with bimodal, audio-visual versions of a story, where visual speech gestures were presented to the subject via a visual projector and the voice of the speaker was heard through earphones. BOLD signal changes to bimodal stimulus presentations were compared between a congruent and incongruent condition, i.e. in one condition visual speech gestures matched the auditory speech signal and in the other condition the auditory stimulus timeline was shifted to a sufficient degree to disrupt audio-visual synchrony. The congruent conditions, which presumably lead to more multisensory integration, activated significantly more the STS and the inferior parietal lobule, as well as areas that are usually regarded as unisensory, such as occipital visual cortex and temporal auditory cortex.

In a study using a combination of unimodal and bimodal speech stimuli, (Calvert et al., 1999) found a large network of areas preferentially activating to bimodal vs unimodal stimulation. Findings included increased activation in secondary auditory cortex (BA 42) when contrasting $AV_{con} > A$, and increased activation in secondary visual cortex (BA 18) when contrasting $AV_{con} > V$, indicating a crossmodal recruitment of sensory specific cortex during AV speech perception.

Applying criteria from single-unit electrophysiology (Stein and Meredith, 1993) to fMRI BOLD data, Calvert et al. (2001) found super-additive activation to bimodal AV input compared to the linear sum of the uni-modal inputs ($AV_{con} > (A+V)$). This study used simple, non-speech stimuli and found evidence for cortical AV interactions in STS, visual cortex (BA 17/18), intraparietal sulcus, insula, the frontal lobe. All of these areas, except visual cortex, also showed response depression to incongruent (i.e. temporally asynchronous) bimodal presentation. Using the same experimental strategy with speech stimuli, Calvert et al. (2000) identified an area that showed superadditive responses with concurrent response depression to incongruent stimulation which was located to the left STS.

However, others report conflicting results. For example, using conditions of unimodal silent lipreading and bimodal synchronous and asynchronous AV speech (Olson et al., 2002) found that speechreading activated STS. A region of interest analysis of the STS and parietal areas, however, did reveal a difference between bound and unbound AV speech in the right claustrum but not in STS. Jones and Callan (2003) measured brain responses with fMRI to AV speech integration during a phoneme categorization task and found that visual stimulation had a strong influence on speech perception. This effect was positively correlated with activity of visual motion processing areas in the left occipital-temporal region, but not STS. This led the authors to propose that auditory input modulates visual motion processing to affect perception. Lewis et al. (2000) measured BOLD signal changes with fMRI

during an AV visual motion discrimination task. They found active regions in the intra-parietal sulcus (IPS) and fronto-parietal network, but not in the STS. Similarly, using non-speech stimuli. Bushara et al.(2001) used PET to detect neural correlates of AV asynchrony detection and found that cross-modal binding was associated with higher activity in insula/frontal operculum, dorsolateral and medial prefrontal cortex, posterior parietal cortex, SC, and thalamus. In summary, the neural substrates of auditory-visual crossmodal processing and integration, particularly between voice and lip movements and concerning the role of STS and primary sensory cortices, remain controversial.

It is worth noting that all of these studies employed block designs, where long periods of stimulation are contrasted with equally long, 'passive' periods (Bandettini and Wong, 1997). This type of design is by far the most commonly used in fMRI research and consists of ON/OFF periods lasting between 20s and 1 min. This means that subjects are continuously presented with stimuli of the same trial class (e.g. auditory alone) for long durations and that the final result is an averaged response over this time window. This has several undesirable implications for the interpretation of the results. As a contrast, event-related (ER) studies measures the neural activity of interest to short and discrete stimuli, with relatively short inter-stimulus intervals of events from different trial classes (Buckner et al., 1996). The advantage of ER designs is the random intermixing of events of different types, as is standard in psychological and electrophysiological studies. This means that the response to any one event is not systematically influenced by prior events, nor confounded by differences in the subject's perceptual and cognitive state. Such "state effects" are not trivial: Johnson et al. (1997), for example, showed that ERPs to "old" and "new" words in a memory test differed according to whether the old and new words were blocked or intermixed. This can also have effects on multisensory experiments involving block designs, where the subject can predict the stimulus modality or experimental manipulation

already after the first trial and can therefore ‘adjust’ to the task quite quickly, e.g. focus attention to only one stimulus modality. A further advantage of ER designs for fMRI is that their design is similar to that used in ERP and MEG studies (Regan, 1989; Hämäläinen et al., 1993) and therefore are better suited for comparing and integrating these imaging methodologies (Dale and Halgren, 2001). For these reasons we are using an ER design in the present investigation.

There is also no clear consensus on the criteria for defining areas of multisensory integration with fMRI (Calvert and Thesen, 2004; Laurienti et al., 2005). This may in part be a reason for the discrepancies in the literature. In this section we will systematically investigate the effects of different statistical criteria for identifying multisensory integration. In a later section, these results will be compared to outcomes from whole-head MEG and direct intracranial recordings from the putative and still somewhat controversial AV integration regions of the superior temporal lobe. This approach offers more face validity compared to similar attempts using fMRI results only (Beauchamp, 2005).

The aim of this section is to identify, with fMRI, cortical areas involved in speechreading, auditory speech perception and auditory-visual speech perception, with a special focus on describing multisensory integration of AV speech. The present investigation builds on the studies of Calvert and colleagues (Calvert et al., 1997; Calvert et al., 1999; Calvert et al., 2000) and uses event-related fMRI methods for increasing the external validity of the stimulation and to minimize attentional confounds that arise due to the use of blocked designs. A further attempt is made to keep the experimental design as similar as possible between the different imaging methodologies, but with the aim to optimize the design parameters for each imaging modality (Dale and Halgren, 2001).

3.2 Methods

3.2.1 Stimuli

The stimuli employed in this initial fMRI experiment were designed to be applicable to the forthcoming planned MEG and EEG studies. Since we were especially interested in probing the earliest crossmodal influences of AV speech perception we choose non-semantic phonetic stimuli whose integration is likely to happen at a pre-lexical, phonetic categorization stage (Summerfield and McGrath, 1984; Summerfield and Assmann, 1987; Summerfield, 1991) (MacSweeney et al., 2000).

The same 3 audio stimuli /da/, /gi/, /tu/) and two visual stimuli tokens (/da/, /tu/) were used throughout the experiment. Stimuli consisted of audio-visual consonant-vowel (CV) speech utterances of various congruent and incongruent combinations, as well as their unimodal parts. Stimuli were divided into targets and non-targets and occurred with different frequencies (/da/ [non-target (standard): 80 % probability]) or /tu/” [target: 20 % probability]).

Depending on condition, stimuli were presented either in the form of mouthed lip movements (V) or auditory speech sounds (A) alone, or with the congruent combination of the audio-visual components (AVcon). We also used incongruent AV combinations since they allow for a better assessment of the speech specificity of crossmodal effects. For a detailed list of non-target stimulus conditions see Figure 3-1. Subsequent data analysis was performed only on non-targets.

Condition	Auditory	Visual
Unimodal		
A	/da/	-none-
V	-none-	/da/
Bimodal congruent		
AVcon	/da/	/da/
Bimodal incongruent		
AVincon(V)	/da/	gurning
AVincon(A)	/gi/	/da/

Figure 3-1: Stimulus conditions of the frequent standards (80 % occurrence) to which data were analyzed and their auditory and visual content. Infrequent target (20 %) auditory and visual consisted of /tu/ tokens and responses to these were not analyzed.

Visual stimuli depicted the lower face of a real female person and were recorded and played on digital medium at 29.97 fps. The face stimulus was restricted to the lower face to ensure attention was focused the articulators in AV speech (Summerfield, 1992) and to minimize the influence of possible confounding factors such as gaze and facial identity (Campbell et al. 1996).

Audio signals were digitised at 44.1 kHz at 16 bit and adjusted for average RMS power. The audio stimulus /da/ consists of the aspirated plosive coronal consonant /d/ with place-of-articulation, i.e. the point of contact, where an obstruction occurs in the vocal tract between an active (moving) articulator (part of the tongue or lips) and a passive (stationary) articulator, being the apical part of the tongue with the superior alveolar ridge on the upper roof of the mouth between hard palate and upper teeth. The consonant is followed by the open vowel [a] with a very low position of the

tongue in the mouth. See Figure 3-2 for a spectrogram showing the time-frequency characteristics of the CV stimulus /da/.

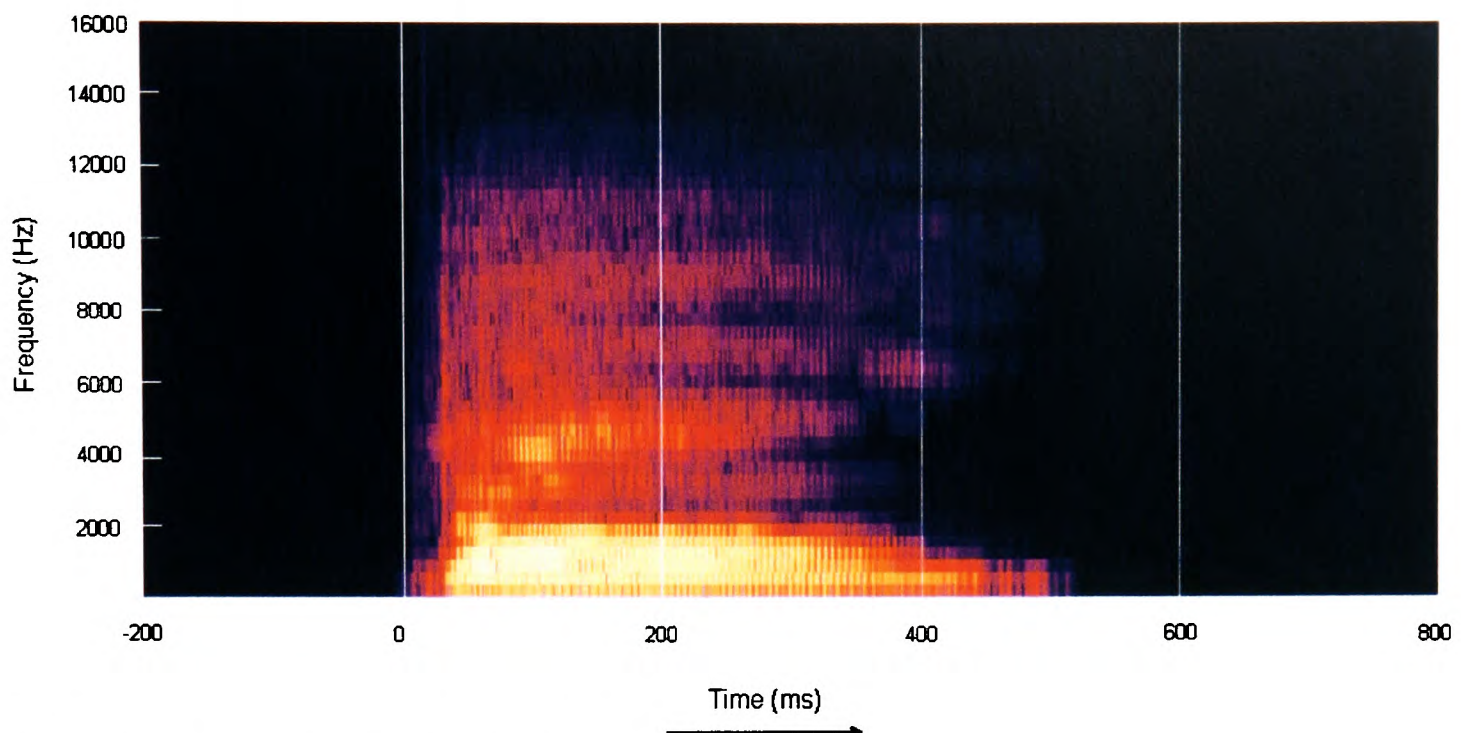


Figure 3-1: Spectrogram of CV sound /da/.

The image shows power in the frequency range from 1 Hz to 16000 Hz over time. Time point zero marks the onset of auditory stimulation. Power amplitude is colour-coded with higher amplitudes being represented by lighter (yellow) colours.

Incongruent combinations of articulatory sound and gesture consisted of visual /da/ or a face producing mouth and jaw movements of similar extent and rhythm, but without opening the mouth and representing articulatory content (gurning) (Campbell et al., 2001), AVincon(V) and speech sound (/gi/ or /da/) (see Figure 3-2 for a detailed description of the stimulus conditions). These CV syllables were chosen partly because they are easily discriminated both acoustically and visually. Incongruent stimuli pairs were perceived as conflicting and not fused at the phonetic level.

Integration of auditory and visual CV syllables is assumed to be based on common features that mediate integration, such as location, temporal coincidence and the correspondence of phonetic content (Stein & Meredith, 1993; Treisman, 1996; Treisman, 1998). Simultaneously presented auditory and visual stimuli coming from

the same spatial location are naturally bound to AV percepts (Stein and Meredith, 1993). The integration of stimuli with complex features, such as those present in AV speech are likely to depend on other mechanisms and to be dependent on their AV congruence on the phonetic dimension. Therefore, the integration mechanisms and properties of AV speech stimuli is likely to feature both those of simple AV stimuli and those related to phonetic matching. Therefore, we hypothesized that differences between congruent and incongruent pairs reflect audiovisual interactions at the phonetic level, whereas contrasts involving unimodal conditions also reflect non-phonetic modulations between acoustic and visual stimuli.

Natural AV speech is an ecologically valid stimulus and is characterized by the temporal precedence of visual speech, as the movement of the facial articulators usually precedes the onset of the acoustic stimulus by tens to hundreds of milliseconds. Each stimulus featuring a visual component had a duration of 2000 ms. At the beginning of each presentation a static face was presented for 734 ms before the lip started to move. The sound was presented 1167 ms after face on-set (Figure 3-3). Between stimuli a luminance and colour-matched uniform background with a black fixation cross was visible on the screen. The still face period was chosen to avoid contamination of the evoked components of interest by face-evoked event-related potentials (ERP). In order to define the exact onset of events of interests, the start of the lip movement was recorded through a photo-sensitive diode and a white marker in the upper left corner of the relevant video frame (unseen to the subject). Furthermore, the audio waveform of each stimulus presentation was recorded from the audio channel of the sound presentation system. A post-hoc detection algorithm was applied to retrieve the onset timing of the auditory stimulus.

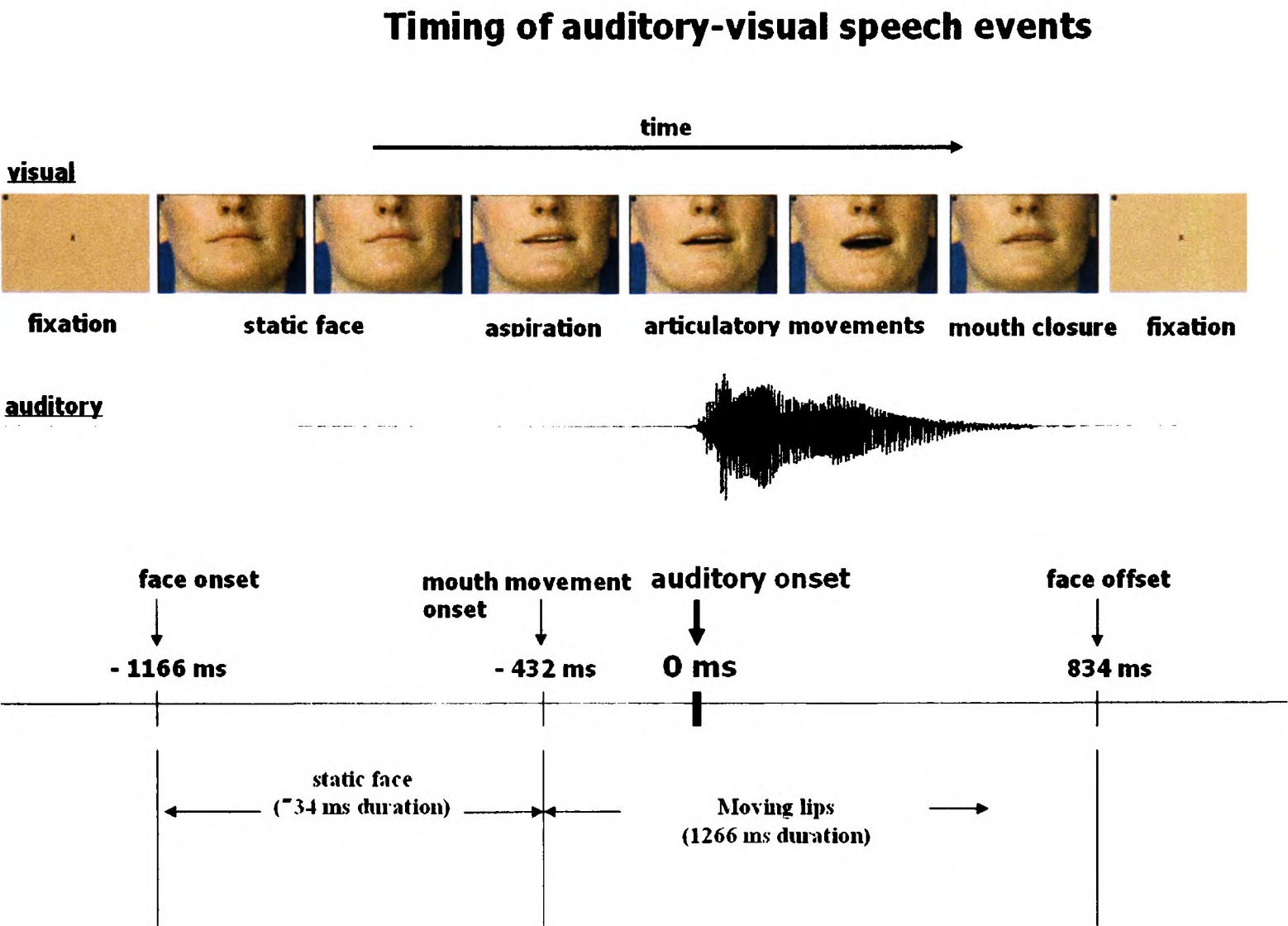


Figure 3-2: Time-course of stimulus presentation.

Timing is relative to the onset of the auditory stimulus. Between trials subjects saw a luminance matched, skin-colored blank screen with a fixation cross. Then, a static face appeared and after 734 ms (432 ms before auditory onset) the mouth began to move. 834 ms after the auditory onset the face disappeared and was replaced by a fixation screen until the beginning of the next trial.

3.2.2 Task

Subjects were instructed to maintain central fixation between trials and to detect the occurrences of the infrequent /tu/ in either modality by pressing a response

button with the right index finger as quickly as possible. This procedure guaranteed that subjects monitored and paid attention to channels of both modalities. In subsequent analyses of the data, only responses to presentations of the frequent /da/, which are free of button-response related activation, are considered.

3.2.3 Subjects

Data were recorded from 21 (13 female, mean age = 24 years, stdev = 2.3 years) native English speakers. All subjects were right-handed, without previous history of psychological or neurological disorders and had normal or corrected to normal eyesight. The study received ethical approval from the Central Oxford Research Ethics Committee and informed consent was obtained from all subjects.

3.2.4 Procedure

Subjects lay inside the scanner bore in a supine position. A projector displayed the visual stimuli onto a back-projection screen inside the scanner room while the subject was wearing prism glasses. The sound was presented with a specially engineered, MR-compatible sound system that delivers high-quality sounds through electrostatic headphones combined with standard industrial ear-defenders for passive attenuation of the scanner sound (Palmer et al., 1998). At the beginning of each experiment the sound intensity was adjusted to a comfortable level where subjects were able to discriminate the stimuli over the ambient scanner noise. It needs to be noted that the presence of scanner noise throughout the whole experiment has influences on the interpretation of the activation results. The T2* weighted voxel intensities measured by the MR scanner represent arbitrary units which have only a relative connection to the BOLD contrast, namely when comparing two states. Here, the passive auditory state contains scanner noise only and the active auditory state contains scanner noise plus an auditory speech signal. Therefore, the resulting contrast represents auditory activation related to speech signal processing alone

independent of scanner noise. Approaches circumventing this problem and allowing that the auditory stimulus is presented during silence exist (e.g. ‘sparse sampling’ (Amaro et al., 2002)) but were not chosen because it would have increased the total time of the experiment considerably.

During the experiment, 40 stimuli of each standard condition (/da/; 80 % frequency) and 8 stimuli of each target condition (/tu/; 20 % frequency) were presented in random sequence across 4 runs in an event-related fashion. Inter-stimulus intervals were pseudo-Poisson distributed and ranged from 6 s to 15 s in 1 s steps and to sample the BOLD signal with a 1 temporal resolution.. Condition-to-ISI assignment was counter-balanced. The total duration of each run was 11 min 12 s. Parameters of the Poisson-approximated distribution of ISI’s used in the event-related fMRI experiment were: min = 6s, max = 15 s, lambda = 9s, n = 240.

3.2.5 Scanning Protocol

Subjects heads were scanned using echo-planar imaging (EPI) on a 3T system. The scanning parameters were: repetition time (TR) = 3 s with a field-of-view of 192 x 256 mm and an in-slice resolution of 3 x 4 mm, voxel size of 3 x 4x 5 mm. A total of 896 T2* weighted EPI volumes were taken over 4 presentation runs, each lasting 11 min 12 s (224 volumes/run). After the functional image acquisition a T1-weighted volume [IR 3D TurboFLASH with a voxel-size 1x1x1.5 mm (128 slices)] was acquired from each subject to aid in anatomical definition and co-registration (TR = 15 ms, TE = 5 ms, TI = 500 ms, flip angle = 12°, FOV = 192 x 256).

3.2.6 Data Analysis

Analysis of the time-series data was carried out on 3 levels using FEAT (FMRI Expert Analysis Tool) Version 5.04, part of FSL (FMRIB's Software Library). All responses to the infrequent target were modelled together as a separate explanatory variable; however, only responses to the frequent target /da/ were

subjected to further statistical analysis. Event-related modelling was based on the onset timing of the auditory stimuli, or in case of V stimulation, the diode marker of the frame at which the auditory stimulation would have happened. At the first level, pre-statistics processing steps on the individual runs (minus 5 dummy volumes) included slice-timing correction using Fourier-space time-series phase-shifting and motion correction using MCFLIRT (Jenkinson et al., 2002). A Gaussian kernel of FWHM (full width half maximum) of 5mm was used to spatially smooth the data, which was subjected to a mean-based intensity normalisation of all volumes by the same factor and highpass temporal filtering (Gaussian-weighted LSF straight line fitting, with $\sigma=7.5s$). The high-resolution structural and all functional images were registered to the MNI152 standard space (Montreal Neurological Institute) (Collins et al., 1994) using FLIRT (Jenkinson et al., 2002).

Statistical analysis was carried out using FILM (FMRIB's Improved Linear Model) with local autocorrelation correction (Woolrich et al., 2001). Each of the 5 standard event types was modelled as a separate explanatory variable. At higher levels a mixed effects analysis employing FLAME (FMRIB's Local Analysis of Mixed Effects) (Woolrich et al., 2004; Woolrich et al., 2004) was first used to combine conditions across runs for each subject. Then, the 2nd level results were combined to generate group mean statistical images for every defined contrast. Z statistic images were thresholded using clusters determined by $z > 2.3$ and a corrected cluster significance threshold of $P = 0.05$ (where not explicitly stated otherwise) (Worsley et al., 1992) (Friston et al., 1994; Forman et al., 1995). Thresholded activation images were then rendered onto a high-resolution structural image in MNI standard space for visualization of the group activation data.

3.2.6.1 *Statistical criteria for defining multisensory integration with fMRI*

Previous studies have used different strategies to identify multisensory integration sites with fMRI. For reviews, see (Calvert. 2001; Calvert and Thesen, 2004).

Criterion:

1. An area activates to both unimodal conditions ($A \cap V$)
2. Response enhancement, where an area responds more to bimodal stimulation than to either of the unimodal stimulations. $(AV_{con} > A) \cap (AV_{con} < V)$.
3. $AV_{con} > (A + V)$ super-additivity
4. $AV_{con} < (A + V)$ sub-additivity
5. Mean criterion: $AV_{con} > \text{mean}(A+V)$ (Beauchamp et al., 2004)
6. $AV_{con} \neq AV_{incon}$
7. An area that activates only to multisensory stimulation (AV_{con}) but not to unisensory stimulation (A and V).

Since there is so far no gold-standard for identifying multisensory sites with fMRI (Calvert and Thesen, 2004 ; Beauchamp, 2005), we have calculated multisensory effects based on all individual criteria, except for criterion 7. This can be readily dismissed since activation can easily be a thresholding phenomenon and may not have neurophysiological reasons, a claim that can also be true for criterion 1. The assumptions inherent in each contrast and their implications are discussed in this section and results are compared and validated with electrophysiological methods in section 8.

3.3 Results

3.3.1 Behavioural

Subjects correctly identified 96.9 %, plus/minus 1.5 % (mean $\pm$ SEM) of the target stimuli, indicating that they maintained attention to the CV syllables of both modalities and were able to discriminate between /da/ and /tu/ vowels within the fMRI testing environment.

3.3.2 Auditory speech perception

A widely distributed set of brain regions in occipital, parietal, temporal and frontal lobes showed a significant modulation of the MR signal time-series time-locked to the onset of each trial. The upper left quadrant of Figure 3-4 shows activation to the auditory condition minus baseline. Auditory speech activated large parts of bilateral temporal cortex, including the inferior & posterior STG (BA 22), MTG (BA 21), and auditory cortex, incorporating at lower thresholds of $z > 2.3$ both primary (A1) and secondary (A2) zones in the medial and lateral portions of Heschl's gyrus (BA 41 & 42). Further bilateral activation clusters included the inferior frontal gyrus (BA 44 & 45), insula, thalamus, occipito-temporal regions (BA 37, MT), and in the occipital lobe parts of the lingual gyrus including activation in primary visual cortex (BA 17). See Table A (Appendix) for a detailed listing of activated areas in each condition with a volume measure of the extent of the activation and the peak z -value within the given structure and its MNI coordinates. By examining the MR response from different regions of interest (ROI) we can assess the sensitivity of different brain regions to different trial types.

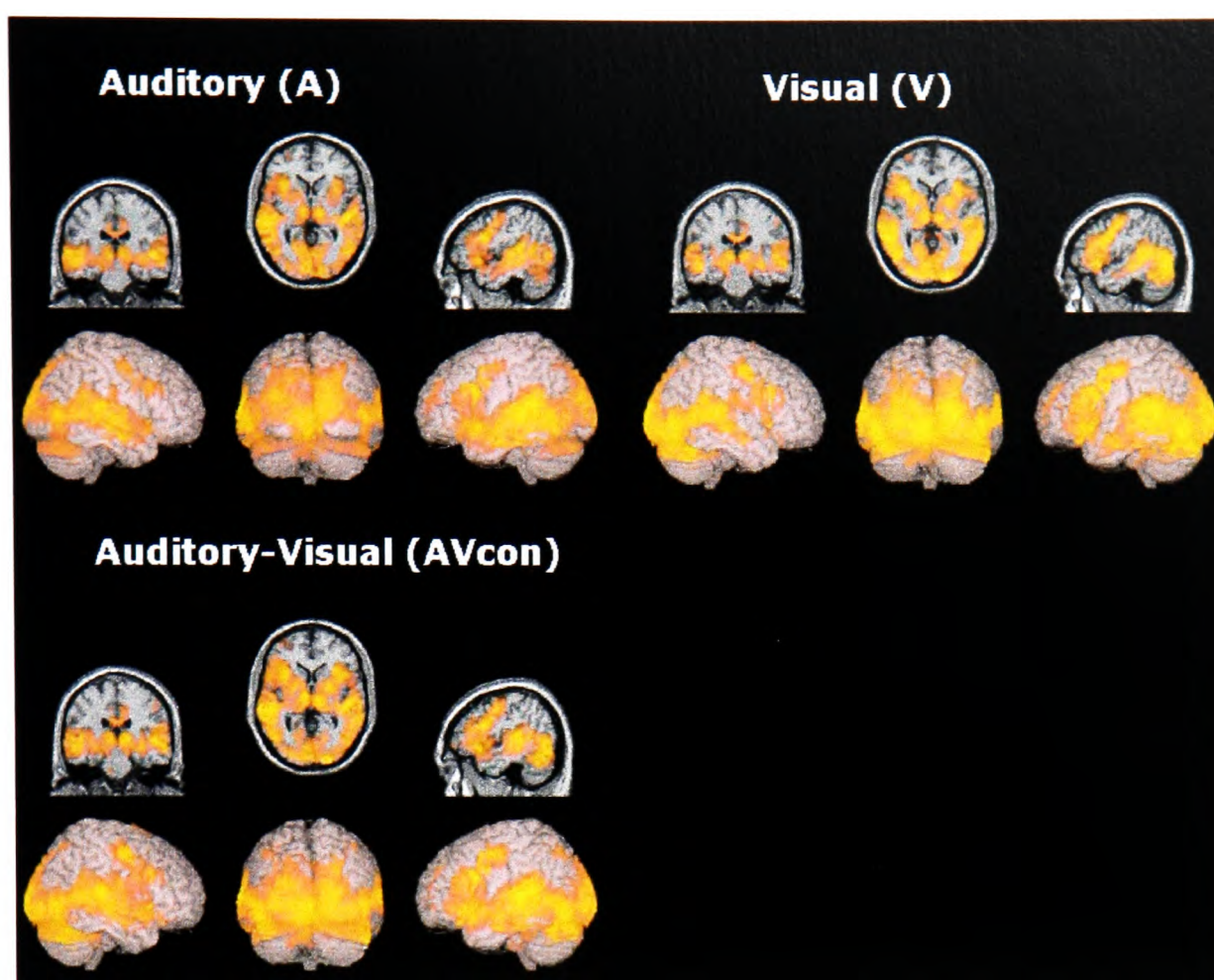


Figure 3-3: FMRI activation to A, V & AVcon

Activation compared to baseline in 21 subjects, $z > 2.3$, $p = .05$. (See Table a, Appendix for a detailed listing of activated areas, z-values and coordinates)

3.3.3 Speechreading

The upper right quadrant of Figure 3-4 shows activation to the visual only condition minus baseline. Silent lipreading activated large parts of the occipital cortex including the lingual gyrus, cuneus and occipital pole, encompassing a large number of visual areas including the primary and secondary visual cortex (V1 & V2), and area MT+. Silent lipreading also activated a large bilateral network of areas in the temporal cortex involving the superior, middle and inferior temporal gyri (STG, MTG, ITG), and at a lower threshold of $z > 2.3$ primary and secondary auditory cortex on HG.

Activation was also detected in the inferior frontal gyrus (Broca's area), the precentral gyrus, thalamus and insula (see Table 1 in Appendix for a complete list of activated areas).

3.3.4 Multisensory speech perception

Bimodal AV speech activated exactly the same cortical areas as A and V combined together, and did not activate any additional areas (Figure 3-4). This result was obtained by intersecting the contrast $AV_{con} > A$ and $AV_{con} > V$. Even though no additional area was activated during the AVcon condition, quantitative differences between responses to silent lipreading, auditory and bimodal speech were observed in a number of areas.

3.3.5 Multisensory integration

3.3.5.1 *Non-phonetic interactions*

A number of criteria have been proposed for identifying cortical sites as multisensory in neuroimaging studies (Calvert & Thesen, 2004; Calvert, 2001) but there is so far no general agreement in the literature. Results obtained from these strategies can vary considerably and can lead to different interpretations from the same data set. We present fMRI data which have been analyzed using various different statistical criteria for identifying multisensory areas, thereby allowing a direct comparison of the results obtained from each strategy. A detailed understanding of the effects of different statistical criteria on the data then allows researchers to evaluate multisensory imaging experiments more correctly and helps to synthesize results obtained with different approaches.

3.3.5.2 $[A \cap V]$

(Criterion 1)

Defining multisensory areas through intersection is achieved by identifying only those areas as multisensory which are commonly activated by the two unimodal stimulations (Bremmer F. et al., 2001).

We first calculated the intersection ($A \cap V$) from the unimodal contrasts and contrasted the unimodal conditions A vs V thresholded at $z > 2.3$ ($p < 0.05$) to determine areas which are preferentially responsive to A and V conditions. It should be noted that the division of cortex into unisensory responsive and bisensory responsive according to this criterion depends to a large degree on the threshold chosen. A low threshold of $z > 2.3$, for example, categorizes most of the occipital and temporal lobes as bisensory responsive. Also, areas indicated in blue (Figure 3-5) may still be responsive to visual stimulation; however the area is significantly more sensitive to auditory stimulation. As such, this contrast is best seen as an indicator of the preference or dominance of a certain area for processing A, V or both A and V information. (See Appendix 1 for a detailed table of activated regions and their coordinates).

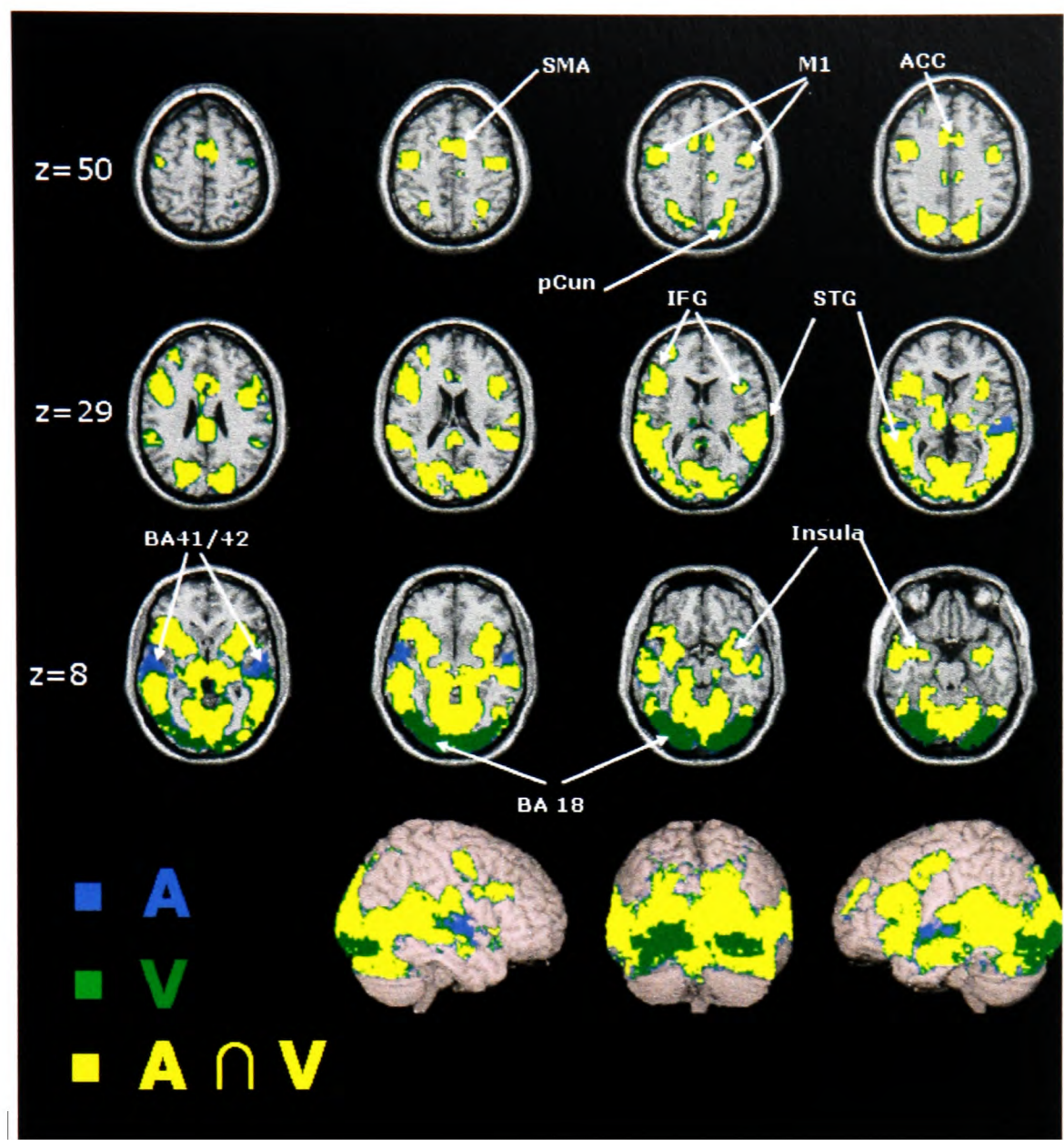


Figure 3-4: FMRI activation to A, V & A∩V
Binary activation images are color coded and categorize cortical areas based on their response dominance to auditory speech only (A>V), to speech reading only (V>A), and to both (A ∩ V). Arrows indicate brain areas: Supplementary motor area (SMA), motor cortex (M1), anterior cingulate cortex (ACC), pre-Cuneus (pCun), inferior frontal gyrus (IFG), superior temporal gyrus (STG), auditory cortex (BA41/42), and visual cortex (BA 18).

Areas commonly activated by silent lipreading and auditory speech included a wide variety of brain areas with different functional correlates. We found bimodally responsive regions on the precentral gyrus, which constitutes the primary motor cortex (M1) (Penfield and Rasmussen, 1952) and SMA. SMA is located in BA 6 and is largely involved in controlling many aspects of motor behaviour, especially motor

planning and sequencing (Tanji, 1996). Also involved in speech production, SMA has been shown to activate during silent lipreading (Paulesu et al., 2003) suggesting a role of the motor system in both speech production and perception (Fiez, 2001). Activity was also seen in the anterior cingulate cortex (ACC), located on the medial wall of the cerebral hemispheres, an area that is involved in cognitive and motor functions, especially attention and reward expectation (Isomura and Takada, 2004), and which is often reported in multisensory studies (Laurienti et al., 2003). An activation cluster was also seen in posterior cingulate cortex (PCC), an area that is related to visuo-spatial orientating and that has been shown to be involved in facial and word recognition (Bernstein et al., 2002). Both auditory and visual speech activates Broca's area on the pars opercularis and pars triangularis (BA 44&45) in the IFG in both hemispheres, but much more so on the left. Broca's area is classically considered a speech-production region also is involved in much wider language related tasks and communication behaviours (Bookheimer, 2002). Broca's area has previously been shown to be active during A and V speech perception (Campbell et al., 2001; Callan et al., 2003), including the matching of phonetic AV speech segments (Ojanen et al., 2005). The precuneus is located on the medial surface of the parietal lobe of the cerebrum, bounded posteriorly by the medial part of the parieto-occipital sulcus and anteriorly by the paracentral lobule. This structure is involved in memory processing (Schmidt et al., 2002), especially retrieval (Shannon and Buckner, 2004), irrespective of the modality of memory encoding (Buckner et al., 1996).

3.3.5.2.1 $A > V$

Areas more responsive to auditory than visual stimulation included the middle and anterior part of the superior and middle temporal gyri in both hemispheres.

3.3.5.2.2 $V > A$

Areas more responsive to visual than auditory stimulation were located bilaterally in the occipital lobe including parts of the lingual and middle/inferior occipital gyrus, cuneus and fusiform gyrus.

3.3.5.3 $[AVcon > A \cap AVcon > V]$

(Criterion 2)

This criterion defines those voxels as multisensory which respond positively to auditory and visual stimulation alone and activate significantly more to auditory-visual stimulation (Beauchamp et al., 2004). To find this pattern, we calculated the contrasts ($AVcon > A$) and ($AVcon > V$) and then identified voxels that survived a z score of 2.3 ($p=0.05$) or higher in both contrasts (Figure 3-7).

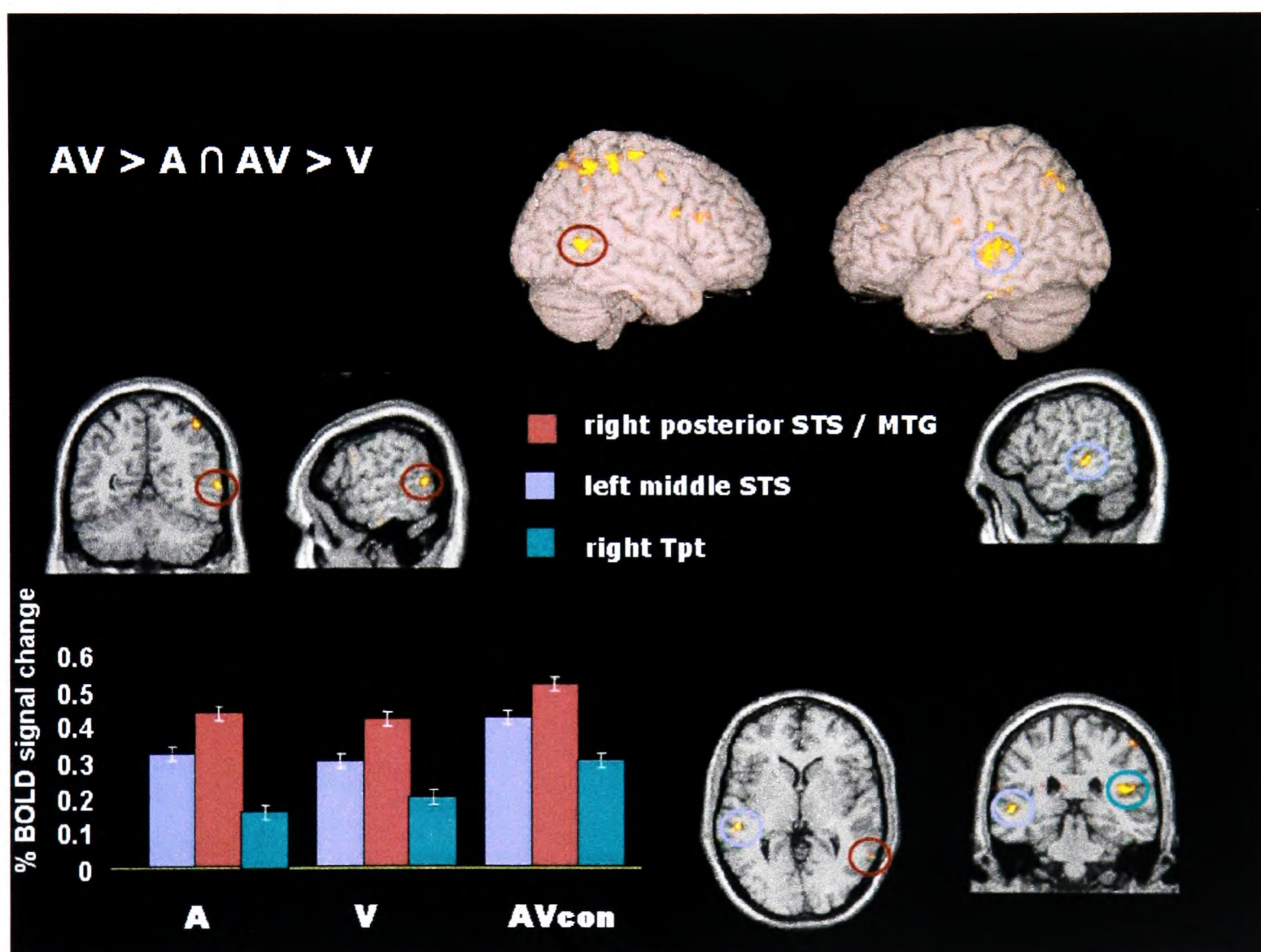


Figure 3-7: FMRI activation to $[Avcon > A \cap Avcon > V]$

Coloured circles depict the left posterior STS/middle temporal gyrus (MTG), right posterior part of the supra-temporal plane (Tpt) and the right STS. Percent BOLD signal changes for these areas (colour-coded red, green and blue respectively) are shown for each of the three stimulus conditions (chart in lower right corner, A = auditory, V = visual, AVcon = auditory-visual).

Table of [AVcon > A ∩ AVcon > V] activation (z> 2.3)				
Area	Volume (cm3)	Talairach coordinates		
		x	y	z
Inferior Parietal Lobule (R)	1.39	45	-49	63
Insula (R)	0.9	41	-23	22
Postcentral Gyrus (R)	0.47	45	-29	67
Precentral Gyrus (R)	0.45	39	-21	69
Superior Temporal Gyrus (L)	0.27	-66	-31	20
Superior Temporal Gyrus (R)	0.25	59	-31	20
Middle Temporal Gyrus (R)	0.23	57	-57	8
Superior Parietal Lobule (R)	0.21	27	-61	69
Middle Temporal Gyrus (L)	0.15	-57	-31	4
Inferior Frontal Gyrus (R)	0.15	51	1	33
Fusiform Gyrus (R)	0.1	51	-27	23

Table 3-1: Areas of activation for [AVcon > A ∩ AVcon > V]

Activation to bimodal speech exceeded both A and V responses in the left superior temporal gyrus, the left inferior parietal lobule, right claustrum and right posterior STG/MTG, as well as in the right posterior part of the supra-temporal plane (Tpt).

3.3.5.4 *AVcon > (A + V)*

(Criterion 3)

Super-additive responses, where the multisensory response is greater than the sum of the unimodal responses (*AVcon > (A+V)*), have been observed at the cellular level in the superior colliculus (Stein & Meredith, 1993). It requires activation to both unimodal conditions and more than their sum during the bimodal condition. The mean criterion and the [AVcon > A ∩ AVcon > V] criterion all need to be met as well for a voxel to fulfil the superadditive criterion. As such, it is the most stringent of the criteria presented here. In

fMRI general linear models, this effect is calculated by defining the contrast -1 -1 1 (for A, V, AVcon respectively) (Figure 3-8).

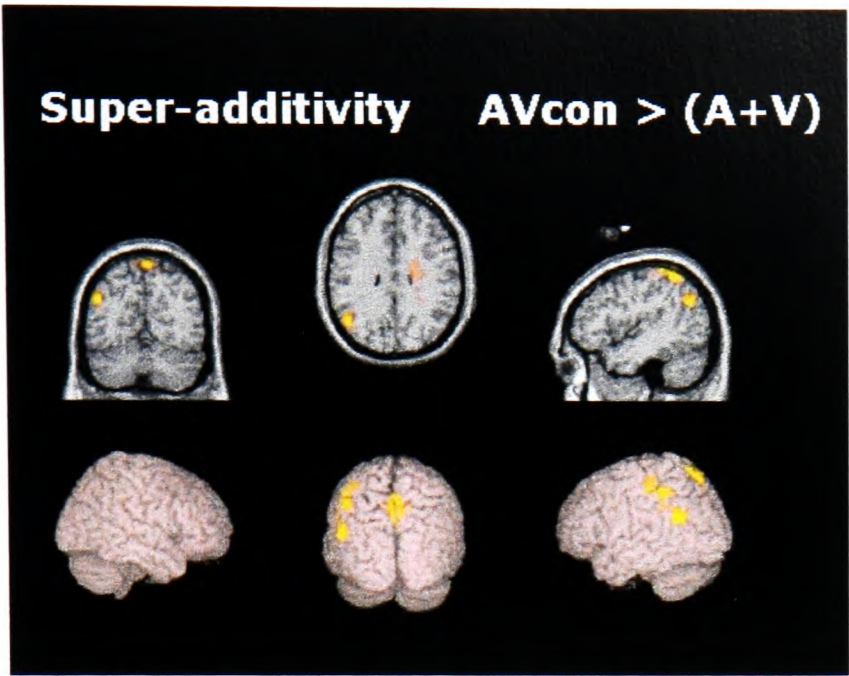


Figure 3-8: FMRI activation to [AV > (A+V)]
Activation calculated through the contrast -1 -1 1 (for A, V, AVcon respectively) without the criterion that all contrast have to show a significant activation relative to baseline.

However, inspecting the BOLD signal changes in these AG and SMG (Figure 3-9), it becomes apparent that the interaction effect is due to a higher response in the bimodal condition vs the unimodal conditions but is a result of an overall signal decrease in all conditions (see Figure 3-9). When we applied the additional criterion of $A > 0$ and $V > 0$, requiring that the uni-modal responses activate (positive z-values), the super-additive criterion did not produce any significant activation clusters.

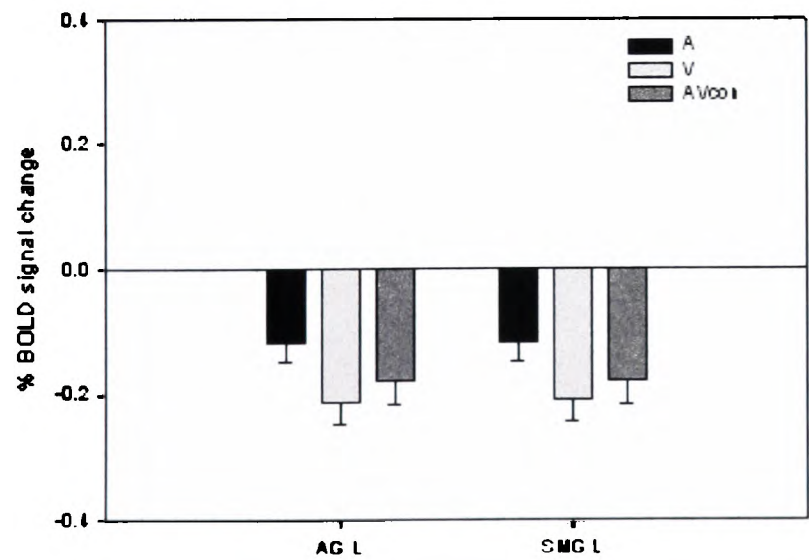


Figure 3-9: BOLD signal changes during interaction effect

For A, V and AVcon conditions in left angular gyrus (AG) and left supra-marginal gyrus (SMG).

3.3.5.5 *Sub-additivity*

(Criterion 4)

A sub-additive response shows a negative interaction effect where bimodal stimulation results in less activation than the sum of the unimodal response or AVcon $< (A + V)$. The unimodal responses added together exceed the response to the bisensory stimulation (Calvert et al., 2000). This effect is calculated by defining the contrast $[1\ 1\ -1]$ (for A, V, AVcon respectively) (Figure 3-10).

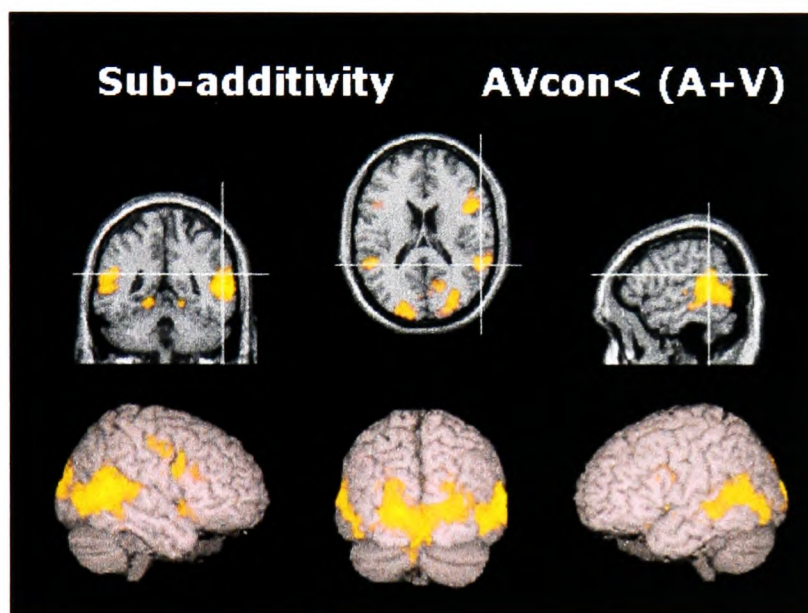


Figure 3-10: FMRI activation to $[AV < (A+V)]$

Subadditive responses, where the multisensory response is smaller than the sum of the unimodal responses.

Table of [AVcon< (A+V)] activation (z> 2.3)					
Area	peak z-value	Volume(cm3)	Talairach coordinates		
			x	y	z
Middle Temporal Gyrus (L)	5.4	6.39	-63	-45	2
Superior Temporal Gyrus (L)	5.7	5.75	-61	-45	4
Lingual Gyrus (R)	5.2	4.3	7	-83	-5
Superior Temporal Gyrus (R))	4.9	3.33	53	-47	19
Lingual Gyrus (L)	4.9	3.26	-13	-75	-1
Cuneus (R)	5.2	3.19	15	-99	15
Inferior Frontal Gyrus (L)	5.0	2.13	-45	10	23
Cuneus (L)	4.8	1.87	-1	-87	3
Middle Temporal Gyrus (R)	4.5	1.64	57	-45	6
Middle Occipital Gyrus (R)	5.4	1.58	13	-100	11
Middle Occipital Gyrus (L)	4.8	1.46	-47	-71	0
Insula (L)	4.7	0.82	-31	18	0
Culmen (R)	4.5	0.76	5	-65	-7
Inferior Frontal Gyrus (R)	5.0	0.62	35	21	-7
Inferior Temporal Gyrus (L) !	4.7	0.60	-49	-65	-4
Insula (R)	4.6	0.56	35	13	-1
Inferior Temporal Gyrus (R)	4.3	0.20	43	-65	-4
Thalamus (L)	4.3	0.19	-9	-19	8
Posterior Cingulate (L)	4.2	0.16	-10	-69	8
Lingual Gyrus (C)	4.9	0.11	0	-87	0

Table 3-2: Areas of activation for [AVcon< (A+V)]

Sub-additive responses were observed in the lateral and posterior STS bilaterally (BA 22 & BA 42). Large activation clusters were found in the primary and secondary visual cortex (BA 17 & 18), including bilateral cuneus and pre-cuneus (BA 31) and culmen. The contrast also showed bilateral thalamic activation, as well as midbrain structures, left and right parahippocampal gyrus and the enthorinal cortex. Smaller activation was evident in post and precentral gyrus of the frontal lobe, as well

as the medial frontal gyrus and medial cingulate gyrus (BA 24). The most anterior activation was seen in bilateral anterior insula (BA 13) and inferior frontal.

3.3.5.6 $AVcon > mean(A, V)$

Mean Criterion

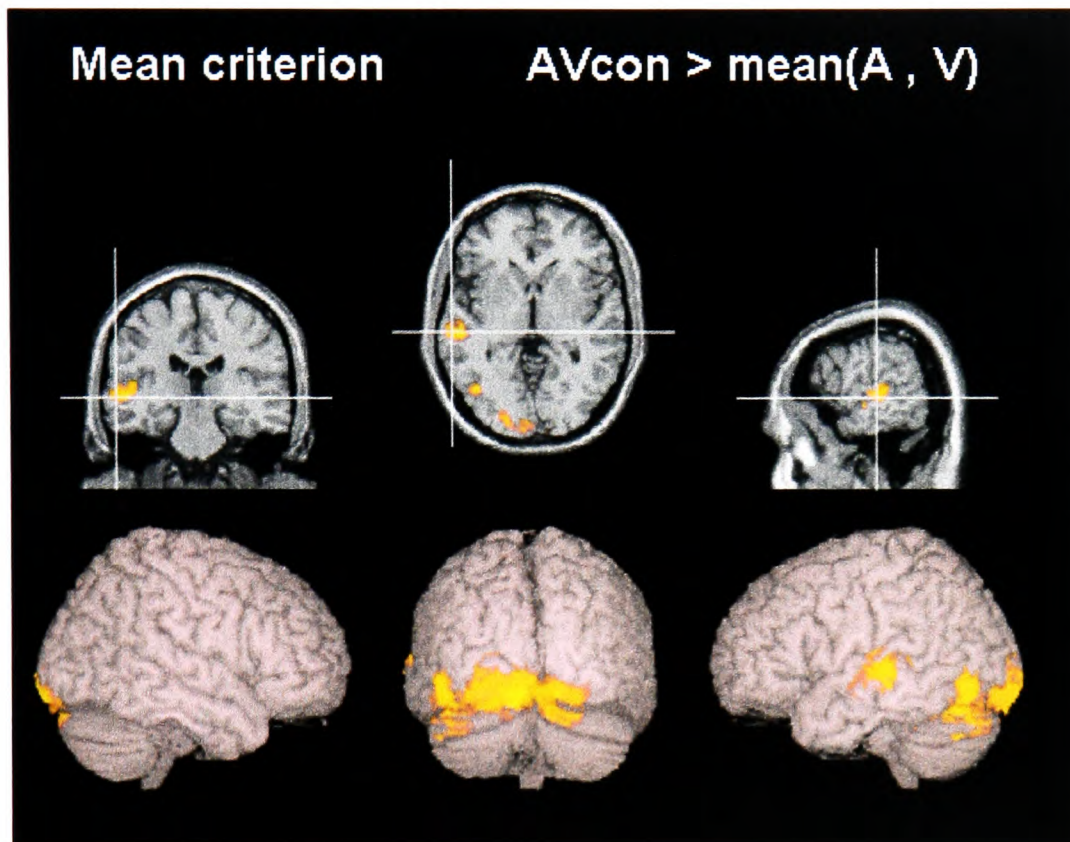


Figure 3-11: FMRI activation to $[Avcon > mean(A, V)]$

This contrast shows activation based on criterion 5. Crosshair on left STS.

Activity to the mean criterion was observed in the lingual gyrus of the occipital lobe, both in primary and secondary visual cortex (BA 17 & 18) bilaterally with more activation in the left hemisphere. A large activation cluster was also observed in the left STS region, but did not extend into HG. A left occipo-temporal region (area MT) also showed response characteristics of the mean criterion.

Table of AVcon > mean(A,V) activation, (z< 2.3)					
Area	peak z-value	Volume (cm3)	Talairach coordinates		
			x	y	z
Lingual Gyrus (L)	3.7	4.16	-9	-95	-11
Lingual Gyrus (R)	3.3	2.54	17	-97	-9
Superior Temporal Gyrus (L)	3.3	2.02	-57	-29	3
Middle Occipital Gyrus (L)	3.2	1.72	-45	-71	-11
Inferior Occipital Gyrus (L)	3.4	1.17	-11	-93	-13
Cuneus (L)	3.4	1.16	-9	-97	-8
Fusiform Gyrus (L)	3.1	0.74	-43	-69	-19
Cuneus (R)	3.0	0.62	17	-101	-7
Inferior Temporal Gyrus (L)	3.1	0.51	-49	-71	-5
Middle Temporal Gyrus (L)	3.3	0.49	-59	-29	0
Inferior Occipital Gyrus (R)	2.6	0.39	23	-95	-13
Fusiform Gyrus (R)	3.0	0.27	21	-85	-24

Table 3-3: Areas of activation for [Avcon > mean(A, V)]

3.3.6 Phonetic interactions

(Criterion 5)

No increases in BOLD signal were observed for congruent vs incongruent contrast (i.e. AVcon > AVincon(A); AVcon > AVincon(V)). No signal enhancement was found for the opposite contrast AVincon(V) > AVcon, but we found a network of areas showing an increase in activity for AVincon(A) > Avcon.

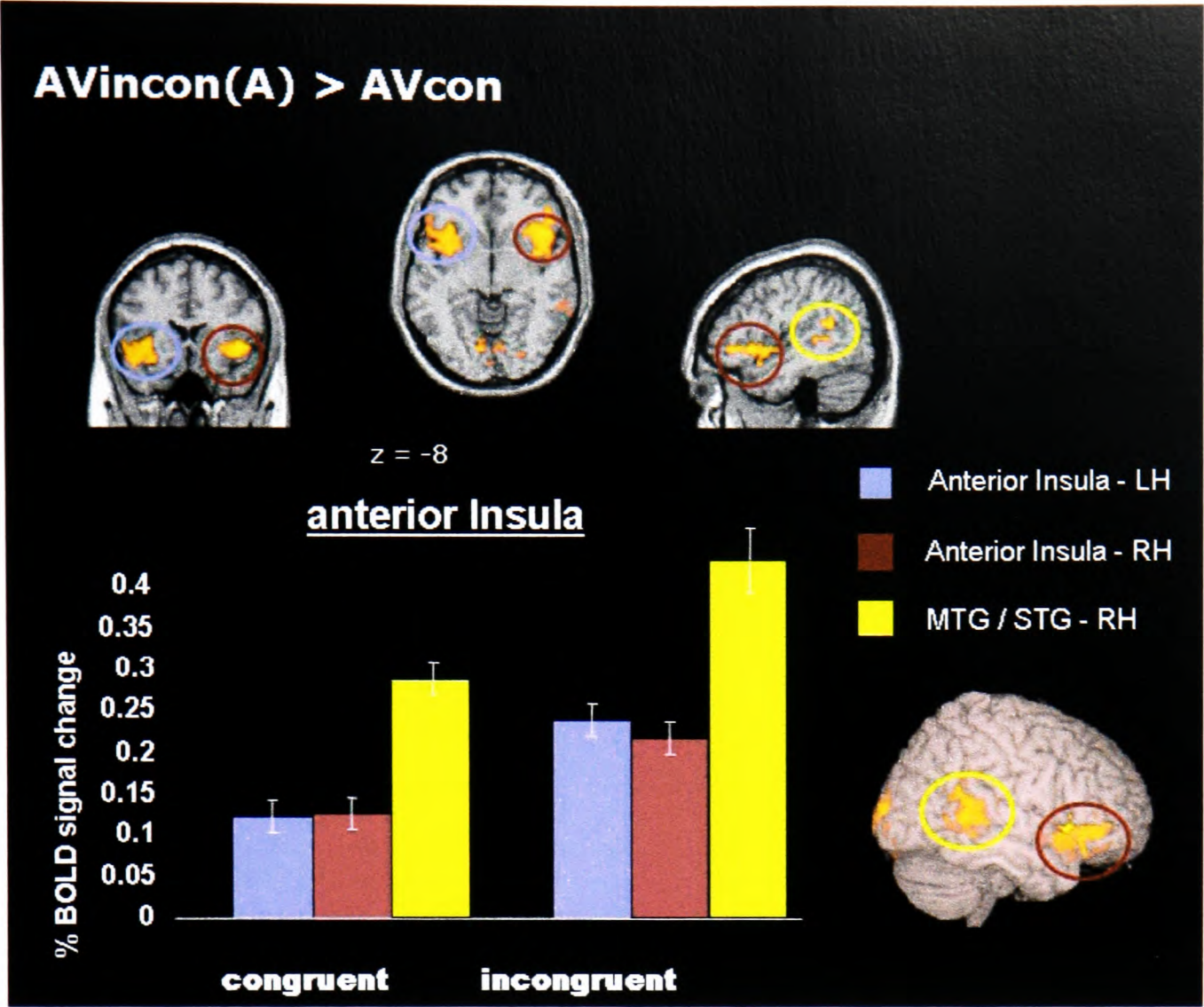


Figure 3-12: FMRI activation to AVincon(A) > AVcon
Activation where AVincon(A), (A = /gi/, V = /da/) activates more strongly than AVcon. Cortical activations are color-coded. Blue = left insula, Red = right insula, yellow = right MTG/STG. The graph shows % BOLD signal changes for left and right anterior insula during bimodal congruent and incongruent AV stimulation.

Contrasting AV incongruent stimulation (A = /gi/, V = /da/) vs congruent AV stimulation we see an increased activation during incongruent stimulation in the occipital lobe in BA 18 and the Cuneus, as well as in the posterior portion of the right STG. The spatially largest differences in activation during occurred bilaterally in the anterior insula (Figure 3-12).

Table for [AVincon(A) > Avcon] activation (z> 2.3)					
Area	peak z-value	Volume (cm3)	Talairach coordinates		
			x	y	z
Superior Temporal Gyrus (R)	3.6	4.41	49	-47	16
Inferior Frontal Gyrus (L)	3.5	3.83	-29	11	-12
Inferior Frontal Gyrus (R)	3.8	3.59	43	19	-4
Insula (L)	3.7	2.52	-37	15	0
Middle Temporal Gyrus (R)	3.5	2.26	55	-43	0
Lingual Gyrus (R)	3.5	1.94	13	-83	13
Insula (R)	3.4	1.75	41	16	-1
Cuneus (R)	2.9	1.54	3	-79	3
Superior Temporal Gyrus (L)	3.1	1.15	-47	9	-7
Middle Occipital Gyrus (R)	3.0	1.1	29	-95	8
Cuneus (L)	2.8	0.72	-19	-91	15
Lingual Gyrus (L)	3.0	0.65	-11	-77	-7
Middle Occipital Gyrus (L)	2.8	0.48	-19	-91	14
Clastrum (L)	3.6	0.26	-34	5	0
Inferior Parietal Lobule (R)	3.1	0.22	49	-47	21
Middle Frontal Gyrus (R)	3.1	0.19	49	37	-4
Parahippocampal Gyrus (L)	2.6	0.11	-29	-1	-18

Table 3-4: Areas of activation to for [AVincon(A) > Avcon]

3.3.7 Crossmodal activation of primary sensory cortex

A large area of visual cortex was found to respond to both auditory and visual speech input.). To understand the nature of this unexpected finding we looked at the temporal development of activity in this region over the whole scanning session (Figure 3-14). We found a gradual decrease for all conditions over time, probably due to habituation factors, except for the auditory condition. Here, auditory speech signals activated visual cortex at first only marginally, however, over time the response to the auditory only condition almost doubled in magnitude (A condition RUN 1 vs RUN 4(t

(1, 20) = 2.75, $p=.012$). This effect was only observed in the visual cortex and not in any other brain area. Conversely, activity in the auditory cortices showed a decrease over time but did not differ systematically between conditions (see Figure 3-14 for a representative graph of left auditory cortex activity).

We specifically looked into the signal changes in HG by creating region-of-interests (ROI) based on anatomical and functional criteria. The boundaries of HG were delimited on a normalized, high-resolution structural image based on the average of all subjects, using gyri and sulci morphology as structural guidelines, as well as Talairach coordinates based on functional mapping (Belin et al., 1999). Visual speechreading alone (V) significantly activated HG, especially the lateral part comprising secondary auditory cortex. Compared to conditions where an actual sound was presented, V activated HG significantly less. Student t-tests revealed significant differences for the contrasts V vs A ($t(1, 20) = 4.94$, $p=.0001$), V vs Avcon ($t(1, 20) = 3.92$, $p=.001$), V vs Avincon(V) ($t(1, 20) = 5.4$, $p=.0001$), V vs Avincon(A) ($t(1, 20) = 7.53$, $p=.0001$).

Figure 3:13 shows BOLD signal changes in HG to the different conditions, along with activation images of different statistical thresholds. This illustrates that HG activates at lower but not higher thresholds, and that superior temporal activation is highest in lateral anterior areas. Figure 3:14 shows that there are no significant changes over time in HG activation.

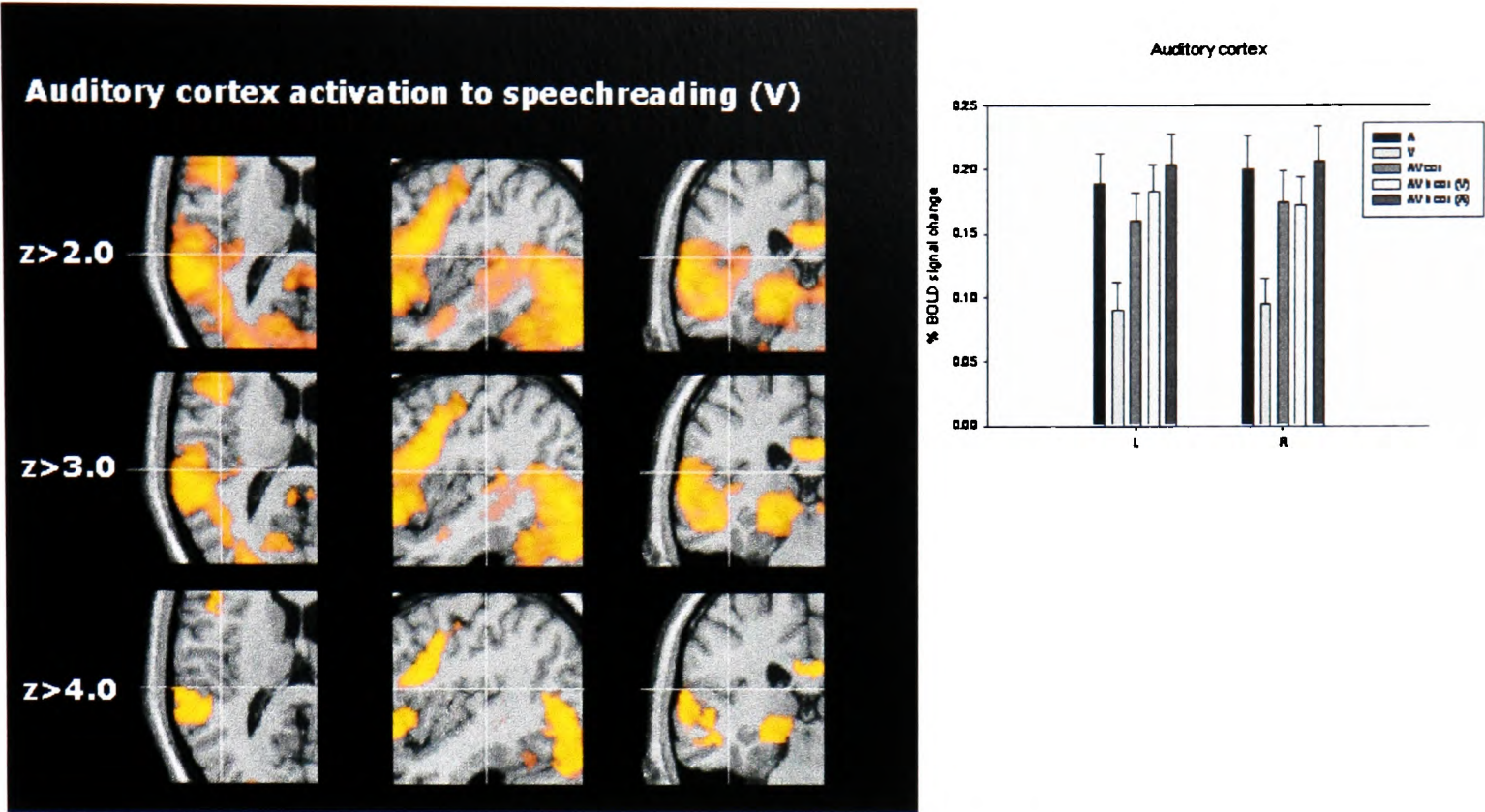


Figure 3-13: Auditory cortex activation to silent speechreading. V activation is depicted on the left image at different statistical thresholds. Crosshair is located at coordinates showing highest signal changes in medial HG after functional mapping of primary auditory cortex (Belin et al., 1999). Right image: BOLD signal changes in HG (both medial and lateral part) for each condition.

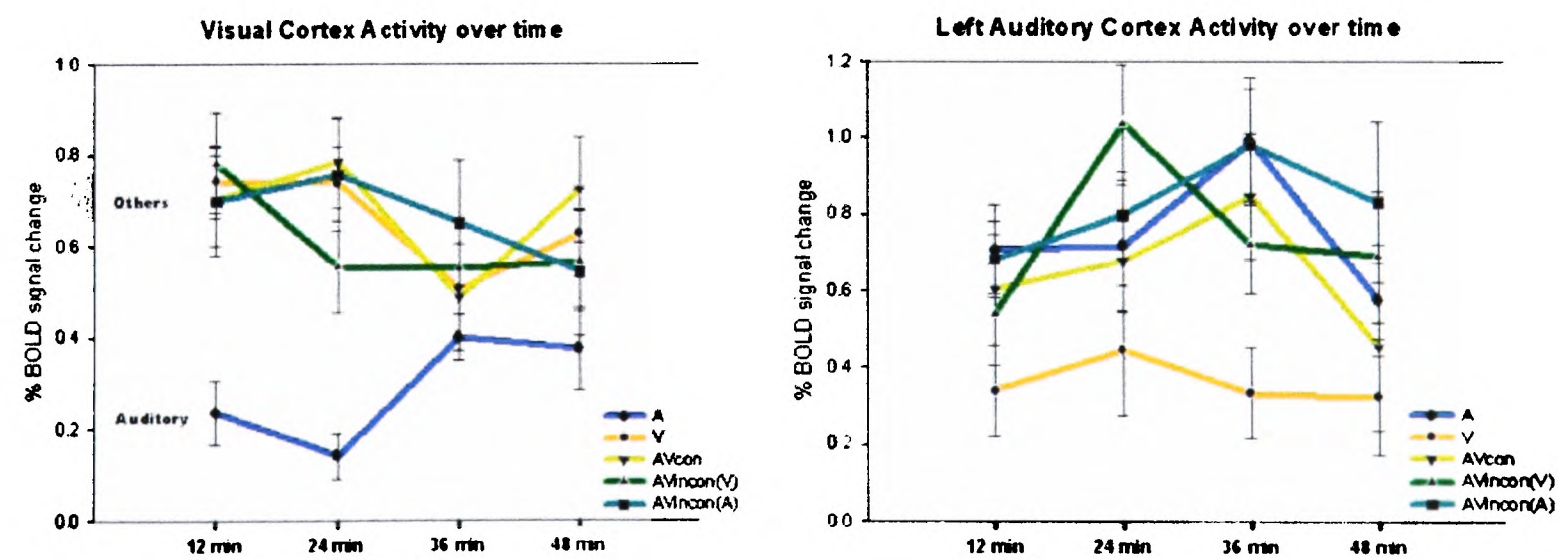


Figure 3-14: BOLD signal changes in primary sensory cortices over time Plotted by scanning session for all stimulus conditions. (Error bars = SEM)

3.3.8 Laterality

3.3.8.1 *ROI analyses*

Laterality is a defining feature of speech and language processing (Frost et al., 1999). Little is known about lateralization of speech function during multisensory processing. We therefore assessed laterality of brain activation during A, V and AV speech processing to see if the distribution of task-related activity changed as a result of multisensory integration.

Voxel-wise comparisons allow an unbiased assessment of language lateralization in the brain. However, significant variance can be expected at voxel level due to gyral and sulcal variations within and between subjects. Comparisons at the larger, regional level may be more sensitive to hemispheric differences of large-scale neuronal assemblies by minimizing sensitivity to local and individual anatomical variability. We therefore choose to assess laterality based on the extent of activation measured in predefined cortical regions of interest.

The structural and functional activation images of 21 subjects were normalized into MNI space (Collins et al., 1994) using the FLIRT algorithm (Jenkinson et al., 2002). Then functional images were thresholded at a cluster-level of $z > 4.0$ ($p < 0.05$). An automated procedure based on the coordinate system of Talairach & Tournoux (1988), adjusted to MNI space (Brett et al., 2002), was used to label anatomical brain areas at the gyral level (Lancaster et al., 1997). This procedure has shown excellent reliability and label matching with activation coordinates from a large number of functional imaging studies, as well as expert-derived Talairach labelling (Lancaster et al., 2000). The functional activation volumes of each subject were then quantified for each ROI in the A, V and AV conditions.

The automated Talairach labelling procedure was chosen because the Talairach reference system is a widely used reference system in neuroimaging research (Fox, 1995), although one also has to acknowledge its limitations (Brett et

al., 2002). Furthermore, the approach is reproducible and treats areas of both hemispheres equally, making it suitable for laterality assessment and comparison between laboratories (Lancaster et al., 2000). An interhemispheric difference analysis was conducted by comparing activation volumes of left and right-sided cortical areas (see Figure 3-16). Statistical analyses to determine hemispheric lateralization and to compare laterality index (LI) scores across conditions were performed using paired Student's t-tests with an alpha level of .05 (Carpentier et al., 2001).

3.3.8.2 *Laterality results*

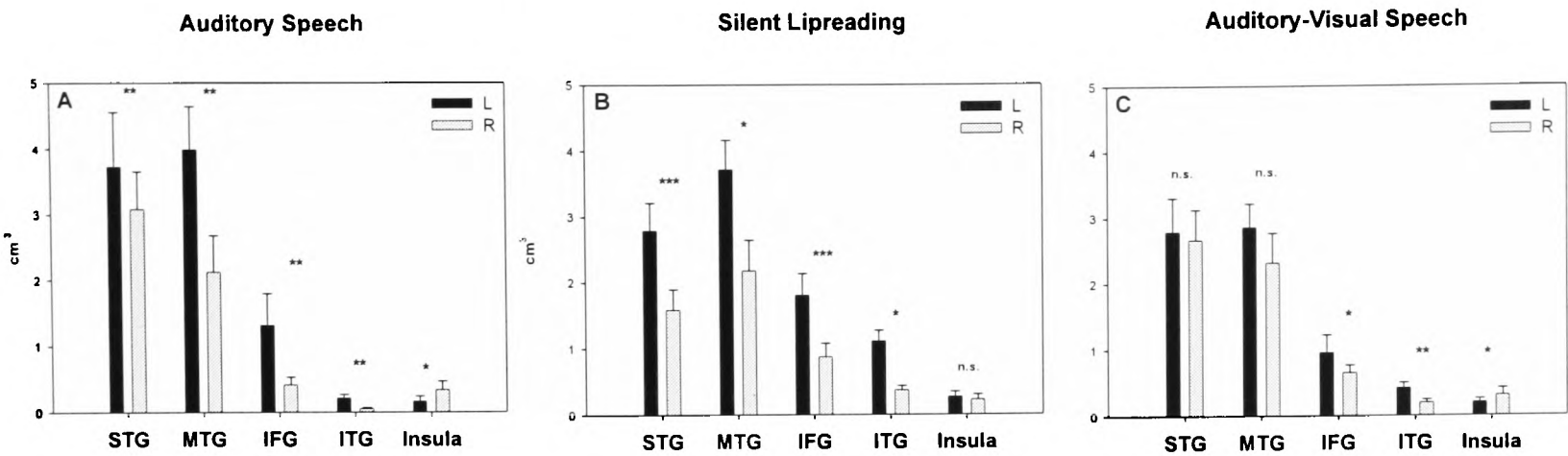


Figure 3-15: Shows mean activation volume (in cm³) for 21 subjects in different cortical areas for the left (L) and right (R) hemispheres (error bars = SEM), separately for auditory only (A), visual only (B) and audio-visual (C). Asterisks denote significance level of L vs R comparison (***: p<.001; **: p<.01; *: p=.05; n.s. = not significant, p>.05)

In the auditory speech condition left-lateralization was observed in the STG ($t(1, 20) = 3.59, p=.002$), MTG ($t(1, 20) = 3.56, p=.002$), ITG ($t(1, 20) = 2.99, p=.007$), and IFG ($t(1, 20) = 3.05, p=.006$). Insular activation was right lateralized ($t(1, 20) = -2.72, p=.013$). Responses to silent lipreading were also stronger on the left for STG ($t(1, 20) = 4.25, p=.001$), MTG ($t(1, 20) = 4.25, p=.047$), ITG ($t(1, 20) = 2, p=.05$) and IFG ($t(1, 20) = 5.38, p=.001$), whereas no significance was reached in the insula ($t(1, 20) = .58, p=.56$).

In the AVcon condition, ITG ($t(1, 20) = 2.87, p=.009$), IFG ($t(1, 20) = 2.1, p=.045$) and Insula ($t(1, 20) = -2.23, p=.047$) showed significant laterality responses, whereas no significant differences between left and right STG ($t(1, 20) = .32, p=.74, p=.11$) and MTG ($t(1, 20) = 1.85, p=.08$) were detected.

To express laterality as a function of experimental condition, a laterality index was calculated using the following formula: $LI = (L-R)/(L+R)*100$, where L is the activation extend of the left and R that of the right hemisphere. Expressed as a

percentage, an LI of 100 means complete left-sided laterality, a score of -100 laterality completely to the right, and zero indicates no evidence of laterality at all (Hinke et al., 1993; Nagata et al., 2001); (Frost et al., 1999) see Figure 3-16.

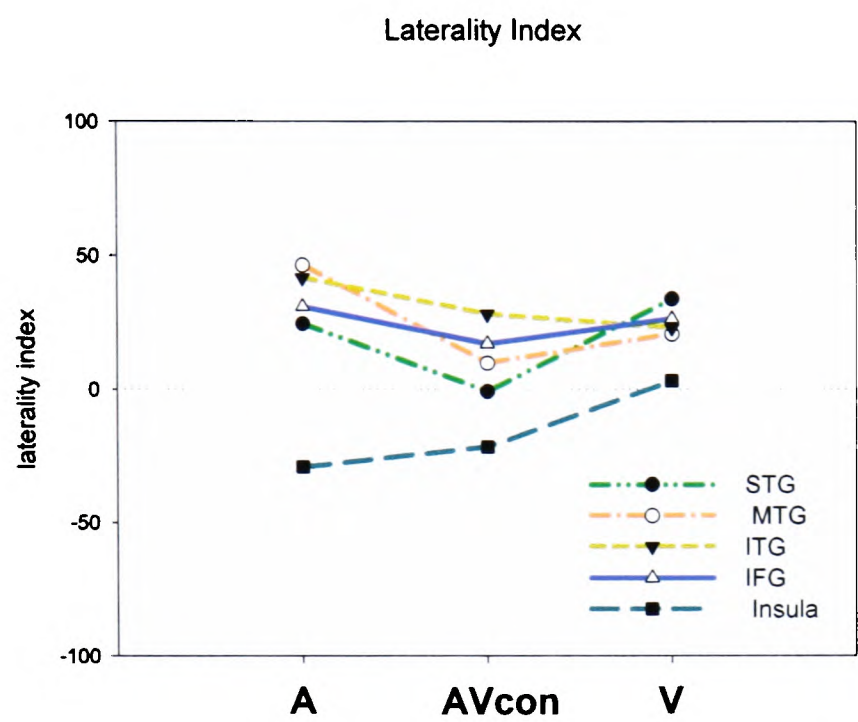


Figure 3-16: Laterality indexes across areas and conditions
Left (L) and right (R) hemispheric activation volume for different cortical areas to A) auditory speech stimulation, B) silent lipreading and C) auditory-visual speech.

Here, significant differences were observed between the A and V condition for MTG ($t(1, 20) = 2.33, p=.03$) and the insula ($t(1, 20) = -2.08, p=.05$). Significant differences between auditory and bimodal speech were observed in the STG ($t(1, 20) = 3.92, p=.001$), MTG ($t(1, 20) = 3.2, p=.004$), and IFG ($t(1, 20) = 2.74, p=.012$), and differences between lipreading and bimodal speech in the STG ($t(1, 20) = 4.81, p=.001$) and IFG ($t(1, 20) = 2.27, p=.034$).

3.4 Discussion

3.4.1 Summary

Using an event-related fMRI paradigm with a randomized trial sequence, we were able to record auditory-visual speech responses while minimizing attentional

confounds present in previous block design studies. The behavioural data gathered during the scanning sessions confirmed that subjects were paying equal attention to both modalities throughout the experiment. We found that auditory and visual speech activated, together with bimodal auditory-visual speech, largely the same areas in the occipital, temporal, parietal and frontal lobes of the brain. Upon closer inspection of the data, we found areas with a clear activation preference for a specific modality, and that within the areas showing strong bimodal responses, signal magnitude differed between stimulus conditions. Looking at crossmodal influences of primary sensory cortex, we found strong evidence for an involvement of auditory cortex on Heschl's gyrus in silent speechreading. Equally, our data show that auditory speech alone can activate visual cortex. However, we found that these crossmodal recruitments follow different rules depending on the input modality. Visual speech modulates auditory cortex immediately beginning with the first trials, whereas visual cortex is modulated by auditory speech only after an association, through repeated pairing, has been formed between the visual speech signal and its auditory counterpart.

To test for cortical areas mediating multisensory integration, we systematically applied different statistical criteria from the literature for detecting multisensory brain areas and compared their results. Most notably, we found no super-additive responses in our data, but less stringent criteria, such as the mean criterion, produced reliable results in areas which have been previously implicated in multisensory integration. For determining the appropriateness of a specific criterion in detecting multisensory integration based on hemodynamic data, we will use these fMRI results for comparison with direct recordings of neuronal activity from electrophysiological methods in the subsequent sections.

3.4.2 Detailed discussion

3.4.2.1 *Auditory speech*

Auditory speech processing elicited activation in both the left and right temporal cortices, especially the STS, auditory cortex (both A1 and A2), the middle and inferior temporal gyri and Broca's area in the inferior frontal gyrus. These findings are consistent with previous reports that these regions are involved in processing complex auditory stimuli, such as speech and human voice (Specht and Reul, 2003; Frost et al., 1999; Price et al., 1996; Belin et al., 2000; Belin et al., 2002). Functional imaging studies have shown activation of these regions in the processing of both linguistic and non-linguistic sounds (Alavi et al., 1981; Petersen, Fox, Snyder, & Raichle, 1990; Petersen et al., 1990; Binder, 1997; Binder, Frost, Hammeke, Rao, & Cox, 1996). Posterior temporal regions, including Wernicke's area and the supramarginal gyrus, have been implicated in phonological processing (Petersen et al., 1989; Demonet, 1994; Zatorre, 1996) and have been found to play a role in the transformation of visual orthographic characters into phonemic representations (e.g. Xu, 2001). Previous studies have also shown that anterior superior temporal responses are greater for intelligible than unintelligible speech, and that the posterior parts of the STG responds to speech stimuli that are both intelligible and unintelligible (Scott et al., 2000) and that areas of right anterior temporal cortex are activated by speech whereas the posterior portion of the STG is not (Price et al., 2005). Furthermore, results suggest that IFG and the middle part of the STG are part of a modality independent network for speech perception since they are activated equally by visual and auditory speech.

3.4.2.2 *Criteria for identifying multisensory areas*

The method of defining an area as 'multisensory' clearly effected the subsequent results to a high degree. It became apparent that a careful examination of

different criteria and of the underlying data is necessary to understand the full extent and nature of these crossmodal effects.

3.4.2.2.1 $A \cap V$

The intersection of A and V produced large activation and shows those voxels that respond to both unimodal visual and auditory stimulation. In the present data all of the voxels activating to this criterion activated to bimodal AVcon stimulation as well. A brain region that activates to this criterion is sure to contain auditory and visual responsive neurons. However, these neurons might exist in separate pools within the same voxel and thus may not be bimodal neurons or show any other neuronal characteristic of multisensory integration. Therefore, this criterion separates unimodal from heteromodal areas involved in AV speech processing, but not necessarily areas involved in integrating these signals.

3.4.2.2.2 $AVcon > A \text{ AND } AVcon > V$

The maximum criterion ($AVcon > A$) AND ($AVcon > V$) (Beauchamp et al., 2004; van Atteveldt et al., 2004; Beauchamp, 2005) is more strict than criterion 1 and identified focal multisensory responses in the right posterior temporal regions, left middle temporal gyrus and superior and inferior parietal areas, regions which have also shown super-additive responses in other studies (Calvert et al., 2001; Calvert and Thesen, 2004) This contrast also activated right Tpt on the posterior part of the planum temporal, which is distinct from Wernicke's area and has been found to respond to both auditory and somatosensory input (Leinonen et al., 1980). One shortcoming of the maximum criterion in evaluating crossmodal influence is that not all unimodal responses contribute in an equal, linear fashion, i.e. only the stronger of the responses determines this criterion. Also, this criterion does not offer conclusive proof that a particular area is involved in multisensory integration. Here the same problem as with criterion 1 is acute because of summation from unimodal neurons in one voxel

during bimodal stimulation. However, this criterion is more stringent than $A \cap V$ and requires an increase in signal magnitude during AVcon processing.

3.4.2.2.3 *Super-additivity*

The super-additive interaction criterion ($AVcon > A + V$) is the most stringent and logically plausible of the ones tested here. It is especially attractive since it uses a criterion that has been successfully applied to single-unit studies (Stein and Meredith, 1993; Calvert et al., 2001). In a recent study, Calvert et al. (Calvert et al., 2000) observed an increase in BOLD signal in the posterior part of the STS during the perception of synchronous and congruent auditory-visual speech. This increase was significantly higher in voxels that activated to A and V than the sum of responses to the acoustic and visual stimulus features presented in isolation. The same area also showed a decrease in activity when the auditory speech signal did not match the lip movements. Neither superadditivity nor congruency effects were observed in the present study. It is possible that some neurons in the direct vicinity of super-additive neurons also show sub-additive properties (Stein et al., 2004). Since BOLD fMRI measures the summed response of many neurons in a voxel it may not have the necessary spatial resolution to distinguish between these units. Consequently, the summed activity reflected in the BOLD signal may fail to detect multisensory regions and could therefore be overly conservative.

In fact, a recent estimation of neuronal population responses and their relationship to BOLD measures, (Laurienti et al., 2005) showed that superadditive responses at the neuronal level, as known from single-unit studies, are unlikely to lead to superadditive responses in the BOLD signal. This findings strays doubt on the validity of using the super-additive criterion as a measure in fMRI research. However, these estimations are based on data from action potential recordings, which represent processing in a given area, whereas recent evidence suggests that the BOLD signal is

a better correlate of synaptic activity, i.e. representing input into a given area rather than output from that area (Logothetis et al., 2001). But as there is also evidence that the BOLD signal correlates well with spiking activity (Mukamel et al., 2005) and the exact coupling between hemodynamic and neural response is poorly understood (Logothetis, 2002), the nature of the exact relationship between fMRI BOLD measures and multisensory integration at the neuronal level remains open. It therefore becomes important to compare and validate multisensory integration criteria with and between different methodologies.

From our data it also emerges that when using super-additivity as a criterion it is necessary to consider the individual signal changes of the activated areas closely. For example, if a uni-sensory condition shows major deactivations instead of signal increases, the physiological interpretation of the results will invariably be different than when the superadditivity criterion was reached (e.g. both (AVcon, V, A = 3%, 1%, 1%) vs. (AVcon, V, A = 1%, 1%, -1%) would show an interaction effect, but suggest a difference in the underlying neuronal processes). Therefore, to calculate super-additive responses with fMRI, an additional criterion is needed. One option to guard against the eventualities is to calculate the interaction effects solely with positive activations.

The failure to detect any super-additive responses, as opposed to previous studies who have reported such findings (Calvert et al., 2000; Calvert et al., 2001), could also be a result of the specific task and recording parameters of this study (i.e. event-related auditory-visual speech discrimination at 3 T). However, the reason is unlikely to be due to SNR issues related to statistical power (which are lower in ER studies), since the sample size was relatively large for an fMRI study (n=21) compared to the earlier studies. An exploratory analysis of the superadditive contrast revealed that even very low statistical thresholds did not show 'activation' in areas of the superior temporal lobe. In fact, these areas showed a clear and significant

subadditive response, supporting the view of Laurienti et al. (2005) that putative bimodal neurons in these areas may not sum up sufficiently to produce superadditive changes in the hemodynamic BOLD signal. Recent studies of tactile-visual and auditory-visual multisensory processing confirm that super-additivity often fails to detect multisensory responses (Amedi et al., 2001; Amedi et al., 2002; Beauchamp et al., 2004; van Atteveldt et al., 2004).

3.4.2.2.4 *Subadditivity*

Multisensory suppression is a common phenomenon observed at the neural level (Stein and Meredith, 1993) and can be assessed with fMRI using the sub-additive criterion $AV_{con} < A + V$. However, this statistical test is prone to classify neurons as multisensory which show the same % signal change in all 3 conditions and therefore might be overly liberal by including supra-modal processing areas, such as those related to motor responses or stimulus expectation (Teder-Salejarvi et al., 2002; Jiang and Stein, 2003). In the present investigation it was indeed the criterion which showed the most widespread activation including motor areas.

3.4.2.2.5 *$AV_{con} > \text{mean}(A, V)$*

The mean criterion, championed by some investigators (Beauchamp et al., 2004; Beauchamp, 2005), compares bimodal stimulation with the mean of the unisensory responses ($AV_{con} > (A+V)/2$). Compared with criterion 2, the mean criterion behaves linearly and both unisensory signals contribute equally. Its advantage is that it is not sensitive to task related activations that appear to the same degree in all 3 conditions. It is more stringent than the intersection criterion 1 and functionally closer related to multisensory integration than the subadditive criterion 4, but more liberal than the superadditive criterion 3.

3.4.2.3 *Phonetic interactions*

Congruent AV speech did not cause any increases in BOLD signal. But contrasting congruent auditory-visual speech activation with responses to a stimulus in which the auditory and visual components did not match (AVincon(V), revealed higher signal changes in the left STS and bilateral anterior insula during the incongruent stimulation (Figure 2-14). The insula is a functionally heterogeneous structure receiving input from both the thalamus and cortical regions (Bamiou et al., 2003). Whereas the left STS has been shown to discriminate various speech sounds (Jancke et al., 2002), anterior insula has been previously identified as neuronal structure mediating cross-modal matching (Hadjikhani and Roland, 1998; Banati et al., 2000) and the detection of auditory-visual temporal asynchrony (Bushara et al., 2001). Even though the present experiment does not provide conclusive evidence, the results suggest that the insula region is also involved in the comparison of speech features across the senses and may serve as a detector of inter-sensory incongruence. However, another interpretation unrelated to multisensory processing is also possible. The auditory /gi/ in the AVincon(A) condition could also be categorized as an infrequent auditory stimulus among the more frequent auditory /da/ present in conditions A, AVcon, and AVincon(V). In effect, this would render the AVincon(A) condition a part of an odd-ball task where an infrequent stimulus is presented amongst frequent distractors (Polich et al., 1996). Data showing that the P3 amplitude, a characteristic of the ERP in oddball paradigms, correlated well with hemodynamic activity recorded from the insula (Brazdil et al., 2005) support this interpretation. Because of this confound, results from the AVincon(A) condition is hard to interpret.

3.4.2.4 *Laterality*

We found significant laterality effects towards the left hemisphere in response to auditory speech processing in the large gyri of the temporal lobe (STG, MTG, ITG)

and in the IFG. Visual speechreading showed very similar effects in the temporal regions and in the IFG, albeit somewhat reduced in magnitude. Bimodal speech, however, caused a shift in the STG and MTG from being left lateralized to equal representations across hemisphere. The finding that auditory speech is pre-dominantly processed in left temporal lobe structures has been well documented in the literature (see section 1). While speech is also processed in right hemisphere networks, these areas tend to specialize in non-speech auditory analysis, such as music and voice recognition for example (Zatorre et al., 1992; Belin et al., 2000). Seen from the perspective of sound parameter processing, the left hemisphere specializes in rapid spectro-temporal processing of sound changes (such as in CV formant transitions) whereas the right hemisphere is more tuned to processing changes in pitch.

Recent theories of speech perception argue against the presence of speech-specific cortical modules (Price et al., 2005). They suggest that different auditory areas specialize in certain aspects of auditory analysis, and that the functional connectivity between these areas changes as a result of different sensory tasks. Following this line of reasoning, our data suggest that the functional connectivity between specialized auditory areas of the left and right hemispheres changes during multisensory speech perception. The behavioural consequences of multisensory integration are faster reaction times and lower error rates, thereby suggesting that multisensory integration serves an adaptive function. Speech perception in busy social situations is a multisensory task which often requires one to quickly determine who said what. This process involves both content-related analysis (of speech information) and object-related analysis (of voice). Increases in performance then would require changes the functional connectivity between speech and voice analysis and a more efficient allocation of processing resources between left and right auditory areas.

3.4.3 Crossmodal influence of primary sensory cortex

The traditional view holds that primary sensory areas such as V1 and A1 are unisensory and only process information from one sensory modality. However, in agreement with our findings, it has been shown that visual speech stimuli can activate primary auditory cortex ("silent lipreading"; Calvert et al., 1997; Ludman et al., 2000). Our results also show that auditory speech during the absence of visual input has the ability to activate primary visual areas (see Fig. 6).

3.4.3.1 *Speechreading*

The current study demonstrated that silent speechreading recruits bilateral auditory cortex including both medial and lateral portions of HG, and that activation is left lateralized. So far the neural substrates underlying speechreading are not well understood. A study with patients who suffered from acquired cortical deafness suggests that visual speech serves the function of an extra phonetic resource that supports the speech recognition system at the phonological level (Campbell et al., 1990), a finding supported by the McGurk effect (McGurk and MacDonald, 1976). Evidence from various neuroimaging methods and from both monkeys and humans show that visual articulation can modulate auditory cortex. Sams et al. (1991) were the first to show such an effect with MEG, followed by Calvert et al. (1997) with fMRI. Recent work using single-unit recordings in ferrets showed primary auditory field responses to light stimuli (Bizley et al., 2004). Calvert et al. (1997) showed that even areas which had previously been thought of as purely unimodal and related to auditory speech processing are part of the neural circuitry for both auditory and visual speech perception. Silent lipreading of digits from 1 to 10 activated the posterior parts of the STS and bilaterally the primary auditory cortex on the lateral surface of the planum temporale (BA 41/42). The finding that speechreading recruits regions of the STG have been replicated many times (Calvert & Campbell, 2003; Bernstein et

al., 2002. MacSweeney et al, 2002; Campbell et al, 2001; Calvert et al. 2000).

However, the involvement of primary auditory regions in such tasks has not always been replicated and remains disputed (Bernstein et al, 2002; Macsweeney et al 2000).

Our findings are in agreement with a very recent study by Pekkola et al. (2005) who used anatomically-defined mapping of HG on an individual subject basis aided by high-resolution structural MRIs and separating medial and lateral parts of HG.

Scanning 10 subjects, they showed convincing evidence for both primary (in 7 subjects) and secondary (in 9 subjects) auditory cortex involvement in speechreading. The authors also found that HG activation was less for visual speech than is usually observed for auditory speech.

Auditory cortex probably does not respond to visual changes of all kinds all the time, and it seems plausible that an auditory equivalent of the visual stimulus has to exist and that this relationship has to be task relevant, as is the case during an AV vowel discrimination task. Interestingly, Pekkola et al. (2005) found that even moving circles activated primary auditory cortex in 3 out of their 10 subjects. In an auditory imagery experiment by Jaencke & Shah (2004), subjects were asked to imagine auditory CV syllables during the presentation of a visual cue. One group of subjects was trained to associate the visual stimulation with the speech sounds through concurrent AV presentations, whereas a control group had not. When subjects were instructed to perform auditory imagery each time the visual stimulus was presented, only the trained group showed activation in bilateral A2 and STS, whereas the control group did not show any responses in the temporal lobe. Even though the authors of this study did not interpret their results in a multisensory framework, their findings support the view that modality specific sensory cortices can be recruited during the processing of stimuli from another modality if a cross-modal association between the stimuli pair exists. As such, prior experience and context seem to play an important role in crossmodal recruitment of primary sensory cortex.

3.4.3.2 *Auditory speech*

Our data on the visual cortex also support the learning hypothesis in respect to crossmodal recruitment of primary sensory cortex. When analyzing the fMRI data on a session basis (subjects were presented with 4 consecutive sessions of 12 min each, see fMRI Methods) the signal changes for auditory only stimulation in the visual cortex showed an increase with session number (Figure 3-14). This increase is not seen for any other stimulus condition and is specific to visual cortex. This suggests that visual cortex activation is not automatic and that maybe a learning mechanism is responsible for this crossmodal activation. McIntosh et al. (1998), using simple stimuli, found occipital cortex activity to an auditory stimulus after learned association with a visual stimulus. When they subjected their data to a connectivity analysis, they deduced that visual cortex activation was mediated by auditory areas in the temporal lobe, and not by frontal attention networks. This suggests a stimulus-driven learning mechanism that's independent of attentional modulation.

Based on the effects we observed we propose that the crossmodal recruitment of visual cortex is a learning effect resulting from repeated pairing of the AV stimuli throughout the experiment, hence the increase in signal over time. However, the question remains whether the cross-modal recruitment of sensory cortices in our study is a result of top-down attentional modulation or a stimulus dependent mechanism to facilitate the integration of AV speech signals. The attention hypothesis would postulate a context dependent effect that is due to the cognitive set of the subject. Previous AV speech studies that did not show occipital activation in response to A stimulation all used block designs during which stimuli from the same modality class were presented during the long ON phase (~30 s). Therefore, subjects in a block design study are learning quickly to expect repeated presentations of stimuli from the same trial class. In the current event-related study, however, the presentation of stimulus conditions is randomised. While subjects are required to constantly monitor

both sensory channels to detect a target stimulus, visual areas are recruited even during auditory only trials. Indeed, work by Kastner et al. (1999) has shown that endogenous visual attention activates visual areas even in the absence of a visual stimulus. As we see from Figure 3-14, this attentional mechanism is present even during the first session. However, the increase over subsequent sessions can not be explained by attentional factors alone since one would rather expect a decrease over time. Therefore, it seems likely that the visual cortex activity to an auditory speech stimulus reflects mechanisms of both attention and crossmodal learning.

3.4.3.2.1.1 Different rules for cross-modal recruitment of primary cortices

We found that cross-modal influences on primary sensory areas do not follow the same recruitment rules in both directions. As we have shown here, auditory speech activates visual cortex only after an auditory-visual association between stimuli of the two modalities has been formed. Silent lipreading, however, recruits auditory areas from the first trial without requiring learning. This suggests a difference in the underlying neural mechanisms of these effects. Whereas we interpret the changes in visual cortex as related to attentional processes, we suggest that visual influence over auditory cortex reflects processing related to the sensory analysis of the incoming auditory signal and hence has the capability to change the resulting auditory percept at an early stage in the processing hierarchy, as evident in the McGurk illusion (McGurk and MacDonald, 1976).

One possibility for the visual signal to achieve such drastic perceptual changes is rooted in the nature of an AV speech event. Vocalization is, among other things, a result of position and movement of articulators, including the tongue, palate, lips and jaws, all of which are visible and active before sound onset. So the onset of the auditory stimulus is usually preceded in time by the relevant visual stimulus, in our case by around 400 ms. This gives the visual system ample of time to prepare auditory

areas for the arrival and subsequent interpretation of the auditory speech signal. Thus, through a contextual modulation of sensory analysis the visual input could cause a 'priming' or biasing of auditory areas after a visemic analysis towards the representation of the visual input.

In this model, timing of crossmodal influence is of utmost importance. Questions about the timing of neuronal events, however, cannot be answered by the present fMRI experiment because of limitations in the temporal resolution of the method. To specifically answer questions about the timing of events during this AV speech task, we conducted studies with MEG and iEEG using the same stimuli and, where possible, the same subjects, using a similar experimental design. These experiments will be presented in the forthcoming sections.

3.5 Conclusion

In summary, we found evidence for a large network of brain areas mediating speechreading, involving mainly the occipital, frontal and temporal lobes. More specifically, our results did provide further evidence for the involvement of auditory cortex in silent lipreading. We also found that auditory speech activated primary visual cortex, an effect that is suggested to be mediated by crossmodal learning and attention. We also found evidence for changes in brain lateralization during AV speech perception, namely the release from left lateralization during multisensory processing, a finding that warrants further investigation.

Evaluating different statistical criteria for investigating multisensory processing, we found that the super-additivity criterion did fail to show multisensory effects, a finding reported by others and which was likely to be the result of coupling between neural activity and the BOLD signal. The criteria $AV_{con} > \text{mean}(A, V)$ and $(AV_{con} > A \text{ and } AV_{con} > V)$ were most sensitive for detecting specific putative multisensory integration structures (STS/STG). These criteria are most consistent with

our previous findings (Calvert et al., 2000, also see (Beauchamp, 2005; Calvert et al., 2001) and suggest that, amongst other areas, the left STS in the superior temporal lobe is chiefly involved in AV speech integration. However, an appropriate criterion for detecting the neuronal correlates of cross-sensory integration should not only rely on the plausibility of such a result alone. We therefore will attempt to circumvent the issue of hemodynamic coupling and superadditive effects by testing the superadditive criterion with direct measures of neural activity, such as MEG and EEG, using the same stimuli and, if possible, the same subjects. A spatial match of electrophysiological superadditive effects with the fMRI results would lend strong support to the appropriateness of the above criteria for use with fMRI, whereas a mismatch would strongly question the use of these criteria. Such findings, if complementary, will validate and strengthen the results from the present fMRI study and can serve as a guide for future investigation of multisensory processing using fMRI.

4 The time-course of multisensory speech perception

The previous section used fMRI to identify a widely distributed network of brain areas associated with multisensory processing and integration. The results from the fMRI study offer high spatial resolution and images represent average brain activity over several hundred milliseconds. But the brain processes information faster, and therefore in this section we are using MEG and dipole source analysis to describe the exact time course of multisensory processing and integration on a millisecond time scale.

4.1 Introduction

Timing is a very important factor in neuronal processing. During sensory and higher order processing the sequence in which cortical areas are activated is an essential element in computing perceptual input and action (Foxy and Simpson, 2002; Halgren et al., 1994; Dhond et al., 2005). The timing of multisensory responses and their integration has so far remained elusive as many investigations have employed imaging methods with very poor temporal resolution. For example, Calvert et al. (2000) showed blood flow increases to congruent AV speech in posterior STS compared to the sum of unimodal activations (super-additive effect). They also observed that incongruent AV speech caused a decrease in blood flow in this area. They suggested that AV signals integrate in STS and that the auditory cortex activation increase observed during AV speech processing (Calvert et al., 1999) is reaching auditory cortex via feedback projections from this area, an explanation that has become known as the ‘late’ integration model of multisensory speech perception (Calvert, 2001). However, ‘early’ integration is also a possibility in that visual input could modulate auditory cortex via direct connections between the two sensory

cortices, or after elaborate visual processing but still before auditory stimulus onset by taking advantage of the timing difference between A and V onset in natural speech. Investigations using high temporal resolution imaging methods have yielded conflicting results for both speech and non-speech stimulation, with some showing early interactions (<60 ms) and others late interactions (>150 ms) (Giard and Peronnet, 1999; Raij et al., 2000; Fort et al., 2002; Fort et al., 2002; Molholm et al., 2002; Molholm et al., 2004; Mottonen et al., 2004).

Behavioural studies show that when auditory and visual speech stimuli are presented synchronously within a time-window of 200-250 ms, speech comprehension increases compared to auditory stimulation alone, especially in noisy surroundings (Sumbly and Pollack, 1954; Summerfield, 1979; Summerfield and Assmann, 1987; Summerfield, 1992). The interaction of auditory and visual speech inputs on perception is further illustrated in the McGurk effect (McGurk and MacDonald, 1976) where an auditory /pa/ is dubbed onto a visual /ka/, resulting in the perception of /ta/. Sams and colleagues (1991) were the first to investigate the neural manifestation of this multisensory effect by using MEG recordings over the left hemisphere. When the illusion was perceived, a modulation was found starting at 180 ms which located to the auditory cortex in the superior temporal lobe. In a more recent study, incongruent auditory-visual speech activated the temporal auditory areas more strongly than congruent bimodal speech at 140-160 ms after auditory onset and 200-300 ms in the left hemisphere and 345-375 ms in the right hemisphere (Mottonen et al., 2002). Together, these magnetoencephalographic studies suggest that auditory processing in temporal auditory cortex is influenced by visual speech input. Using fMRI, Calvert and colleagues (1997) confirmed these findings by showing that silent speechreading activates auditory cortex.

However, the question remains at which stage of information processing the acoustic and visual speech inputs are integrated. The late integration model suggests

that the sensory signals are first processed in a modality-specific fashion independently in their respective unisensory cortices and are then relayed to multisensory integration areas for integration, such as the STG. Multisensory effects observed in primary sensory cortices are explained through feedback projection from multisensory integration areas. The early integration model on the other hand suggests that multisensory integration already occurs in primary sensory areas, probably mediated through direct connections between these cortical modules. These models were largely constructed based on fMRI data in humans. The epistemological limitation of this technique does not allow to actually test these models accurately, since the temporal resolution in fMRI is too coarse to map the temporal sequences of cortical processing. We therefore employed MEG with its high temporal resolution to study the timing of auditory-visual speech processing and integration in cortex.

The AV interaction effect for determining multisensory integration has been successfully employed in electrophysiological research, and unlike fMRI, has yielded consistent results over a wide range of studies (e.g. see (Giard and Peronnet, 1999; Fort et al., 2002; Fort et al., 2002; Molholm et al., 2002; Molholm et al., 2004). To improve the MEG source localization accuracy and spatial resolution, we used results from the fMRI study in section 3 to inform the MEG source reconstruction.

The strategies employed for a better understanding of neurocognitive processes by using fMRI/PET for localization information and MEG/EEG for timing information have ranged from simple juxtaposition of results to integrated analyses (Dale and Halgren, 2001). Simple juxtaposition of results has the advantage that the analysis does not assume that the signals from both imaging modalities are generated by the same neuronal sources. However, such analyses often use this assumption in the interpretation of results. This approach has been successfully used in the study of somatosensory processing (Rossini and Pauri, 2000), optic flow (Greenlee, 2000), novelty processing (Opitz et al., 1999) and feature binding (Schoenfeld et al., 2003).

The most common method for localizing the cortical sources of MEG activity is to search iteratively for the position, orientation and strength of an individual dipole that explains best the observed magnetic field (Scherg and von Cramon, 1985). However, if reasonable evidence suggests that multiple generators are contributing to the magnetic signal, a multi-dipole model can be used to characterize the response in these regions. One example are the generators of the N400 response for faces and words where iEEG and fMRI have separately shown multiple focal generators (Halgren et al., 2002). Here, the multi-dipole approach offers a more accurate representation of the cortical activity than do single dipole fits. The model solution depends strongly on the number of assumed sources. However, the actual number of ECDs cannot be estimated by the MEG data alone since the inverse problem is fundamentally ill-posed (Hämäläinen et al., 1993). That is, for any MEG data set there are an infinite number of current source distributions which can explain the same observed magnetic field pattern. Therefore, investigators often use plausible constraints for modelling the MEG data, such as those obtained from structural and functional physiology and anatomy (Hari, 1996). Incorporating fMRI data into the model offers the advantage of using more biologically plausible constraints than the independent dipole model. There is large evidence for a close coupling between the hemodynamic BOLD signal and neuronal activity (Logothetis, 2002), and between the particular neural generators of the MEG signal (Hari, 1996) and those thought to be the main contributors to the fMRI BOLD signal (Logothetis et al., 2001; Logothetis, 2002). Current dipoles can be used to represent the center of gravity of an extended brain region (Hari, 1996). Therefore, using fMRI to constrain the MEG source localization makes sense from a modelling and physiological point of view. Indeed, this approach has been successfully applied to elucidate spatio-temporal patterns of brain activity (Hillyard et al., 1997; Menon et al., 1997; Opitz et al., 1999; Woldorff et al., 2002). Besides from providing converging evidence from separate

experiments by juxtaposition of the final results, functional imaging data have also been combined through direct data fusion (Ahlfors et al., 2002; Vitacco et al., 2002). The main assumption, as discussed by (Dale and Halgren, 2001), is that the MEG or EEG data are generated by a few, localized dipoles and that the local maxima of the fMRI results are reflections of the same underlying neuronal processes. The fMRI results can thus be used as a plausible, biological constraint in the MEG/EEG source reconstruction (Ahlfors et al., 1999).

Detection of magnetic activity in primary and secondary auditory cortex by MEG is aided by the anatomical structure of these areas. They are generally identified as BA 41 and BA 42, respectively, and are located on the anterior transverse temporal gyrus, i.e. HG. They are located on the superior surface of the temporal lobe and their currents are orientated perpendicular to the cortical surface in respect to the inner surface of the skull and thus have a tangential orientation, to which MEG is most sensitive (Hämäläinen et al., 1993). In fact, Scherg et al. (1989) observed a clear dipolar pattern of the magnetic field at 19 ms which located to the auditory cortex. A series of subsequent investigations have confirmed that the first evoked cortical response to auditory stimulation appears at ~20 ms (P20m) and is located on the posterior/medial aspect of HG, i.e. primary auditory cortex (Hashimoto et al., 1995; Kuriki et al., 1995; Rupp et al., 2000; Borgmann et al., 2001; Lutkenhoner et al., 2003).

4.2 Methods

4.2.1 Design

Data were recorded from 9 (3 female, mean age = 25 years, stdev = 2.3 years) native English speakers all of which have participated in the fMRI experiment. All subjects were right-handed, without previous history of psychological or neurological disorders and had normal or corrected to normal eyesight. The study received ethical

approval from the Central Oxford Research Ethics Committee and informed consent was obtained from all subjects

Visual stimuli were displayed through a CRT monitor placed outside the magnetically shielded room. The sound was presented with a Samson G5 headphone amplifier and an Ear Tone 3A sound pressure converter. At the beginning of each experiment the sound intensity was adjusted to a comfortable level at which subjects were able to discriminate the CV stimuli.

During the experiment, 80 stimuli of each standard condition (/da/) and 20 stimuli of each target condition (/tu/) were presented in random sequence across 4 blocks in an event-related design. The inter-stimulus intervals varied randomly from 1500 ms to 3500 ms in 500 ms steps.

4.2.2 MEG Data Acquisition

MEG data were acquired on a 151 channel magnetometer (CTF Systems, Vancouver, Canada) at a sampling rate of 625 Hz and A/D hardware filter of 0 – 208 Hz (Nyquist limit = 104 Hz (Krauss and Webber, 2005)). After data collection, a 3D digitizer (Polhemus Isotrack) was used to digitise the shape of the subject's head to allow the co-registration of the MEG sensor space with the structural MRI of the subject (Adjamian et al., 2004). A more detailed account of this procedure is given in section 8.

4.2.3 Data analysis

4.2.3.1 *Source analysis*

The spatio-temporal distribution of neural activity as measured by the MEG sensors was estimated in terms of ECDs, which are models for localized electrical activity at the macroscopic scale in the brain (Hämäläinen et al., 1993). The conductivity distribution in the brain was modelled as a sphere. The preprocessed data were used to generate isocontour maps for visual inspection of dipolar field patterns.

Channels showing maximum amplitude changes were selected for illustration of the activation time course (Figure 4-1). The source localizations were estimated based on the same field patterns.

4.2.3.2 *Independent MEG dipole source localization*

Dipole source localization of the MEG data was first achieved without reliance on any information obtained by the functional MRI experiment and was based on commonly used multi-dipole modelling strategies (e. g. (Nishitani and Hari, 2002)).

Analyses were performed on averaged waveforms of each subject individually. It was assumed that the AVcon condition contains all sources that are potentially active during the unimodal A and V conditions. Therefore, the initial dipole modelling was first performed on the congruent bimodal condition (AVcon) before the resulting ECD location parameters were used for modelling the source amplitude of the remaining conditions.

First, we determined the most prominent deflections in the MEG signal (see Figure 4-2) and plotted the isocontour field pattern for a 20 ms time-window around the highest amplitude. For each dipolar field pattern an ECD was estimated using a subset of the sensor data that contributed to a distinct dipolar field pattern in the original data. The source localization algorithm employed an iterative minimization by which the error term is estimated through reduced chi-square tests using the plus-minus average as variance (Stok et al., 1986). We accepted only ECDs with a goodness-of-fit above 85 % and with a confidence volume $< 1 \text{ cm}^3$. This procedure was repeated for all major signal deflections up to 400 ms post-auditory onset. Co-localized ECDs were removed from the model and the analysis period was extended to a larger measurement time window (-1500ms to 1000 ms) and to all 151 sensors. Then a spatio-temporal multi-dipole model fit was applied in which the strengths of

the previously identified moments were allowed to vary over time while the position and orientation were kept constant. The spatial coordinates of the ECDs were first overlaid onto the structural MRI of the subject for visualization of the cortical locations and then transformed into MNI space for comparisons across individuals.

4.2.3.3 *fMRI-constraint dipole modelling*

The coordinates for fMRI-constraint dipole fit were obtained from the AVcon condition in section 3 by using a high threshold ($z > 4.0$, $p = .05$) and selecting the peak activation voxels in highly activated clusters. This ensured spatial separability between fMRI activation centres and inclusion of only the most robustly activated areas in the model. The fMRI coordinates were then transformed from MNI space into the MEG sensor space of the individual subjects.

In the MEG source analysis, fMRI-based dipoles placed at regions that are not contributing to MEG signals (i.e. that are visible with fMRI but not with MEG; “fMRI extra-dipoles”) are not a problem. Simulation studies have shown they produce little activity in the MEG inverse solution (Liu et al., 1998). For the initial fit of the evoked data from the AVcon condition, dipole locations were kept constant while the orientation were restricted to be tangential to the cortical surface and the strength allowed to vary. Then the orientations of the ECDs were kept fixed and the model was applied to the data of the remaining conditions to estimate the dipole strength over time for each condition. The analysis was performed separately for each subject and an average source model was generated for visual display.

4.3 Results

4.3.1 Neuromagnetic responses to AV speech

Figure 4-1 shows magnetic evoked fields in response to bimodal speech input. Largest signals were observed at temporal sensors (Fig. 1 A). Prominent deflections peaked at several latencies between 70 and 400 ms (Fig. 1 C). The varying spatio-

temporal pattern of the averaged sensor data indicated the presence of multiple neural generators with different time-courses (Fig. 1 B/C). The relatively flat pre-stimulus baseline and a prominent bilateral M100 component at around 100 ms post-stimulus in the temporal region indicate an adequate signal-to-noise ratio and are in accordance with previous auditory MEG experiments (Hari, 1991; Hari, 1996).

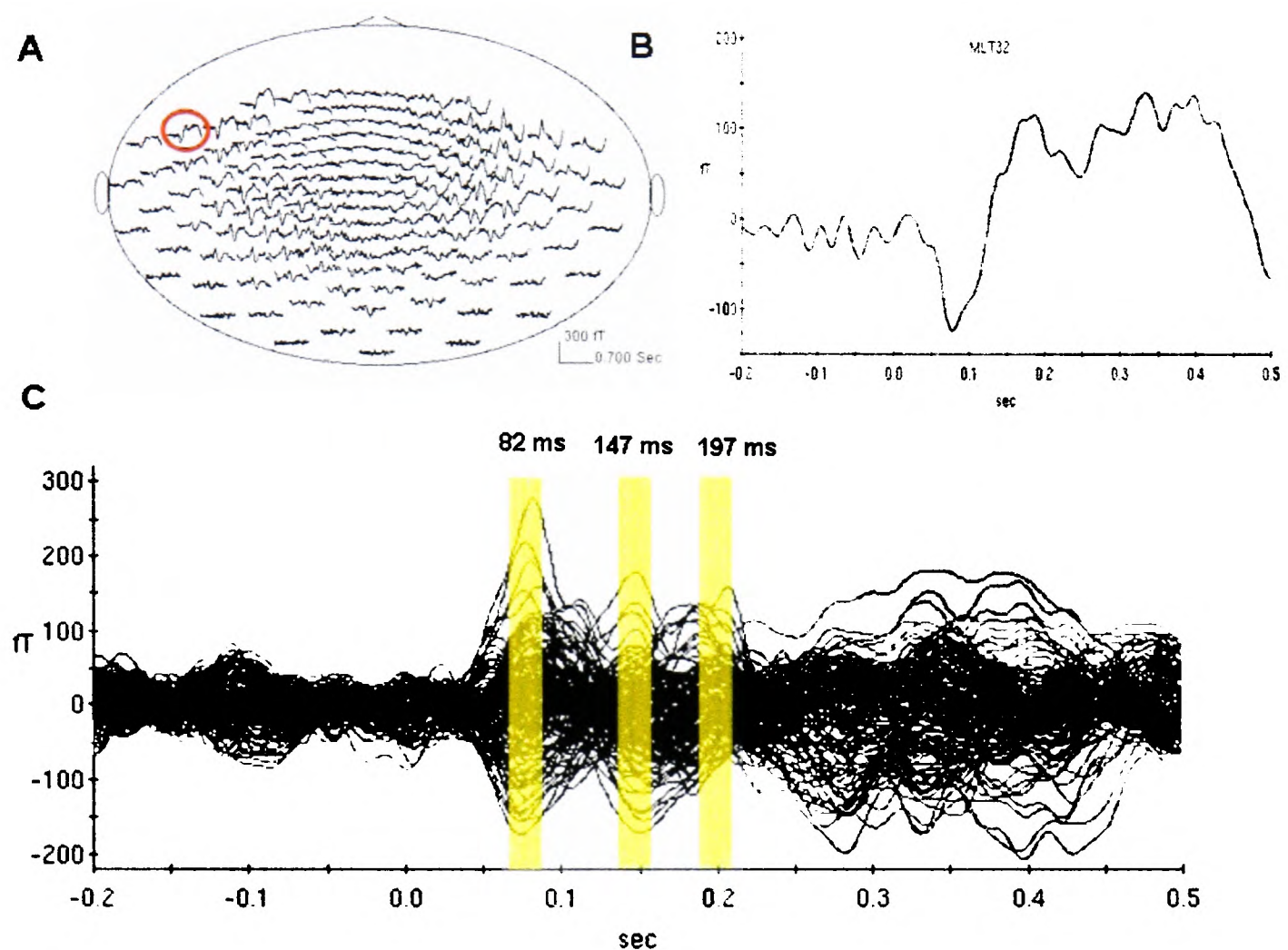


Figure 4-1: Averaged MEG data

Shown are 700 ms (– 200 to 500 ms) of subject PR in response to bimodal speech (AVcon). A shows all 151 sensors and their approximate positioning relative to the head. Red circle: sensor depicted in B, which shows in greater detail the waveform of this left temporal sensor. C depicts an overlay of all 151 channels, showing major deflections at 82 ms, 147 ms and 197 ms (yellow) for which dipoles were fitted.

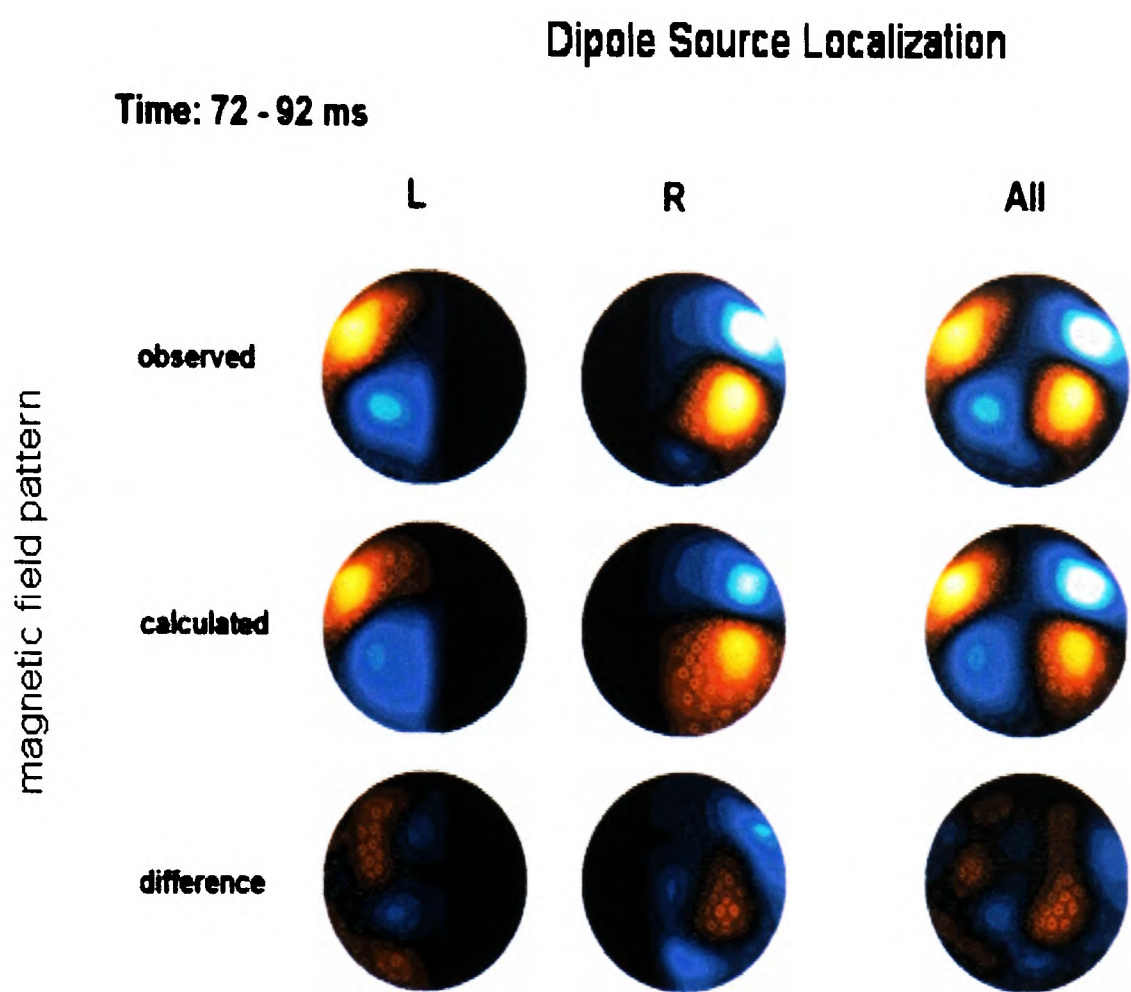


Figure 4-2: Strategy for dipole source localization

Illustration of the strategy for identifying dipolar magnetic field pattern and fitting equivalent current dipoles. Interpolated isocontour maps are shown for one subject (PR) in the time-window of 72-92 ms, corresponding to the average 82 ms time-period (yellow) in Figure 4-1 (same data). At this time interval we observe a bilateral dipolar pattern (see “observed ALL”) which can be modelled with at least two separate dipoles, one in each hemisphere. We first use only the data from the sensors of one hemisphere (see “L” and “R”) to fit one dipole which explains the observed magnetic field best. Based on this dipole fit, a forward solution is computed and visualized, showing the field pattern that such a dipole would produce (see “calculated”). Then the difference between calculated and observed magnetic field is computed and visualized (see “difference”). If any residual activity that remains unexplained, further dipoles are added. The same procedure is then repeated for the other hemisphere and for other prominent deflections in the averaged MEG signal.

4.3.2 Magnetic sources to AV speech

Figure 4-3 illustrates the results from the unconstrained MEG multi-dipole model. In subject PR we identified 7 ECDs based solely on the MEG data (see Table

4-1 for corresponding areas in the Talairach atlas (Talairach and Tournoux, 1988).

Visual analysis revealed an early response to auditory speech in the left auditory cortex (dipole 1) at around 40 ms, followed by the right auditory cortex at around 50 ms (dipole 5). Both activation peaks were followed by a larger deflection at around 180 ms, this response being more pronounced in the left auditory area. A dipole located in the left IFG (dipole 4) showed initial activity after 80 ms. Further dipoles were located in bilateral middle occipital regions (dipoles 2 & 6), in the bank of the precentral sulcus (dipole 3) and the supramarginal gyrus (dipole 7). The right-most column shows the fMRI activation of that subject in response to the bimodal auditory-visual (AVcon) stimulus (see section 3). Dipole locations and fMRI activation show a considerable degree of spatial overlap, especially in temporal and frontal areas of the left hemisphere. However, some fMRI activations were present that could not be explained by the ECD model, especially in the occipital regions. This may be because averaged evoked responses were not very pronounced over this area during the analysis period. This can be explained by the fact that the visual onset preceded the auditory stimulus by about 1166 ms (see Figure 3-2) and that evoked responses are particularly sensitive to discrete stimulus onsets and not to sustained activity over a long time window.

Unconstrained Modelling					
Dipole	L/R	Area	MNI coordinates		
<u>Subject OD</u>			x	y	z
1	L	Primary Auditory Cortex	-54	-18	16
2	L	Middle Occipital Gyrus	-24	-78	18
3	L	Pre-Central Gyrus	-36	-4	44
4	L	Inferior Frontal Gyrus	-44	10	16
5	R	Primary Auditory Cortex	48	-26	10
6	R	Post. Superior Temporal Gyrus	52	-48	10
7	L	Post. Superior Temporal Gyrus	-60	-36	4
<u>Subject PR</u>					
1	L	Auditory Cortex	-55	-18	8
2	L	Middle Occipital Gyrus	-28	-86	12
3	L	Pre-Central Gyrus	-57	-16	35
4	L	Inferior Frontal Gyrus	-50	11	20
5	R	Auditory Cortex	60	-26	11
6	R	Middle Occipital Gyrus	28	-86	6
7	R	Supramarginal Gyrus	36	-44	46

Table 4-1: Unconstrained dipole fit (dipole locations)

This table shows the locations and time-courses of ECDs based on an unconstrained multi-dipole model in two subjects. The first discernible evoked response was also seen in the left auditory cortex albeit somewhat later at around 70 ms (dipole 1), followed in time by a smaller response in the right auditory cortex at around 78 ms (dipole 5) and in the left occipital region at 84 ms (dipole 2). At around 136 ms the left IFG (dipole 4) became active. The ECD model in this subject was able to detect separate sources of activity in the primary auditory cortices (dipoles 1 & 5) from those in the posterior part of the STS (dipoles 6 & 7). In the posterior right and left STS activity was first seen in at 86 ms and 96 ms, respectively.

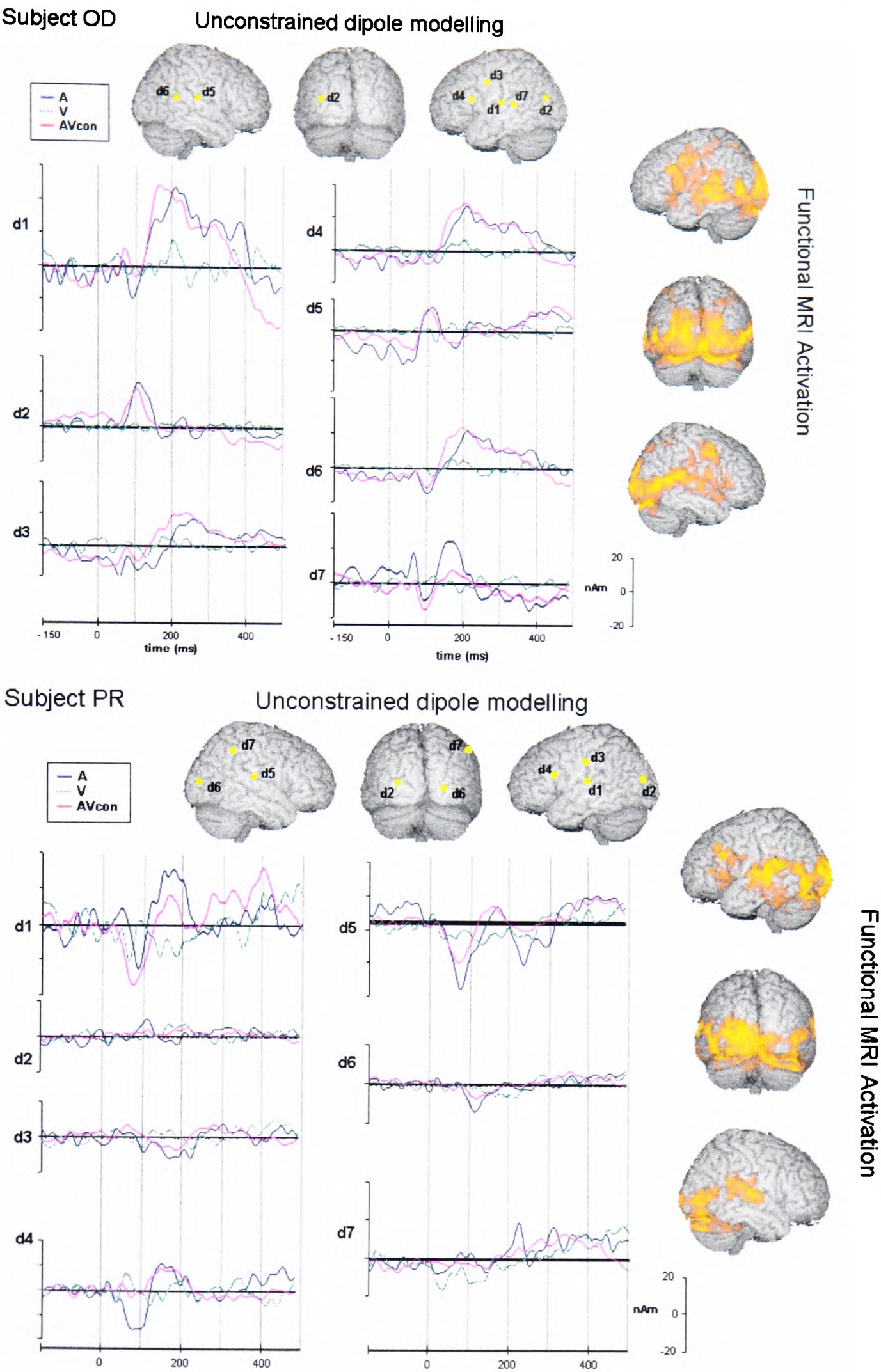


Figure 4-3: Unconstrained dipole fit (dipole moments)
Source activity of the MEG response to auditory (A), visual (V) and auditory-visual (AVcon) speech perception in individual subjects (PR & OD). These

independent and unconstrained MEG dipoles were derived from the MEG data only. Top: Projections of the MEG dipoles onto the scalp surface of a standardized brain (dipoles are labelled d1, d2, etc.). Bottom: Amplitudes (nAm) of the dipole moment as a function of time (0 = auditory stimulus onset). Right: Functional MRI responses for the individual subject to the same multisensory speech (AVcon) stimuli as used in the MEG experiment ($z>2.3$, $p=.05$).

4.3.3 FMRI constrained modelling

A total of 8 dipoles were fixed at cortical locations determined with fMRI by selecting voxels of peak signal changes in highly activated clusters (Figure 4-4).

fMRI constraint					
Dipole	L/R	Area	MNI coordinates		
1	L	Primary Auditory Cortex	-44	-19	10
2	L	Post. Superior Temporal Gyrus	-68	-39	8
3	L	Inferior Frontal Gyrus	-50	17	27
4	L	Middle Occipital Gyrus	-51	-77	3
5	R	Primary Auditory Cortex	46	-16	6
6	R	Post. Superior Temporal Gyrus	71	-29	6
7	R	Middle Occipital Gyrus	53	-76	-3
8	8	Primary Visual Cortex	6	-97	8

Table 4-2: FMRI constrained dipole fit (dipole locations)

Locations of fixed ECD sources in the fMRI constraint model with respective MNI coordinates. Area label have been identified based on the stereo-taxic atlas of Talairach & Tournoux (Talairach and Tournoux, 1988).

The grand-averaged source waveforms and dipole locations rendered onto a standardized structural MRI are shown in Figure 4-4. The data shown for MEG and fMRI are derived from the same 9 subjects. Highest signal changes were observed in the left hemisphere, especially in auditory areas of the STG (dipoles 1 & 2). Signal

changes were also seen in the left IFG (dipole 3) and occipital regions (dipoles 4, 7 & 8).

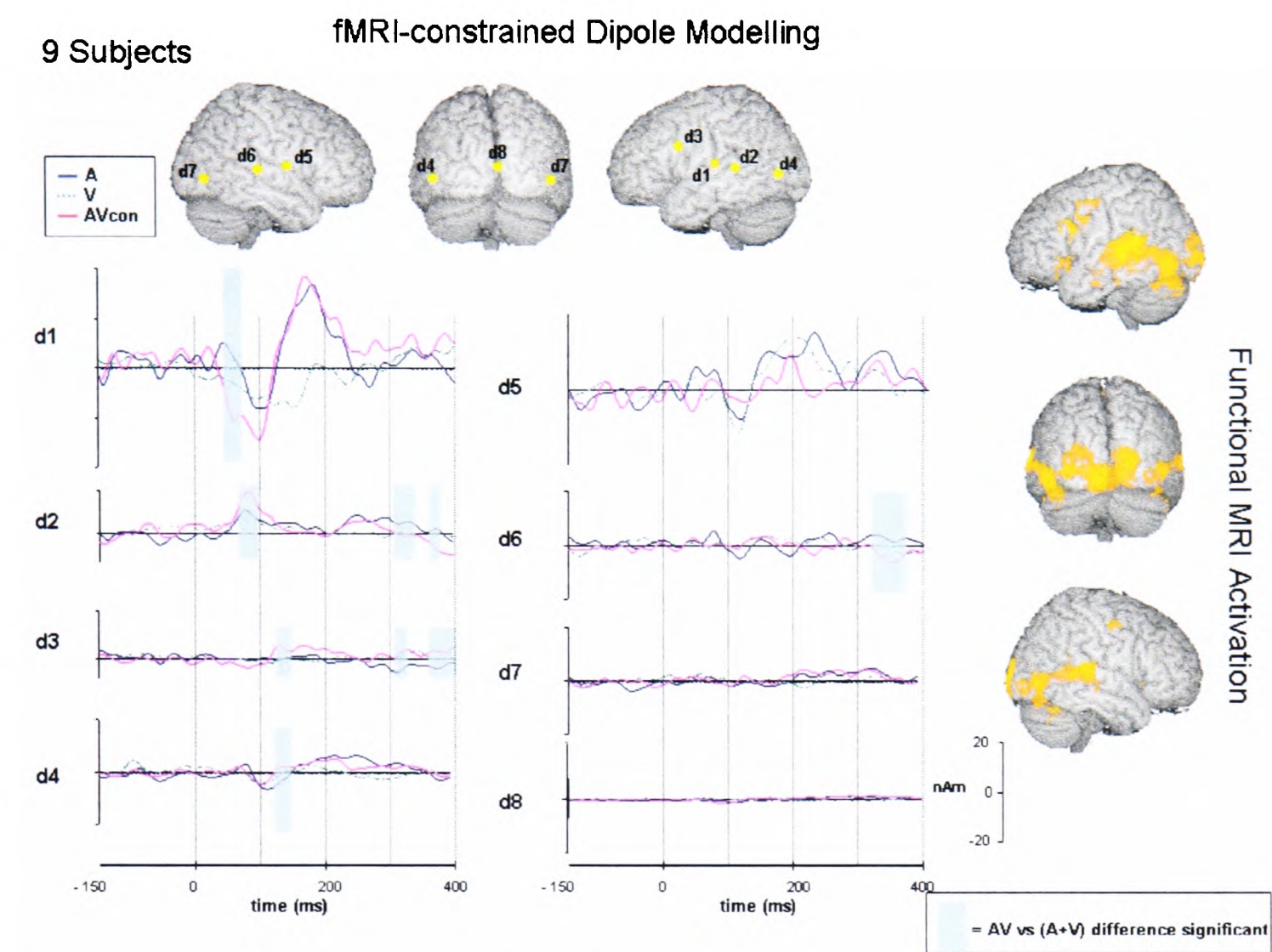


Figure 4-4: fMRI constrained dipole fit (dipole moments)
Group source activity of MEG responses to auditory, visual and auditory-visual speech from 9 subjects, derived by fMRI-constrained dipole modelling. Top: Projections of the MEG dipoles onto the scalp surface of a standardized brain (dipoles are labelled d1, d2, etc.). Bottom: Amplitudes (nAm) of the dipole moment as a function of time (0 = auditory stimulus onset). Right: Functional MRI responses from the group maps (same 9 subjects) to the same multisensory speech (AVcon) stimulus as used in the MEG experiment ($z > 4$, $p = .05$). Blue shaded areas reveal time-windows of interactions; AVcon vs (A+V) $p > .01$.

4.3.3.1.1 *The timing of cortical events*

In order to describe the timing of events in the brain areas involved in auditory and visual speech processing we needed to determine the exact onset of activity in

dipoles at different locations. The MEG time-series, with its positive and negative deflections, inevitably contains noise which makes defining the exact onset of activity not straightforward. To solve this problem we used a statistical criterion which takes the system noise into account for determining the significance of activation. For this purpose we calculated the root square mean (RSM) value of each time point of the dipole moments of all individual subjects and used the RSM time-series to compare a passive pre-auditory stimulus baseline (assumed to contain the system noise) with an active period after auditory stimulus onset using a paired Student's t-test ($p < 0.01$, requiring at least 20 ms of consecutive significance to reduce the likelihood of Type-I errors (Guthrie and Buchwald, 1991)). Results are shown in Figure 4-5.

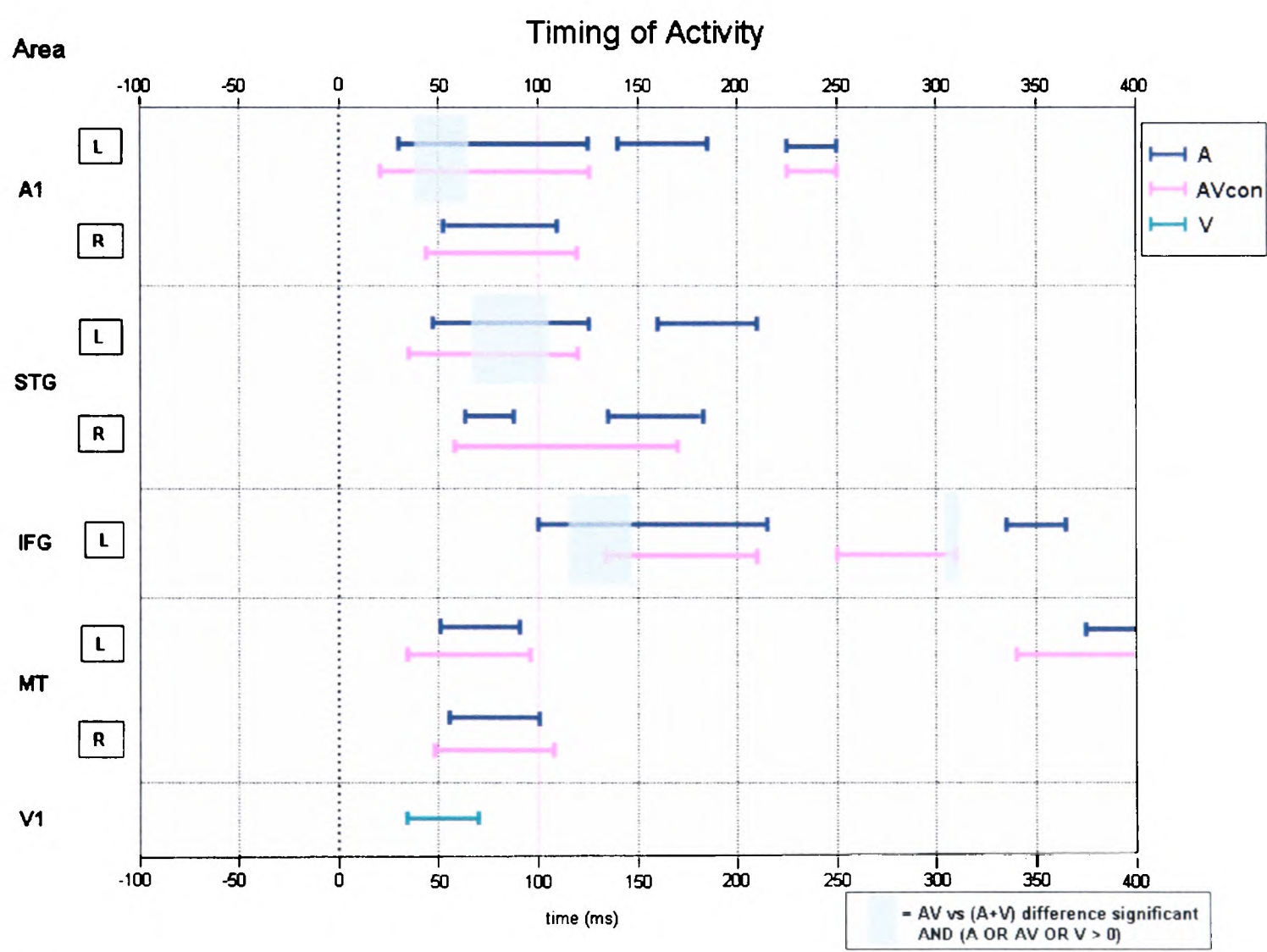


Figure 4-5: Timing of dipole activation

Onset and offset of activity in different cortical areas during auditory speech (A), visual speech (V) and bimodal speech (AVcon) perception. A1: primary auditory cortex; STG: superior temporal gyrus; IFG: inferior frontal gyrus; MT: occipito-temporal region; V1: primary visual cortex. Time point zero: auditory

stimulus onset. The blue shaded areas show time-windows of significant difference between AVcon vs (A+V) given that at least one of the conditions activates significantly during this time-period.

4.3.3.1.1.1 **Timing of auditory speech**

By using statistical criteria we were able to estimate the onset of responses in different cortical areas and to describe the activation sequence of auditory and auditory-visual speech (Figure 4-5). Auditory speech activated the left auditory cortex 30 ms after auditory stimulus onset, followed by the left STS at 45 ms. Activity in the right hemisphere developed more slowly, but followed a similar profile with the first activity being detected in A1 (52 ms) and the STG (62 ms). Magnetic signal changes to auditory speech in area MT were detected at 51 ms (L) and 56 ms (R). The IFG did not activate until 102 ms after auditory onset and no statistically significant activation to auditory speech was detected in the primary visual cortex. The addition of a congruent visual stimulus speeded up the response time in various areas, including the left A1 (20 ms) and STG (36 ms) and right A1 (35 ms) and STG (58 ms). Only in the IFG was the activity delayed as a result of bimodal speech input.

4.3.3.1.2 ***Timing of crossmodal convergence***

When we calculated the sum of responses to A and V and compared it to the responses in the AVcon condition, we found that the AVcon response could not be solely explained by the superimposition of the unimodal responses. This was indicative of an interaction effects, suggesting underlying bimodal integrative processes. The interaction effect was defined as $AVcon - (A+V)$ (Giard and Peronnet, 1999; Molholm et al., 2002; Molholm et al., 2004). Source waveforms from the bimodal condition AVcon and the summed unimodal conditions (A+V) were compared using a paired t-test. To avoid an inflation of Type II errors due to multiple comparisons an additional ‘cluster-like’ criterion was used by requiring at least 12

consecutive time-points to pass a significance threshold of $p < 0.01$ before a time-window was judged significant (Guthrie and Buchwald, 1991). Blue shaded areas in Figure 4-4 indicate differences of statistical significance for the interaction effect ($p < 0.01$).

AV speech interacted first in the left A1 (38-64 ms), then left STS (83-113 ms), followed by the left IFG (118-139 ms). Several other areas also showed a significant difference between the two conditions. However, when we applied the additional criterion of requiring significant activity (as determined according to Figure 4-5) in at least one of the three conditions, only A1, STS and the IFG of the left hemisphere showed crossmodal effects (see blue shaded area in Figure 4-5). The multisensory effect was manifested as an increase in moment amplitude when a visual stimulus was present.

4.3.3.1.3 *Timing of phonetic convergence*

Figure 4-6 shows the grand averaged results of the fMRI constrained dipole fit for bimodal stimulus pairings AVcon, AVincon(A) and AVincon(V). For this model an additional two dipoles were fixed in the insulae bilaterally to account for changes seen in the fMRI data between congruent and incongruent AV pairings. An incongruent auditory stimulus (AVincon(A)) caused initial activity changes in the auditory cortex at 72-104 ms (L) and 88-110 ms (R). In the right insula statistically significant activity between AVcon and AVincon(A) were observed at 91-113 ms, 171-193 ms and 320-361 ms, whereas the left insula showed a difference only at 320-361 ms. Late differences were shown in the left STS (355-400 ms), left IFG (304-326 ms) and left MT (320-342 ms) and right MT (329-362 ms).

Incongruent pairing of the auditory stimulus with visual gurning (AVincon(V)) generally affected neural processing at a later stage in left STS (112-

136 ms), left auditory cortex (243-313), left IFG (132-195 ms & 238-259 ms) and right insula (217-241 ms).

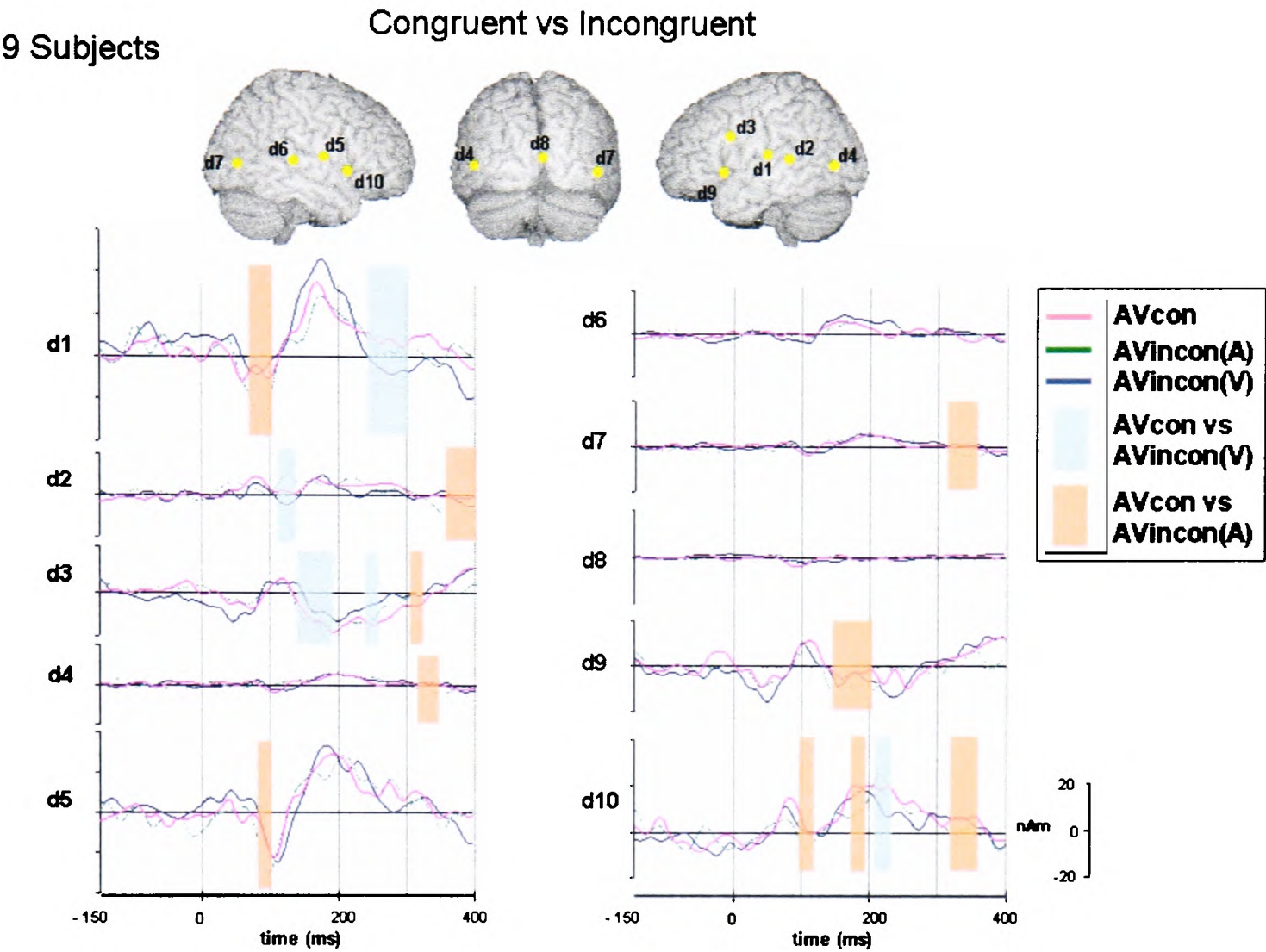


Figure 4-6: Timing of phonetic interactions

Group source activity of MEG responses to congruent (AVcon) and incongruent (AVincon(A), AVincon(V)) auditory-visual speech from 9 subjects, derived by fMRI-constrained dipole modelling. Top: Projections of the MEG dipoles onto the scalp surface of a standardized brain (dipoles are labelled d1, d2, etc.). Bottom: Amplitudes (nAm) of the dipole moment as a function of time (0 = auditory stimulus onset). Shaded areas reveal time-windows during which the incongruent activity differs significantly ($p < .01$) from congruent auditory-visual speech.

4.4 Discussion

4.4.1 General findings

The results of the MEG study reveal three major characteristics of auditory-visual speech processing. First, bimodal speech appears to interact at an early stage of auditory processing in primary and secondary areas of auditory cortex, and only subsequently in the posterior STG and IFG. These crossmodal speech effects appear to be restricted to the left hemisphere. Second, the results show the progression and timing of the effect through speech-related cortical areas ($A1 \rightarrow STG \rightarrow IFG$) and reveal that the presence of a visual stimulus seems to speed up the neural response to an auditory speech stimulus, even though this was not statistically tested. Third, phonetic integration occurs at a later stage of cortical processing than integration based on spatial and temporal coincidence. To explain our findings of very early interactions during non-phonetic integration, which are already evident at the stage of primary sensory cortex, we suggest a mechanism by which visual speech modulates activity in auditory cortex before the onset of acoustic stimulation. Phonetic integration is explained by hierarchical processing of speech sounds and merging with the visual stream after auditory speech feature extraction.

4.4.2 Detailed discussion

The earliest cortical response was observed to auditory-visual speech in auditory cortex at 20 ms. Auditory activity in the cortex has been reported previously for components of that latency from vertex recordings (Ponton et al., 2000) and their generator has been suggested to locate to the primary auditory cortex (Scherg and Von Cramon, 1986). This finding is also consistent with data from intracranial recordings in humans, where primary auditory field responses during this latency were localized to the dorso-postero-medial part of Heschl's gyrus (Liegeois-Chauvel

et al., 1994). They also found that components with onset latencies of around 50-100 ms were localized in more lateral parts of the HG and in posterior parts of the STG. These iEEG results are in agreement with the latency results from the present MEG experiment and thereby support the assumption that the fMRI-informed ECD model is able distinguish activity of two independent dipoles in A1 and STG regions. This was not always the case in the unconstrained model.

Evidence for multisensory integration was also observed in auditory cortex from 38-64 ms, followed by interactions in STS at 83-113 ms and IFG at 118-139 ms.

Early interactions between 40-60 ms have been previously observed with simple AV stimuli (Giard and Peronnet, 1999; Fort et al., 2002; Molholm et al., 2002). Together with our results the findings from these studies support our assumption that effects based on the interaction metric largely reflect general, non-phonetic integration processes. Using scalp EEG with 30 electrodes and a similar set of AV stimuli, (Klucharev et al., 2003) found earliest interactions starting at 50 ms and peaking at 85 ms, which localized to extra-striate areas in occipital cortex. Only at 125 ms did they observe first integration responses in temporal areas, which were right-lateralized. The reason for finding similar early latencies but at different cortical sites than in our study might be based on the type of stimulus chosen by (Klucharev et al., 2003) and the study design. Whereas we used an event-related design with AV stimulus asynchrony of ~432 ms, they used a block-design and AV stimuli with onset asynchronies of 95 ms. This suggests that both the context and particular differences in AV asynchrony may be important variables in AV speech integration.

The IFG is traditionally associated with motor aspects of speech, and lesions in this regions lead to disorders of speech production (Broca, 1961). However, there is also evidence for an involvement of IFG in perceptual speech functions (Zatorre et al., 1996). There is also ample evidence for a role of IFG in processing visual speech (Campbell et al., 2001; Nishitani and Hari, 2002; Nishitani et al., 2005) and in the

processing of auditory-visual speech (Watkins et al., 2003). Ojanen et al. (2005) found an increase in BOLD signal to phonetically/visemically mismatching AV vowels in left IFG. Even though we did not find any such evidence in our fMRI study, we found a modulation by gurning (AVincon(V) vs Avcon at 132 ms in the MEG data, suggesting that IFG compares visemic and phonetic gestures via a common code manifested in motor representations.

Contrasting congruent AV stimuli with AVincon(V) where gurning replaced visemic information in the visual signal, earliest difference were observed at 112 ms in left STS. Further differences were observed in left auditory cortex (243-313), left IFG (132-195 ms & 238-259 ms) and right insula (217-241 ms). To detect a difference between gurning and articulatory mouth movement in the context of AV integration, the brain needs to process the input first based on its phonetic representation. Thus, the present result further support the notion that left STS is involved in the phonetic analysis and categorization of speech signals (Jancke et al., 2002). The results suggest that both primary sensory and sensory association areas are involved in phonetic and non-phonetic AV feature extraction and binding, but that they do so at different times in the processing sequence, supporting the view that not only space, but also timing is an important component for understanding brain function.

The presence of an incongruent auditory speech stimulus was detected early at the level of the bilateral auditory cortex compared to congruent bimodal stimulation. This is in agreement with previous findings in the literature since auditory cortex seems to be necessary for computing and representing acoustic properties of stimuli, such as those differentiating between /da/ and /gi/ (Griffiths et al., 2004). As such, these findings provide further evidence for integrative processing of auditory-visual speech information in the auditory cortex.

Even though the role of the insula cortex in processing auditory information is still poorly understood, there is ample evidence that shows that this structure is involved in a variety of auditory tasks, such as sound detection, temporal processing, and motion detection (Bamiou et al., 2003). Indeed, the right anterior insula has been shown to be a multimodal area responsive to visual, tactile and auditory stimulation (Downar et al., 2000) and involved the detection of auditory-visual onset asynchrony (Bushara et al., 2001). Similarly, Calvert et al. (2001) used fMRI and white noise and a visual chequerboard as stimuli to assess the areas responsible for non-speech AV integration. Here, insular activation was seen in the auditory alone condition and in response to the strict multisensory criterion of requiring a super-additive response during congruent bimodal stimulation and a suppressive response during incongruent bimodal stimulation.

Together with previous findings, the present results suggest that the right insula is involved in the detection of auditory-visual incongruence. However, the present experimental design does not allow any definite conclusions to be drawn on this matter, as the insula is also recruited during phonological processing, such as rhyming and verbal short-term memory (Paulesu et al., 1993), and auditory sequencing (Griffiths et al., 1997). Auditory /gi/ stimuli in this experiment can not be purely treated as components of an incongruent AV pair, but also as an infrequent deviant stimulus amongst more frequent auditory /da/ stimuli. Therefore, without a condition to control for changes in the auditory sequence, such interpretation becomes difficult.

A note of caution is needed at this point in regard to multimodal imaging using the spatial resolution of fMRI and the temporal resolution of MEG. Hemodynamic measures may not capture all elements of the neurophysiological phenomena recorded with MEG. For example, latency variations in specific ERP peaks are often associated with different perceptual consequences, peaks may be too small and transient to be

noticed as hemodynamic activations (Vitacco et al., 2002) and some features of neuronal processing, such as oscillatory or spectral changes are not detected by hemodynamics, at least in no way we presently know (Horwitz and Poeppel, 2002). Functional MRI results may prove to better constraints in MEG modelling and provide better confidence in the inverse solution; however, more basic research is needed on the coupling between neuronal dynamics, hemodynamic response and MEG signal.

Previous MEG studies using mismatch-negativity (MMN) paradigms showed that incongruent AV pairings could elicit a MMN response (Sams, 1991; Colin et al., 2002; Colin et al., 2002; Mottonen et al., 2002; Colin et al., 2004). MMN responses are elicited by infrequent deviant auditory stimuli which are presented within a sequence of frequent standards and are thought to reflect the difference between the memory trace of the preceding sensory stimulus and that of the current stimulus (Naatanen, 2001). Being capable of localizing different analysis stages of auditory processing, MMN responses are less apt at determining the timing of the first interactions between bimodal stimuli as they usually peak between 100-250 ms. Therefore, these studies may show that the auditory cortex is involved in AV speech integration, but the exact timing of the integration of these responses remains unclear.

Previous fMRI studies have reported findings of auditory-visual integration in the auditory cortex and STG regions using a variety of different criteria, but the exact temporal development of the synergistic effects have remained controversial (see Calvert, 2001; Calvert and Thesen, 2004 for reviews). The present results suggest a network of cortical areas which sequentially evaluate and integrate inputs from different modalities and show that the auditory and visual speech signals interact first at the level of the auditory cortex (A1) at ~ 40 ms and that further integrative processing takes place in the left STG (~ 85 ms) and finally in the left IFG (~ 120 ms). Integration of multisensory speech signals based on their phonetic features

occurs later and follows a different spatio-temporal pattern. Integration starts in STS, proceeds to IFG and then later enters auditory cortex.

The evoked response around 40-50 ms has been localized to auditory cortex by many investigators (Pelizzone et al., 1987; Woldorff et al., 1993; Borgmann et al., 2001; Yvert et al., 2001). Finding a similar temporal profile for the A1 dipoles in our model lends further support to appropriateness of the fMRI constraint dipole fit and that auditory cortex and STS regions can be dissociated.

Not all previous studies concur with our results. In one event-related MEG study using synthetic AV CV speech stimuli, Mottonen et al (Mottonen et al., 2004) were able to fix a dipole to the N100 response occurring at around 100 ms, which located to the auditory cortex, in a location slightly more lateral than the A1 dipole in our fMRI constrained fit. A statistical significance between AVcon vs (A+V) was observed between 150-200 ms from the dipoles in both hemispheres, but no earlier interactions were evident. The discrepancy in findings could be explained by the fact that their model did not differentiate between A1 and STG activity and therefore the effect did not reach statistical significance when pooled over a larger area. A minimum current estimate (MCE) analysis conducted on their data also averaged together activity within an ROI encompassing all activity in super-temporal area and revealed multisensory interaction effects in the right hemisphere only. Another interesting speculation to explain the fact that the study by (Mottonen et al., 2004) did not detect early interaction effects relates to our hypothesis that AV asynchrony is an important factor in bimodal speech processing, especially for early modulation of auditory cortex. Their synthetic AV stimuli were perfectly synchronized and thus did not offer the advantage of visual modulation of auditory cortex before auditory onset. Synchronous stimulus onset would have made an early priming of auditory cortex by visual speech impossible as far as timing is concerned (Scherg and Von Cramon, 1986; Foxe and Simpson, 2002).

A recent EEG study by Wassenhove et al. (2005), found early modulatory effects of visual speech on auditory processing at ~ 50-100 ms for AV stimuli with onset asynchrony. Even though this study did not use any signal source projection technique to infer about the source of the activity, it does provide supporting evidence for our finding of early integration in auditory cortex. Consistent with the present findings, the study by Wassenhove et al, (2005) failed to find any evoked response to visual stimulation alone after the time-point where the auditory stimulus would have occurred.

Unlike artificial AV pairings often used in experimental research (Fort et al., 2002; Molholm et al., 2002), natural auditory and visual speech stimuli are rarely synchronized in respect to their start. Stimulus onset asynchronies can often be several hundred milliseconds long and were 432 ms for the stimuli used in the present study. Therefore, models concerned with the temporal aspect of multisensory integration of speech have to be constructed differently from those where AV pairings are synchronous.

It is tempting to suggest that during the period between lip movement and auditory onset the visual system ‘primes’ or prepares auditory areas for the arrival and processing of the auditory stimulus counterpart (van Wassenhove et al., 2005). Lip movements and speech utterances are inherently correlated and this association has been learnt throughout development. Since visual speech inputs precede the auditory signal and our fMRI data show that lipreading alone activates the cortical speech system, the preceding visual information can therefore potentially act as a predictor for the acoustic input and influence its subsequent processing. This mechanism is particularly attractive insofar that it might explain the facilitating effects of audio-visual speech seen behaviourally (Welch and Warren, 1986) and the neuronal integration at early stages of sensory processing reported sporadically in previous studies (Sams, 1991; Mottonen et al., 2002) and confirmed in the present

investigation. Such a mechanism would not require a direct cortico-cortical connection between primary sensory areas. Even though recent tracer studies in monkeys have shown these connections exist from auditory parabelt regions to areas in the primary visual cortex coding for peripheral space (Rockland and Ojima, 2002), they are very sparse and connections of the reverse projection have not been reported. In order to test this ‘visual predictor hypothesis’, one would have to show a direct modulation of visual speech on activity in auditory cortex, occurring after the start of lip movements and before the arrival of the acoustic utterance. An evoked potential analysis as used above is not particularly suited to test this question since ERPs, magnetic and electric, are most sensitive to drastic points of change in a stimulus (e.g. onsets) and tend to decrease rapidly in amplitude after this. In fact, the averaged evoked response of the MEG data shows no discernible activation for the 700 ms time period immediately preceding auditory onset. And in the absence of an evoked response, dipole modelling becomes impossible.

The next section, therefore, will address this question by using an analysis method which detects stimulus-related changes that are not necessarily phase-locked and using the frequency information contained in the MEG signal to infer about the nature and timing of the neuronal correlates of crossmodal influences during speech perception.

5 The superior temporal lobe during bimodal speech perception

In a previous section we have demonstrated with MEG averaging and dipole source analysis techniques that the superior temporal lobe processes and integrates auditory and visual speech information in a specific fashion. We found that first interactions take place in auditory cortex and subsequently in STS. Since MEG source solutions inherently contain a degree of uncertainty, this section will test these findings using direct recordings from the surface of the brain of behaving humans using intracranial EEG.

5.1 Introduction

Results from both our fMRI and MEG experiments suggest a strong involvement of the left superior temporal lobe in the integration auditor and visual speech signals (see section 3&4 of this thesis). Dipole modelling of the MEG data with fMRI-constraints suggests that these multisensory effects in the temporal lobe start early at around 40 ms after auditory onset.

However, the certainty of these assumptions is restricted mainly by the technical limits that the individual neuroimaging methods impose, namely the low temporal resolution of fMRI and the ambiguity associated with the MEG source localization. Intracranial-EEG does not suffer from any of these restrictions and is therefore an ideal method to test specific spatio-temporal hypotheses about the functional coding and temporal processing of a circumscribed cortical region.

A further advantage of using a patient population undergoing surgery for temporal lobe epilepsy (see Methods) is the availability, and hence exactness, of speech lateralization estimates obtained from invasive WADA testing. A disadvantage of iEEG is that the field-of-view is comparably small and in our case extended to ~ 5 cm², which was, however, still sufficient to cover a substantial area on the superior

temporal lobe. Results from iEEG can also be used to validate fMRI and MEG analysis approaches and thereby help strengthening the confidence in results obtained over the whole cortex by these methods (Halgren, 2004). Section 8 of this thesis will discuss this topic in more detail and present a comparison of the results obtained from MEG, fMRI and iEEG.

Both the present results and previous studies suggest that areas of the STS and STG are involved in the integration of auditory-visual speech signals. Superior temporal areas are consistently activated by bimodal speech signals and often show an enhanced response to bimodal versus unimodally induced stimulation (Calvert et al., 2000; Callan et al., 2003; Wright et al., 2003). It has been shown that visually, the STS responds best to complex stimuli, such as biological motion and different facial expressions (Puce et al., 1995; Puce et al., 1998; Hasselmo et al., 1989; Oram and Perrett, 1994), and that it receives extensive projections from visual and auditory regions (Seltzer and Pandya, 1978), including a region containing neurons shown to respond selectively to different species-specific calls (Rauschecker et al., 1995). Single-unit physiology in macaques has shown that individual cells in the STS can code both auditory and visual information (Benevento et al., 1977; Bruce et al., 1981; Hikosaka et al., 1988). A recent single-unit recordings in monkeys revealed that a subset (23 %) of neurons in the STS, that code for the sight of actions, show a modulated response to auditory-visual stimulation (Barraclough et al., 2005). In about half of the multisensory neurons the visual response decreased (by about 47%), whereas the other half showed an increase (by about 86%) that was contingent upon the congruence of the visual and auditory actions.

There is also a large body of evidence suggesting a role of auditory cortex in multisensory speech processing. For example, in fMRI studies visual speech gestures have repeatedly been shown to activate secondary areas of auditory cortex in the absence of acoustic speech input (Calvert et al., 1997; MacSweeney et al., 2000;

Campbell et al., 2001). In a recent study, Pekkola et al. (2005) performed an ROI analysis of anatomically defined HG on BOLD signal changes in response to visual speechreading and visual perception of moving circles. Visual speech vs baseline activated left HG in 9 out of 10 subjects, and showed higher signal changes when compared to moving circles. Interestingly, in 7 subjects the activation extended into the medial part of HG, i.e. into primary auditory cortex. Further support of the idea that, in addition to the STS, auditory cortex is involved in the integration of congruent AV speech signals comes from MEG. Here, timing information suggests that integration in auditory cortex precedes that of the STS (Mottonen et al., 2004).

In summary, the available data suggest that the auditory cortex and superior temporal sulcus in humans are involved in the integration of auditory-visual signals, and of bimodal speech signals in particular. To test this hypothesis with unambiguous, high spatial and high temporal resolution we used an almost identical paradigm as in section 3 & 4 and recorded EEG responses from subdurally implanted electrode grids on the pial surface of the superior temporal lobe of awake and behaving human subjects. Based on our MEG results we hypothesized that modulation of auditory speech processing by a congruent visual stimulus would occur at an early stage (< 40 ms) in the auditory cortex and that further integrative processes occur subsequently in the STS region

5.2 Methods

5.2.1 Human subjects

Recordings were performed on 2 patients undergoing diagnosis and, subsequently, surgery for medically intractable epilepsy. Multi-contact electrodes were implanted over the perisylvian cortex to monitor seizure activity continuously for a period of up to two weeks (Figure 5-1). Research recordings were obtained during this time while subjects were reclining comfortably in their bed with calibrated

ear insert phones in their head and a monitor screen in front of them. Informed consent was obtained from both subjects and the study protocol was approved by the University of Iowa Institutional Review Board. Presurgical audiometric tests were performed and showed that for both subjects’ pure-tone thresholds (500-4000 Hz) and language comprehension were within normal limits. Intra-arterial amobarbital (WADA) testing showed that for both subjects the left cerebral hemisphere was dominant for speech.

Subject	S100	S106
Age	31	35
Gender	F	M
Years of Education	18	11
VIQ	114	115
FSIQ	113	103
Digit Span	15	13
Age of Seizure Onset	high school	30
Language Dominance (WADA)	LH	LH
Laterality of Recording	LH	LH
Seizure Focus	Left Medial Temporal	Left Medial Temporal

Table 5-1: IEEG subject demographics

Including results from neuropsychological and WADA testing, and type of anticonvulsant medication. VIQ: verbal IQ, FSIQ: full scale IQ, These scores were prorated from scores of part of WAIS-R subtests.

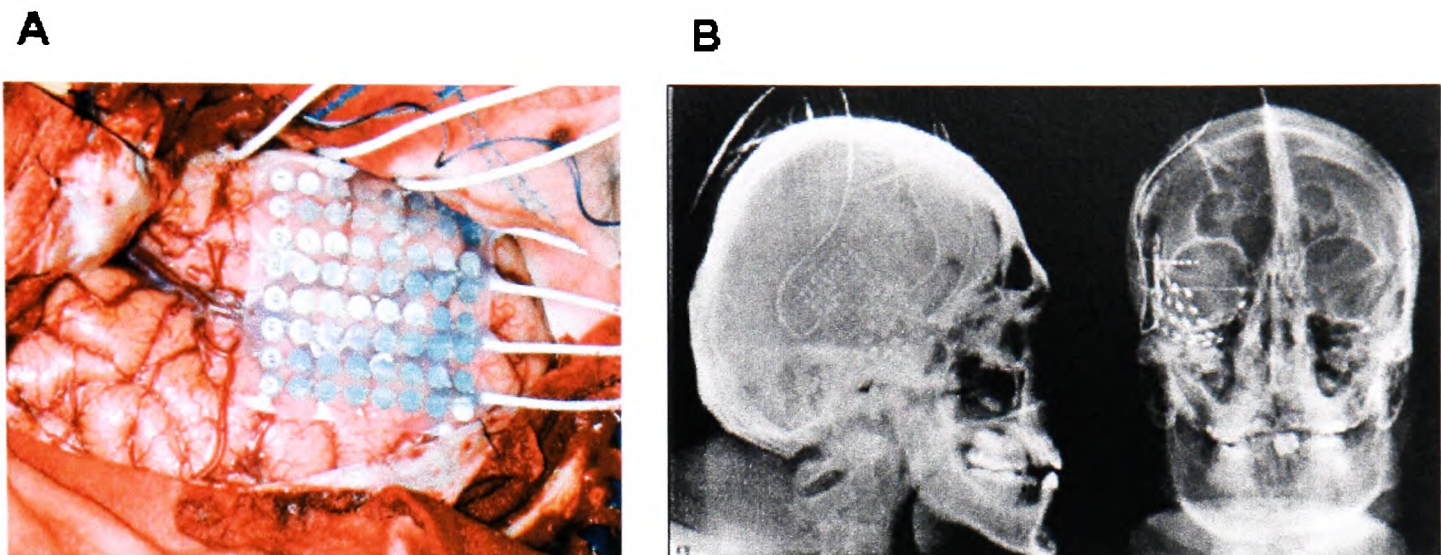


Figure 5-1: Electrode grid placement during surgery

A. Intraoperative photograph of electrode grid placement on the brain surface

B. X-ray image of the skull of one subject with the implanted recording electrodes on the temporal lobe.

5.2.2 Acoustic calibration and stimulus presentation

Auditory stimuli were delivered binaurally over miniature earphones (ER-4; Etymotic Research, Elk Grove Village, IL) integrated into custom fitted ear moulds. Acoustic calibration was achieved by presenting maximum length sequences (Golay codes, Zhou et al., 1992) and recording their responses through a clinical probe microphone (ER-7C: Etymotic Research) inside the ear canal. Equalizer filters were created for sound-path electrodes to the left and right ears and were used to pre-compensate the auditory stimuli for each ear. Auditory-visual AVI files were played with a commercial software (Real Player, Real Networks Inc.) at a comfortable sound level (typically 35-40 dB above threshold) from a dedicated PC platform with a standard CRT monitor. An experimenter sat at the patient's bedside to monitor for fixation and alertness. As in the preceding fMRI and MEG experiments, patients had to monitor both sensory modalities (auditory and visual) for the infrequent occurrence

of a /tu/ stimulus. The requirement of making button responses to the target provided a means for assessing the patient's attention and ability to perform the task.

5.2.3 Electrophysiological recordings

In each subject, an array of 64 (8 x 8) platinum-iridium disc electrodes (1.6 mm diameter, 4-5 mm inter-electrode distance) embedded in a silicon membrane was implanted on the pial surface over the peri-Sylvian regions of the temporal lobe. Reconstruction of the electrode grid on rendered structural MRIs in respect to gyral landmarks was aided by intraoperative photographs (Figure 5-1). The distribution of recording sites varied between subjects, as the placement of the electrode grid depended entirely on clinical considerations. Intracranial EEG data were acquired (2k Samples/sec; bandpass 1-1000 Hz) simultaneously from the 64 surface-electrode recording grids implanted on the left hemisphere of two subjects. Separate platinum dish electrodes located subcutaneously near the vertex were used as references and ground. ERPs obtained with audio-alone and with auditory-visual stimuli were time-locked to the onset of the auditory stimulus, which was derived from a digitized copy of the delivered audio waveform. For the V condition, ERPs were timed with respect to the video frame from which the auditory stimulus was removed (nominally 432 ms after the onset of lip movement). This fiducial timing signal was marked by the appearance of a small white circle (see Figure 3-3) in the top left corner of the screen (unseen by the subject). A TTL signal from the button box was digitized on another A/D channel. These multiple time markers were needed to synchronize audio and visual events, as the operating system of a PC could interrupt playback at unpredictable times. Electrophysiological recordings were screened for epileptiform activity, and only data that were free of abnormal discharges were analyzed.

5.2.4 Data analysis

For ERP analysis of the iEEG data a low-pass filter of 40 Hz was applied to the raw data trials before averaging. A total of 40 trials were acquired for S100 and a total of 70 trials for S106. Further preprocessing steps included detrending and baseline correction (baseline= -1500 to -1300 ms). In order to make inferences about the effects of the different conditions on ERP waveforms, a two-sided pair-wise t-test was employed on the individual trials for every time-point at each electrode. A cluster-level randomization procedure was used to control for inflated Type-II errors using ClusterRandAnalysis, a MATLAB based function which is part of the FieldTrip analysis package (<http://www2.ru.nl/fcdonders/fieldtrip>). For clustering, a cluster-finding algorithm finds all clusters connected via the temporal and spatial dimensions. Then a cluster-level statistic is computed by calculating a single t-statistics for every spatio-temporal cluster. The Type-I error rate for the spatio-temporal data matrix was controlled by evaluating the cluster-level statistics under the null distribution of the maximum cluster-level statistics. Thus, the actual test statistic in this case was the maximum of the cluster-level statistics. The multiple comparison problem is solved by controlling the Type-I error rate for this single test statistic (Maris, 2004; (Maris and Oostenveld, in press). The randomization null distribution was obtained by a Monte-Carlo randomization with 500 draws. A conservative confidence interval of $p < 0.001$ was selected for both the cluster-level threshold and randomization tests.

5.2.4.1 *Spatio-temporal maps*

In order to make topographic comparisons of activity and effects over time, ERPs and statistical effects were visualized on a two-dimensional matrix equivalent to the spatial distribution of the sensors on the 8x8 electrode grid using MATLAB (MathWorks Inc.). Grid Data were interpolated and then colour-coded for polarity and amplitude (Figure 5-4).

5.2.4.2 *Heschl's gyrus*

The transverse temporal gyrus of Heschl's (HG) contains the human auditory cortex. It is located in the depth of the Sylvian fissure on the temporal plane. Primary auditory cortex corresponds to BA 41, which is surrounded caudolaterally by area 42 (secondary auditory cortex), rostromedially by area 22 (lateral posterior STG), and medially by area 52 (insula). The morphology of HG makes it possible to identify this structure on structural MRI scans of even medium resolution (e.g. 1x1x1.5 mm), (Rademacher et al., 2001). Structural MRI scans were acquired from both patients at a resolution of 1x1x1 mm. Heschl's Gyrus was identified based on these scans similarly as in Pekkola et al. (2005), and as the most anterior of the transverse temporal gyri, if more than one was present, bounded caudally by Heschl's sulcus and laterally by the lateral tip of the transverse temporal gyrus. Since HG cannot be seen directly from the cortical surface, where the grid lays, the trajectory of HG from posterior to anterior was marked by a yellow line which was projected onto the surface rendering of the structural MRI (Figure 5-2). This surface projection allows the specification of electrodes immediately above HG.

5.3 Results

5.3.1 Electrode grid locations

5.3.1.1 *S106*

The grid covered much of the middle and posterior aspects of the STG, and extended on to the parietal/frontal lobes dorsally and the middle temporal gyrus ventrally (Figure 5-4). Some electrodes also extended into the ventral aspect of supramarginal gyrus (BA 40).

5.3.1.2 *S100*

In S100 the electrode grid was placed over the left, speech dominant hemisphere, and slightly more posterior/superior than that of S106 (Figure 5-5).

Electrodes covered some of the middle and much of the posterior aspects of posterior STG. Dorsally, the grid expanded beyond the sylvian fissure, covering inferior parts of the central and precentral sulci and extending into anterior parts of the supramarginal gyrus.

5.3.2 Event-related potentials

Intracranial ERPs revealed a prominent auditory evoked response recorded from the lateral surface of the STS after auditory stimulation. The responses mostly had the character of a polyphasic potential with major peaks at around 40-60 ms, 80-90 ms, and 130-150 ms. No evoked response was visible before auditory onset (e.g., to face onset) and no evoked response was detected in the V only condition (Figure 5-3). Responses in the incongruent bimodal conditions were of similar morphology as those in the AV congruent condition.

Location of Heschl's Gyrus

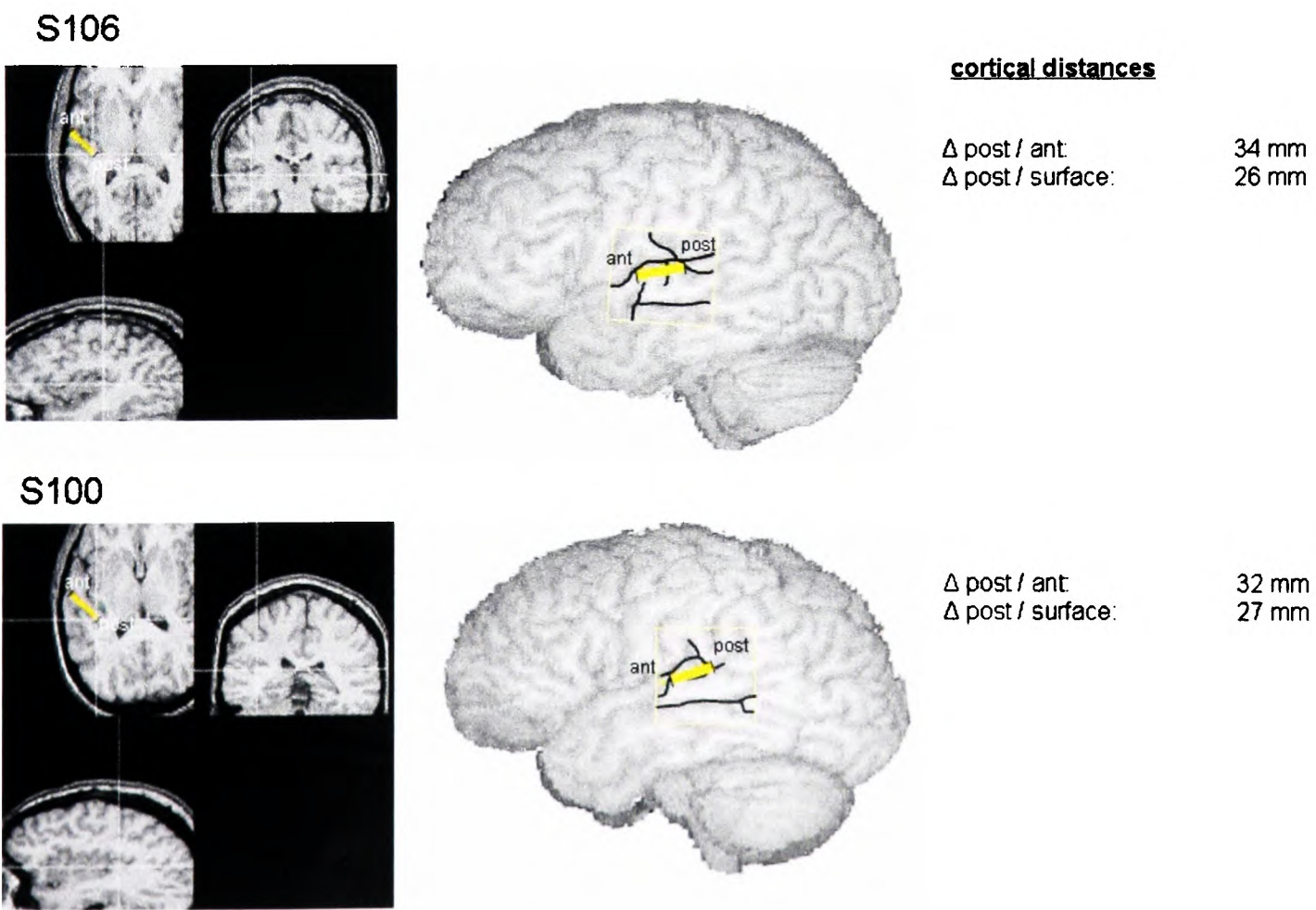


Figure 5-2: Location of Heschl's gyrus

Figure shows the location of Heschl's gyrus (HG) in relation to the surface electrode grid in both subjects. The yellow line in the sagittal MRI slices of the left column runs the length of HG (see text for convention used in anatomical definition). Black bars mark the posterior/ medial (post) and anterior/lateral (ant) boundaries of HG. B: The markers of HG were projected on the lateral surface of the rendered MRIs of the individual subjects, showing also the boundary of the electrode grid (yellow square) and the position of schematic drawings of anatomical landmarks (black lines) that are used in the visualization of all activation maps (e.g. Figure 5-4). Cortical distances are given for the length of HG (delta post/ant) and the distance between the posterior-medial end of HG and the nearest surface electrode (delta post/surface).

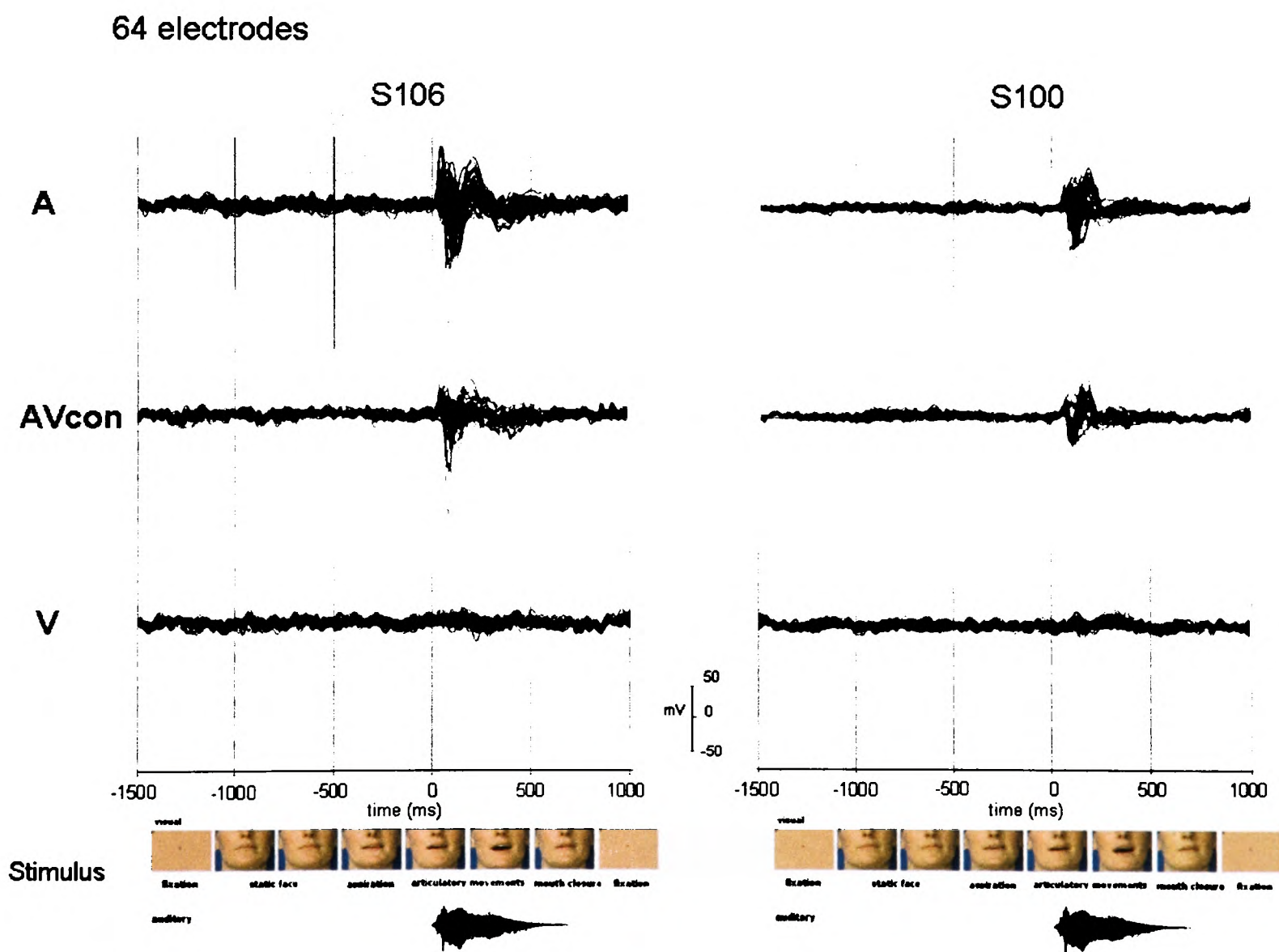


Figure 5-3: Averaged iERPs

ERPs from all 64 channels of the electrode grid overlaid onto each other for three conditions. A= Auditory; AV= Auditory-Visual; V= Visual. Bottom:

Timeline of visual and auditory stimuli in respect to the EEG recordings. Note the asynchrony of visual onset (of both face and mouth movement) with the auditory signal.

5.3.2.1 *Responses to auditory speech*

Intracranial auditory evoked potentials in both subjects were characterized by a complex succession of multiple components spread over a number of electrodes. Figures 5-4 and 5-5 show the topography of the evoked auditory potential on the grid over time. The evoked response waveforms are shown in Figures 5-6 and 5-7.

5.3.2.1.1 *S106*

Figure 5-4 shows the evoked potential topography over time for S106 to auditory only stimuli (A). The earliest response to auditory speech was a positive deflection starting at ~ 18 ms and peaking at ~45 ms at electrode sites on the middle section of the PLST (posterior lateral superior temporal area) (electrodes 20 & 28). This was followed by a polarity reversal in the same cortical region at ~ 75-90 ms with the centre of activity located slightly more anterior (electrodes 21 & 20). These electrodes also showed the highest amplitude response. After ~ 90 ms activity started to spread into more posterior regions while the number of simultaneously active sources increased. Another major focus of activity was located dorsally of the sylvian fissure between the precentral gyrus and inferior frontal gyrus (electrodes 32, 40 and 48) which peaked first at ~80 ms and then again at ~120 and ~ 220 ms.

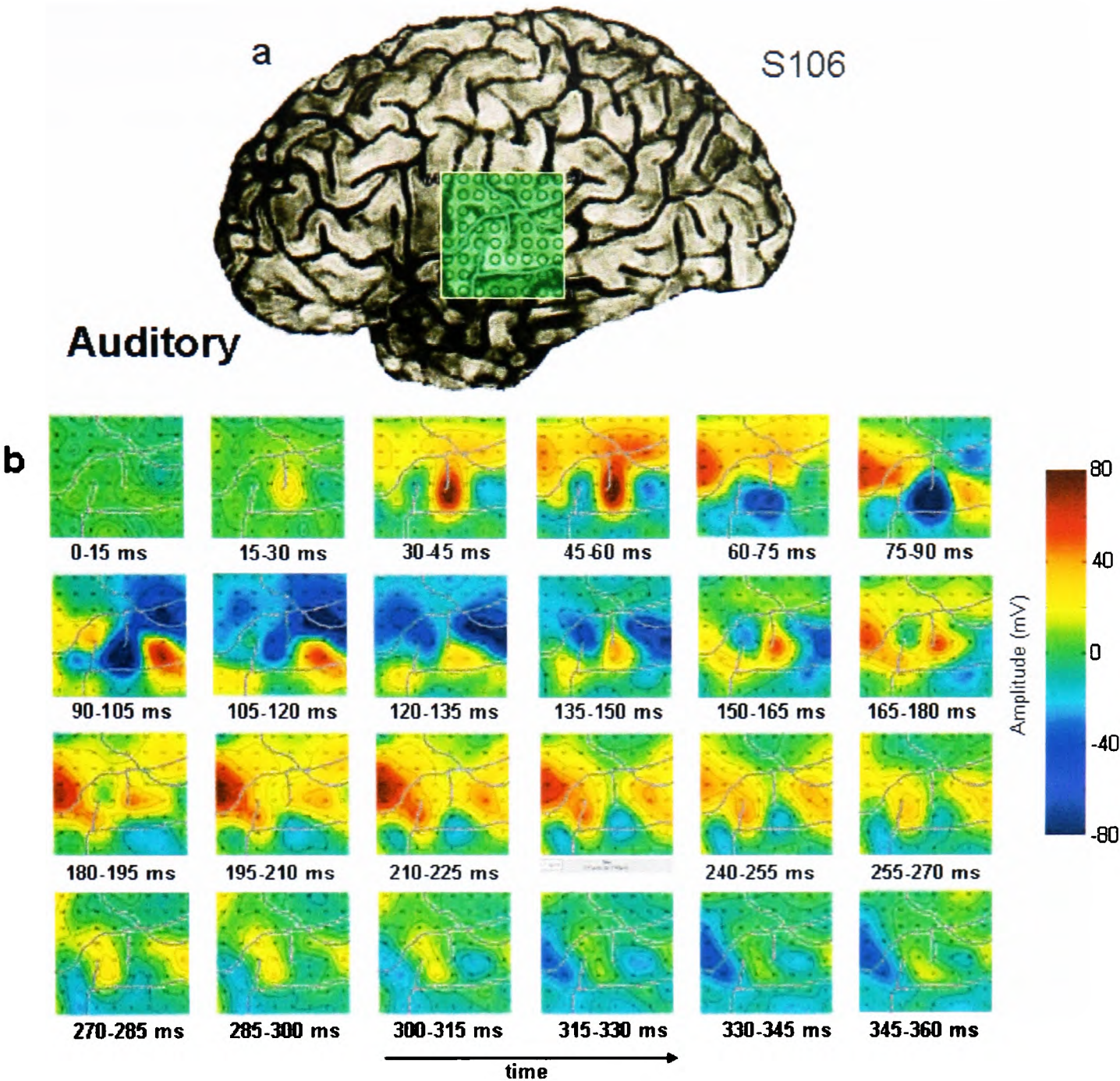


Figure 5-4: Activation to auditory speech in S106

A: Position of the 64 contact electrode grid on the surface of the left hemisphere of S106 (green square). Dark circles correspond to the location of electrodes 1-64. Electrodes are numbered sequentially in vertical order from bottom to top: electrode 1 is in the lower right corner, electrode 8 is in the lower left corner, electrode 9 is on the right-most side of the 2nd vertical electrode row from the bottom, electrode 57 is in the top right corner and electrode 64 in the top left corner.

See Figure 1 in Appendix for a high-resolution example of the numbering of electrodes on the grid.

B: Surface rendering over time of the averaged evoked response to auditory only stimulation. Plotted are averages of a 15 ms time interval in 15 ms steps from zero to 360 ms after auditory onset. Color-coding indicates the amplitude level of the evoked response in microvolt (mv). Red corresponds to a positive

potential, blue to a negative potential (see colour bar on the right). Grey lines on the grid correspond to anatomical landmarks derived from A. to help locating activation to cortical structures on the rendered MRI (also see Figure 5-2).

5.3.2.1.2 *S100*

Figure 5-5 shows the evoked potential topography over time for S100 to auditory only stimuli. Responses in this subject appeared later than in S100. The earliest response was a positive deflection starting at ~ 35 ms and peaking at 55 ms followed by a negative peak with a maximum at ~ 85 ms and a subsequent positive peak at ~145 ms at electrode sites on the middle section of the PLST (electrodes 30 & 31). At 65 ms another major source appeared, peaking at ~115 ms and located more inferior at the border between STG and MTG (electrodes 20 and 21). Further sources were located in the dorsal tip of the middle PLST (electrode 44) with peaks at ~120 ms and ~180 ms.

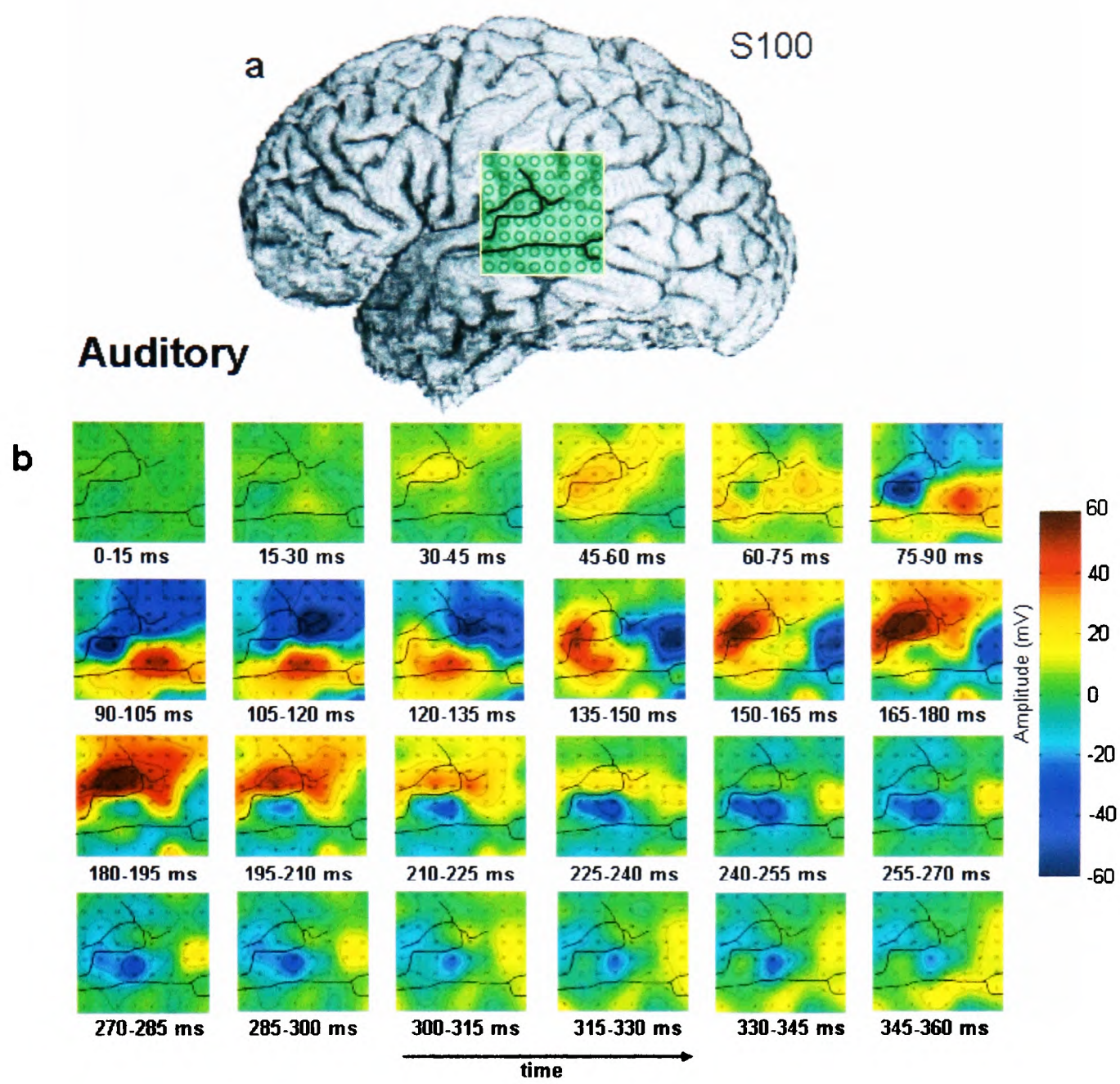


Figure 5-5: Activation to auditory speech in S100
(for a detailed figure legend see Figure 5-4)

5.3.2.2 *Interactions between auditory and visual speech*

Cluster-randomization analysis was performed to show the statistical difference between the responses in the auditory and auditory-visual stimulus conditions. Several significant clusters were found at different grid locations and temporal windows.

5.3.2.2.1 *S106*

Figure 5-6 shows the results of the AV vs A+V contrast for S106. An early significant effect was observed at 26 ms at electrodes on the middle section of the PLST (electrodes 20 & 28), the area directly above HG where the earliest and most prominent ERPs were observed during in the auditory only condition, and above area Tpt. A second cluster showed the same early temporal pattern of significance at a more posterior location on the PLST bordering on the supramarginal gyrus (BA 40). Directly on the posterior STG a significant modulation was evident starting at 85 ms (electrodes 18 and 19).

A significant ‘wave’ of a large spatial extent spread from the most posterior part of the STG covered by the electrode grid in an anterior direction along the inferior border of the sylvian fissure starting at 83 ms (electrode 41) and ending at ~164 ms (electrode 56). At 86 ms a cluster at the most inferior part of the STG showed a significant difference between A and AV from 86 ms to 119 ms after auditory onset. Finally, after ~195 ms significant clusters appeared dorsally of the sylvian fissure between the precentral gyrus and inferior frontal gyrus (electrodes 40 and 48) and on the anterior-middle part of STG bordering inferiorly on the MTG lasting until ~250 ms. As can be seen from the averaged time-series of the individual electrodes, the effect was mainly a result of a decrease in amplitude in the bimodal condition.

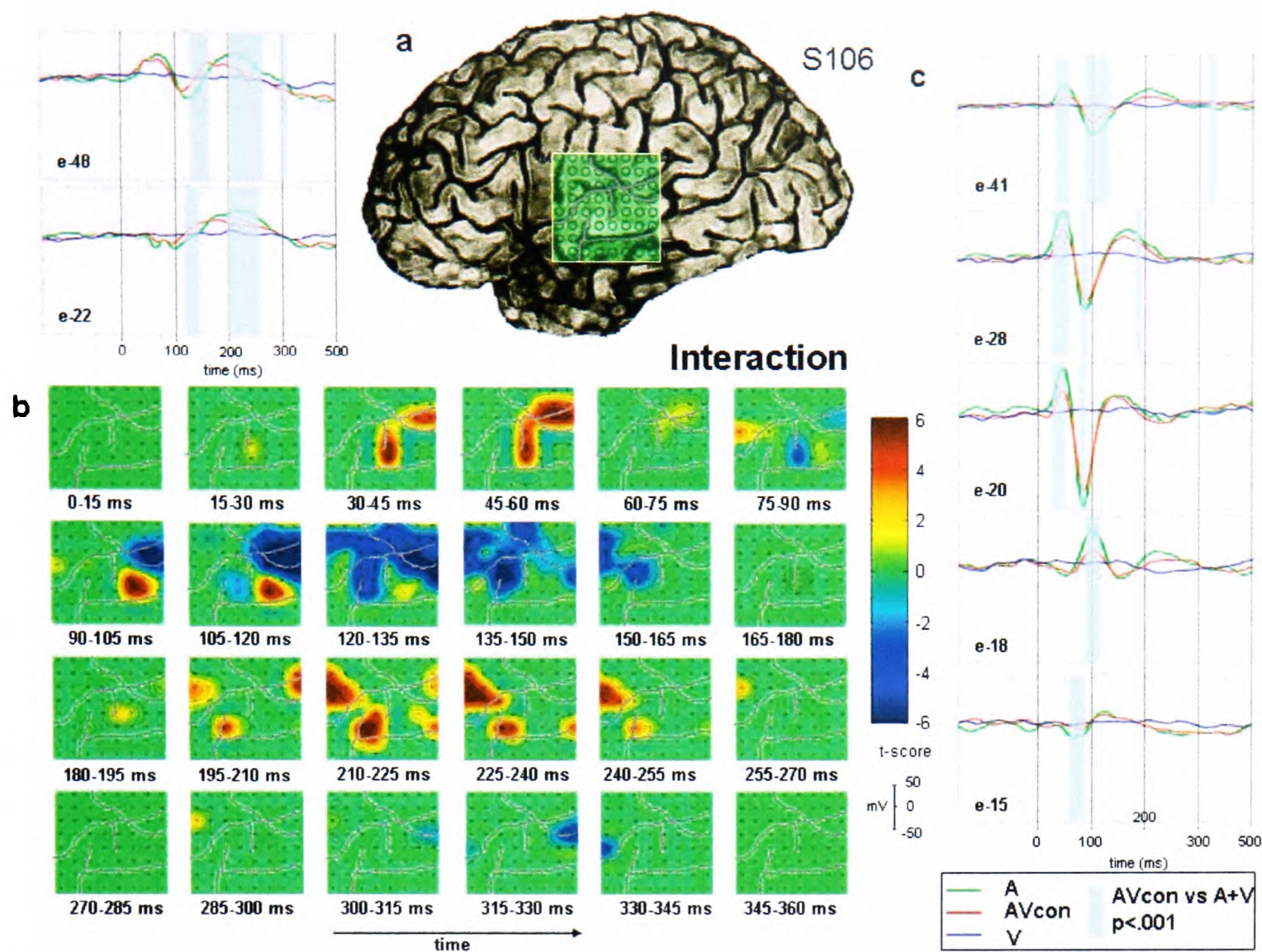


Figure 5-6: Statistical comparisons of A vs AV in S106

A: Position of the 64 contact electrode grid on the surface of the left hemisphere of S106 (green square).

B: Results of cluster statistics for the statistical contrast (p<.001) rendered on the grid surface across time. Color-coding indicates the value of the cluster t-statistics (see colour bar).

C: Plotted are averaged evoked responses to auditory (A), visual (V) and auditory-visual (AV) stimulation plotted for selected electrodes (e.g. "ch 15"). Blue shading indicates time window of significant difference for AVcon vs (A+V) (p<.001), corresponding directly to the surface renderings of B.

5.3.2.2.2 S100

The results of the interaction effect for S100 are shown in Figure 5-7. The major effect was largely restricted to one area at the posterior end of the sylvian fissure ventral to the temporal-parietal junction, area Tpt. The earliest significant effect ranged from 58 ms to 93 ms after auditory onset (electrodes 44 and 36), which

then continued to spread caudally along the STG (electrodes 34 and 25) until 106 ms. The same effect continues from 108 ms to 170 ms by means of a polarity reversal over the same cortical area. Similar to S106 the effect was largely a result of a decrease in amplitude during the bimodal condition.

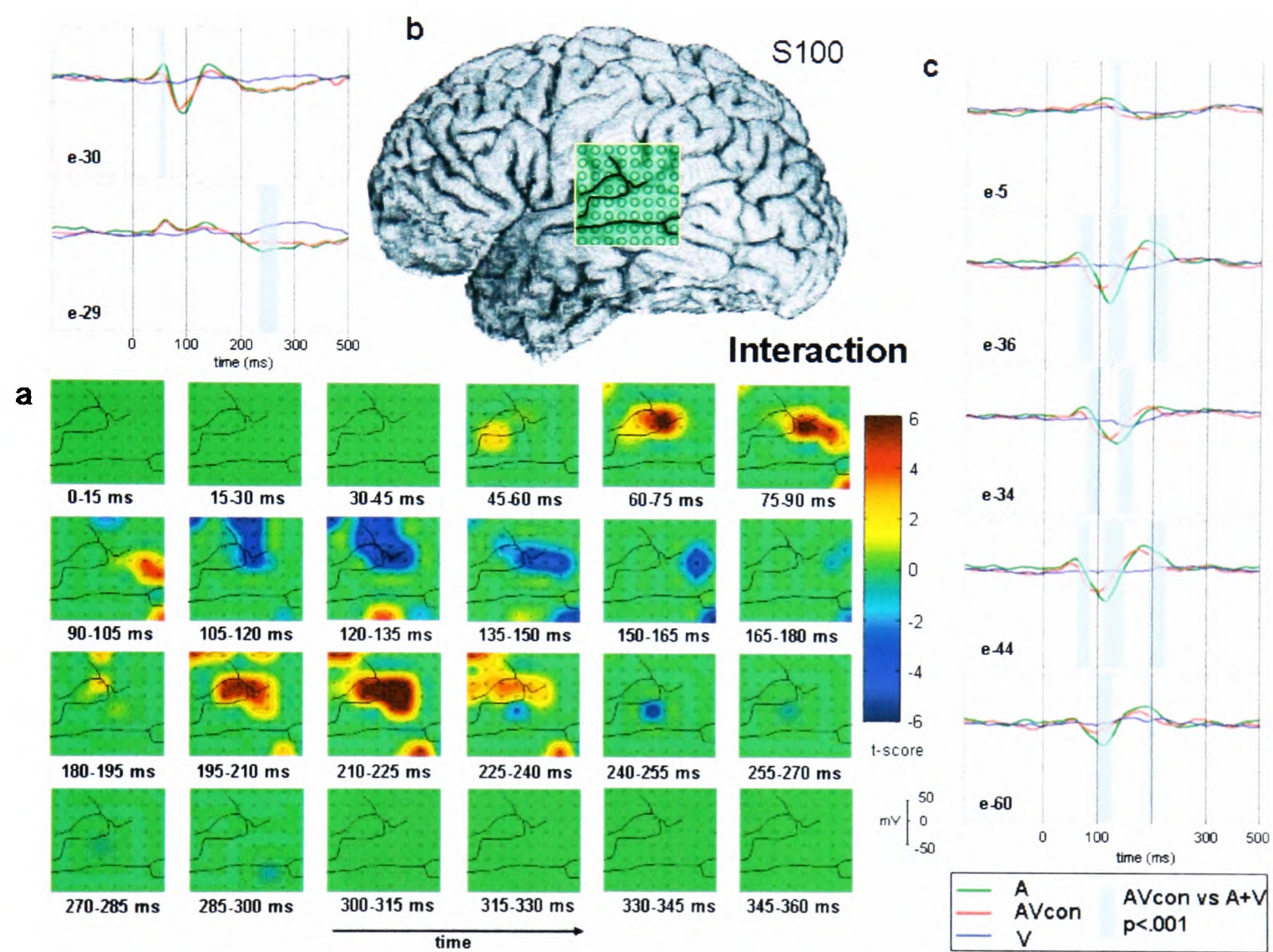


Figure 5-7: Statistical comparisons of A vs AVcon in S106 (for a detailed figure legend see Figure 5-6).

5.3.3 Phonetic interactions

5.3.3.1 AVcon vs AVincon(V)

5.3.3.1.1 S106

The contrast AVcon vs AVincon(V) showed a significant difference at 72 ms on electrodes of the lateral posterior superior temporal lobe (electrodes 29 and 30). These sites were situated immediately above anterior HG. A cluster at a more posterior location showed a similar latency response profile (electrodes 33 and 34). Interestingly, a significant cluster appeared also at supra-sylvian sites in the inferior postcentral gyrus (electrodes 40, 48 and 56), ranging from 174 ms to 256 ms. Interestingly, a significant difference cluster was found at electrodes over the central sulcus directly between the borders of the pre- and postcentral gyri (electrodes 48 and 56).

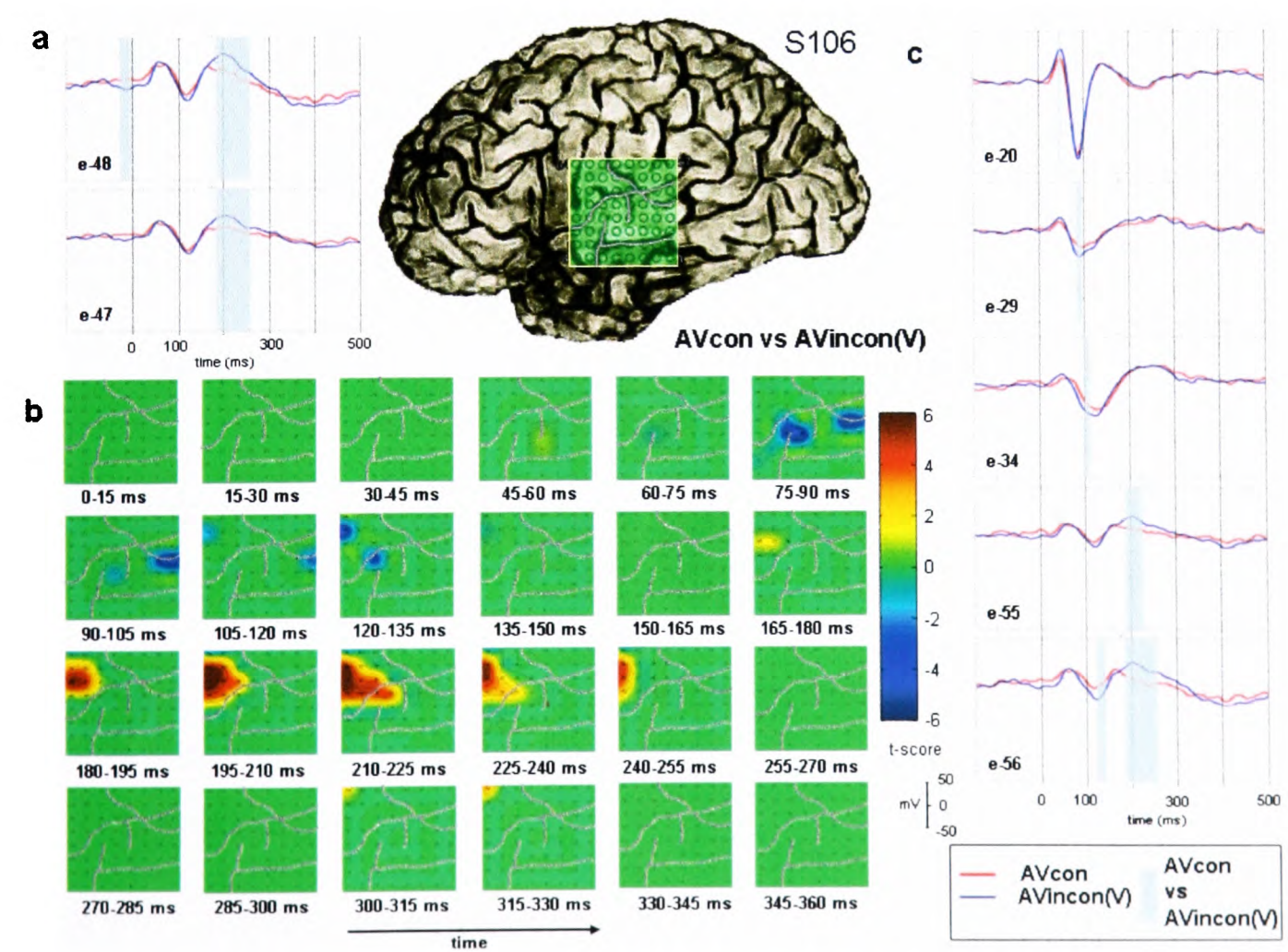


Figure 5-8: Statistical comparisons of AVcon vs AVincon(V) in S106

(for a detailed figure legend see Figure 5-7).

5.3.3.1.2 *S100*

In S100, the earliest difference between AV con and AVincon(V) appeared at 100 ms, stayed on until 125 ms and was located to sites at the posterior end of the sylvian fissure ventral to the temporal-parietal junction (electrodes 35 and 36). A large cluster at supra-sylvian sites on the inferior post- and precentral gyri (electrodes 46, 48, 54, 62) activated from 200 to 270 ms. Finally, a significant difference was detected at 283 ms and 305 ms at electrode sites on the middle STG (electrodes 30 and 31) and the dorsal MTG (electrodes 7 and 15), respectively.

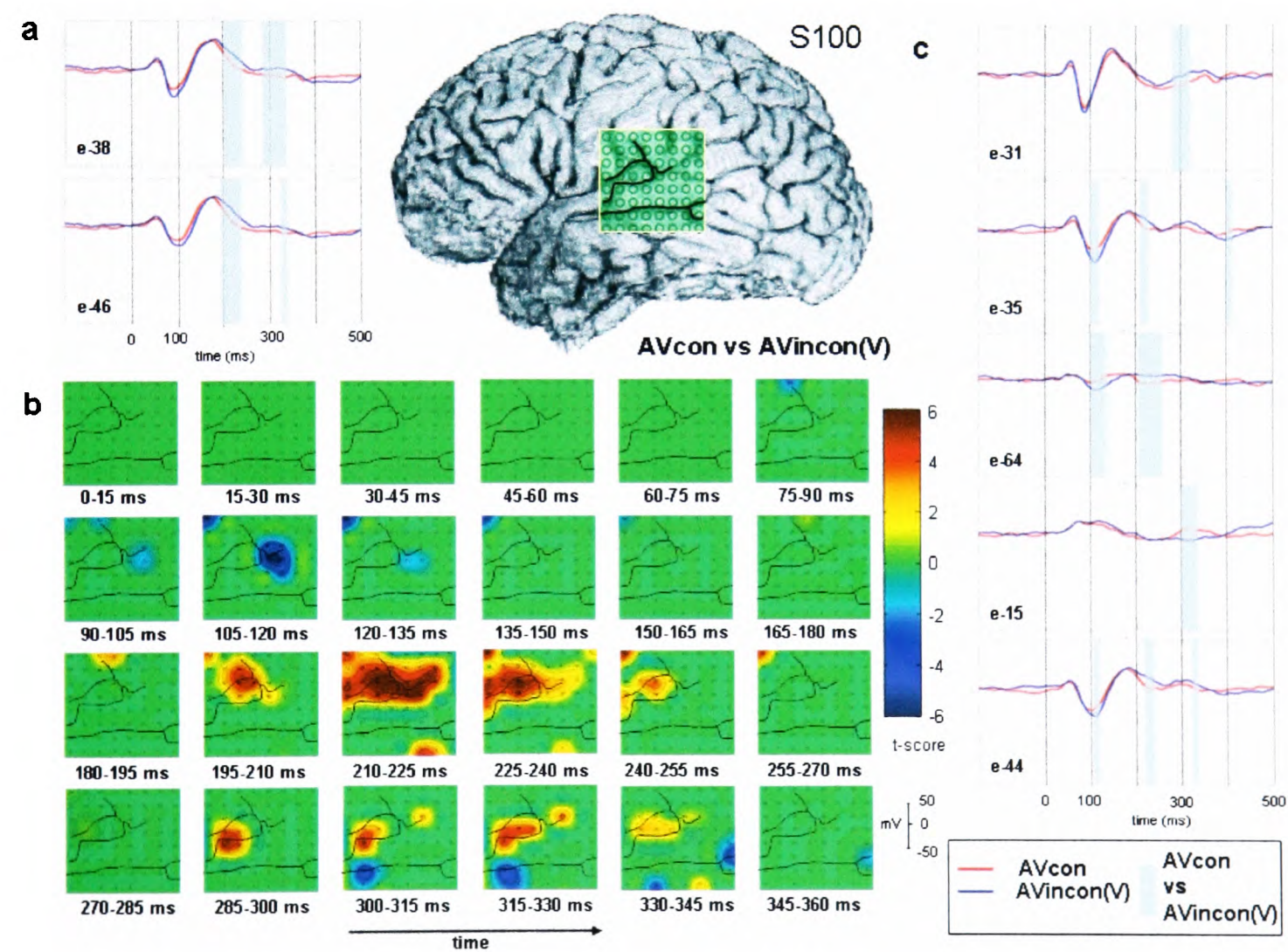


Figure 5-9: Statistical comparisons of AVcon vs AVincon(V) in S100
(for a detailed figure legend see Figure 5-8)

[added sulci]

5.3.3.2 *AVcon vs AVincon(A)*

5.3.3.2.1 *S106*

Figure 5-10 shows the results of contrast AVcon vs AVincon(A). A significant difference at 77 ms was detected on electrodes on anterior STG (electrodes 15 and 16). Several clusters at more posterior locations (electrodes 28 and 29, 26 and 27, 36 and 44, 42 and 41) activated between 91 ms and 360 ms.

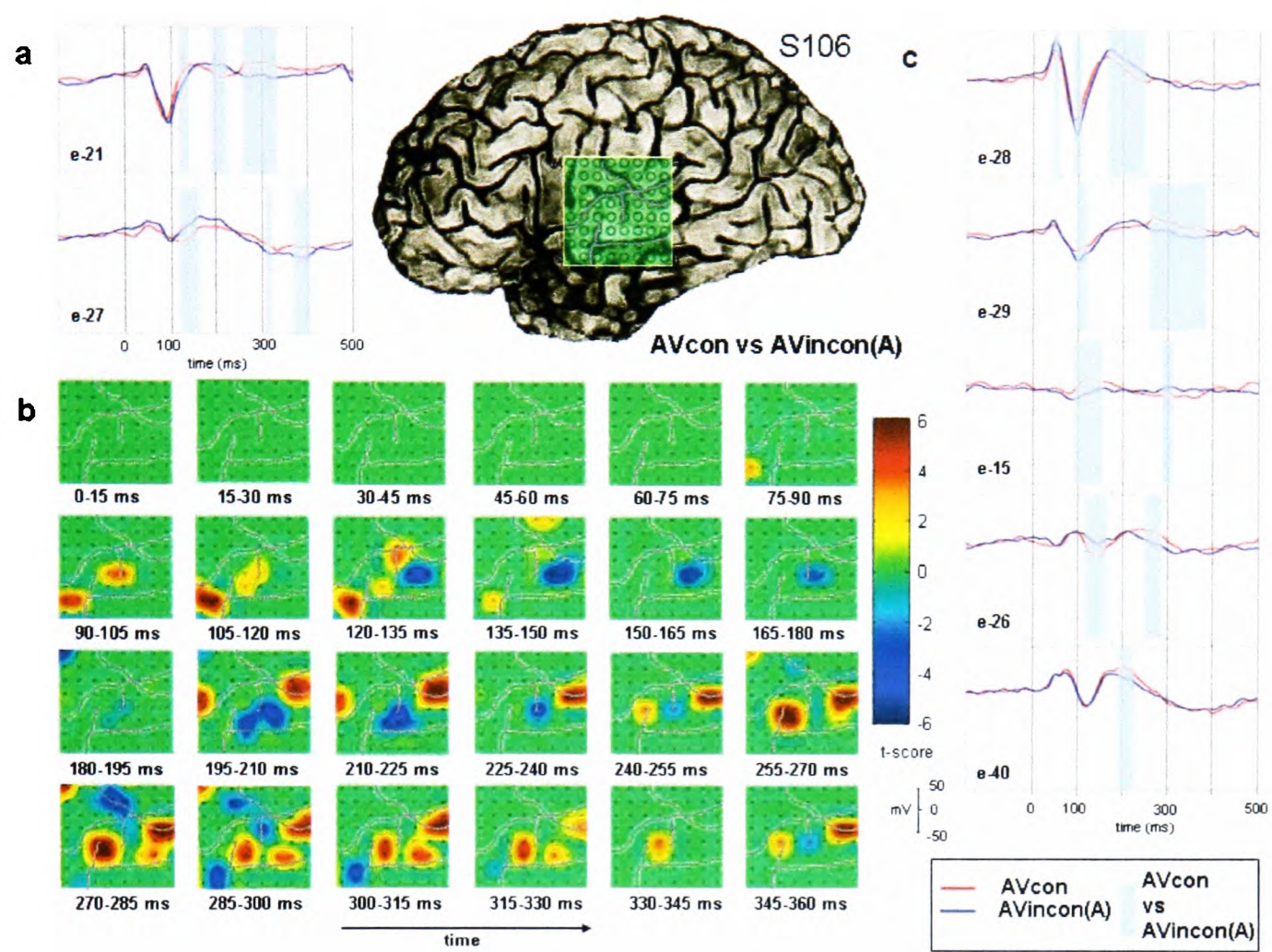


Figure 5-10: Statistical comparisons of AVcon vs AVincon(A) in S106 (for a detailed figure legend see Figure 5-9)

[added sulci]

5.3.3.2.2 S100

In S100, the earliest difference between AV con and AVincon(A) started at 42 ms, stayed on until 64 ms and was located to sites at the superior part of the middle STG, (electrodes 38 and 39) (Figure 5-11). At 85 ms a significant cluster appeared on the posterior part of the STG (electrodes 34 and 34) and remained significant until 133 ms. Finally, a large cluster starting from supra-sylvian sites on the inferior post-

and precentral gyri (electrodes 64, 48, 54, 62) at around 195 ms spread to areas of the middle STG where it remained active beyond 360 ms after auditory onset (electrodes 29, 30, 35 and 46).

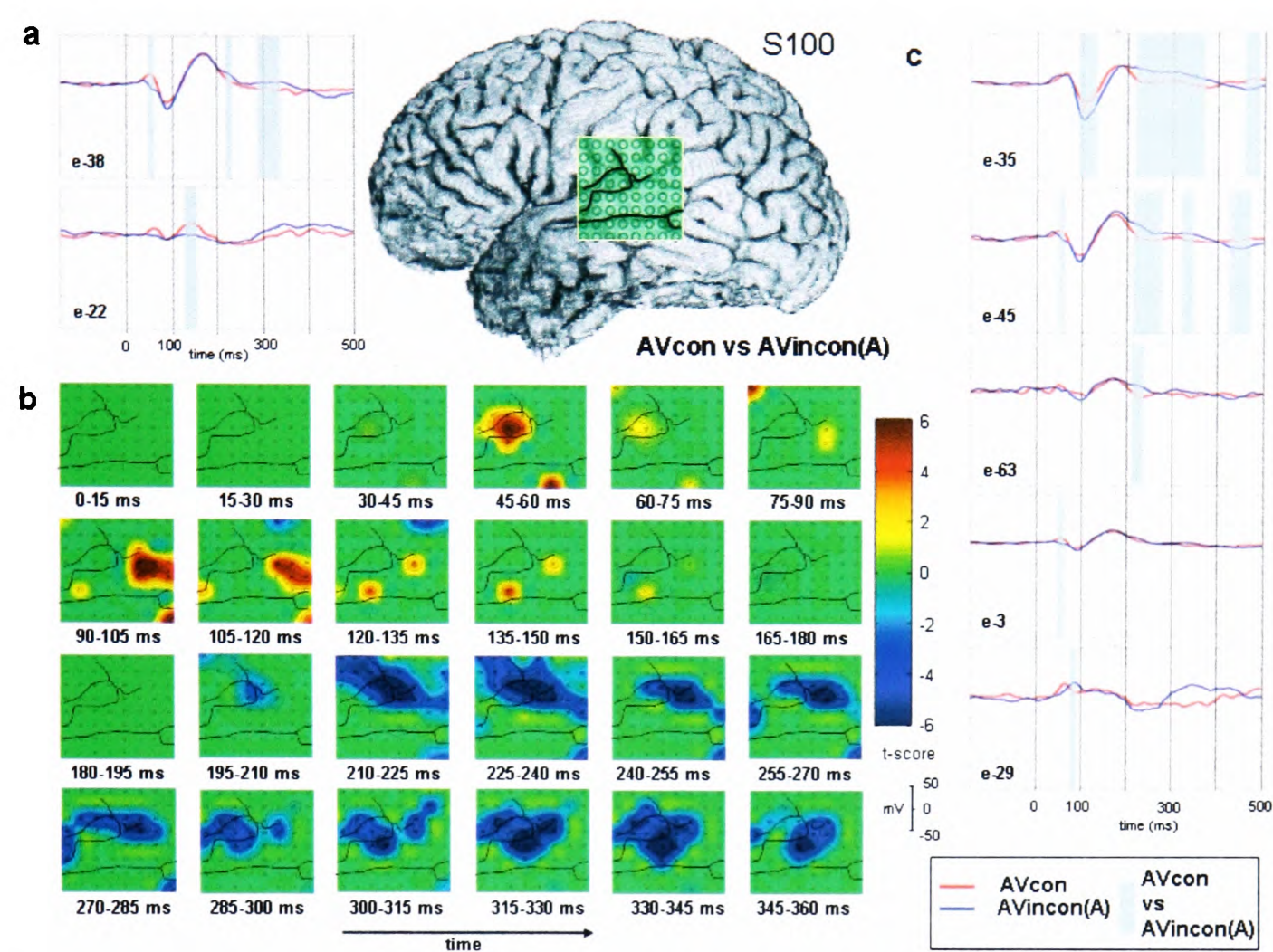


Figure 5-11: Statistical comparisons of AVcon vs AVincon(A) in S100 (for a detailed figure legend see Figure 5-10)

5.4 Discussion

Auditory activity arrived at the cortex by ~18 ms after stimulus onset at electrodes closest to the postero-medial HG, suggesting that the response originated in the primary auditory cortex, and that these electrodes indeed reflect HG activity. This is consistent with findings of an early auditory evoked magnetic field at 20 ms (P20m), which has been located to primary auditory cortex (Lutkenhoner et al., 2003). Activity then spread to more anterior regions of HG, starting at ~35 ms. The recruitment of lateral/anterior regions of HG at this latency with simultaneous activity

in primary auditory cortex is consistent with the magnetic P30m component, for which evidence exists that it consists of at least two distinct cortical areas, one of which is A1 and the other is presumed to be A2 (Lutkenhoner et al., 2003). While MEG sensors cannot differentiate between such close sources unambiguously, the results from the electrode grid confirm the differentiation of P30m into responses from two distinct cortical areas, A1 and A2. It should be noted, however, that the activity recorded from sensors immediately above HG does not necessarily have to be a result of volume conduction from auditory cortex, but could also reflect activity of sources directly located on the lateral cortical sheet immediately underneath the grid (Howard et al., 2000). However, the short latency of response onset (~18 ms) on electrodes closest to HG suggests that they are sensitive to primary auditory field responses, and that at least the initial response at these electrodes is likely to represent activity from HG sources. Next, a posterior part of STG became active at a latency of ~60 ms, slightly later than the range of 40-60 ms observed by Yvert et al. (2001) who used short tone bursts as stimuli.

Multisensory interactions of auditory-visual processing started as early as 26 ms at electrode sites immediately above HG, suggesting that the auditory and visual signals integrate already at the first cortical stages of auditory processing, i.e. in auditory cortex. At later latencies (~85 ms) posterior region of the STG also became involved. This sequence of integrative responses confirms findings of MEG studies of AV speech (Mottonen et al., 2004). Over the remaining time course the superior temporal lobe showed a complex pattern of interactions at multiple locations and different latencies.

Even though topographic EEG studies suggest that visual information has spread to all cortical lobes by ~80 ms, the first cortical response arrives in visual cortex only at ~56 ms (Foxe and Simpson, 2002). This shows that the visual modulation we observe in auditory cortex cannot be mediated by a visual stimulus

whose onset coincides with the start of the auditory signal. Therefore, these data provide further evidence that visual speech modulates auditory cortex before auditory onset.

Visual modulation of auditory speech processing was also seen in suprasylvian areas, namely the inferior tip of the postcentral sulcus (i.e. of the somatosensory cortex or SI and the caudal edge of the ventral precentral gyrus (i.e. the primary motor cortex or MI). The modulated areas were in regions involved in the movement and sensation of speech related muscle groups, such as those influencing the tongue, jaw and lips (Penfield and Rasmussen, 1950), suggesting an involvement of the mirror neuron system in the integration of auditory-visual speech. Motor regions, for example, are consistently activated during the perception of actions, such as hand or mouth movements (Fadiga et al., 1995; Fadiga et al., 2000). In monkeys, several studies have identified neurons in ventral premotor cortex that respond both during the production of an action and the visual observation of that action (Gallese et al., 1996; Rizzolatti et al., 1996). Similar visual “mirror neuronal” responses have also been shown with fMRI in humans (Rizzolatti et al., 1999; Buccino et al., 2001). More recently, evidence in monkeys suggests that the mirror neuron system is also involved in the perception of sounds relating to a particular action. Kohler et al. (2002) and Keysers et al. (2003) found auditory-visual mirror neurons in premotor cortex that responded to the performance, visual observation and auditory perception of an action. Evidence for an involvement of the motor system during both auditory and visual speech perception was reported by Watkins et al. (2003), who used TMS to stimulate the lip area of the motor cortex. Concurrent presentation of auditory and visual speech stimuli increased the motor-evoked potential recorded from the lips compared to adequate control conditions. This effect was only observed after stimulation of the left hemisphere and suggests that the left MI is part of a larger

network of cortical areas involved in the processing and, possibly, integration of auditory-visual speech.

The human somatosensory cortex (S1) has also been implicated in the mirror neuron system in a similar fashion (Avikainen et al., 2002; Rossi et al., 2002). Moettoenen et al. (2005) stimulated the lips and recorded the associated neuromagnetic fields on the level of the cortex while subjects were engaged in auditory or visual perception of speech stimuli. Concurrent speechreading specifically increased the source amplitude of a dipole in the left somatosensory cortex, implicating this area as part of the mirror neuron system for speech processing. However, no modulation by auditory speech was observed in left somatosensory mouth cortex, raising doubts as to whether S1 is also involved in multisensory speech processing. In a recent fMRI study, Tettamanti et al. (2005) showed that listening to action-related sentences activated a network of areas including ventral premotor cortex, Broca's area, and, amongst others, also the posterior STG, suggesting the mirror neuron system might be part of a general speech circuit involving multiple distributed areas.

In the present study, further evidence for a flow of information between areas involved in the perception and execution of speech is given by the modulation of activity in left ventral motor and somatosensory cortex by incongruent AV pairings. This effect was most evident in the AVincon(V) condition, where intelligible articulatory mouth movement was replaced by non-meaningful mouth movement (gurning). Considered in the light of previous findings on action perception and execution, our results suggest that the mirror neuron system in SI and motor cortex (MI), with its observation-execution matching function, is used as a computational unit for comparing auditory and visual speech signals based on their phonetic features.

6 Neuromagnetic frequency responses

The previous section demonstrated that auditory and visual speech processing interacts very early in the auditory stream at the level of auditory cortex, and then subsequently in STS and IFG of the left hemisphere. Such early interactions suggest that the visual stimulus modulates auditory cortex before the arrival of the auditory speech signal. In this section, we are analyzing the frequency spectrum of the MEG data in order to describe the oscillatory behaviour of neuronal networks associated with multisensory speech processing. This technique does not rely on phase-locking of the neuronal response and allows for the detection of crossmodal modulation of the primary sensory areas before and after auditory stimulus onset.

Swiftly the brain becomes an enchanted loom, where millions of flashing shuttles weave a dissolving pattern-always a meaningful pattern-though never an abiding one.

-Sir Charles Sherrington

6.1 Introduction

In the previous section we have applied ERP and dipole modelling techniques to MEG data and found visual effects on auditory stimulus processing in auditory areas of the superior temporal lobe. ERPs are constructed by averaging over many individual trials so that the signal related to neural processing of the stimulus can be delineated from the background noise. This approach assumes that individual trial data are composed of time-invariant phase-locked responses (i.e. evoked responses) and background noise. However, the individual trials also contain stimulus induced rhythmic EEG activity which is time-locked but not phase-locked to the stimulus, and therefore remains undetected by the ERP technique. This happens because time-domain averaging, such as linear averaging to generate ERPs, removes non-phase locked EEG activity through phase cancellation (Makeig, 1993; Pantev, 1995). This

can lead to an incomplete picture of the neuronal dynamics associated with the respective sensory and cognitive processing. Therefore, in this section we employ methods to analyze the frequency content of the MEG signal.

Oscillatory brain responses have been shown to be involved in information processing and can be modulated by external and internal events. Since Berger's discovery that visual stimulation blocks alpha band activity (8-12 Hz) over occipital areas (Berger, 1930) various investigators have reported correlations between EEG frequency responses and various cognitive processes such as attention, memory, and language (e.g. (Krause et al., 1995; Krause et al., 1996; Basar et al., 1999).

For example, neural activity in the hippocampus seems to be reflected as oscillations within the theta frequency band (4-8 Hz) (Burgess and Gruzelier, 1997); (Niedermeyer, 2005) while the alpha rhythm is generated mainly by corticocortical and thalamocortical neural networks (Cotillon et al., 2000; Klimesch et al., 2005). During sensory and cognitive stimulation, the responses in different EEG/MEG frequency bands differ from each other (Klimesch et al., 1990; Pfurtscheller and Lopes da Silva, 1999; Klimesch et al., 2000), and reflect different cognitive and/or mental processes or states (Pfurtscheller and Klimesch, 1992; Klimesch et al., 1994; Klimesch et al., 1998; Klimesch et al., 2001) e.g., working*memory processes seem to be reflected as oscillations within the EEG theta frequencies. 8-10 Hz alpha activity seems to be modulated as a function of attentional demands (Klimesch et al., 1998) and oscillations around 10-12 Hz are modulated mainly by stimulus-related aspects and/or semantic memory processes (Klimesch et al., 1993; Klimesch et al., 1994). These stimulus-related changes in the power spectrum are often termed event-related desynchronization (ERD) if the power decreases and event-related synchronization (ERS) if power increases compared to a passive baseline (for a review, see (Pfurtscheller and Lopes da Silva, 2005).

Relatively few studies have described event-related oscillations in the auditory modality and most of these relate to auditory memory processing (see work by (Krause et al., 1996); (Krause et al., 1994) found a significant increase in alpha band power to linguistic auditory stimulation compared to baseline. Some investigators have reported phase-locked early gamma responses to auditory stimuli using scalp EEG and MEG (Pantev et al., 1991; Ribary et al., 1991; Basar et al., 1999). This transient gamma response occurs early (from 25 ms after stimulus onset, lasting about 100 ms) and has a spectral peak between 30 and 40 Hz. Responses to simple auditory stimuli (sine wave sounds) show a large increase in power in the alpha and beta bands peaking at around 100 ms and have been related to the N100 evoked response of the auditory ERP (Makinen et al., 2005). With a few exceptions, speech-related oscillatory brain responses to auditory stimuli have not been characterized much.

Herdman et al. (2004) used synthetic aperture Magnetometry (SAM) to analyze oscillatory activity in response to incongruent and congruent AV stimuli. The authors presented subjects with matching and mismatching pairs of Japanese graphemes of vowels and the corresponding voiced representations. The authors found a significant ERS in HG and the planum temporale in the 8-16 Hz frequency band, together with ERD of the visual cortex. These findings suggest that neuronal networks in the two primary sensory cortices use different oscillatory strategies during sensory processing of AV stimuli.

ERPs are most sensitive to strong changes in stimulation, such as visual stimulus onsets, and decline rapidly if stimulation remains stable (Regan, 1989). Therefore, the averaged evoked response to visual stimulation long after visual stimulus onset (~600 to 1100 ms) does not produce an ERP waveform. However, the MEG/EEG signal in that time range may still contain specific information related to neuronal processing that is not detectable by the ERP averaging technique. An advantage of ERD/ERS frequency analysis is that it does not rely on phase-locking of

the event-related response (Varela et al., 2001). That means that the signal during time ranges which show no clear evoked response i.e. immediately before auditory onset in temporal or frontal areas, can be analyzed regarding its information content related to stimulus processing. This information is important in order to determine if there are cortical areas that may visually ‘prime’ the auditory system before the arrival of an auditory stimulus. In the present section, therefore, we will try to characterize the oscillatory cortical response to auditory-visual speech processing across a range of frequencies, and to detect early influences of the visual signal on the neuronal network involved in speech processing.

6.2 Methods

6.2.1 Subjects

Data were recorded from 9 (female, mean age = 25 years, stdev = 2.3 years) native English speakers all of which also participated in the fMRI experiment. The data used for this analysis is identical to the data used for dipole analysis in section 4. All subjects were right-handed, without previous history of psychological or neurological disorders and had normal or corrected to normal eyesight. The study received ethical approval from the Central Oxford Research Ethics Committee and informed consent was obtained from all subjects. Exactly the same stimulation and recording parameters as in section 4 (MEG dipole analysis) was used here. Please refer to the Methods of this section for a more detailed account.

6.2.2 SAM analysis

The MEG data were analyzed using SAM (Synthetic Aperture Magnetometry) which is a source localization method for identifying active sources represented in the MEG measurements using an adaptive beam-former technique for spatial filtering (Van Veen and Buckley, 1988). Beamformers have originally been developed in fields of signal array processing, such as radar, sonar, and seismic explorations and

have been recently applied to the study of source localization of brain dynamics (Van Veen et al., 1997; Robinson and Vrba, 1999; Singh et al., 2002; Singh et al., 2003). Using a minimum-variance beamformer, each cortical area of interest (“virtual electrode” (VE)) was linked to the detection array via an optimal spatial filter using the weighted sum of all the MEG sensors. The data were then passed through the spatial filter to return a measure of the current density over time in the target voxel.

Various methods have been employed to evaluate cortical oscillations over time. Simple methods include the application of various kinds of digital filters to the signal in either the time or frequency domain (Pfurtscheller and Andrew, 1999). Time-frequency approaches such as moving window short time Fourier transform (Haig et al., 1999) and wavelet transforms (Mallat, 1989; Tallon-Baudry et al., 1997) have also been employed. In the present analysis we used Morlet wavelets to compute 3D Spectrograms for each virtual electrode location. Wavelets were used because of their superiority to Fourier Transform in decomposing time-domain signals with non-stationary frequency content such as EEG data. A major advantage of wavelets is the variable time resolution involving shorter temporal windows at higher/faster frequencies and longer time windows for lower/slower frequencies. Wavelets can be parameterized such that they are well-localized in time and frequency and thus allow an optimal discrimination of shortly spaced events over time and between frequencies.

Average spectrograms were produced for virtual electrodes in bilateral A1 and STS, left Broca’s area and right primary visual cortex. (see Figure 6-1) within a frequency range of 1 - 80 Hz across a 3500 ms time window (-1500 to 2500 ms). To show ERD and ERS the spectrograms were then converted into percent power changes relative to a passive baseline (-1600 ms to -1400 ms).

6.2.3 Virtual electrode locations

Virtual electrode locations were determined based on Talairach coordinates derived from previous imaging studies. STS coordinates (Left: $x = -53$, $y = -48$, $z = 9$; Right: $x = 48$, $y = -55$, $z = 20$) were taken from (Calvert et al., 2000) who identified activation clusters in bilateral STS showing super-additive enhancement to congruent audio-visual inputs. Auditory cortex locations (Left: $x = -41$, $y = -28$, $z = 13$; Right: $x = 44$, $y = -22$, $z = 11$) were taken from an fMRI mapping study which used synthetic speech as stimuli and functional and anatomical criteria for mapping primary auditory cortex (Belin et al., 1999). Left Broca's area (Left: $x = -45$, $y = 21$, $z = 6$) was defined based on coordinates from an fMRI study involving a language task (Musso et al., 2003). The visual cortex location was based on right primary visual cortex coordinates from a retinotopic mapping study of visual field representations (Dougherty et al., 2003). Figure 6-1 shows VE location on a rendered standard brain.

6.3 Statistical analysis

For statistical comparison of conditions, a two-sided pair-wise t-test was used on the individual trials for every time bin and frequency bin at each electrode (-0.2 ms to 1000 ms, 1-80 Hz). A cluster-level randomization procedure was used to control for inflated Type-II errors using ClusterRandAnalysis, a MATLAB based function which is part of the FieldTrip analysis package (<http://www2.ru.nl/fcdonders/fieldtrip>). (see Methods, section 3)(Maris and Oostenveld, in press). The randomization null distribution was obtained by a Monte-Carlo randomization with 500 draws. A conservative confidence interval of $p < 0.05$ was selected for both the cluster-level threshold and randomization tests.

Virtual electrode locations

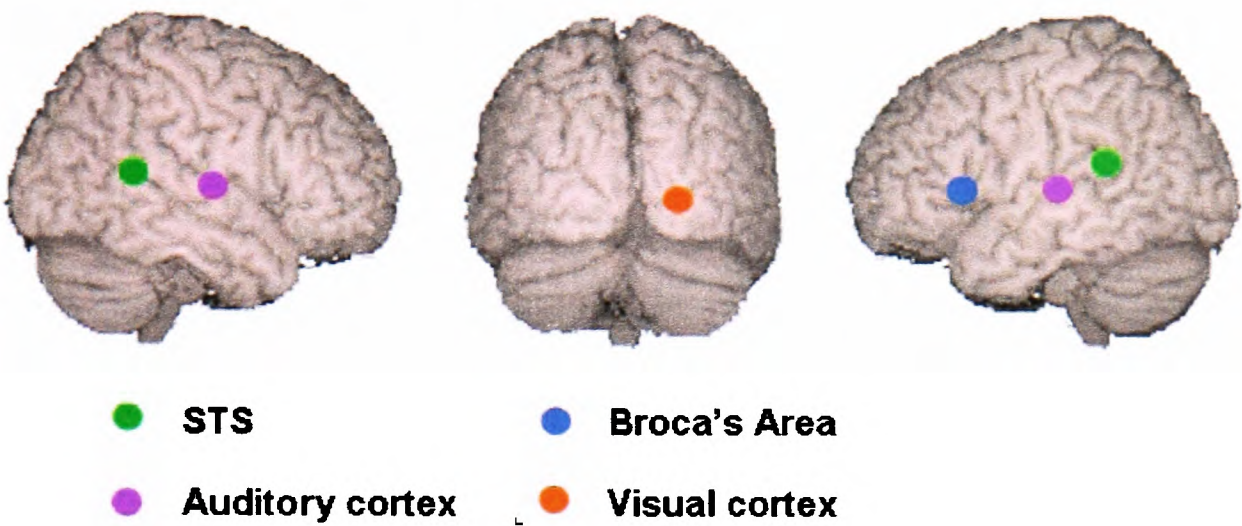


Figure 6-1: Locations of virtual SAM electrodes
Rendered on a standard brain in MNI space

6.4 Results

Stimulus-related power changes (ERD & ERS) were observed over a number of VE sites and frequency bands. The overall magnitude of ERD/ERS changes during sensory stimulation (compared to baseline) was highest in the alpha and beta frequency bands and then decreased with higher frequencies, an observation consistent with previous studies showing that natural phenomena often have spectra with a one-over-frequency ($1/f$) shape (Mandelbrot, 1998), including MEG and EEG data (Niedermeyer, 2005). Figure 6-2 show the time-frequency representation of the power grand-averaged across subjects in bilateral A1, STS, left Broca's area and right A1 after auditory stimulus onset for conditions A, AVcon, and V. The bottom row shows the results of the statistical comparison of AVcon vs A at $p < .05$ and corrected for multiple comparisons.

6.4.1 Cortical power changes in A1

Activity in response to auditory only stimulation in A1 was marked by a large ERS in the theta (4-7 Hz) and alpha (8-13 Hz) bands immediately after auditory onset lasting approximately 300 ms. A transient ERS burst was observed in the beta (14-30

Hz) and gamma (31-60 Hz) bands immediately after stimulus onset. These ERS were followed by an ERD in the alpha and beta frequency bands. A concurrent congruent visual stimulus caused a significant decrease in the alpha ERS responses and led to an ERD in the lower beta range. Visual only stimulation in form of articulatory mouth movements lead to an ERS in the alpha and beta range immediately after auditory onset, and to an ERS in the beta range immediately before auditory onset. Furthermore, statistical comparisons show an omega ERD (61-80 Hz) during the presence of a concurrent visual stimulus. The presence of a visual stimulus also markedly reduced the transient gamma burst (ERD). Figure 6-3 shows the time course of changes at 12 Hz and 33 Hz for conditions A, V and AVcon.

6.4.2 Cortical power changes in STS

Oscillatory activity in the STS follows a similar pattern as A1 and shows an overall reduction in ERD/ERS amplitude. A transient gamma ERS was also observed in this cortical area. AVcon vs A shows a significant difference in the theta and alpha bands, but fails to show a significant difference in the beta and omega frequencies. Figure 6-3 shows the time course of changes at 12 Hz for the different conditions.

6.4.3 Cortical power changes in Broca's area

Visual only stimulation caused a beta ERS before and after the supposed onset of the auditory stimulus in Broca's area. The post-auditory ERS is not as pronounced in the auditory only condition and absent before auditory onset in that condition. This is reflected in a significant difference in the lower beta frequencies for the AVcon vs A contrast. Right A1 and STS show a similar response pattern as the left hemisphere, but with an overall reduction in response amplitude in all conditions. The number of significant time-frequency clusters for the AVcon vs A contrast is markedly reduced, with the right STS showing only very few significant differences.

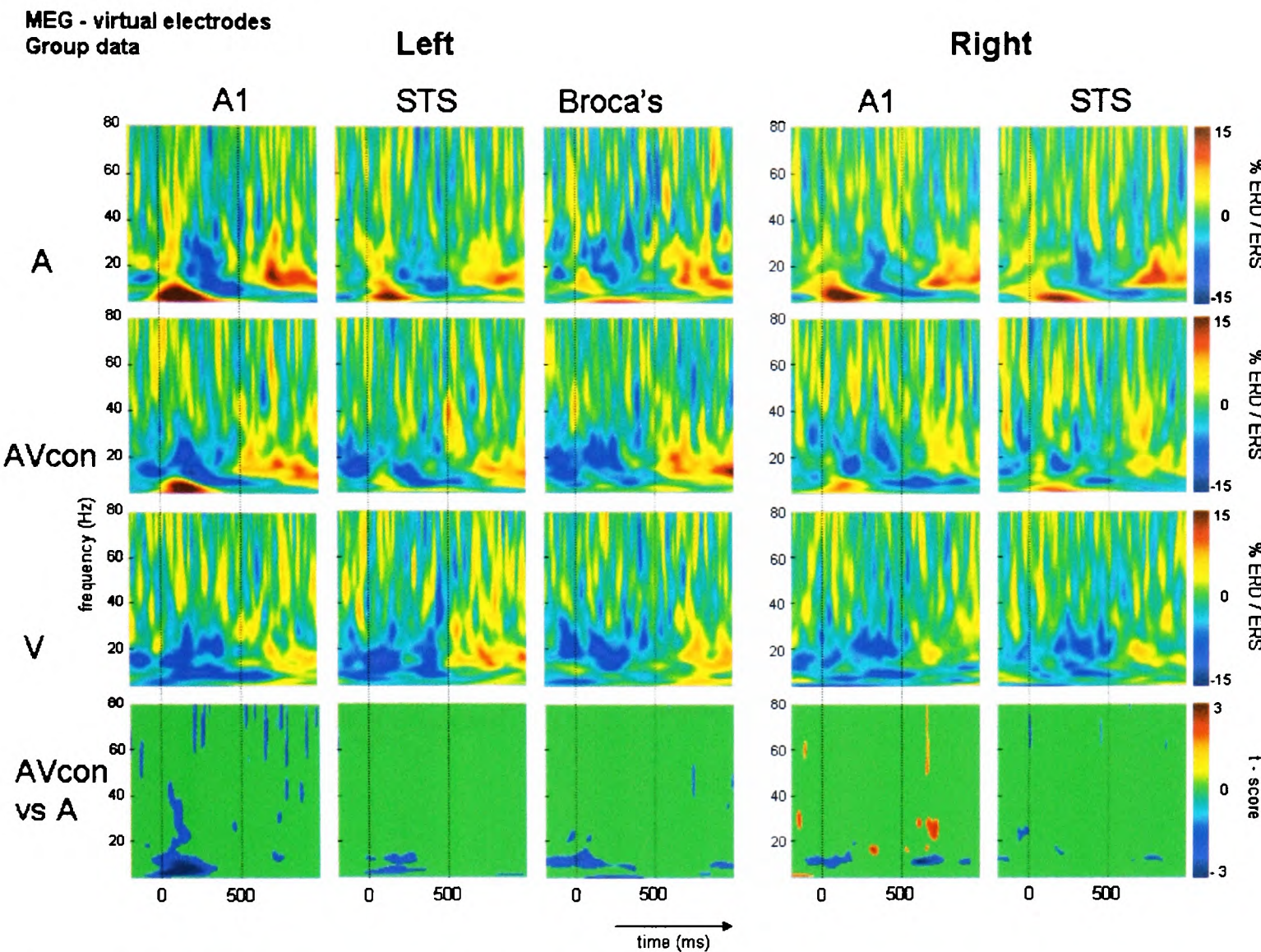


Figure 6-2: ERD/ERS group data for A, AVcon and V

At virtual electrode locations for the time period from – 150 to 1000 ms relative to auditory onset and for frequencies from 4 to 80 Hz. ERD/ERS percent values are color-coded (see scale). Bottom row shows the results of the statistical comparison between AVcon and A, corrected for multiple comparison (color scale indicates t-value).

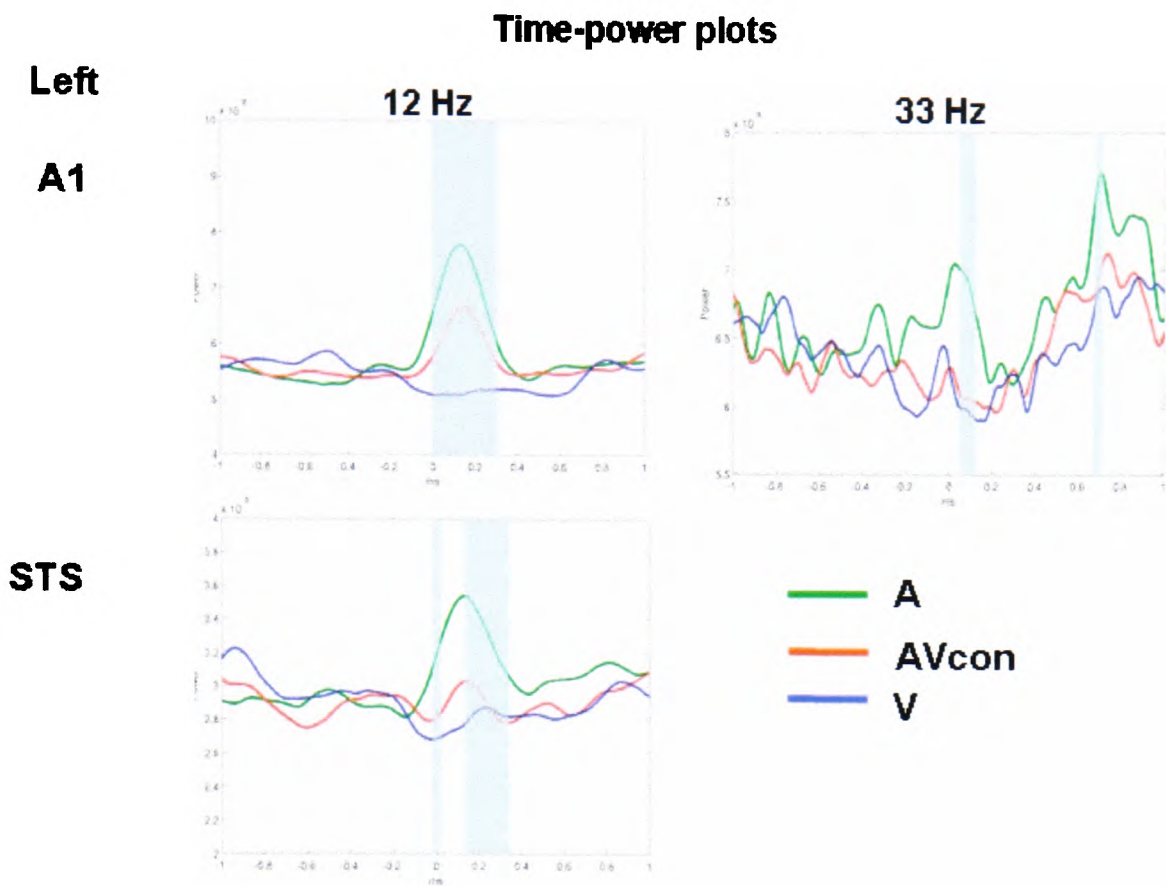


Figure 6-3: Time-power plots of left A1 & STS
A1 (top row) at 12 Hz and 33 Hz and left STS at 12 Hz (bottom row) comparing A, AVcon and V conditions. X-axis shows time from -1 s to 1 s relative to auditory onset. Blue-shaded areas show time segments of significant difference between AVcon and A.
[Modified figure]

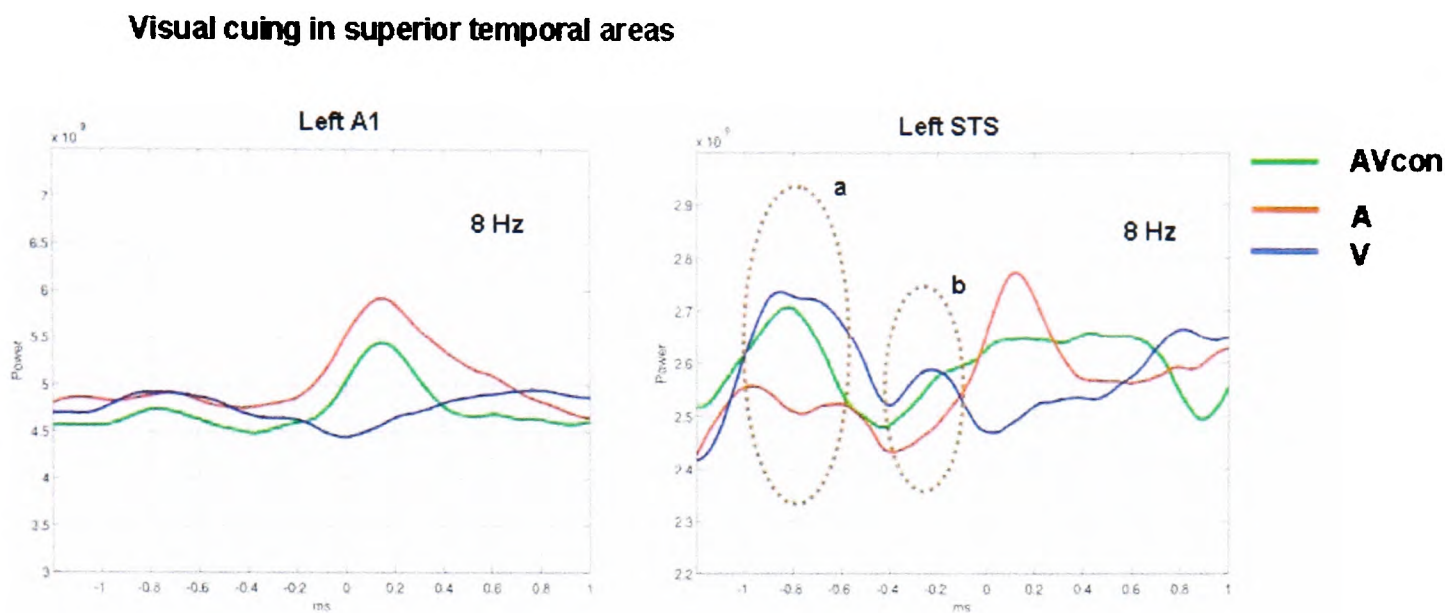


Figure 6-4: Visual cuing in left A1
A1 (top row) and left STS in the lower alpha range (8 Hz) comparing A, AVcon and V conditions. Circle a. and circle b. highlight visual cuing effects in left STS. Note the low temporal resolution of the lower alpha band (125 ms per cycle). X-axis shows time from -1 s to 1 s relative to auditory onset.

[FIGURE modified]

We found evidence for a visual influence on STS regions before auditory onset (Figure 6-4). Power in the lower alpha band increased in STS first after face onset ($\sim$ 900 ms) and then again, less pronounced, after the onset of visual mouth movement ($\sim$ 300 ms). This effect was not seen in A1.

6.4.4 Phonetic interactions

Presenting incongruent AV pairs where the visual stimulus shows non-articulatory mouth movements (gurning) causes widespread changes in the left STS and right A1, but not in left A1 and right STS (Figure 6-5). Left STS changes include the beta and omega bands. Right A1 shows a strong modulation in the beta band before and after auditory onset. Figure 6-6 shows this effect in more detail. Broca’s area showed moderate modulations in the alpha band (before auditory onset) and in the beta band (after auditory onset). At $\sim$ 200 ms in the lower beta band we see

evidence for a differential processing of visual mouth movement in right auditory cortex. Non-speech mouth movements show a reduction in power (ERD) compared to articulatory mouth movements which lasts until the onset of the auditory stimulus.

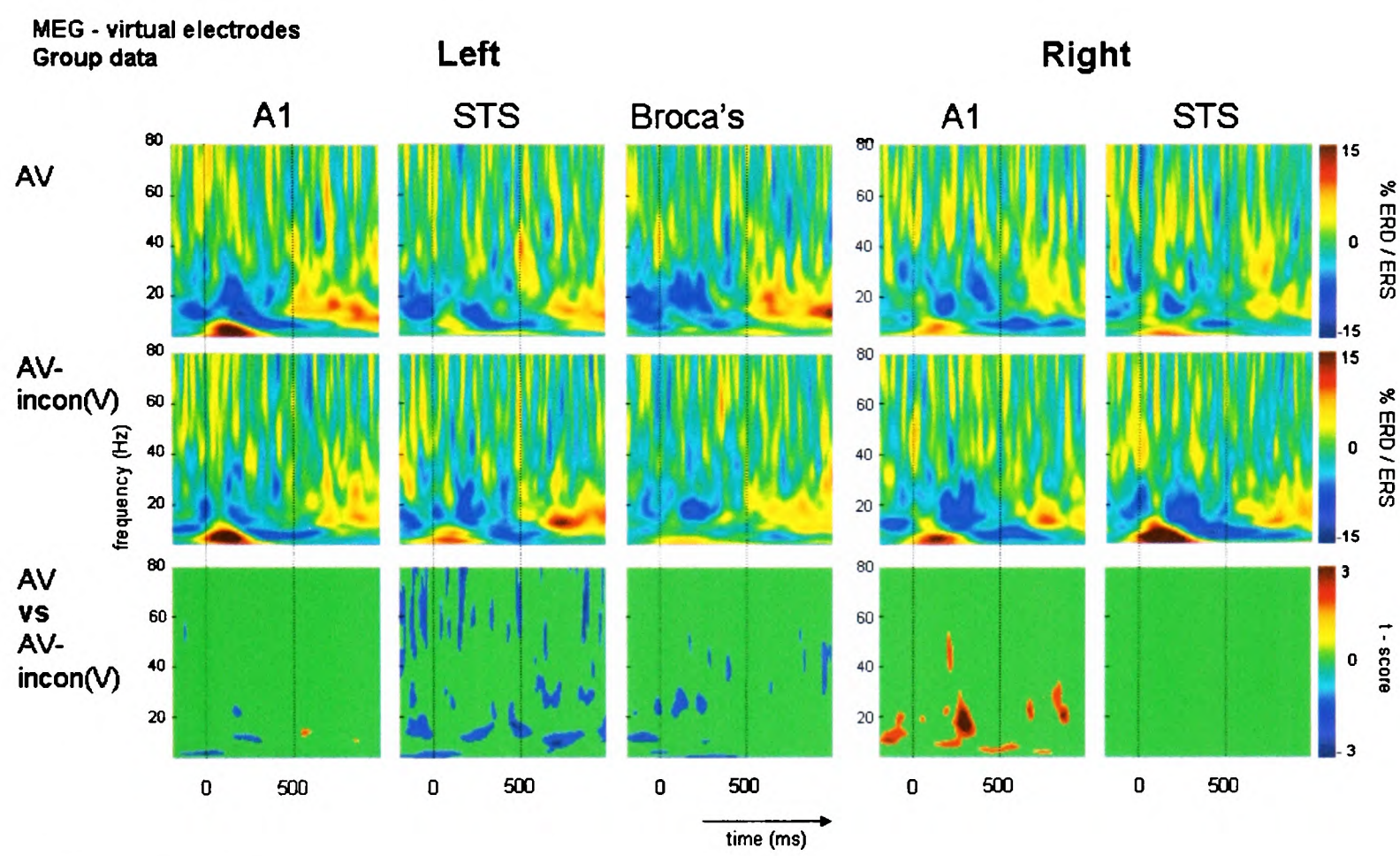


Figure 6-5: ERD/ERS group data for AVcon and AVincon(V)
At virtual electrode locations for the time period from - 150 to 1000 ms relative to auditory onset and for frequencies from 4 to 80 Hz. ERD/ERS percent values are color-coded (see scale). Bottom row shows the results of the statistical comparison between AVcon and AVincon(V), corrected for multiple comparison (color scale indicates t-value).

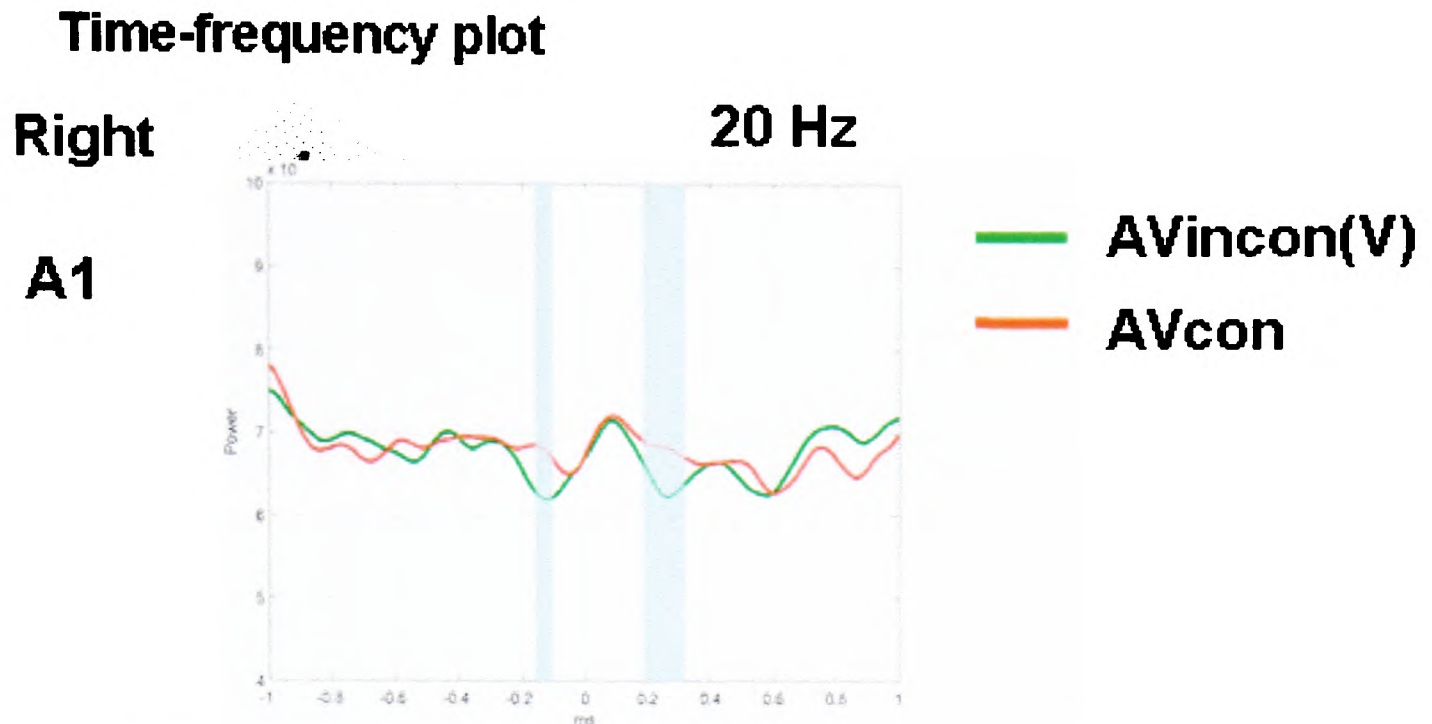


Figure 6-6: 20 Hz oscillations in right A1

Time-frequency plot showing differential processing of visual mouth movement in right auditory cortex at 20 Hz before and after auditory onset for AVincon(V) and Avcon. X-axis shows time from -1 s to 1 s relative to auditory onset.

6.4.5 Cortical power changes in visual cortex

A VE located in primary visual cortex showed large signal modulation in response to uni- and bimodal stimulation (Figure 6-7). Visual stimulation caused a large ERD in the alpha and lower beta band. We also observed a modulation of visual cortex by auditory stimulation. This occurred in the alpha and beta bands around 200 ms after auditory onset. Figure 6-7 (right column) shows these effects in the 15 Hz frequency band in more detail. The presence of a visual stimulus leads to a subsequent ERS in alpha/beta bands ~600 ms after auditory onset, a response difference which is statistically significant. These data suggest that this process might reflect a release from task, i.e. a result of having finished with the task, since the presence of an auditory stimulus seems to decrease the latency, whereas in the V condition, where uncertainty over task completion is more ambiguous, shows a longer latency of the 15 Hz ERS.

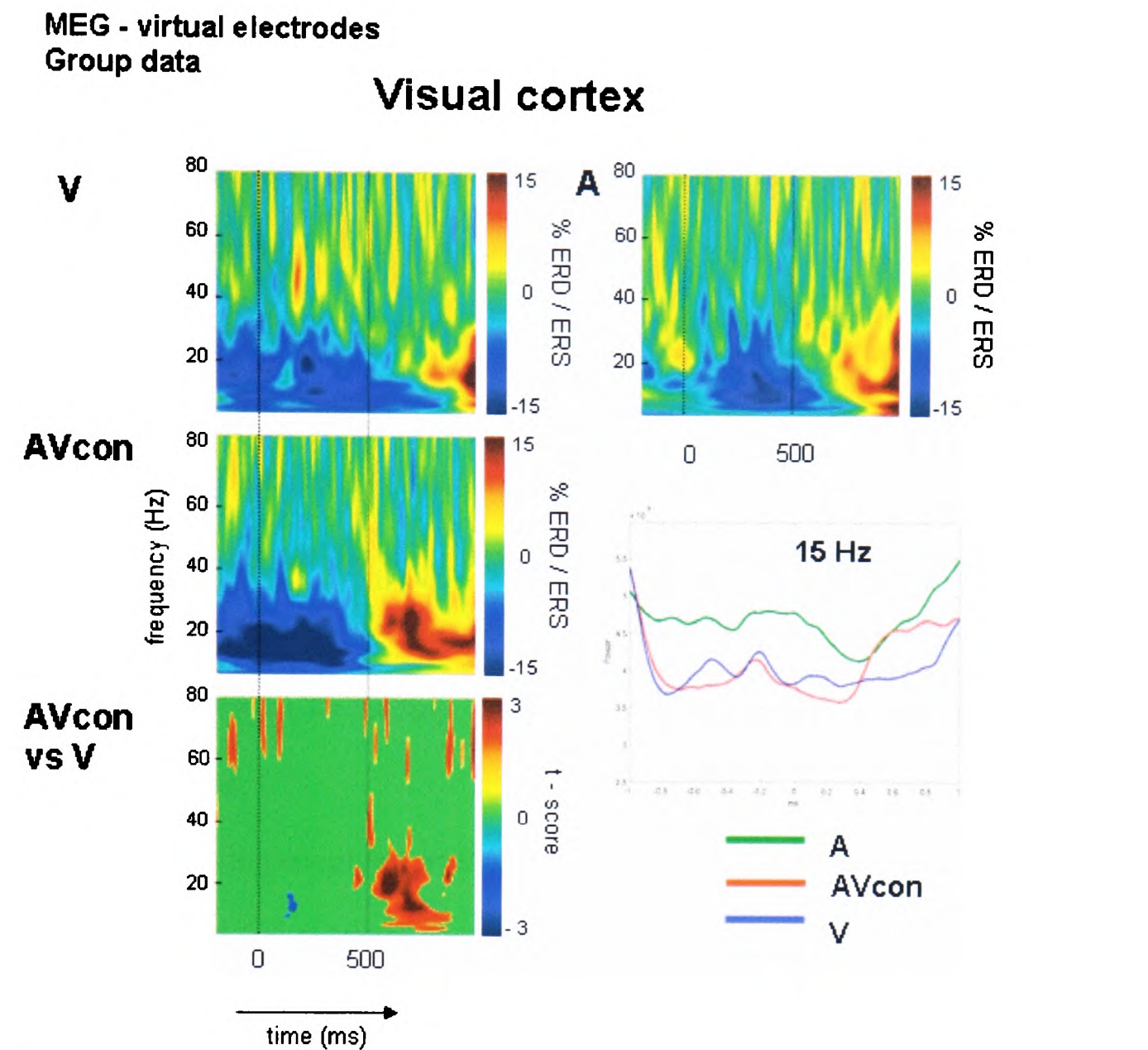


Figure 6-7: ERD/ERS in visual cortex
Spectrograms showing ERD and ERS values for V, A and AVcon conditions in visual cortex for the time period from - 150 to 1000 ms relative to auditory onset and for frequencies from 4 to 80 Hz. ERD/ERS percent values are color-coded (see scale). Bottom left spectrogram shows the results of the statistical comparison between AVcon and V, corrected for multiple comparison (color scale indicates t-value). Time-frequency plot shows activity in the lower beta band (15 Hz) for the conditions A, Avcon and V. Note the decrease in power after auditory only stimulation.
[Modified figure]

6.5 Discussion

We identified a dominant response, in terms of relative power change, to auditory stimulus processing in auditory cortex in the form of an alpha ERS immediately after stimulus onset. The response in this frequency band was sensitive

to the presence of a visual stimulus alone. Presentation of a visual stimulus together with a congruent auditory stimulus (AVcon) caused an ERD, indicating a crossmodal influence on auditory stimulus processing in auditory cortex which is mediated through oscillation in the 8-12 Hz frequency range.

Our findings are consistent with those of Herdman et al. (2004) who found an ERS in the 8-16 Hz frequency band in response to AV phoneme and grapheme presentations. This response was located to the left superior temporal lobe, encompassing HG and the planum temporale and was confined to an analysis time window ranging from 0-250 ms after auditory onset. These authors also reported an ERD in visual cortex for the time window 250-500 ms, which is also consistent with our observations.

Contrasting congruent vs incongruent graphemes and phonemes (Herdman et al., 2004) found that response power in the 4-8 Hz band, located to HG, was greater to congruent AV pairings. Furthermore, voxels in the visual cortex showed higher ERD to congruent stimulation in both frequency bands.

Visual influence on both STS and A1 was modulated by an alpha ERD which started before and increased after auditory onset. However, we show that visual influence on both areas is systematically different. Our data suggest that this crossmodal modulation is likely to be mediated through cortico-cortical synchronization of large-scale networks oscillating in the lower alpha frequency range (~ 8 Hz). There was also evidence for an involvement of faster beta oscillations in modulating auditory cortex through visual input.

Broca's area showed a modulation by visual input in the form of an alpha desynchronization before auditory onset, suggesting a possible early modulation route for visual speech to influence auditory cortex. A possible route could be via the arcuate fasciculus which sends heavy projections to STS (Matsumoto et al., 2004) which in turn has reciprocal connections with A1.

Our results show that during auditory only (A) stimulation transient gamma band bursts (ERS) were observed immediately after stimulus onset. These ERS were not present during visual only stimulation (V) and were reduced by the presence of a visual stimulus in the congruent bimodal condition (AVcon). In previous studies, this gamma ERS was shown to be enhanced in case of detected versus undetected auditory stimuli (Makeig and Jung, 1996) and when stimuli are attended (Tiitinen et al., 1994), showing that this response may be a modulator of selective attention or reflect a mechanism of increased attention (Aoki et al., 1999). The observed decrease of this response during a bimodal task supports the role of attention effects in multisensory processing and suggests that visual modulation of auditory cortex is achieved by top-down processing. But assuming a limited attentional capacity, a concurrent visual stimulus may reduce the attentional resources that are allocated to the auditory modality, which would explain the reduction in the attention-related gamma response. Indeed, when attentional factors are controlled for by bimodal contrasts, as is the case in the bimodal incongruent contrast (AVcon vs AVincon(V)), a modulation in the gamma frequency band does not occur.

Non-phonetic modulations (AVcon vs A) were greater in left A1 and Broca's area, whereas modulation based on phonetic level congruency were largest in left STS and right A1. This is consistent with the view of a role for STS in the analysis of phonetic features (Jancke et al., 2002) and the involvement of right superior temporal areas in the automatic detection of dissonant tones in melodies (Kuriki et al., 2005). Feature extraction during melody perception involves matching of an expected with an actual auditory percept within a sequence of events. The present findings suggest that this processing step may be equally related to the matching of expected auditory input during AV speech perception with actual input, the only difference being the modality of the preceding cue (i.e. viseme vs tone). Furthermore, we found beta activity in auditory cortex at around 20 Hz showing a differential sensitivity for

articulatory vs non-articulatory mouth movements. Interestingly, this difference emerged already before auditory onset and may represent a mechanism for cuing auditory cortex for the arrival of meaningful and speech-relevant auditory input.

The results also show a crossmodal modulation of visual cortex. We see an alpha ERD, an indicator of stimulus processing in visual cortex (Berger, 1930; Pfurtscheller, 2001), after auditory only stimulation. This reduction is larger and more wide-spread during visual stimulation, but still clearly visible in responses to auditory speech alone. Visual imagery can cause similar changes in 10 Hz alpha rhythms during visual imagery (Salenius et al., 1995). These findings are in agreement with the increased BOLD activity in visual cortex to auditory only stimuli we see in section 3, and support an attention or top-down modulated account of crossmodal influence

Together, these findings challenge the traditional view that primary sensory cortices only process information from one modality, and call for a reassessment which takes into account the increasing number of studies showing crossmodal effects in primary sensory cortices. Regarding models of AV speech perception, the present results highlight the importance of natural asynchrony between A and V stimuli during AV speech integration. Our results suggest that auditory cortex is activated by visual speech before the arrival of auditory input, a mechanism which explains our findings in section 4 of early auditory cortex modulation at 38 ms and might explain the perceptual changes seen in the McGurk illusion.

7 Brain oscillations on the superior temporal lobe during AV speech perception

In a previous section we have demonstrated with MEG frequency analysis techniques specific patterns of event-related synchronization and desynchronization during auditory and visual speech processing and their integration. We found modulations of auditory cortex and STS by the presence of a visual speech stimulus, and found evidence for a visual modulation of STS immediately before auditory stimulus onset. This is suggestive of a visual ‘priming’ effect of auditory areas which explains our findings from MEG dipole analysis and iEEG ERP analysis of early auditory-visual interactions in auditory cortex. In this section we will test these findings using the superior spatial resolution of intra-cranial EEG recordings on the superior temporal lobe.

7.1 Introduction

In section 5 we have applied event-related averaging techniques (ERP) to iEEG data and found visual effects on auditory stimulus processing in auditory areas of the superior temporal lobe. Section 6 used frequency analysis on MEG data to detect cortical power changes across a network of brain areas and found evidence for a visual modulation of superior temporal areas including auditory cortex before auditory stimulus onset, suggesting a role for visual stimulus precedence in AV speech integration. The present section will use a similar frequency analysis on the iEEG data to confirm the MEG results and to allow for a higher spatial differentiation of synchronization of cortical oscillatory activity on the superior temporal lobe.

Synchronization of neuronal activity has been described as an important feature of dynamic functional integration in the brain. Phase synchronization between oscillating units in the neuronal network is thought to be an essential mechanism

driving synchronization (Varela et al., 2001). A large body of research has established a role for phase synchronization during the binding of visual features. Stimulus properties, such as edges, motion, colour depth, texture, etc. are processed in specialized cortical modules in the visual system. A solution to this visual binding problem (Treisman, 1996) is offered by phase synchronization, where visual objects are coded by neuronal assemblies that fire synchronously (Roskies, 1999). As such, the coordinated behaviour of local neural assemblies through synaptic interactions seems to be the most relevant and appropriate level of analysis for studying integrative cortical functions, such as multisensory integration.

From an anatomical point of view it is worth noting the difference between large-scale and local synchronization. Large-scale synchronization refers to oscillatory power changes in neuronal assemblies that are far apart, for example between occipital and frontal lobes or across the hemispheres. Here, the synchronization is based on distant, poly-synaptic connections (of either cortico-cortical or reciprocal thalamo-cortical pathways) between related units in the networks (Bressler, 1995). Local integration, on the other hand, occurs on a smaller spatial scale. In the dense ordered cytoarchitectonic structure of cortical columns for example, clusters of excitatory and inhibitory interneurons on the scale of 2mm synchronize together into a common resonance mode (Gray, 1999). However, 'local' integration also occurs on a slightly larger scale. Primary visual cortex columns that are separated by 2-7 mm and have non-overlapping visual fields but share similar feature properties, tend to synchronize as well (Gray, 1999), probably mediated by lateral connections. Even though properties of local networks are likely to differ between cortical regions and functional specifications, as a rule of thumb local integration occurs over an area of ~ 1 cm and is mediated by monosynaptic connections with conduction delays less than 7 ms (Girard et al., 2001). This suggests that ERD and ERS observed from electrode grids which record LFPs with an inter-

electrode distance of 4-5 mm are best interpreted as synchronization of a local, interconnected network of neurons. Whether the synchronization is mediated by a distant area belonging to the same network or generated by a local oscillator cannot be said with certainty (Varela et al., 2001). Using iEEG and frequency analysis, in the present section we attempt to confirm findings from the MEG frequency analysis in section 6 and to determine ERS and ERD patterns of AV speech processing on the superior temporal lobe with greater spatial resolution and certainty.

7.2 Methods

7.2.1 Subjects

The analysis in this section is performed on the same data as presented in section 5. A detailed account of the methodology and recording parameters is given in the Methods section of section 5. For a description of stimulus condition see section 3.

[Removed repetition]

7.2.2 Data analysis

7.2.2.1 *Time-frequency spectrograms*

In an exploratory spectral analysis Morlet wavelet analysis was used to compute 3D Spectrograms for all electrodes within a frequency range of 1 - 80 Hz across a 5000 ms time window (-2500 to 2500 ms relative to the onset of the auditory stimulus) at 1 Hz frequency resolution and 10 ms temporal resolution. Wavelet width was held constant across frequencies ($g=5$). To produce ERD/ERS representations the spectrograms were converted into percent power changes relative to a passive time window with no stimulation immediately before visual stimulus onset (-1400 ms to -1200 ms) and averaged across trials (Figure 7-1). This method has been used previously to identify reactive frequency bands from EEG scalp recordings (Pfurtscheller and Lopes da Silva, 2005).

For statistical comparison of conditions, a two-sided pair-wise t-test was used on the individual trials for every time bin and frequency bin at each electrode (-1400 ms to 1000 ms, 1-80 Hz). The statistical analysis was performed for each frequency and time point without any prior assumption about specific modulation related to our comparisons. Therefore, a cluster-level randomization procedure was used to control for inflated Type-II errors using ClusterRandAnalysis, a MATLAB based function which is part of the FieldTrip analysis package (<http://www2.ru.nl/fcdonders/fieldtrip>). (see Methods, section 5) (Maris and Oostenveld, in press). The randomization null distribution was obtained by a Monte-Carlo randomization with 500 draws. A conservative confidence interval of $p < 0.001$ was selected for both the cluster-level threshold and randomization tests.

In order to make topographic comparisons of activity and effects over time, time-frequency averages and statistical effects were visualized on a three-dimensional matrix equivalent to the spatial distribution of the sensors on the 8x8 electrode grid using MATLAB (MathWorks Inc.). Grid Data were interpolated and then colour-coded for effect direction (ERD/ERS) and amplitude (Figure 7-1). For visual presentation we created spatio-temporal maps for the traditional EEG frequency ranges Alpha (8-13 Hz), Beta (14-30 Hz), lower Gamma (31-60 Hz) and higher Gamma (Omega, (61-80 Hz)(Niedermeyer, 2005). These were chosen to allow comparison with the existing literature and because an exploratory visual analysis showed a somewhat similar distribution of clusters of reactive signal changes within these frequency boundaries. This visual analysis also suggested different reactivity patterns between low and high gamma responses. We therefore adopted a nomenclature proposal by Curio (2000) and divided the higher frequency range into Gamma (31-60 Hz) and Omega (61-80 Hz).

7.3 Results

Stimulus-related power changes were observed over a number of electrode sites and frequency bands. The overall magnitude of oscillatory power was highest in the alpha and beta frequency bands and then decreased with higher frequencies, an observation consistent with previous studies concerning the spectra of natural phenomena (Mandelbrot, 1998), including MEG and EEG data .

7.3.1 Sensory responses and non-phonetic modulations

7.3.1.1 *Alpha band*

7.3.1.1.1 *S106*

In the auditory only condition (A) a large alpha EDS was detected over electrode sites on the middle section of the PLST (electrodes 20 & 28) around stimulus onset and then spread over the entire length of the middle STS section covered by the electrode grid, including the ventral part of the post-central sulcus (Figure 7-1, b). This ERS was followed by an ERD with a similar spatial distribution after 200 ms post-auditory onset. Visual speech alone (V) influenced regions along the entire anterior, middle and posterior sections of the STG, an effect that was manifested by ERD. Bimodal AV speech (AVcon) showed a similar, but reduced response as the A condition (Figure 7-1, c). Similar to the visual only condition (V), an ERD was observed on the posterior and anterior parts of the STS, surrounding a large ERD immediately above HG (electrodes 20 & 28). After 200 ms most STG electrode sites showed ERD. The bottom row of Figure 7-1 shows the spatio-temporal distribution of the statistical difference between A and AVcon. Most significant effects happen along the STG between -100 and 200 ms.

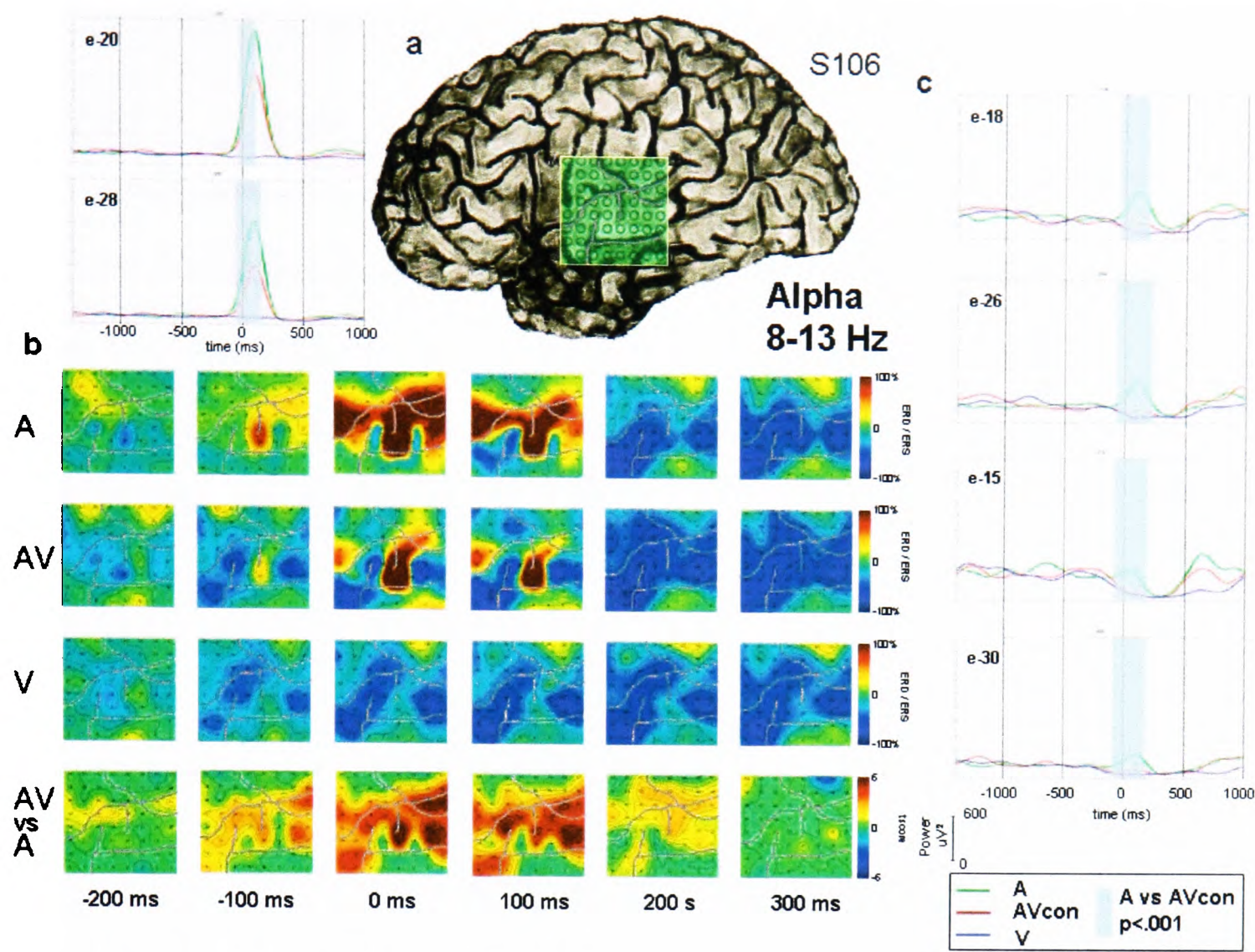


Figure 7-1: Alpha ERD/ERS in S106

a: Position of the 64 contact electrode grid on the surface of the left hemisphere of S106 (green square

See Figure 1 in Appendix for a high-resolution example of the numbering of electrodes on the grid.

b: Surface rendering over time of the averaged evoked response to A, AVcon and V condition. Lower row shows the results of the statistical comparison AVcon vs A. Plotted are averages of a 100 ms time interval in 100 ms steps from -200 ms to 300 ms after auditory onset. Color-coding indicates the amplitude level of the ERD/ERS in percent. Red corresponds to an increase in power relative to baseline (ERS), blue to a decrease in power (see colour bar on the right). Grey lines on the grid correspond to anatomical landmarks derived from A. to help locating activation to cortical structures on the rendered MRI (also see Figure 7-1).

c. C: Plotted are averaged frequency responses to auditory (A), visual (V) and auditory-visual (AV) stimulation plotted for selected electrodes (e.g. "ch 18"). Blue shading indicates time window of significant difference for A vs AV ($p < .001$), corresponding directly to the surface renderings of B.

7.3.1.1.2 *S100*

S100 showed a very similar response profile as S106 in the alpha band. Auditory stimulation induced a large ERS on superior temporal areas around stimulus onset. The visual ERD observed in S106 was not clearly visible in this subject, however, a similar reduction in alpha power was observed during AV presentation in both subjects (contrast AV vs A).

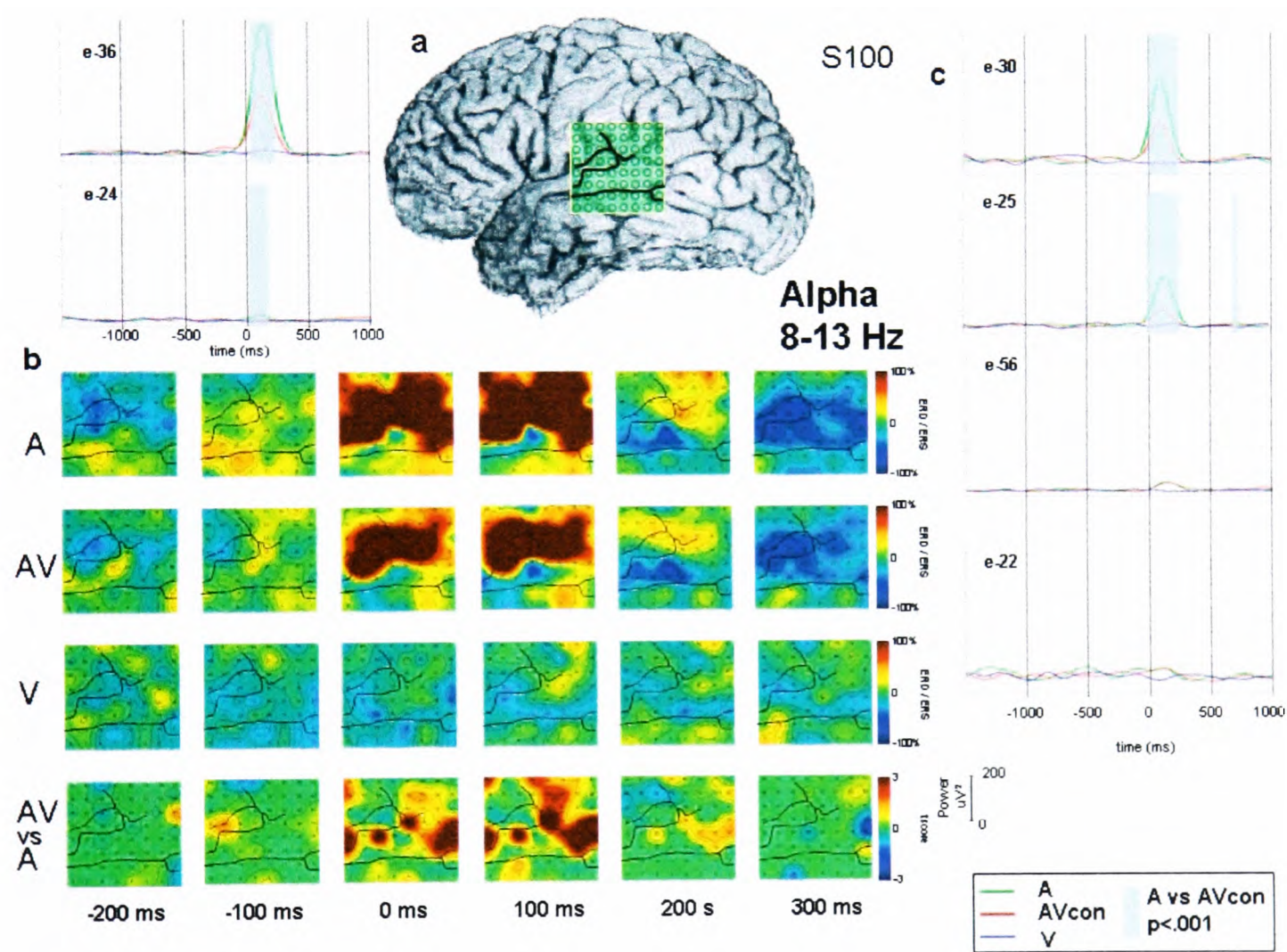


Figure 7-2: Alpha ERD/ERS in S100
(for a detailed figure legend see Figure 7-1)

7.3.1.2 *Beta band*

7.3.1.2.1 *S106*

Beta oscillations were less pronounced than alpha, but did show a dominant peak (ERS) over HG after auditory stimulation. Visual only stimulation, however,

modulated superior temporal activity through an ERD over most of the STG and the ventral part of the post-central sulcus. The difference contrast between A and AVcon shows that visual modulation effects ventral somatosensory cortex and large parts of STG up to until 100 ms after auditory onset. This modulation is expressed as an ERD during the bimodal condition as a consequence of visual stimulation.

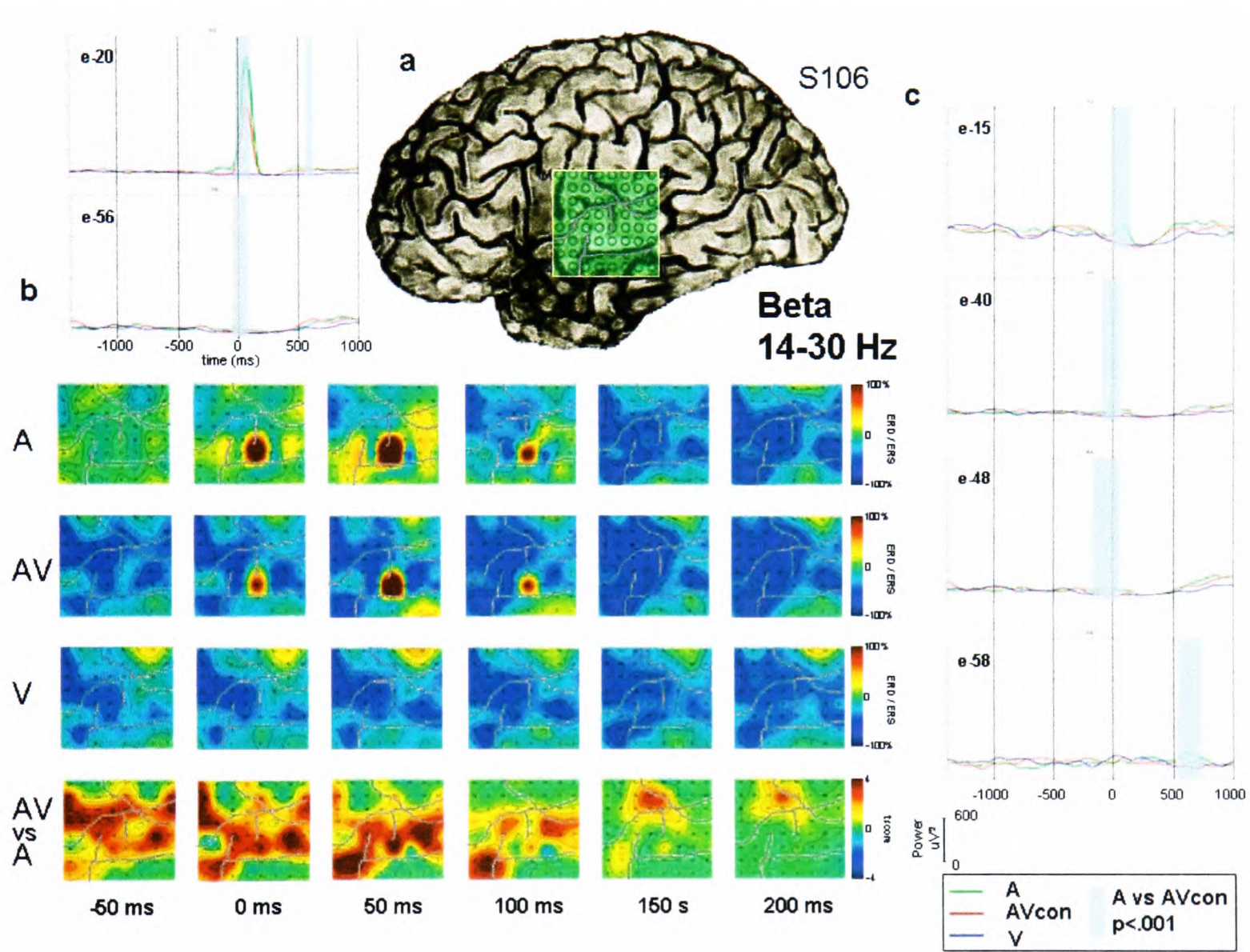


Figure 7-3: Beta ERD/ERS in S106
(for a detailed figure legend see Figure 7-1)

7.3.1.2.2 S100

This subject also showed a prominent beta ERS immediately after stimulus onset over middle STG during the presence of an auditory stimulus. The contrast AV vs A showed similar, albeit less pronounced differences between these conditions.

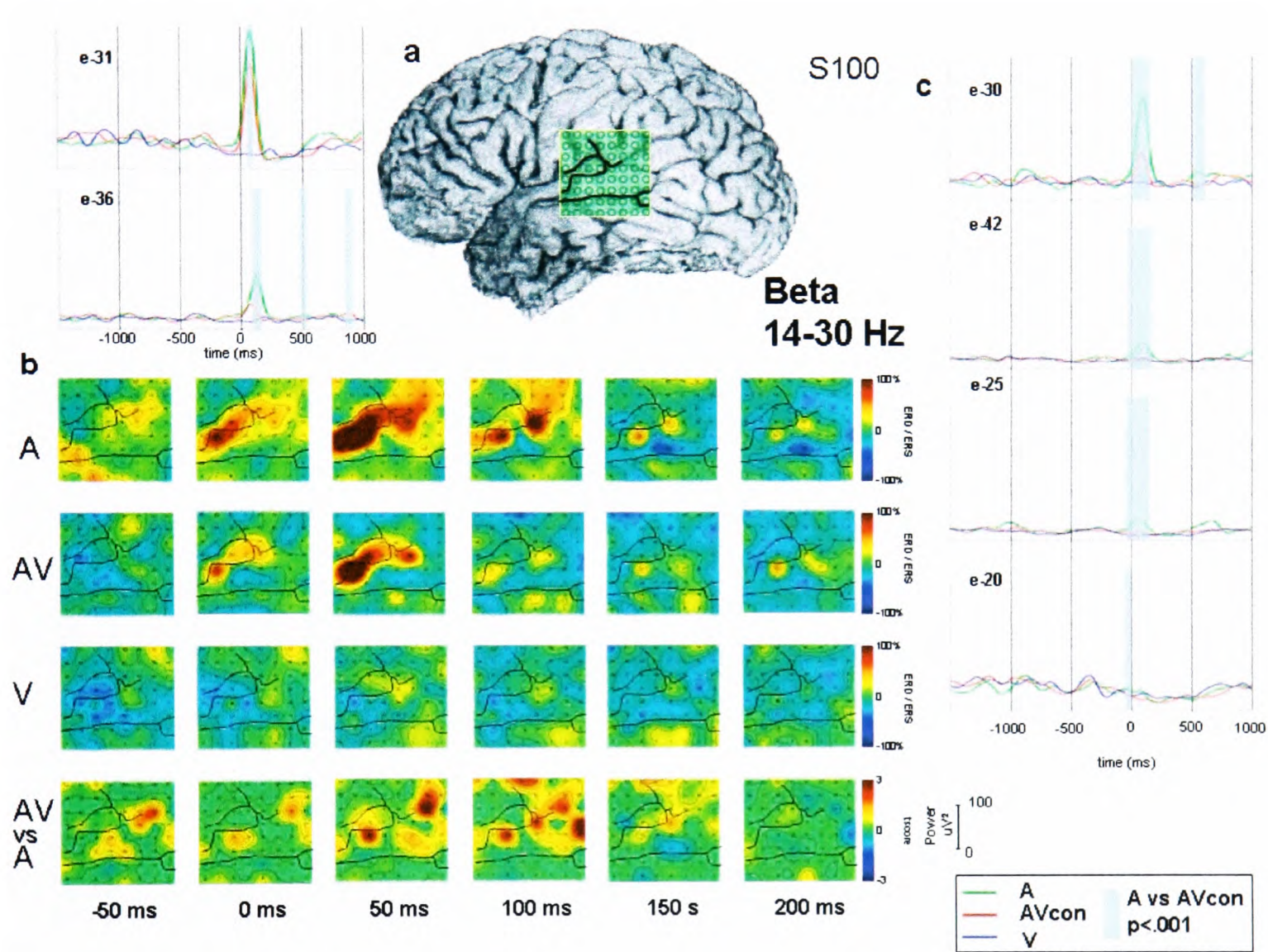


Figure 7-4: Beta ERD/ERS in S106
(for a detailed figure legend see Figure 7-1)

7.3.1.3 *Gamma band*

7.3.1.3.1 *S106*

Gamma band ERS occurred around 50 ms after auditory onset on the posterior STG and was related to the auditory stimulus in condition A and Avcon. No gamma band changes were observed in response to visual only stimulation after auditory onset. However, we observed a small but significant modulation of visual input before auditory onset at electrode sites on the posterior STG.

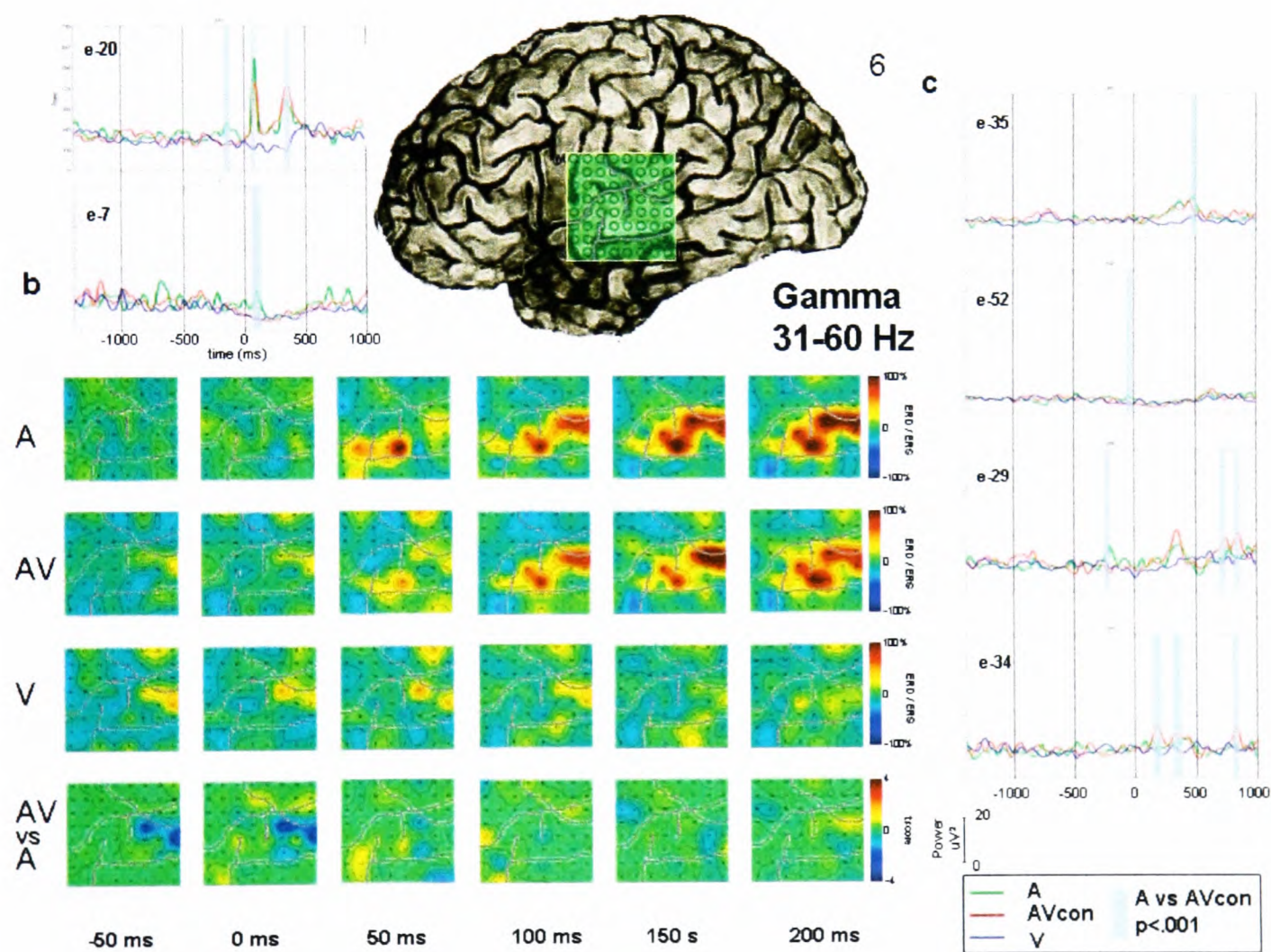


Figure 7-5: Gamma ERD/ERS in S100
(for a detailed figure legend see Figure 7-1)

7.3.1.3.2 *S100*

A gamma band ERS was observed during auditory presentation on the middle STP. However, no significant differences between AV and A were observed in this subject.

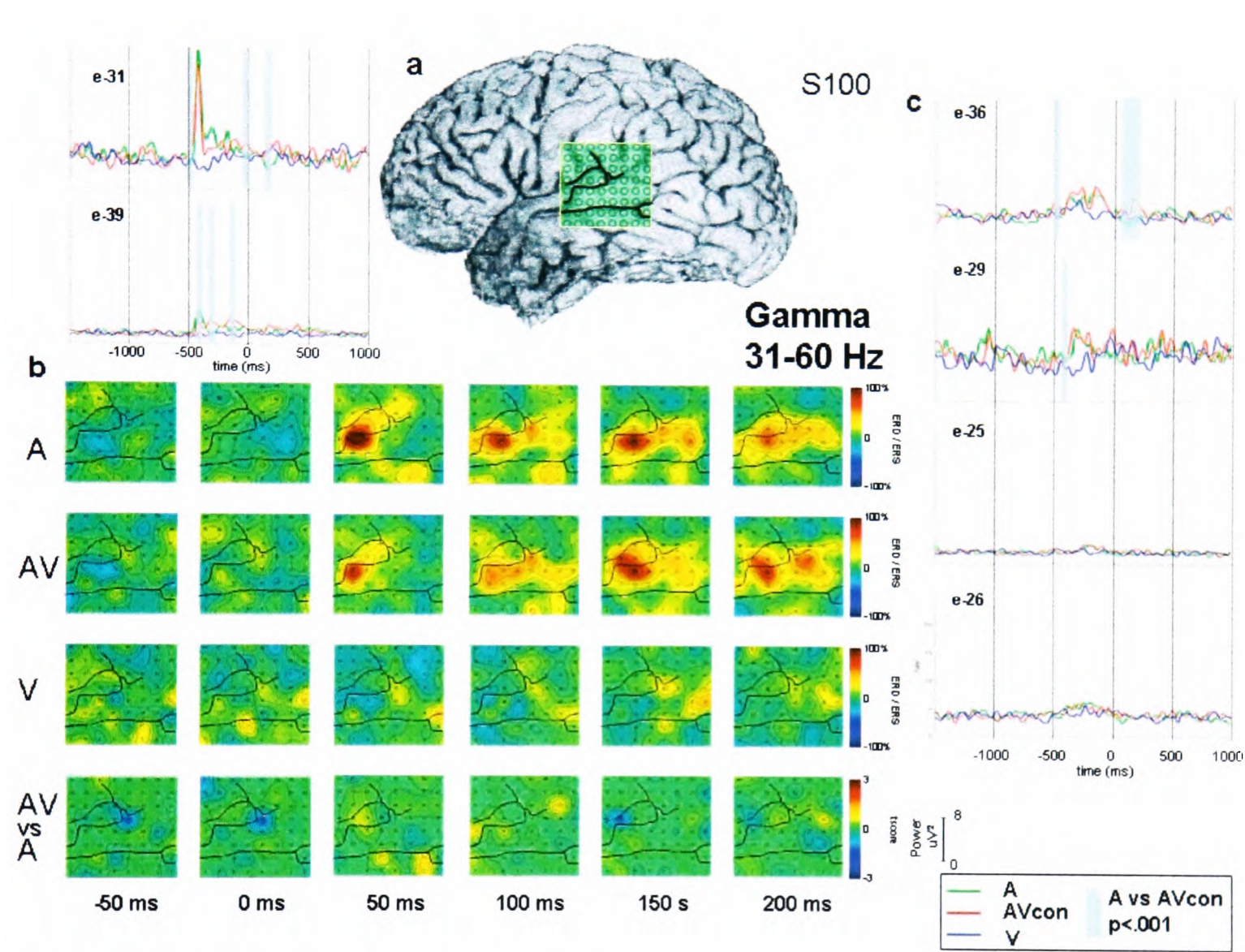


Figure 7-6: Gamma ERD/ERS in S100
(for a detailed figure legend see Figure 7-1)

7.3.1.4 *Omega band*

7.3.1.5 *S106*

The omega frequency band showed high reactivity to both auditory and visual speech input. The largest response is an ERS to auditory stimulation over the middle and posterior bank of the STG starting over HG and progressing caudally. Visual stimulation accessed the anterior STG before auditory onset through an increase in omega power during the V and AVcon condition (electrode 25). This effect was also seen when calculating the statistical contrast Avcon vs A.

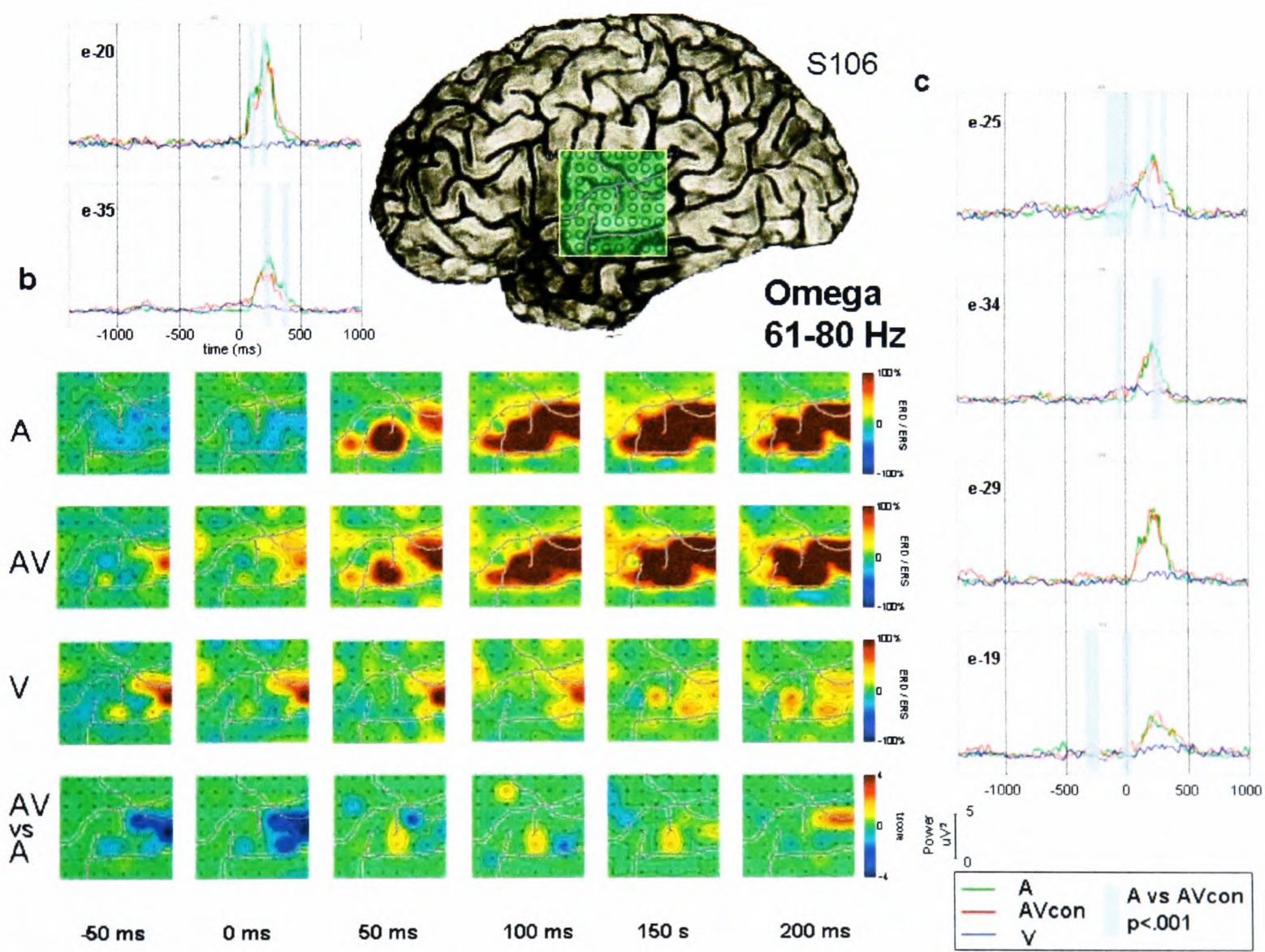


Figure 7-7 Omega ERD/ERS in S106
(for a detailed figure legend see Figure 7-1)

7.3.1.5.1 *S100*

S100 showed also larger ERS changes in the high frequency omega band on the middle STG. An area on posterior STG also showed modulation by visual input before auditory stimulus onset.

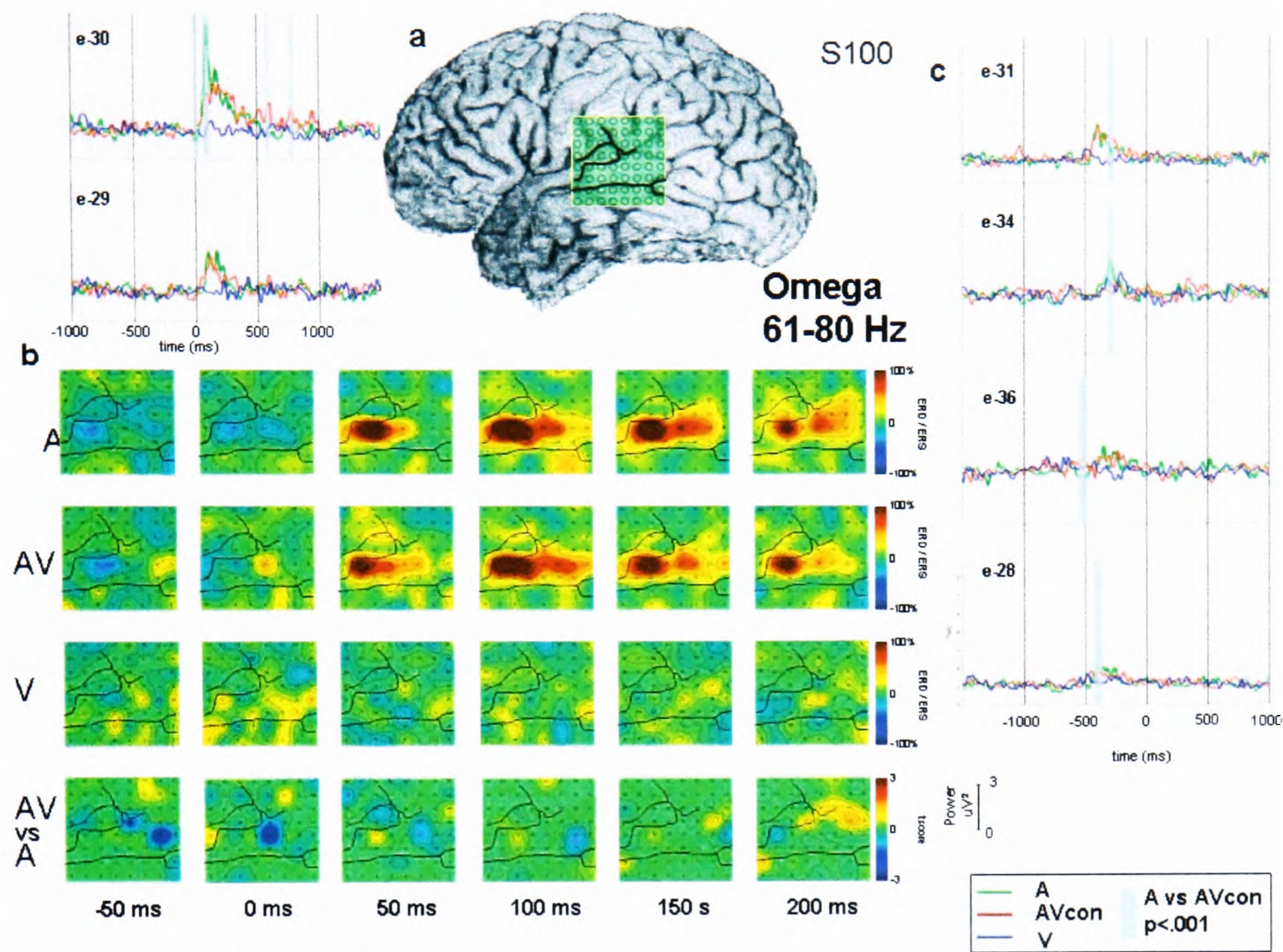


Figure 7-8: Omega ERD/ERS in S106
(for a detailed figure legend see Figure 7-1)

7.3.2 Phonetic modulations

7.3.2.1 AVcon vs AVincon(V)

7.3.2.1.1 S106

The effect of incongruent and non-linguistic facial movement on auditory-visual speech processing was tested by contrasting AVcon vs AVincon(V). In the alpha band this effect was significant around auditory onset on the anterior part of the STG, whereas significant effects in the beta frequencies showed a less focussed spatial distribution with significant differences on the anterior STG and ventral somatosensory cortex. No large effects were observed in the gamma band, whereas the high frequency omega band showed a clear effect cluster over the posterior STG 200 to 100 ms before auditory onset.

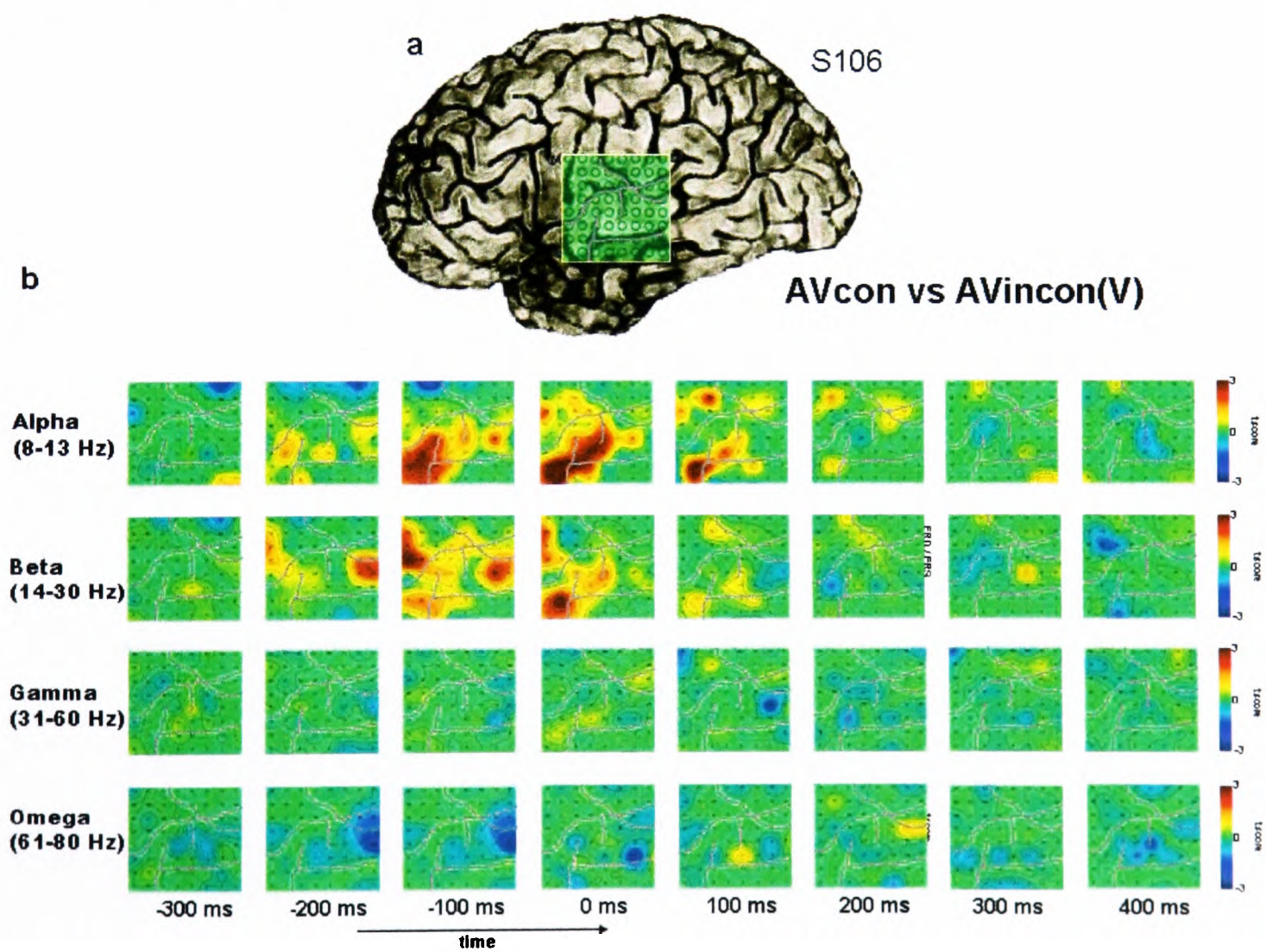


Figure 7-9: AVcon vs AVincon contrast for S106

- a: Position of the 64 contact electrode grid on the surface of the left hemisphere of S106 (green square). See Figure 1 in Appendix for a high-resolution example of the numbering of electrodes on the grid.
- b: Surface rendering over time of the statistical result for each frequency band (alpha, beta, gamma, omega) comparing congruent AV stimulation with incongruent AV stimulation where the visual signal is replaced by gurning. Plotted are averages of a 100 ms time interval in 100 ms steps from -300 ms to 400 ms after auditory onset. Color-coding indicates the t-score of the statistical test.

7.3.2.1.2 S100

In S100 the contrast Avcon vs AVincon(V) also showed differences in alpha power after stimulus onset, however the higher frequencies showed fewer significant differences between congruent and incongruent AV speech.

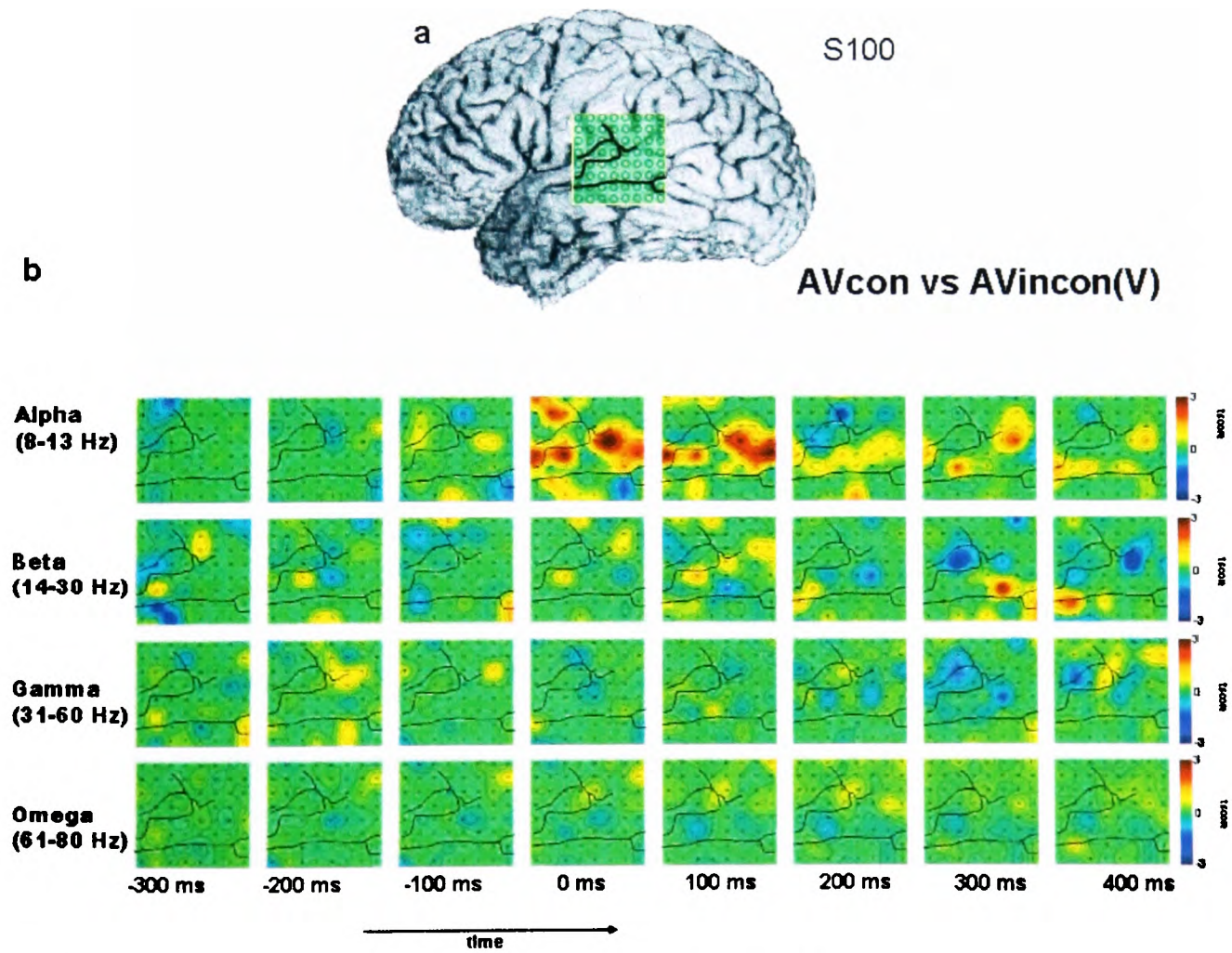


Figure 7-10: AVcon vs AVincon contrast for S106

(for a detailed figure legend see Figure 7-1)
[Figure modified]

7.3.3 Visual cuing in the Omega band

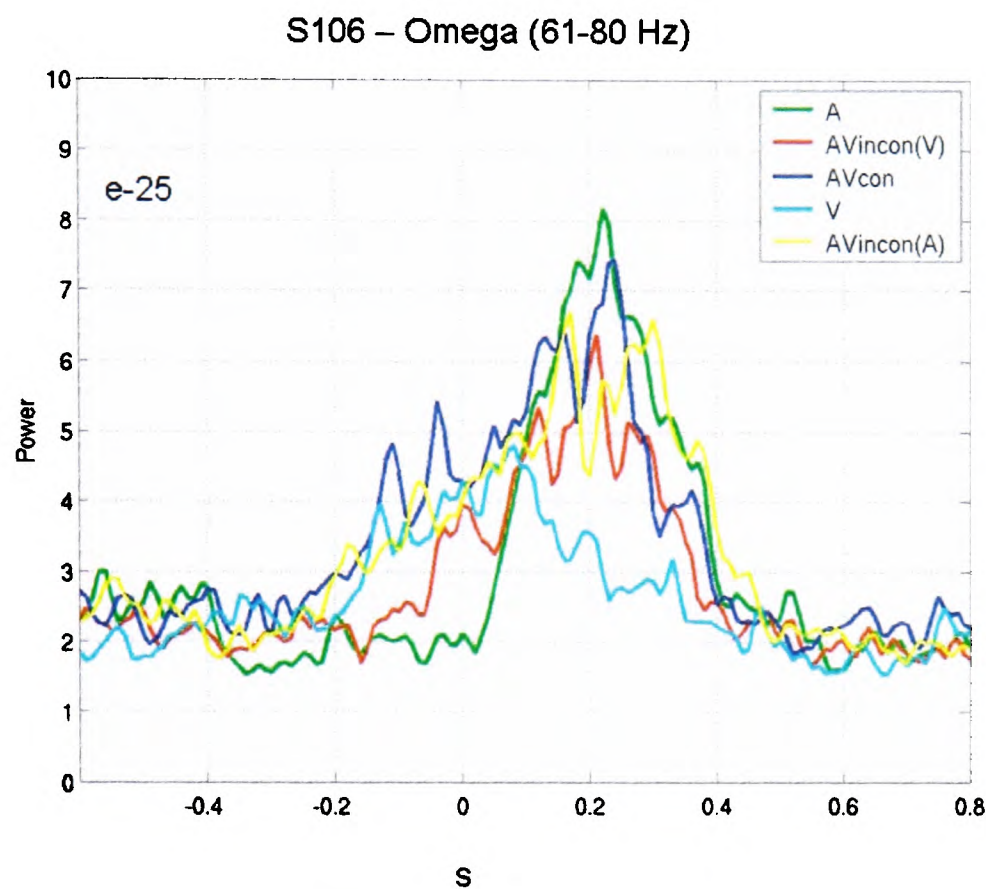


Figure 7-11: Visual cuing in the omega band
Averaged Omega activity at one electrode on posterior STG on S106 (for detailed location see Figure 1 in Appendix) for each condition. 0 = auditory onset
[Modified figure]

In S106 we found an area over the posterior STG which showed a clear responsiveness to visual input before auditory onset (Figure 7-11). This area responded with an ERS in the fast omega range (13-16 ms cycle length) to visual face movements starting as early as ~ -200 ms. This area on posterior STG was also able to discriminate meaningful articulatory mouth movements (AVcon, V, AVincon(A)) from non-meaningful gurning (AVincon(V) through a latency delay in ERS of about 120 ms. Showing auditory co-responsiveness, the presence of an auditory stimulus caused a further increase in high-frequency cortical oscillatory power compared to

visual only (V) stimulation. However, a similar response pattern was not found in S100.

7.3.4 Coherence

A preliminary analysis of the phase relationship between electrodes (Pfurtscheller and Andrew, 1999; Schoffelen et al., 2005) showed that highest coherence values were observed between immediately neighbouring electrodes, but not between distant electrodes on the grid. This suggests that the recorded activity stems from focal and independent local neuronal networks.

7.4 Discussion

Recording directly from the surface of the brain, we found transient episodes of synchrony in various neuronal populations across the superior temporal lobe and adjacent areas in response to auditory, visual and various combinations of auditory-visual speech stimuli. Time-frequency analysis revealed a dynamic and complex pattern of enhancements and reductions in local energy over various frequency bands. We were able to describe the development of these oscillatory behaviours of specialized neuronal assemblies subserving auditory and visual processing over space and time, and get a first idea of how they might be involved in the crossmodal processing and integration of audio-visual speech.

7.4.1 Quantification of synchronization in the superior temporal lobe

7.4.1.1 *Alpha band*

The most prominent frequency signal to auditory stimulation was observed in the alpha band, showing an ERS of up to 200 % compared to baseline at the highest electrode sites. Generally, the amplitude of the oscillatory components is inversely related to the frequency, meaning higher amplitudes for slower frequencies (Elul, 1971), a finding which is supported by the present results. Alpha band oscillations

seem to be an important mechanism in the auditory system. In the low frequency domain, 10 Hz “afterdischarges” were observed in mammalian auditory cortex immediately after presentation of tones (Maldonado and Gerstein, 1996) and click trains (Eggermont, 1992). At the same time, sections of the auditory thalamus show rhythmic discharges in the 8-12 Hz frequency range (Galambos et al., 1952; Bordi and LeDoux, 1994). Pacemaker neurons found in thalamus display oscillatory behaviour *in vitro* in the frequency range of 6-10 Hz, even after blockage of synaptic transmission (Jahnsen and Llinas, 1984; Jahnsen and Llinas, 1984). The intrinsic oscillatory properties of such neurons are likely to play an important role in shaping the oscillatory behaviour of the network to which they belong. These findings suggest that alpha oscillations observed in auditory cortex may reflect thalamo-cortical auditory activity related to bottom-up sensory processing (Cotillon et al., 2000; Schurmann et al., 1997).

In a series of cross-modality experiments Schurmann et al. (1997) measured responses to auditory and visual stimulation with three different imaging modalities. Using chronically implanted electrodes in cat primary auditory and visual cortex, they found high amplitude ERP responses to what they called “adequate” stimulation, i.e. high amplitudes in visual cortex to visual stimulation and high amplitudes in auditory cortex to auditory stimulation. However, a residual response to “inadequate” stimulation was also observed in both auditory and visual cortices. Interestingly, the frequency response in the auditory cortex showed the largest difference between auditory and visual stimulation in the alpha band, whereby figures presented by (Schurmann et al., 1997) suggest an alpha suppression (ERD) during “inadequate” visual stimulation and an alpha increase (ERS) during “adequate” auditory stimulation. Extending the paradigm to human subjects, very similar results were found for electrodes/sensors over auditory and visual cortex during EEG and MEG

recordings. Even though the authors interpreted their finding in a strict uni-sensory framework, the results are in agreement with our findings from iEEG recordings over human auditory cortex. If the alpha response indeed reflects thalamic input, the cortical modulation we observe from electrodes over auditory cortex in this frequency band is likely to be a result of crossmodal integration at the subcortical level, and likely to be related to the effects observed in the superior colliculus of the cat (summarized in (Stein and Meredith, 1993)). This effect then seems to be speech non-specific and a result of a more general integration of bimodal auditory-visual input at pre-cortical stages of information processing.

This view is supported by our results from comparing bimodal speech conditions which vary based on the congruency of the visual input while the auditory input is held constant, an integration process that can only occur after phonetic analysis at the cortical level. Here, congruency effects arise mainly in the regions of the STG and not at electrode sites immediately over HG. Electrodes over HG, however, show the largest difference when comparing unimodal and bimodal stimulation (AVcon vs A). This suggests that auditory and visual input signals are integrated at several cortical and subcortical structures depending on their stimulus features. Furthermore, concurrent bimodal input seems to interact already in subcortical structures irrespective of the nature and compatibility of the signal content. This can be explained by the evolutionary advantages offered by a faster reaction time and orienting response that is observed after bimodal stimulation with temporally, and sometimes even spatially, correlated arbitrary signals (Welch and Warren, 1986; Molholm et al., 2002). Integrating input signals based on a higher-order evaluation of the input signals, such as phonetic congruency, requires the initial step of decoding based on an analysis of stimulus properties. This type of complex sound and visual feature analysis requires the cortical processing architecture (Scott

and Johnsrude, 2003; Griffiths et al., 2004) such as that of the STS /STG (Binder et al., 2000). Therefore, oscillatory rhythms coding for content-related matching, such as anterior STG alpha, ventral post-central gyrus beta and posterior STG omega oscillations are likely to reflect information processing at a cortico-cortical level.

7.4.1.2 *Gamma band*

Local field potential recordings from monkey auditory cortex have shown a two-fold increase in oscillatory gamma amplitude (ERS) induced by simple auditory stimuli (Brosch et al., 2002). However, this response was dependent on the frequency of the stimulus and in the present study we did not find large gamma changes in response to AV stimulation.

7.4.1.3 *Omega band*

High gamma band activity is rarely observed with scalp EEG or MEG, perhaps because of the poor sensitivity of these methods to these frequencies. Power estimates generally decreased with frequency, from a peak signal of ~2000 mV² in the 8-13 Hz band to ~ 14 mV² in the 61-80 Hz band, and might thereby have decreased the signal-to-noise ratio. Furthermore, skull attenuates cortical high frequency activity in EEG (Pfurtscheller and Cooper, 1975) while at the same time preventing the iEEG from being contaminated by cranial muscle activity in the gamma band (Barker et al., 1989).

Most previous studies on electrophysiological high frequency activity have focussed on frequencies around 40 Hz (Pfurtscheller and Neuper, 1992; Pulvermuller et al., 1995). More recently, recordings from subdural electrodes have also shown stimulus related changes in higher frequency ranges. Modulations in the 75-100 Hz band in sensorimotor cortex were found during a visually paced (Crone et al, 1998) and self-paced motor task (Ohara et al 2000).

Oscillatory power changes to auditory stimuli in the omega range have been previously reported in a study of auditory phoneme discrimination (Crone et al., 2001). Recording from electrodes on the posterior superior temporal gyrus, this study found an omega ERS starting at around 100 ms after auditory onset for both phonemes and simple tones. Omega augmentation was generally higher for phonemes on posterior STG sites which lead the authors to suggest that a possible functional role of these high frequency oscillations might be cortico-cortical information transfer and higher-order integration of neuronal populations serving specialized functions. Similarly, gamma ERS were observed at iEEG electrode sites on the posterior STG during word reading when the input modality was either auditory or visual (sign-language) (Crone et al., 2001).

7.4.1.4 *Individual differences*

The frequency response profile was very similar between individuals and showed a higher degree of matching compared to the ERP results. The reason for this was mainly the latency difference in response between the subjects, which will have larger effects on ERP analysis compared to frequency analysis, where information is basically processed through a temporal filter. However, the spatial information remains preserved. Comparing individuals it becomes apparent that there is less variability in the spatial location of cortical responses, but that speed of neuronal processing is a more idiosyncratic marker of cortical processing which discriminates between individuals.

7.4.1.5 *Visual cuing*

Early arrival of modality-specific input into a cortical area can convey a processing advantage for that modality. For example, early input may have preferential access to local cellular excitability and thus, be in a position to modulate

(by inhibiting or enhancing) the processing of a later arriving input. The finding of visual input into posterior STS regions, mediated by fast omega oscillations and occurring before auditory onset, might be a mechanism by which later-arriving auditory input is modulated, perhaps through changes in local excitability. The fact that phonetic and non-phonetic visual input could be distinguished further supports a functional role of this process in AV speech processing.

In the present study we characterized the ERD and ERS profile of AV speech processing on the superior temporal lobe. We found evidence for a complex pattern of local oscillatory activity which was reproducible between subjects. We also found evidence that vision modulates large parts of the superior temporal lobe, including activity in HG at the first stages of auditory processing. We also found evidence supporting our hypothesis that visual speech activates STG regions immediately before auditory stimulus arrival, and that this modulation differentiates between meaningful and non-meaningful visual speech.

8 Comparing results from different imaging modalities

8.1 Introduction

The technical development of non-invasive neuroimaging techniques has advanced rapidly over the last decades. Structural MRI now allows for the exact anatomical visualization of cortical structures of individual subjects and for quantitative comparisons of brain structures across groups, including patients. fMRI, with its ability to measure changes in blood oxygenation and perfusion, has revolutionized the field of cognitive neuroscience and enabled researchers to perform non-invasive experiments on behaving humans with a spatial resolution that has been previously restricted to animal studies. At the same time, technological and theoretical advances in EEG and MEG methods have made it possible to estimate changes in brain activity on a millisecond time scale and with increasingly higher spatial precision. However, there are also significant limitations associated with each technique. Whereas fMRI offers detailed spatial resolution of hemodynamic changes associated with neuronal activity, it does not capture the temporal dynamics of electrophysiological activation, a key feature of neuronal coding (Friston, 2000). MEG and EEG, with their high temporal resolution, provide an excellent picture of the dynamics of neuronal populations, but due to the nature of the inverse problem, source localization is always uncertain (Hämäläinen et al., 1993). During a cognitive task many brain areas are active simultaneously and for such ‘real source’ constellations modelling the MEG data as ECDs allows identification of the ‘most probable’ areas. But applied together, anatomical source ambiguity can be reduced

considerably and thereby these imaging modalities offer the potential for great insight into neuronal function and the functional organization of the brain.

However, there are also a number of epistemic challenges posed by neuroimaging techniques. The observations of brain function are mediated through complex instruments and sophisticated research techniques, which in effect alter the phenomena under study to make them detectable to us. For example, in the case of MEG, the phenomenon ‘neuronal activity’ passes through several stages of transformation (e.g. neuronal current, magnetic field, magnetic sensors, digitization, signal averaging, source reconstruction, statistics, amongst others) before a decision about the original phenomenon under observation is made. This poses the question in how far does what is taken as evidence (z-values) in actual fact represents the true underlying phenomenon (in this case neuronal activity). This is especially true for new imaging methods, such as fMRI, where not all transformations from one stage to the next are known. The processes by which nuclei align themselves in a magnetic field are reasonably well understood, as is the mechanism that induces them to tilt by a radiofrequency pulse. Even the fact that fMRI measures increased blood flow in brain areas is not much of a problem in this respect. But what still remains an important question is the relation between neuronal activity and increased blood flow (Raichle, 2001).

However, missing links in such a chain of transformations inherent in scientific research techniques have not stopped scientific progress in the past. A good example is the Golgi silver nitrate stain method, which only stains selective neurons (Golgi, 1989). The exact mechanisms by which the method operates and the reasons for why it only stains selective cells are still not understood even 100 years after the introduction of the method. Nevertheless, this method was still instrumental in the discovery that neurons were discrete individual cells. How does one then evaluate the

appropriateness of a new research technique for the subject under study and establishes that it still produces genuine evidence, even if not all stages of transformation are understood? Bechtel and Stufflebeam, (2001) have found 3 common criteria that are used in the scientific community for such evaluations:

1. Does the instrument or technique produce well-defined or determinate results?
2. Is the degree to which the purported evidence coheres with existing theories plausible?
3. Does the degree to which the results from one instrument or technique agree with results generated in other ways?

In the case of fMRI, the usefulness of the method has been particularly controversial (Uttal, 2001; Donaldson, 2004). In this case, however, criterion 1 is well-satisfied as results are often localized very focally (as opposed to being scattered across the brain as single voxels), appear inside the cranium and show high test-retest reliability (Kurland et al., 2004; Wagner et al., 2005). Evidence in favour of fMRI for criterion 2 is given by the current knowledge on MRI physics (Jezzard and Clare, 2001) and its relation to the BOLD signal and, thus, blood flow (Ogawa et al., 1993). What remains disputed is the exact relationship between neuronal activity and blood flow, especially since the BOLD technique relies on blood flow increases more than is required by oxygen consumption by neurons (Raichle, 1998). More work in this field is clearly required. Testing criterion 3 is the purpose of this section.

The aim is to directly compare results from different techniques and analysis methods, and to test if converging evidence is provided by different methodologies and analytic approaches. For researchers it is important to know how far diverse imaging modalities tell the same or different stories of the same events, and how far they complement or maybe even contradict each other. Such knowledge is also key in

evaluating neuroimaging studies in other fields, many of which approach the same questions with a variety of methods, albeit during separate investigations. As such, it is quite common that one laboratory investigating, let's say, language perception, uses MEG as the method of choice, and another laboratory in the same field employs fMRI methods. The implicit assumption often is that both laboratories investigate the behaviour of the same neuronal generators. But to what degree is this true? If these laboratories publish divergent results it could be due to differences in the stimulation paradigm or to differences in what their imaging techniques measure. In order to disambiguate the contributions of these factors, we briefly discuss the nature of the MEG, EEG and fMRI signals, and then compare the results from these imaging technologies under nearly identical stimulus conditions.

8.2 Spatial coregistration

One of the prerequisites of multimodal imaging is to define the spatial relationship between the measures and to co-register the data into a common space. A wide-spread strategy is to use a structural MRI scan for anatomical identification and to align all other measurements to the space of the high-resolution MRI image. The MRI and activation images can be transformed into a standard reference system, such as MNI space, and thus be compared across techniques and studies (Fox et al., 1998). By using the standard coordinate system of MNI152 as a common link (Collins et al., 1994), we will compare in this section imaging results directly by transforming all activation images into the individual space of an iEEG subject.

8.2.1.1 *Functional MRI to structural MRI*

In section 3, we describe the co-registration of the functional MRI scans (EPI) into the space of a high-resolution T1 image by using a 12 parameter affine linear registration algorithm (FLIRT, (Jenkinson et al., 2002)). This allows for a direct

visualization of the BOLD signal changes of an individual subject in relation to the detailed neuroanatomy of that subject. For group fMRI studies it is often necessary to transform functional images into a common reference space to gain higher statistical power and to be able to infer from the results to the population in general. It has to be noted, however, that at the same time the spatial specificity of the measure is reduced if data are averaged in space over many subjects.

The most common standard brain space was developed by Talairach and Tournoux (1988) and is based on geometric parcellation of the cortex according to major anatomical landmarks. The purpose of this development was to provide a common reference framework to express relative anatomical positions in any brain. In most neuroimaging publications, activated brain regions are reported in the x, y and z coordinates of the Talairach system. Another common reference system is based on the average of 152 or 301 MRI scans and was developed by the Montreal Neurological Institute (MNI) (Evans et al., 1993). In our fMRI study the results from all individual subjects were coregistered with the 152 MNI template brain using FLIRT before statistical analysis at the group level was carried out. The standard Talairach and MNI brains differ in their shape and thus the coordinate systems and the relation of anatomical structures therein differ as well. To compensate for this difference MNI coordinates can be transformed into Talairach space before reporting them (Duncan et al., 2000; Brett et al., 2002). Thus, our group fMRI results conform spatially to the MNI152 standard space which will later allow us to compare them to MEG and iEEG results.

8.2.1.2 *MEG sensor space to structural MRI*

When studying brain function with MEG, a model of the head is often used to perform source localization and to visualize active areas in relation to brain anatomy. This information can be obtained from a high-resolution structural MRI and requires

the coregistration with the functional data from the MEG, i.e. transformation of the two sets of data into a single co-ordinate system. There are several possibilities to achieve this. We report here the methodology employed in section 4 and 5.

First, the location of the MEG sensor array was defined in relation to fiduciary markers semi-permanently attached to the skull of the subject during the recording session. Afterwards a magnetic 3D space tracking and digitizing system (Polhemus Isotracker) was used to produce a digitization of the shape of the head and note the relative position of the fiduciary markers. The digitized head shape was then coregistered with the MRI via a surface-matching technique and the resulting transformations combined to align the MRI of each individual subject with the MEG sensor array (Adjamian et al., 2004). By transforming the individual subject's MRI into a common reference system, such as MNI152, we have all the information needed to calculate the inverse transform from standard MNI space to MEG sensor space.

8.2.1.3 *iEEG to structural MRI*

Electrodes implanted on the subdural surface of the brain cause MRI image artefacts, so instead a pre-operative high resolution structural MRI scan was acquired for each subject. Detailed photographs of the exposed surface of the brain were taken during the implantation procedure and localization of the grid was achieved by matching the grid's positions with anatomical landmarks. For comparison with other imaging modalities the structural MRI can be transformed into MNI space. Through the inverse transformation, MEG and fMRI results can be overlaid onto the brain of the iEEG patient in its original space alongside the iEEG electrode grid activation. Thus, activation patterns from all 3 imaging modalities can be expressed and compared in the same coordinate system.

8.3 Results

8.3.1 Auditory (A)

Results from activation to auditory stimulation alone (A) showed a high degree of spatial overlap between fMRI and intracranial ERPs (Figure 8-1, rendered brain in top row). Both showed large activation of the middle portion of the STG. This finding is contrasted and validated by no detection of activation on supra-sylvian areas with either method. MEG dipole analysis also revealed a major activation source in STP on HG (A1), whose spatial location coincided with an earliest and maximally responsive electrode (e-36) located above HG in the iEEG subject (pink dot). Both the MEG dipole source waveform at A1 and the ERP at electrode e-36, which are spatially co-located, show first activity between 20-30 ms with a second peak at around 100 ms (middle row, first two plots). This spatio-temporal agreement is a validation of the ECD model, especially at early latencies in A1.

In A1 the two signals start to differ after ~ 170 ms where the dipole activity shows another major peak and the intracranial ERP waveform slowly fades out. An overall reduction in amplitude at STS sites, compared to A1, is seen in both MEG and iEEG; however, the dipole activity is less pronounced (black dots, first two plots bottom row). This suggests that the ECD moments at later latencies might contain signal from other sources not located on STP.

The ERS/ERD frequency spectrum of both the SAM virtual electrode at A1 coordinates and at iEEG electrode e-36 shows a remarkably similar pattern (pink dots, right-most two spectrograms in middle row). Both show a large alpha band ERS immediately after auditory onset, followed by an ERD in the beta band, which in both instances rebounds at 600 ms. A similar, but reduced pattern is seen in STS locations (red dots, lower row spectrograms). However, there are also differences noticeable in higher frequency ranges. Gamma activity shows an ERS immediately after auditory

onset in the MEG data, whereas no such effect is observed in the iEEG channel. The major difference between the frequency spectra of the two methods is a large ERS in omega frequencies above 60 Hz in the intracranial data. The reason this activity is not seen in data recorded outside the scalp might be that muscles around the temporal cranium emit oscillations at this frequency and thereby lower the signal-to-noise ratio of this frequency band, making cortical activity undetectable outside the skull (Barker et al., 1989).

Comparing Imaging modalities

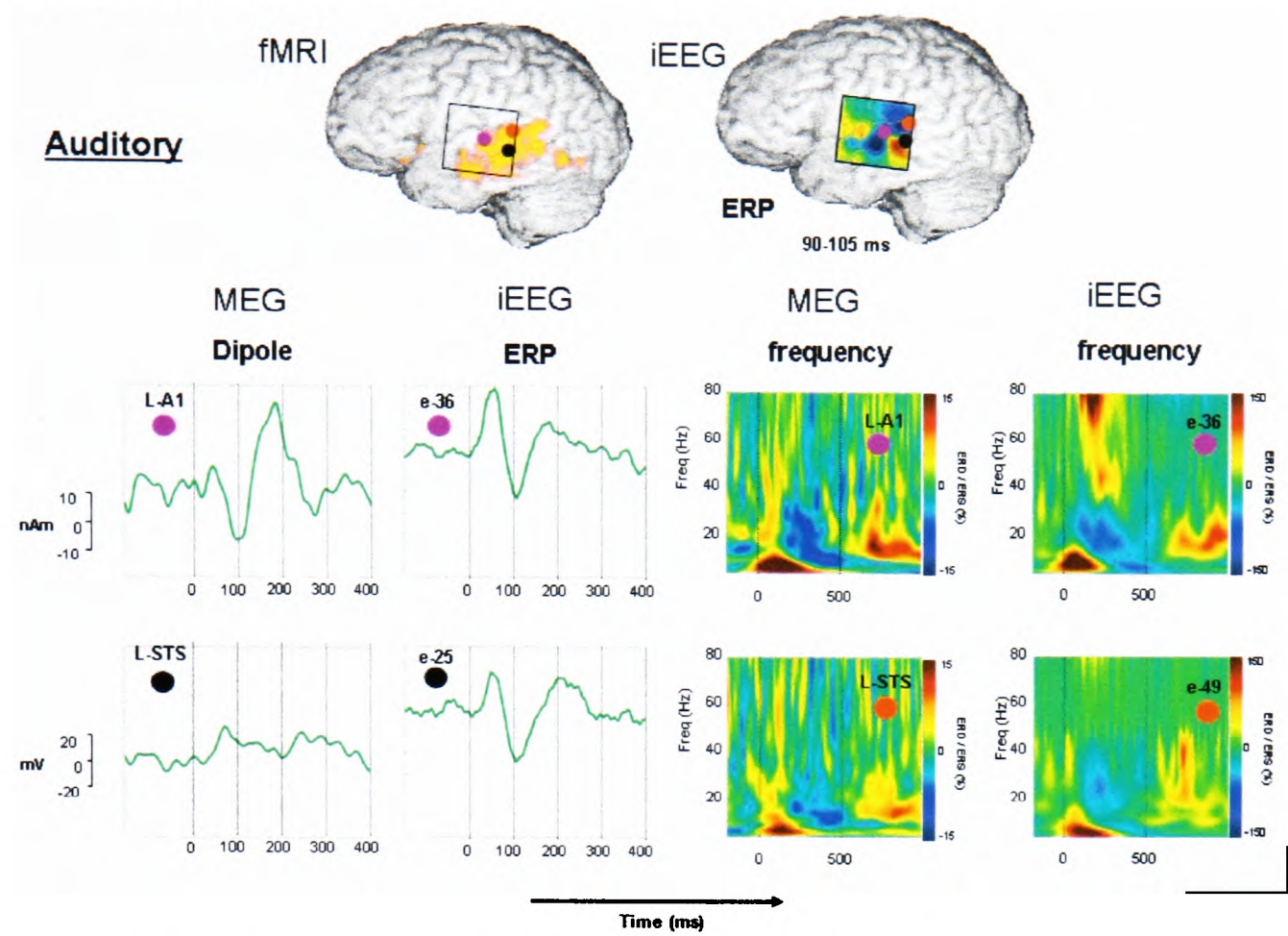


Figure 8-1: Comparing fMRI, MEG & iEEG (S106) results (A)

Comparison of imaging results in response to auditory speech perception.

Top row: Rendered brain of iEEG subject S106 with fMRI activation overlaid from 9 subjects (left) and ERP activity on the electrode grid between 95-105 ms (right). Colored dots refer to the spatial location (transformed into iEEG subject space) of MEG dipole in left auditory cortex (L-A1) and the

corresponding electrode on the grid (e-36, both purple), the location of MEG left STS dipole (L-STs, black), the location of the left STS SAM virtual electrode (L-STs, red) and the corresponding iEEG electrode (e-49, red). The

middle and bottom rows show the time course of MEG dipole moment and iEEG amplitude (first two images) and the frequency spectrogram ranging from 4-80 Hz for MEG virtual electrode locations and iEEG channels.

8.3.2 Auditory-visual (Avcon)

Responses to AVcon stimulation were reduced for all electrophysiological measures but showed a similar response as in the A condition with still good but slightly worse matching of the results from different techniques.

Comparing imaging modalities

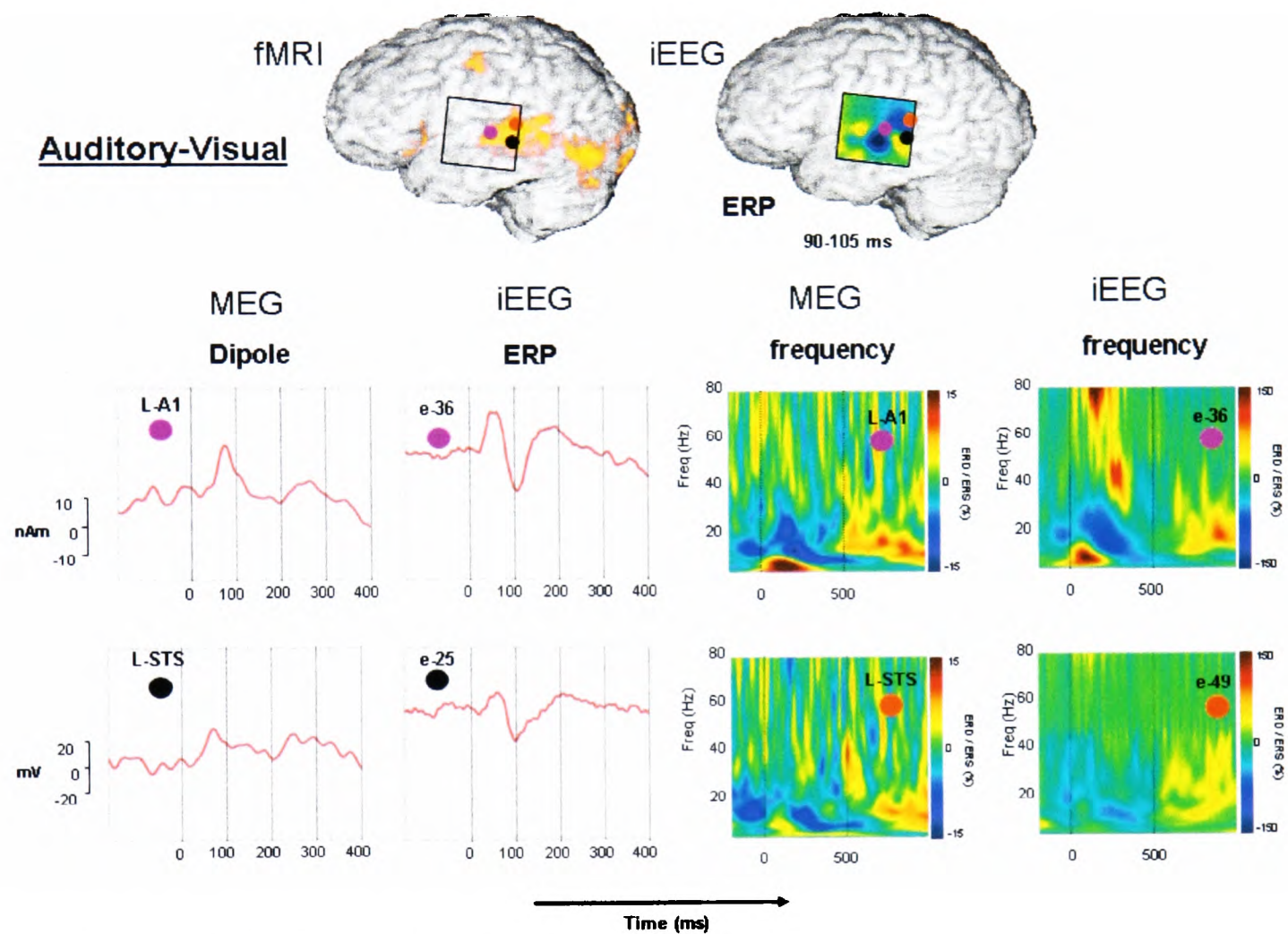


Figure 8-2: Comparing fMRI, MEG & iEEG results (AVcon)
Comparison of imaging results in response to auditory speech perception>
For a detailed description see Figure 8-1.

8.3.3 Electrophysiological validation of hemodynamic integration criteria

Interaction effects are a convincing metric for defining multisensory integration responses. However, recent fMRI studies, including our experiment in section 3, have failed to find superadditive effects to AV stimulation. Evidence

suggests that superadditivity, first used in multisensory electrophysiology, might not be an appropriate measure for detecting multisensory integration with hemodynamic methods (Laurienti et al., 2005). Other investigators have encountered similar problems and have tested a variety of statistical criteria in an attempt to reveal sites of integration (Beauchamp, 2005). However, these investigators had no objective criterion by which to validate their results and to recommend the most appropriate approach. Our data, however, enables us to do this by using intracranial recordings with high spatial specificity to compare electrophysiological interaction effects with different hemodynamic criteria. For this purpose we have spatially transformed fMRI results into the individual space of iEEG subjects and presented the fMRI results alongside the iEEG effects for direct comparison (see Figure 8-3).

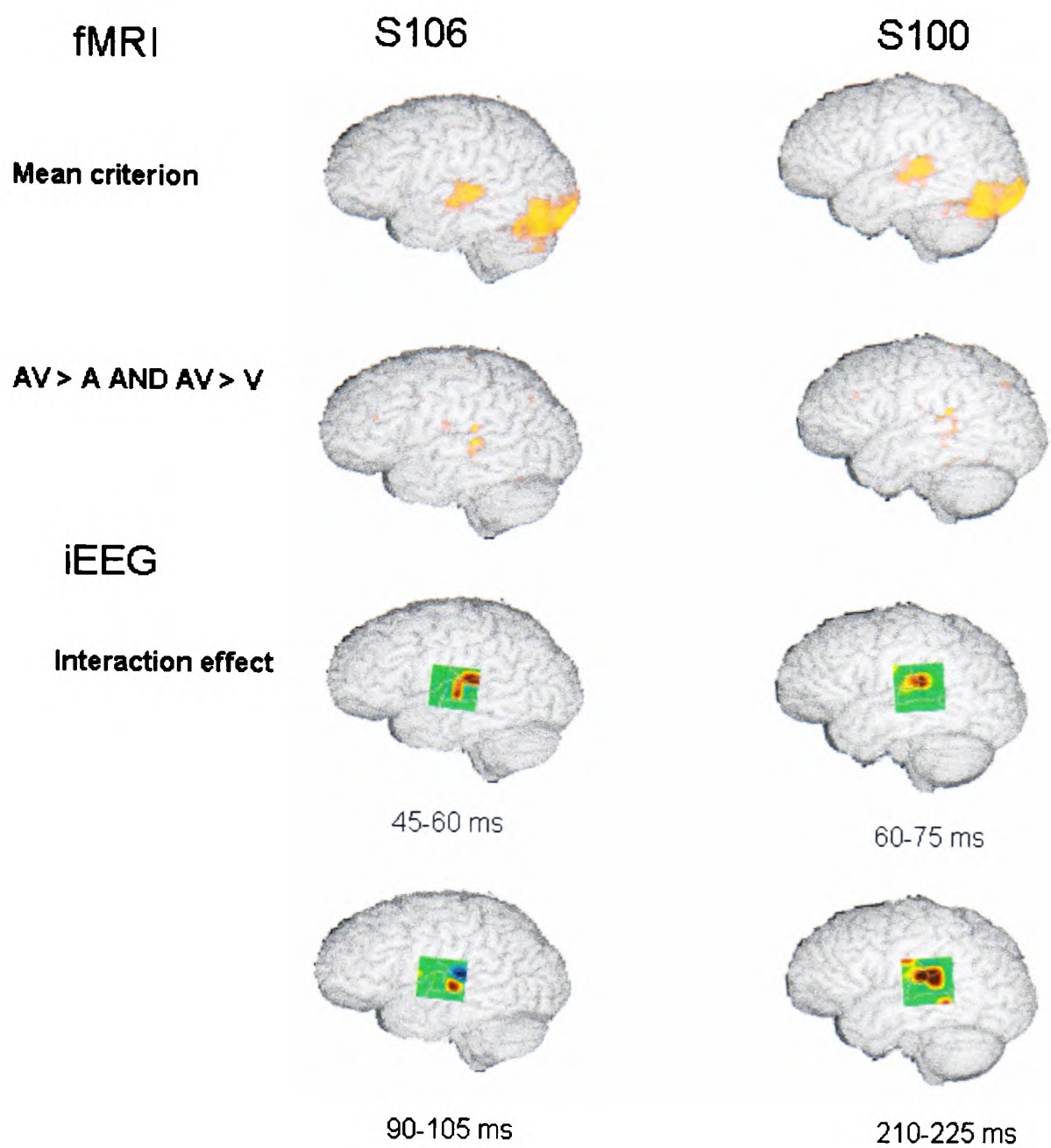


Figure 8-3: Spatial concordance of fMRI & iEEG integration effects
Validating hemodynamic criteria for multisensory integration with iEEG. The two top row shows fMRI responses to the mean criterion ($AV > \text{mean}(A, V)$) and max criterion ($AV > V \cap AV > A$) derived from 21 subjects and overlaid onto the rendered brain of iEEG subjects S100 and S106. Given below are the electrophysiological interaction effect ($AV > (A+V)$) in both iEEG patients at different latencies.

Both the mean and maximum criterion showed activation in the left middle superior temporal cortex and showed good spatial concordance with the electrophysiological interaction effect in both subjects. Essentially, these criteria show better spatial alignment with electrophysiological results than the super-additive effect does, of course. Therefore, we may conclude in favour of the use of the mean and

max criteria for defining multisensory integration based on hemodynamic measurements.

8.4 Discussion

We recorded brain responses to auditory and auditory-visual speech stimulation in 9 subjects using fMRI and MEG. We then transformed the results into the space of an individual iEEG subjects and thus created a means for comparing the spatial relationship between activations from the different neuroimaging methodologies. We found a striking similarity between the location of the cortical hemodynamic response, and the location of evoked activity recorded from an electrode grid over superior temporal sulcus. Moreover, the MEG dipole activity and iEEG ERP waveforms showed a similar temporal pattern; especially during early stimulus processing before 150 ms. The same was true for frequency responses. Even though a different source reconstruction technique was used for frequency analysis than for dipole analysis, the frequency spectrograms still showed a very high degree of similarity with the iEEG spectrograms. A very similar temporal pattern of ERS and ERD was seen in both techniques. These findings validate the fMRI constraint dipole model, especially as far as superior temporal lobe sources are concerned

The comparison between fMRI and iEEG also proved useful in establishing an appropriate metric for detecting brain regions involved in multisensory integration. We found that both the max and mean criterion showed remarkable spatial overlap with electrophysiological interactions derived from both MEG and iEEG. In general, our results support recent findings that the hemodynamic response is tightly linked to underlying neuronal activity (Logothetis et al., 2001), and support the use of combined neuroimaging in the spatio-temporal mapping of neural activity in the brain.

9 Summary & Conclusions

Science is always wrong. It never solves a problem without creating ten more.

-George Bernard Shaw

The work described in this thesis focuses on auditory and visual processing in the brain and the convergence between the two sensory streams during bimodal speech perception. The integration of sensory information is an important issue in understanding how the brain, with its modular structure, fuses information to create the coherent representation of the outside world which we are so accustomed to perceive. We have investigated these neural processes during phonetic and non-phonetic stimulus integration using several invasive and non-invasive neuroimaging techniques and different analysis methods to create a picture as complete as possible. A particular strength of our approach is to tackle the same problem with different imaging techniques and, where possible, doing so with the same subjects. This also allowed us to compare and validate results from the different neuroimaging methods to see in how far their signals tell the same story, an assumption which is often implicit in neuroscience publications but rarely tested. As such, we hope this thesis contributes to the evolving multisensory literature, as well as to the advancement of multimodal neuroimaging, both of which are still relatively novel fields of research.

9.1 Summary of main results

9.1.1 fMRI

The first experimental section (section 3) used whole-brain fMRI to localize hemodynamic changes associated with AV speech processing. The main purpose of this study was to replicate findings from previous block design fMRI experiments on AV speech integration sites obtained in this laboratory, but to do so within an event-related design to allow appropriate integration of the data with MEG and iEEG.

1. Consistent with previous literature, this study found firstly, that visual speechreading activates primary and secondary auditory cortex, albeit less so than auditory speech.
2. Visual cortex was found to be activated by auditory speech alone, a learning effect that increased over time and has not been found before in the context of multisensory speech perception.
3. The areas activated by auditory, visual and auditory-visual speech perception were in the same approximate locations within distinct brain regions, except for auditory-only responsive areas on the anterior STG and visual-only responsive areas in parts of the occipital cortex. However, signal intensities varied between conditions and we found evidence for non-phonetic multisensory integration in left middle STS and right posterior STG and right Tpt based on the max $[AV > A \cap AV > V]$ and mean criteria $[AV > \text{mean}(A, V)]$. Finally, changes in the pattern of functional lateralization in speech-related cortical areas were also observed during AV speech perception.
4. Whereas A and V speech preferentially activated left hemisphere structures, bimodal speech recruited both hemispheres to a similar extend.

9.1.2 MEG dipole analysis:

In order to reveal the time-course of activation in the brain areas involved in multisensory speech perception, the fMRI data were subsequently used to constrain the localisation of electromagnetic sources arising from the same paradigm studied using MEG. FMRI-informed MEG dipole analysis revealed a network of cortical areas which sequentially evaluate and integrate inputs from different modalities.

Results revealed that

1. auditory and visual speech signals interact first at the level of auditory cortex (A1) at ~ 40 ms based on non-phonetic properties and
2. that further integrative processing occurs in the left STG (~ 85 ms) and finally in the left IFG (~ 120 ms).
3. We also found that the same areas mediate phonetic integration starting with STS at 112 ms, IFG at 132 ms and A1 at 232 ms.

9.1.3 MEG frequency analysis:

In order to reveal the underlying oscillatory processes that mediated multisensory convergence, the MEG data were analyzed based on their frequency content. In this section we described a dynamic and complex pattern of oscillatory changes in various areas of cortex over a wide range of frequencies in response to auditory-visual speech perception. We found

1. modulation of a prominent auditory alpha band ERS in A1 by the presence of a visual stimulus, alongside a similar modulation in the gamma band.
2. Visual modulation of alpha band activity was also seen in the STS and IFG, both before and after auditory onset.
3. Evidence for speech-specific integration of AV signals was found in left STS and right A1.

9.1.4 IEEG ERP analysis:

After identifying putative multisensory integration sites and their time-course in the superior temporal lobe, we focussed on this area using the high temporal and spatial resolution of intracranial EEG.

1. Multisensory integration based on non-phonetic stimulus features was observed immediately above HG and area Tpt at the earliest stage of cortical processing around 26 ms, and at 85 ms over STG.
2. Subsequent time windows show a complex pattern of AV interactions on the superior temporal lobe.
3. We also found evidence for an involvement of the somatosensory and motor systems in matching and integrating phonetic auditory and visual speech signals after 170 ms.

9.1.5 IEEG frequency analysis:

Frequency analysis of the iEEG data shows a visual influence on auditory stimulus processing in auditory cortex above HG and area Tpt, and along the length of the middle STG over a wide range of frequencies. Most notably, the prominent alpha ERS after auditory onset was diminished by visual influence, as well as in the beta frequency range. We found large synchronizations in the high frequency omega band which were differentially modulated in an auditory responsive area on STG by articulatory and non-articulatory visual stimuli before auditory onset.

9.1.6 Combining neuroimaging methods

Using a multimethod approach offers many advantages. Functional MRI results provided a plausible and biologically meaningful constraint for the MEG source solution. The appropriateness of the fMRI-informed source model could be confirmed by iEEG. Using frequency analysis of electrophysiological data we were

able to describe oscillatory brain dynamics underlying multisensory processing, and to detect crossmodal modulations that were undetectable by ERP techniques. By comparing different metrics for identifying multisensory integration in hemodynamic data with electrophysiological interaction effects we were able to isolate the mean and max criterion as the most appropriate statistical approaches. We also found support for the implicit assumption that the signals of different imaging techniques reflect neural activity of common sources.

9.2 A spatio-temporal model

Appendix 2 shows a tabular summary of results from the experimental sections of this thesis. It shows the involvement of a selection of cortical areas in AV speech processing based on results from the different neuroimaging methods. Where possible, the timing of activations is given. The table also shows if and when an area showed multisensory effects.

Based on this table and other results from this thesis, and together with prior knowledge of structural and functional brain anatomy we have compiled a spatio-temporal model that, we think, best explains our data on AV speech integration. The model specifically takes into account the differential temporal onset of the audio and visual stimulation and assumes that visual 'priming' is the major source of the early modulation observed in auditory cortex.

9.2.1 Spatio-temporal progression of neural activity

Figure 9-1 shows a model of AV speech processing based on our results. Here, during visual lipreading, activation spreads from primary and secondary visual cortex to middle and inferior temporal areas via visual motion areas and has reached IFG and motor areas at around 250 ms after lip movement onset (Nishitani and Hari, 2002). During bimodal AV speech stimulation with our stimuli, visual information about the

sequence of lip movements enters visual cortex via the thalamus starting at around 350 ms before auditory onset and proceeds via the ventral visual pathways to motion-sensitive area MT from where it is relayed along the visual object recognition pathway to ITG. At around - 250 ms, probably through direct projections via the uncinate fasciculus (Ungerleider et al., 1989; Webster et al., 1994) information is relayed to Broca's area in IFG, as MEG frequency data show. From there information reaches the STS through the arcuate fasciculus (Matsumoto et al., 2004) at around - 200 ms. The STS then projects to HG, where it influences the processing of auditory information that arrives from the thalamus at around 20 ms after auditory stimulus onset, the first integration of A and V signals, as shown with both MEG and iEEG. From HG, the modulated auditory-visual information enters STS where further crossmodal integration occurs. Auditory cortex and STS have strong and reciprocal connections (Kaas and Hackett, 2000) and there is evidence in our iEEG data for sustained functional influence between these two areas. Integrated AV information from HG and STS then, via STS, enters IFG at around 118 ms. From there, information reaches S1/M1 at around 130 ms.

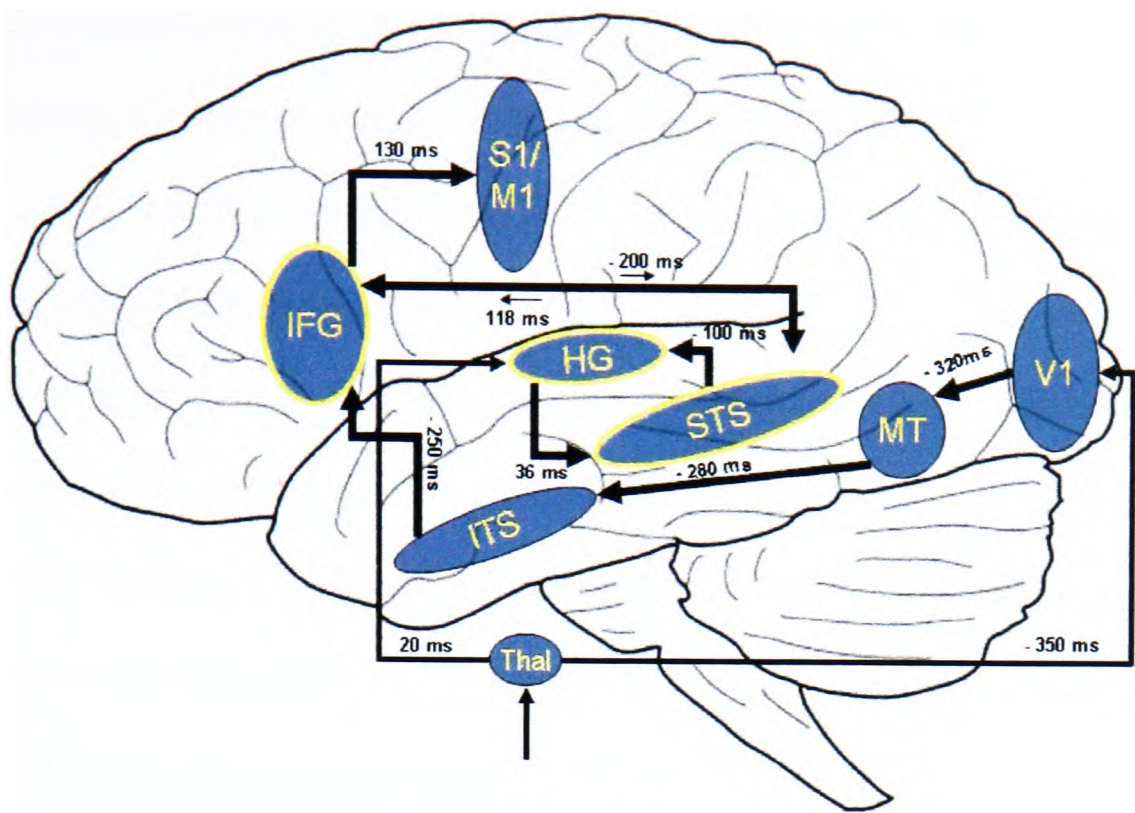


Figure 9-1: A spatio-temporal model of AV processing

Schematic drawing of the brain with brain areas involved in multisensory speech perception: Thalamus (Thal), visual cortex (V1), area MT+ (MT), inferior temporal sulcus, superior temporal sulcus (STS), Heschl's gyrus (HG), inferior frontal gyrus (IFG and primary somatosensory and motor cortices (S1/M1). Timing of activity is given alongside the arrows in milliseconds (ms) relative to auditory stimulus onset. Yellow-rimmed areas are involved in integrating auditory and visual information.

9.2.2 Visual precedence in AV speech processing

The present results suggest that theories of neural speech perception need to consider the influence of preceding visual speech components on auditory speech processing. This finding is complemented by earlier behavioural studies. It has been demonstrated that the perception of an acoustic event can be context dependent, e.g. phonetic identification of CV syllables is modulated by a preceding auditory stimulus, e.g. /ga/ and /da/ syllables are more often identified as /ga/ if preceded by /a/, and more often identified as /da/ if preceded by /a/ (Fowler et al., 2000). Showing a visually moderated context effect, this effect was also seen if modulation of phonetic

identification depended on a preceding visual cue (Fowler et al., 2000; Green and Norrix, 2001). Previous electrophysiological studies have shown early speech interaction effects in response to AV stimuli with asynchronous onsets (Klucharev et al., 2003; van Wassenhove et al., 2005), but not when AV stimuli were artificially synchronized (Mottonen et al., 2004). Our neuronal evidence, that the preceding visual signal primes areas before auditory onset and modulates subsequent auditory processing at early stages of sensory analysis is very much in agreement with such a model. Future studies should investigate the exact role of these visual precedents on neuronal priming and integration during AV speech perception. Of particular interest would be to know if their influence is speech specific or reflects general multisensory integration phenomena.

9.2.3 Auditory cortex during speechreading

Results from virtually all methods and analysis techniques support a role for primary auditory cortex in AV speech integration. Our findings with fMRI show visual recruitment of auditory cortex. These results are complemented by findings of multisensory integration at the first level of auditory cortex based on MEG dipoles in A1 (at 36 ms) and iEEG recordings over medial HG (at 26 ms). The subsequent involvement of STG in synthesizing sensory information at ~85 ms was also supported by both electrophysiological methods, as was the involvement after 120-170 ms of supra-sylvian sites traditionally associated with the motor system. Frequency analysis also revealed that vision modulates STG and IFG immediately before auditory onset.

We also found that previous findings of auditory cortex activation during silent speechreading can be partly explained by a priming role of the visual signal during bimodal speech perception. In addition to early modulation, we also found evidence for multisensory integration in auditory cortex at later stages in processing.

i.e., phonetic categorization. Together with findings from Miller and D'Esposito (2005) who found that hemodynamic responses in HG are higher when AV stimulation results in a perceptual fusion compared to the perception of discrete events, our findings suggest that visual influences on auditory cortex proceed in multiple, sequential steps.

That timing is an important factor in computing sensory signals into perception and action might explain why we failed to see any fMRI changes during incongruent vs congruent AV conditions, but do so with MEG and iEEG. Our hemodynamic results show that congruent and incongruent speech activates the same areas in the brain, and our electrophysiological results indicate that they do so at different times. But since fMRI represents the average signal over several seconds, fMRI only measures the net response and is insensitive to these timing issues. These findings suggest that the sequential nature of multisensory effects, therefore, might constitute an important mechanism in itself. Visual speech activates auditory cortex very early, which might be a mechanism to alert the area to a relevant incoming stimulus and even bias processing towards the visual representation, for example in the McGurk effect. At later stages, auditory cortex, STG and IFG are also involved in more higher-level integrative processes, such as fusing bimodal signals with the same phonological content. We expect that such hierarchical multisensory integration is not only restricted to speech, but that similar mechanisms are at work during the crossmodal integration of all kinds of object features that are represented in two, or more, modalities.

The recruitment of primary auditory cortex by visual cues violates a long-standing concept of how primary sensory areas function, i.e. that they only respond to unisensory inputs. Our results, together with other studies showing multisensory convergence at low-level primary sensory areas (Calvert et al., 1999; Giard and

Peronnet, 1999; Foxe et al., 2002; Schroeder and Foxe, 2002; Pekkola et al., 2005) (Ghazanfar et al., 2005) make it increasingly apparent that the traditional parcellation of cortex into sensory-specific and sensory-association areas is no longer an adequate model of the organisation of sensory information in the brain. The evolutionary advantage offered by such crossmodal convergence in primary sensory cortex might be better spatial localization when vision, the sensory system with higher spatial acuity (Bertelson and Radeau, 1976), influences auditory processing. Or better segmentation of environmental sequences when hearing, the system with better temporal precision, intervenes visual cortex to influence visual processing (Morein-Zamir et al., 2003; Shams et al., 2005). Crossmodal signals could also increase the sensitivity of primary areas, thereby conveying a processing advantage without changing the nature and percept of the incoming signal.

9.2.4 Feedback or feedforward?

Earlier accounts of multisensory convergence in primary sensory areas have assumed this effect to reflect feedback modulations that occur after multisensory processing in higher-order multisensory areas (Calvert, 2001), whereas some recent views also support a direct influence of feedforward connections (Foxe and Schroeder, 2005).

Is the modulation of auditory cortex that we see in our data a feedforward or feedback mechanism? It is not a feedforward mechanism in the sense that there is a direct modulation from primary visual cortex. However, it is a feedforward modulation in the sense that visual information ascends the visual speech processing hierarchy in a feedforward manner first, and then innervates auditory cortex. However, feedback projections into auditory cortex also exist at various later stages in AV processing, and these probably serve different functions. Early feedforward projections may be a general alerting or biasing mechanism to facilitate non-specific

multisensory processing, whereas later feedback innervations may serve more specific functions, such as integration across sensory modalities based on common higher-order features.

9.2.5 A proposal for the categorization of multisensory effects

9.2.5.1 *Different mechanisms serve crossmodal effects*

It is becoming more and more apparent that multisensory integration in the brain is the rule rather than the exception, and that it may represent a fundamental property of brain function. Now that the case for multisensory integration has been made unequivocally in the literature, it is becoming evident that there is a need for multisensory research to become increasingly more specialized and complex. For example, integration effects are most often categorized based on sensory modality and stimulus features (space, object identity, etc.) However, it may be useful to categorize convergence across sensory modalities in terms of general brain organization and functional specificity. As our data have shown there are different routes by which crossmodal influence can exert itself, even within the same task. In order to understand multisensory processing beyond binding of stimulus features, these types of crossmodal modulations have to be categorized and systematically investigated; an effort which is independent of what sensory modality or particular task is studied. Based on our findings we suggest that there are at least 3 different crossmodal mechanisms that differ in their functional role and are mediated by different functional architecture and neuronal behaviour.

The model we are proposing to explain our data is based on the context in which auditory information is processed. It assumes that perceptual analysis of input is not purely stimulus driven but based on expectation derived from previous experience, which acts between sensory modalities. This previous experience has (i)

long-term (ii) short-term and (iii) 'online' consequences and directly influences the architecture and behaviour of cortical networks.

(i) Long-term consequences are based on properties such as spatial localization and temporal coincidence whose crossmodal alignment is shaped early in development and changes only slowly (Lewkowicz and Turkewitz, 1981; Lewkowicz, 1992; Lewkowicz, 2000), such as crossmodal spatial maps in SC (Stein and Meredith, 1993). Here, multisensory input is compared with existing knowledge shaped by experience throughout development. This level of experience is the most robust and acquired over a long period of time and shaped especially during critical periods in development. These consequences can be mediated by 'hard-wired' structural connections built, for example, during stages of activity-dependent synaptic pruning.

(ii) Short-term consequences are seen, for example, when arbitrary crossmodal stimuli are paired and the result leads to multisensory effects, such as visual cortex activated by sound through prior pairing of AV stimuli (McIntosh et al., 1998; McIntosh et al., 1998) (Jaencke and Shah, 2004). Here, multisensory input is compared with existing knowledge shaped by recent experience that is in agreement with immediate intrinsic goals and motivational stages.

(iii) Online consequences are those that are mediated by immediately present or preceding events within the time-range of the task or trial. One example is endogenous attention, where, for example, a preceding cue instructs subjects to shift their attention, which then influences stimulus processing (Coull et al., 2000; Griffin et al., 2002). Here, multisensory input is modulated by transient changes in neural behaviour where a sensory stimulus provides a short-lasting influence on the processing of a stimulus from another modality. This level of experience is the shortest-lasting of all and only survives the length of one trial. It differs from the other levels as it depends partly on crossmodal associations formed at levels (i) and (ii), but

unlike the short-term level, its principle is all-or-none and does not change its quality over time.

In the studies presented in this thesis, long-term consequences are based on shared crossmodal stimulus features such as place and time, as well as speech-specific properties between phonetic and visemic representations that have been formed through associations from early childhood by exposure to face-to-face conversations. We interpret the initial crossmodal recruitment of visual cortex that we found in our data as such an effect. Structural imaging methods, such as diffusion tensor imaging (DTI) (Behrens et al., 2003), may be an appropriate method for testing the anatomical effects of long-term consequences. Of special interest here would be comparing individuals with different auditory or visual ability and training, whose recruitment of the sensory systems and the balance between them might differ from the normal population, such as musicians or people with visual acuity deficits. The age of onset of their sensory training or decrease in sensory ability would be an important variable in such experiments.

The increase in visual cortex activity we see over time (fMRI section, Figure 3-14) is a result of short-term experience. It is the consequence of a progressive strengthening between A and V stimuli association, a process that has immediate task relevance beyond the basic stimulus associations formed at level (i). The underlying functional changes in neuronal network behaviour of such instances of short-term plasticity could be assessed with fMRI, MEG and EEG by using experimental manipulations of crossmodal stimulus associations, such as pairing arbitrary sensory stimuli. An interesting experiment on functional reorganization would involve a short-term readjustment of visual and auditory space. This could be invoked by the use of prism glasses which shift the whole visual field in space and subsequent training of associating the 'new' visual space with the auditory space.

The modulation of auditory cortex by visual speech we saw in our experiments is likely to be an online consequence of AV speech processing. Here, the preceding visual stimulus is analyzed based on its visemic representation and then biases subsequent processing in auditory cortex within the same trial. We showed that areas of STS already distinguish between visemic representation and meaningless facial movements 200 ms before auditory onset, and that auditory processing interacts with visual speech at the earliest stages of processing in auditory cortex. This suggests that the visual context in which the auditory stimulus is presented leads to a reinterpretation of the represented auditory item. This is not a result of a top-down influence in the traditional, cognitive sense as it happens automatically, but rather a direct sensory-sensory influence between two modalities that is achieved without attentional modulation. One possibility of testing the degree and exact mechanisms of visual priming on auditory areas would be to keep visual and auditory onset asynchrony the same, while systematically varying features of the visual stimulus. For example, these could include different visemic stimuli which change the auditory perception to varying degrees. Stimuli could also be constructed where visual onset remains stable while the timing of the defining visual lip movement feature (place of articulation, voice onset time) occurs later at different times. Of special interest here would be visemic stimuli whose modulating characteristics occur after auditory onset (negative voice onset time).

The mechanisms we propose are not only specific to multisensory processing but reflect the functional organization of the brain in general. As such, they highlight the idea that multisensory processing recruits the same structural and functional circuitry and mechanisms as other brain processes. This view emphasizes the idea that the brain is a multisensory processor and that integration of across sensory modalities is the rule rather than the exception.

9.2.5.2 *Changes in neuronal network behaviour*

The interpretation of our data and the finding of online multisensory modulation agree very well with a dynamic network model of brain oscillations and synchrony proposed by (Engel et al., 2001). Their model champions the idea that top-down modulation of sensory information processing is mediated by context-dependent changes in the temporal patterning of neuronal responses, especially their synchronization. The model supports a constructive nature of perception that is driven by prior experience, which is manifested in neuronal architecture and behaviour, and is influenced by context. Contextual modulation is understood as the changes in neuronal response based on intrinsic factors, such as attention, and extrinsic factors, such as prior stimulation (e.g. a preceding crossmodal stimulus). Such factors are dependent on various memory modules and exert their influence mainly through temporal binding within large-scale and local neural networks (Singer, 1994; Varela et al., 2001). Establishing a relationship between the different types of multisensory consequences (especially in ii & iii) and activity patterns in specific frequency bands would be a large step towards deeper understanding of multisensory processing.

9.3 Suggestions for further studies

Many new questions have arisen during the execution and analysis of the present experiments. In the following subsection we want to mention some of the questions that arose that seem to us most important and interesting, and suggest experiments that can provide the answers.

1. In how far is crossmodal activation of primary sensory cortex dependent on attention and crossmodal context? Crossmodal context refers to the particular cognitive set in which the subject is engaged, i.e. “This is a multisensory experiment. There will be auditory, visual and some bimodal signals to which I have to respond. Some of these signals appear immediately after each other and some belong together, and some don’t.” as opposed to “This is an auditory experiment. They are playing sounds to which I have to respond. They told me to keep my eyes on this red cross, but otherwise there is nothing to look at”. Differences between block design and event-related experiments suggest that the crossmodal context has an effect on the neuronal strategies applied by the brain. We hypothesise that both attention and context play a crucial, but distinct role during multisensory tasking. One way to assess effects of attention and crossmodal context is by devising a multiplex design with attention to modality (attend vs non-attend) and different task-related contexts (multisensory vs unisensory) or using mixed-designs (Donaldson, 2004). We hypothesise the following outcome. Visual cortex: Attention causes medium activation whereas non-attended stimulation causes no signal change. In the context of a unisensory experiment the signal amplitude remains steady over time, whereas in a multisensory context the response increases with time. As for the auditory cortex, non-attended visual speech stimuli will nevertheless recruit HG, but more so during

attended trials. The crossmodal context will have a small positive effect on the response, but this will not change as a function of time.

2. What types of multisensory processes are speech-specific? Is there such a thing as speech-specific multisensory integration? These experiments would involve the use of carefully controlled speech and non-speech stimuli, matched by stimulus complexity, to delineate speech-specific and unspecific AV interactions. It would also be useful to include actual measures of perceptual fusion of the two sensory streams which then can be used as a covariate to show areas which mediate the perception of sensory integration (Bushara et al., 2003) (Miller and D'Esposito, 2005).

Bibliography

- Adjamian, P., G. R. Barnes, A. Hillebrand, I. E. Holliday, K. D. Singh, P. L. Furlong, E. Harrington, C. W. Barclay and P. J. Route (2004). "Co-registration of magnetoencephalography with magnetic resonance imaging using bite-bar-based fiducials and surface-matching." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **115**(3): 691-8.
- Ahlfors, S., F. H. Lin, T. Witzel, A. M. Dale, J. W. Belliveau and E. Halgren (2002). Characterization of MEG/EEG signals generated by simultaneous sources. Biomag 2002: 13th International Conference on Biomagnetism, Berlin, VDE Verlag.
- Ahlfors, S. P., G. V. Simpson, A. M. Dale, J. W. Belliveau, A. K. Liu, A. Korvenoja, J. Virtanen, M. Huotilainen, R. B. Tootell, H. J. Aronen and R. J. Ilmoniemi (1999). "Spatiotemporal activity of a cortical network for processing visual motion revealed by MEG and fMRI." J Neurophysiol **82**(5): 2545-55.
- Allman, J. M. and J. H. Kaas (1971). "Representation of the visual field in striate and adjoining cortex of the owl monkey (*Aotus trivirgatus*)." Brain Res **35**(1): 89-106.
- Amaro, E., Jr., S. C. Williams, S. S. Shergill, C. H. Fu, M. MacSweeney, M. M. Picchioni, M. J. Brammer and P. K. McGuire (2002). "Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects." Journal of magnetic resonance imaging JMRI **16**(5): 497-510.
- Amedi, A., G. Jacobson, T. Hendler, R. Malach and E. Zohary (2002). "Convergence of visual and tactile shape processing in the human lateral occipital complex." Cereb Cortex **12**(11): 1202-12.

- Amedi, A., R. Malach, T. Hendler, S. Peled and E. Zohary (2001). "Visuo-haptic object-related activation in the ventral visual pathway." Nature Neuroscience **4**(3): 324-30.
- Aoki, F., E. E. Fetz, L. Shupe, E. Lettich and G. A. Ojemann (1999). "Increased gamma-range activity in human sensorimotor cortex during performance of visuomotor tasks." Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology **110**(3): 524-37.
- Ashmore, J. F. (1987). "A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier." J Physiol **388**: 323-47.
- Aston, U. o. "Introduction to MEG." 2005, from <http://www.aston.ac.uk/lhs/research/facilities/meg/>.
- Avikainen, S., N. Forss and R. Hari (2002). "Modulated activation of the human SI and SII cortices during observation of hand actions." Neuroimage **15**(3): 640-6.
- Bamiou, D.-E., F. E. Musiek and L. M. Luxon (2003). "The insula (Island of Reil) and its role in auditory processing: Literature review." Brain Research Reviews **42**(2): 143.
- Bamiou, D. E., F. E. Musiek and L. M. Luxon (2003). "The insula (Island of Reil) and its role in auditory processing. Literature review." Brain Research. Brain Research Reviews **42**(2): 143-54.
- Banati, R. B., G. W. Goerres, C. Tjoa, J. P. Aggleton and P. Grasby (2000). "The functional anatomy of visual-tactile integration in man: a study using positron emission tomography." Neuropsychologia. **38**(2): 115-124.
- Bandettini, P. A. and E. C. Wong (1997). "Magnetic resonance imaging of human brain function. Principles, practicalities, and possibilities." Neurosurg.Clin.N.Am. **8**(3): 345-371.
- Barker, G. R., D. G. Wastell and A. J. Duxbury (1989). "Spectral analysis of the masseter and anterior temporalis: An assessment of reliability for use in the clinical situation." Journal of Oral Rehabilitation **16**(3): 309.
- Barracough, N. E., D. Xiao, C. I. Baker, M. W. Oram and D. I. Perrett (2005). "Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions." J Cogn Neurosci **17**(3): 377-91.

- Barth, D. S., N. Goldberg, B. Brett and S. Di (1995). "The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex." Brain Res. **678**(1-2): 177-190.
- Basar, E., C. Basar-Eroglu, S. Karakas and M. Schurmann (1999). "Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG?" Neurosci Lett **259**(3): 165-8.
- Basar, E., T. Demiralp, M. Schurmann, E. C. Basar and A. Ademoglu (1999). "Oscillatory brain dynamics, wavelet analysis, and cognition." Brain Lang. **66**(1): 146-183.
- Baylis, G. C., E. T. Rolls and C. M. Leonard (1987). "Functional subdivisions of the temporal lobe neocortex." J.Neurosci. **7**(2): 330-342.
- Beauchamp, M. S. (2005). "See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex." Curr Opin Neurobiol **15**(2): 145-53.
- Beauchamp, M. S. (2005). "Statistical criteria in fMRI studies of multisensory integration." Neuroinformatics **3**(2): 93-114.
- Beauchamp, M. S., R. W. Cox and E. A. DeYoe (1997). "Graded effects of spatial and featural attention on human area MT and associated motion processing areas." J.Neurophysiol. **78**(1): 516-520.
- Beauchamp, M. S., K. E. Lee, B. D. Argall and A. Martin (2004). "Integration of auditory and visual information about objects in superior temporal sulcus." Neuron **41**(5): 809-23.
- Bechtel, W. and R. S. Stufflebeam (2001). Epistemic issues in procuring evidence about the brain: The importance of research instruments and techniques. Philosophy and the Neurosciences: A Reader. W. Bechtel, P. Mandik, J. Mundale and R. S. Stufflebeam. Oxford, Blackwell Publishers.
- Bechtereva, N. P., Y. G. Abdullaev and S. V. Medvedev (1992). "Properties of neuronal activity in cortex and subcortical nuclei of the human brain during single-word processing." Electroencephalogr Clin Neurophysiol **82**(4): 296-301.
- Behrens, T. E., H. Johansen-Berg, M. W. Woolrich, S. M. Smith, C. A. Wheeler-Kingshott, P. A. Boulby, G. J. Barker, E. L. Sillery, K. Sheehan, O. Ciccarelli, A. J. Thompson, J. M. Brady and P. M. Matthews (2003). "Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging." Nature neuroscience **6**(7): 750-7.

- Belin, P., R. J. Zatorre and P. Ahad (2002). "Human temporal-lobe response to vocal sounds." Brain research. Cognitive brain research **13**(1): 17-26.
- Belin, P., R. J. Zatorre, R. Hoge, A. C. Evans and B. Pike (1999). "Event-related fMRI of the auditory cortex." Neuroimage **10**(4): 417-29.
- Belin, P., R. J. Zatorre, P. Lafaille, P. Ahad and B. Pike (2000). "Voice-selective areas in human auditory cortex." Nature **403**(6767): 309-312.
- Belin, P., R. J. Zatorre, P. Lafaille, P. Ahad and B. Pike (2000). "Voice-selective areas in human auditory cortex." Nature **403**(6767): 309-12.
- Benevento, L. A., J. Fallon, B. J. Davis and M. Rezak (1977). "Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey." Experimental Neurology **57**(3): 849-872.
- Benevento, L. A. and J. H. Fallon (1975). "The ascending projections of the superior colliculus in the rhesus monkey (*Macaca mulatta*)." J Comp Neurol **160**(3): 339-61.
- Berger, H. (1930). "Ueber das Elektroenkephalogram des Menschen II." J Psychol neurol **40**: 160-179.
- Bernstein, I.-H., M.-H. Clark and B. Edelstein (1969). "Intermodal effects in choice reaction time." Journal of Experimental Psychology **81**(2): 405-407.
- Bernstein, L. E., E. T. Auer, Jr., J. K. Moore, C. W. Ponton, M. Don and M. Singh (2002). "Visual speech perception without primary auditory cortex activation." Neuroreport **13**(3): 311-5.
- Bernstein, L. E., E. T. Auer and J. K. Moore (2004). Audiovisual speech binding: Convergence or association? Handbook of Multisensory Processing. G. A. Calvert, C. Spence and B. E. Stein. Cambridge, MIT Press.
- Bertelson, P. and M. Radeau (1976). "Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola." Perception and Psychophysics. Vol **19**(6): 531-535.
- Bertelson, P. and M. Radeau (1981). "Cross-modal bias and perceptual fusion with auditory-visual spatial discordance." Percept. Psychophys. **29**(6): 578-584.
- Bilecen, D., K. Scheffler, N. Schmid, K. Tschopp and J. Seelig (1998). "Tonotopic organization of the human auditory cortex as detected by BOLD-FMRI." Hear Res **126**(1-2): 19-27.

- Binder, J. R., J. A. Frost, T. A. Hammeke, P. S. Bellgowan, J. A. Springer, J. N. Kaufman and E. T. Possing (2000). "Human temporal lobe activation by speech and nonspeech sounds." Cereb Cortex **10**(5): 512-28.
- Binder, J. R., S. J. Swanson, T. A. Hammeke, G. L. Morris, W. M. Mueller, M. Fischer, S. Benbadis, J. A. Frost, S. M. Rao and V. M. Haughton (1996). "Determination of language dominance using functional MRI: a comparison with the Wada test." Neurology **46**(4): 978-984.
- Bizley, J. K., F. R. Nodal and A. J. King (2004). Investigating the functional organisation of ferret auditory cortex using multi-electrode recordings. Association for Research in Otorinolaryngolgy.
- Blank, S. C., S. K. Scott, K. Murphy, E. Warburton and R. J. Wise (2002). "Speech production: Wernicke, Broca and beyond." Brain **125**(Pt 8): 1829-38.
- Bookheimer, S. (2002). "Functional MRI of language: new approaches to understanding the cortical organization of semantic processing." Annu Rev Neurosci **25**: 151-88.
- Bordi, F. and J. E. LeDoux (1994). "Response properties of single units in areas of rat auditory thalamus that project to the amygdala. II. Cells receiving convergent auditory and somatosensory inputs and cells antidromically activated by amygdala stimulation." Exp Brain Res **98**(2): 275-86.
- Borgmann, C., B. Ro, R. Draganova and C. Pantev (2001). "Human auditory middle latency responses: influence of stimulus type and intensity." Hearing Research **158**(1-2): 57.
- Borgmann, C., B. Ross, R. Draganova and C. Pantev (2001). "Human auditory middle latency responses: influence of stimulus type and intensity." Hear Res **158**(1-2): 57-64.
- Brandt, M. E., B. H. Jansen and J. P. Carbonari (1991). "Pre-stimulus spectral EEG patterns and the visual evoked response." Electroencephalogr Clin Neurophysiol **80**(1): 16-20.
- Brazdil, M., M. Dobsik, M. Mikl, P. Hlustik, P. Daniel, M. Pazourkova, P. Krupa and I. Rektor (2005). "Combined event-related fMRI and intracerebral ERP study of an auditory oddball task." Neuroimage **26**(1): 285-93.
- Bremmer F., A., N. J. Schlack, O. Shah, M. Zafiris, Kubischik, K. Hoffmann, K. Zilles and G. R. Fink (2001). "Polymodal motion processing in posterior

- parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys." Neuron **29**(1): 287-296.
- Bressler, S. L. (1995). "Large-scale cortical networks and cognition." Brain research. Brain research reviews **20**(3): 288-304.
- Brett, M., I. S. Johnsrude and A. M. Owen (2002). "The problem of functional localization in the human brain." Nature reviews. Neuroscience **3**(3): 243-9.
- Broca, P. (1961). "Perte de la parole; ramolissement chronique et destruction partielle du lobe antérieur gauche de cerveau." Bulletins de la Société d'Anthropologie de Paris **2**: 235-238.
- Brodmann, K. (1908). "Beitraege zur histologischen Lokalisierung der Grosshirnrinde: VI Mitteilung. Die Cortexgliederung des Menschen." Journal of Psychological Neurology **10**(231-246).
- Brosch, M., E. Budinger and H. Scheich (2002). "Stimulus-related gamma oscillations in primate auditory cortex." J Neurophysiol **87**(6): 2715-25.
- Bruce, C., R. Desimone and C. G. Gross (1981). "Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque." Journal of Neurophysiology **46**(2): 369-384.
- Buccino, G., F. Binkofski, G. R. Fink, L. Fadiga, L. Fogassi, V. Gallese, R. J. Seitz, K. Zilles, G. Rizzolatti and H. J. Freund (2001). "Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study." Eur J Neurosci **13**(2): 400-4.
- Buckner, R. L., P. A. Bandettini, K. M. O'Craven, R. L. Savoy, S. E. Petersen, M. E. Raichle and B. R. Rosen (1996). "Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging." Proc Natl Acad Sci U S A **93**(25): 14878-83.
- Buckner, R. L., M. E. Raichle, F. M. Miezin and S. E. Petersen (1996). "Functional anatomic studies of memory retrieval for auditory words and visual pictures." J Neurosci **16**(19): 6219-35.
- Burgess, A. P. and J. H. Gruzelier (1997). "Short duration synchronization of human theta rhythm during recognition memory." Neuroreport **8**(4): 1039-42.
- Bushara, K. O., J. Grafman and M. Hallett (2001). "Neural correlates of auditory-visual stimulus onset asynchrony detection." J Neurosci **21**(1): 300-4.

- Bushara, K. O., T. Hanakawa, I. Immisch, K. Toma, K. Kansaku and M. Hallett (2003). "Neural correlates of cross-modal binding." Nature neuroscience **6**(2): 190-5.
- Callan, D. E., A. M. Callan, C. Kroos and E. Vatikiotis Bateson (2001). "Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study." Cognitive Brain Research **10**(3): 349-53.
- Callan, D. E., J. A. Jones, K. Munhall, A. M. Callan, C. Kroos and E. Vatikiotis-Bateson (2003). "Neural processes underlying perceptual enhancement by visual speech gestures." Neuroreport **14**(17): 2213-8.
- Calvert, G. A. (2001). "Crossmodal processing in the human brain: insights from functional neuroimaging studies." Cereb Cortex **11**(12): 1110-23.
- Calvert, G. A., M. J. Brammer, E. T. Bullmore, R. Campbell, S. D. Iversen and A. S. David (1999). "Response amplification in sensory-specific cortices during crossmodal binding." Neuroreport **10**(12): 2619-23.
- Calvert, G. A., E. T. Bullmore, M. J. Brammer, R. Campbell, S. C. Williams, P. K. McGuire, P. W. Woodruff, S. D. Iversen and A. S. David (1997). "Activation of auditory cortex during silent lipreading." Science **276**(5312): 593-596.
- Calvert, G. A. and R. Campbell (2003). "Reading speech from still and moving faces: the neural substrates of visible speech." J Cogn Neurosci **15**(1): 57-70.
- Calvert, G. A., R. Campbell and M. J. Brammer (2000). "Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex." Curr.Biol. **10**(11): 649-657.
- Calvert, G. A., P. C. Hansen, S. D. Iversen and M. J. Brammer (2001). "Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect." Neuroimage **14**(2): 427-38.
- Calvert, G. A. and T. Thesen (2004). "Multisensory integration: methodological approaches and emerging principles in the human brain." Journal of Physiology, Paris **98**(1-3): 191-205.
- Campbell, R., J. Garwood, S. Franklin, D. Howard, T. Landis and M. Regard (1990). "Neuropsychological studies of auditory-visual fusion illusions. Four case studies and their implications." Neuropsychologia. **28**(8): 787-802.
- Campbell, R., M. MacSweeney, S. Surguladze, G. Calvert, P. McGuire, J. Suckling, M. J. Brammer and A. S. David (2001). "Cortical substrates for the perception

- of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning)." Brain research. Cognitive brain research **12**(2): 233-43.
- Carpentier, A., K. R. Pugh, M. Westerveld, C. Studholme, O. Skrinjar, J. L. Thompson, D. D. Spencer and R. T. Constable (2001). "Functional MRI of language processing: dependence on input modality and temporal lobe epilepsy." Epilepsia **42**(10): 1241-54.
- Cavanaugh, J. R., W. Bair and J. A. Movshon (2002). "Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons." J Neurophysiol **88**(5): 2530-46.
- Cohen, D. (1968). "Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents." Science **161**(843): 784-6.
- Colin, C., M. Radeau, A. Soquet, B. Dachy and P. Deltenre (2002). "Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **113**(4): 507-18.
- Colin, C., M. Radeau, A. Soquet and P. Deltenre (2004). "Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **115**(9): 1989-2000.
- Colin, C., M. Radeau, A. Soquet, D. Demolin, F. Colin and P. Deltenre (2002). "Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **113**(4): 495-506.
- Collins, D. L., P. Neelin, T. M. Peters and A. C. Evans (1994). "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space." J Comput Assist Tomogr **18**(2): 192-205.
- Connors, B. W. and M. J. Gutnick (1990). "Intrinsic firing patterns of diverse neocortical neurons." Trends Neurosci **13**(3): 99-104.
- Cotillon, N., M. Nafati and J. M. Edeline (2000). "Characteristics of reliable tone-evoked oscillations in the rat thalamo-cortical auditory system." Hear.Res. **142**(1-2): 113-130.

- Coull, J. T., C. D. Frith, C. Buchel and A. C. Nobre (2000). "Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts." Neuropsychologia **38**(6): 808-19.
- Crone, N. E., D. Boatman, B. Gordon and L. Hao (2001). "Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **112**(4): 565-82.
- Crone, N. E., L. Hao, J. Hart, Jr., D. Boatman, R. P. Lesser, R. Irizarry and B. Gordon (2001). "Electrocorticographic gamma activity during word production in spoken and sign language." Neurology **57**(11): 2045-53.
- Curio, G. (2000). "Linking 600-Hz "spikelike" EEG/MEG wavelets ("sigma-bursts") to cellular substrates: concepts and caveats." Journal of clinical neurophysiology official publication of the American Electroencephalographic Society **17**(4): 377-96.
- D'Esposito, M., E. Zarahn and G. K. Aguirre (1999). "Event-related functional MRI: implications for cognitive psychology." Psychological Bulletin **125**(1): 155-64.
- Dale, A. M. and E. Halgren (2001). "Spatiotemporal mapping of brain activity by integration of multiple imaging modalities." Current Opinion in Neurobiology **11**(2): 202-8.
- Desimone, R. and C. G. Gross (1979). "Visual areas in the temporal cortex of the macaque." Brain Res. **178**(2-3): 363-380.
- Desimone, R. and L. G. Ungerleider (1986). "Multiple visual areas in the caudal superior temporal sulcus of the macaque." J Comp.Neurol. **248**(2): 164-189.
- Dhond, R. P., T. Witzel, A. M. Dale and E. Halgren (2005). "Spatiotemporal brain maps of delayed word repetition and recognition." Neuroimage.
- Donaldson, D. I. (2004). "Parsing brain activity with fMRI and mixed designs: what kind of a state is neuroimaging in?" Trends Neurosci **27**(8): 442-4.
- Dougherty, R. F., V. M. Koch, A. A. Brewer, B. Fischer, J. Modersitzki and B. A. Wandell (2003). "Visual field representations and locations of visual areas V1/2/3 in human visual cortex." J Vis **3**(10): 586-98.
- Downar, J., A. P. Crawley, D. J. Mikulis and K. D. Davis (2000). "A multimodal cortical network for the detection of changes in the sensory environment." Nat.Neurosci. **3**(3): 277-283.

- Doyle, M. C. and R. J. Snowden (2001). "Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location." Perception and Psychophysics **30**: 795-810.
- Driver, J. (1996). "Enhancement of selective listening of illusory mislocation of speech sounds due to lip-reading." Nature Vol **381**(6577): 66-68.
- Driver, J. and C. Spence (2000). "Multisensory perception: beyond modularity and convergence." Curr Biol **10**(20): R731-5.
- Duhamel, J. R., C. L. Colby and M. E. Goldberg (1991). Congruent representations of visual and somatosensory space in single neurons of monkey ventral intraparietal cortex (Area VIP). Brain and Space. J. Paillard. New York, Oxford University Press: 223-236.
- Duncan, J., R. J. Seitz, J. Kolodny, D. Bor, H. Herzog, A. Ahmed, F. N. Newell and H. Emslie (2000). "A neural basis for general intelligence." Science **289**(5478): 457-60.
- Eggermont, J. J. (1992). "Neural interaction in cat primary auditory cortex. Dependence on recording depth, electrode separation, and age." J Neurophysiol **68**(4): 1216-28.
- Eggermont, J. J. and C. W. Ponton (2002). "The neurophysiology of auditory perception: from single units to evoked potentials." Audiol Neurotol **7**(2): 71-99.
- Elul, R. (1971). "The genesis of the EEG." Int Rev Neurobiol **15**: 227-72.
- Engel, A. K., P. Fries and W. Singer (2001). "Dynamic predictions: oscillations and synchrony in top-down processing." Nat Rev Neurosci **2**(10): 704-16.
- Evans, A. C., D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly and T. M. Peters (1993). "3D statistical neuroanatomical models from 305 MRI volumes." Proc. IEEE-Nuclear Science Symposium and Medical Imaging Conference: 1813-1817.
- Fadiga, L., L. Fogassi, V. Gallese and G. Rizzolatti (2000). "Visuomotor neurons: ambiguity of the discharge or 'motor' perception?" Int.J Psychophysiol. **35**(2-3): 165-177.
- Fadiga, L., L. Fogassi, G. Pavesi and G. Rizzolatti (1995). "Motor facilitation during action observation: a magnetic stimulation study." J Neurophysiol. **73**(6): 2608-2611.

- Felleman, D. J. and D. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." Cereb.Cortex. **1**(1): 1-47.
- Fiez, J. A. (2001). "Neuroimaging studies of speech an overview of techniques and methodological approaches." J Commun Disord **34**(6): 445-54.
- Forman, S. D., J. D. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun and D. C. Noll (1995). "Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold." Magnetic resonance in medicine official journal of the Society of Magnetic Resonance in Medicine **33**(5): 636-47.
- Fort, A., C. Delpuech, J. Pernier and M.-H. Giard-Steiner (2002). "Early auditory-visual interactions in human cortex during nonredundant target identification." Cognitive Brain Research **14**: 20-30.
- Fort, A., C. Delpuech, J. Pernier and M. H. Giard (2002). "Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans." Cerebral Cortex **12**(10): 1031-9.
- Fowler, C. A., J. M. Brown and V. A. Mann (2000). "Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans." J Exp Psychol Hum Percept Perform **26**(3): 877-88.
- Fox, P. T., L. M. Parsons and J. L. Lancaster (1998). "Beyond the single study: function/location metanalysis in cognitive neuroimaging." Curr Opin Neurobiol **8**(2): 178-87.
- Foxe, J. J. and C. E. Schroeder (2005). "The case for feedforward multisensory convergence during early cortical processing." Neuroreport **16**(5): 419-23.
- Foxe, J. J. and G. V. Simpson (2002). "Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing." Experimental brain research; Experimentelle Hirnforschung; Experimentation cerebrale **142**(1): 139-50.
- Foxe, J. J., G. R. Wylie, A. Martinez, C. E. Schroeder, D. C. Javitt, D. Guilfoyle, W. Ritter and M. M. Murray (2002). "Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study." Journal of Neurophysiology **88**(1): 540-3.
- Frassinetti, F., N. Bolognini and E. Ladavas (2002). "Enhancement of visual perception by crossmodal visuo-auditory interaction." Experimental brain

- research; Experimentelle Hirnforschung; Experimentation cerebrale
147(3): 332-43.
- Fries, W. (1984). "Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase." J Comp.Neurol. **230(1): 55-76.**
- Friston, K. J. (2000). "The labile brain. III. Transients and spatio-temporal receptive fields." Philosophical transactions of the Royal Society of London. Series B Biological sciences **355(1394): 253-65.**
- Friston, K. J., K. J. Worsley, R. S. J. Frakowiak, J. C. Mazziotta and A. C. Evans (1994). "Assessing the significance of focal activations using their spatial extent." Human Brain Mapping **1: 214-20.**
- Frost, J. A., J. R. Binder, J. A. Springer, T. A. Hammeke, P. S. Bellgowan, S. M. Rao and R. W. Cox (1999). "Language processing is strongly left lateralized in both sexes. Evidence from functional MRI." Brain a journal of neurology **122(Pt 2): 199-208.**
- Frost, J. A., J. R. Binder, J. A. Springer, T. A. Hammeke, P. S. Bellgowan, S. M. Rao and R. W. Cox (1999). "Language processing is strongly left lateralized in both sexes. Evidence from functional MRI." Brain **122 (Pt 2): 199-208.**
- Galambos, R., J. E. Rose, R. B. Bromiley and J. R. Hughes (1952). "Microelectrode studies on medial geniculate body of cat. II. Response to clicks." J Neurophysiol **15(5): 359-80.**
- Gallese, V., L. Fadiga, L. Fogassi and G. Rizzolatti (1996). "Action recognition in the premotor cortex." Brain **119 (Pt 2): 593-609.**
- Geschwind, N. and A. M. Galaburda (1985). "Cerebral lateralization. Biological mechanisms, associations, and pathology: I. A hypothesis and a program for research." Arch.Neurol. **42(5): 428-459.**
- Geschwind, N. and W. Levitsky (1968). "Human brain: left-right asymmetries in temporal speech region." Science **161(837): 186-187.**
- Ghazanfar, A. A., J. X. Maier, K. L. Hoffman and N. K. Logothetis (2005). "Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex." J Neurosci **25(20): 5004-12.**
- Giard, M. H. and F. Peronnet (1999). "Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study." Journal of Cognitive Neuroscience **11(5): 473-490.**

- Gielen, S.-C., R.-A. Schmidt and P.-J. A. A. G. S.-C. S. U. U. N. Van-den-Heuvel (1983). "On the nature of intersensory facilitation of reaction time." Perception and Psychophysics **34**(2): 161-168 URL-JOURNAL: www.psychonomic.org/prp.htm.
- Girard, P., J. M. Hupe and J. Bullier (2001). "Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities." J Neurophysiol **85**(3): 1328-31.
- Goldman, R. I., J. M. Stern, J. Engel, Jr. and M. S. Cohen (2002). "Simultaneous EEG and fMRI of the alpha rhythm." Neuroreport **13**(18): 2487-92.
- Golgi, C. (1989). "On the structure of nerve cells. 1898." J Microsc **155** (Pt 1): 3-7.
- Gondan, M., B. Roeder, B. Niederhaus and F. Roesler (2002). "Behavior and ERP correlates of redundant visual-auditory stimulus processing." Proceedings of the 3rd Annual Meeting of the International Multisensory Research Forum: 80.
- Gray, C. M. (1999). "The temporal correlation hypothesis of visual feature integration: still alive and well." Neuron **24**(1): 31-47, 111-25.
- Graziano, M. S. and C. G. Gross (1998). "Spatial maps for the control of movement." Curr.Opin.Neurobiol. **8**(2): 195-201.
- Graziano, M. S. A., L. A. Reiss and C. G. Gross (1999). "A neuronal representation of the location of nearby sounds." Nature **397**(6718): 428-430.
- Green, K. P. and L. W. Norrix (2001). "Perception of /r/ and /l/ in a stop cluster: evidence of cross-modal context effects." J Exp Psychol Hum Percept Perform **27**(1): 166-77.
- Greenlee, M. W. (2000). "Human cortical areas underlying the perception of optic flow: brain imaging studies." Int Rev Neurobiol **44**: 269-92.
- Griffin, I. C., C. Miniussi and A. C. Nobre (2002). "Multiple mechanisms of selective attention: differential modulation of stimulus processing by attention to space or time." Neuropsychologia **40**(13): 2325-40.
- Griffiths, T. D., A. Rees, C. Witton, P. M. Cross, R. A. Shakir and G. G. R. Green (1997). "Spatial and temporal auditory processing deficits following right hemisphere infarction: A psychophysical study." Brain **120**(5): 785.
- Griffiths, T. D., J. D. Warren, S. K. Scott, I. Nelken and A. J. King (2004). "Cortical processing of complex sound: a way forward?" Trends Neurosci **27**(4): 181-5.

- Griffiths, T. D., J. D. Warren, S. K. Scott, I. Nelken and A. J. King (2004). "Cortical processing of complex sound: a way forward?" Trends in Neurosciences **27**(4): 181.
- Guthrie, D. and J. S. Buchwald (1991). "Significance testing of difference potentials." Psychophysiology **28**(2): 240-4.
- Hackett, T. A., I. Stepniewska and J. H. Kaas (1998). "Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys." J Comp Neurol **394**(4): 475-95.
- Hackett, T. A., I. Stepniewska and J. H. Kaas (1998). "Thalamocortical connections of the parabelt auditory cortex in macaque monkeys." J Comp Neurol **400**(2): 271-86.
- Hadjikhani, N. and P. E. Roland (1998). "Cross-modal transfer of information between the tactile and the visual representations in the human brain: A positron emission tomographic study." J.Neurosci. **18**(3): 1072-1084.
- Haig, A. R., V. De Pascalis and E. Gordon (1999). "Peak gamma latency correlated with reaction time in a conventional oddball paradigm." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **110**(1): 158-65.
- Halgren, E. (2004). "How can intracranial recordings assist MEG source localization?" Neurol Clin Neurophysiol **2004**: 86.
- Halgren, E., P. Baudena, J. M. Clarke, G. Heit, K. Marinkovic, B. Devaux, J. P. Vignal and A. Biraben (1995). "Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe." Electroencephalogr.Clin.Neurophysiol. **94**(4): 229-250.
- Halgren, E., P. Baudena, G. Heit, J. M. Clarke, K. Marinkovic and M. Clarke (1994). "Spatio-temporal stages in face and word processing. I. Depth-recorded potentials in the human occipital, temporal and parietal lobes [corrected] [published erratum appears in J Physiol Paris 1994;88(2):following 151]." J Physiol.Paris. **88**(1): 1-50.
- Halgren, E., A. M. Dale, M. I. Sereno, R. B. Tootell, K. Marinkovic and B. R. Rosen (1999). "Location of human face-selective cortex with respect to retinotopic areas." Hum Brain Mapp **7**(1): 29-37.
- Halgren, E., R. P. Dhond, N. Christensen, C. Van Petten, K. Marinkovic, J. D. Lewine and A. M. Dale (2002). "N400-like magnetoencephalography responses

- modulated by semantic context, word frequency, and lexical class in sentences." Neuroimage **17**(3): 1101-16.
- Hämäläinen, M., R. Hari, R. J. Ilmoniemi, J. Knuutila and O. V. Lounasmaa (1993). "Magnetoencephalography – theory, instrumentation, and applications to non-invasive studies of the working human brain." Rev. Mod. Phys. **65**: 413–497.
- Hari, R. (1991). "Activation of the human auditory cortex by speech sounds." Acta Otolaryngol.Suppl. **491**: 132-137.
- Hari, R. (1996). "MEG in the study of human cortical functions." Electroencephalogr.Clin.Neurophysiol.Suppl. **47**: 47-54.
- Hari, R., R. Salmelin, J. P. Makela, S. Salenius and M. Helle (1997). "Magnetoencephalographic cortical rhythms." Int J Psychophysiol **26**(1-3): 51-62.
- Hashikawa, T., M. Molinari, E. Rausell and E. G. Jones (1995). "Patchy and laminar terminations of medial geniculate axons in monkey auditory cortex." J Comp Neurol **362**(2): 195-208.
- Hashimoto, I., T. Mashiko, K. Yoshikawa, T. Mizuta, T. Imada and M. Hayashi (1995). "Neuromagnetic measurements of the human primary auditory response." Electroencephalogr Clin Neurophysiol **96**(4): 348-56.
- Hasselmo, M. E., E. T. Rolls and G. C. Baylis (1989). "The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey." Behav.Brain Res. **32**(3): 203-218.
- Helmholtz, H. (1853). "Ueber einige Gesetze der Verteilung elektrischer Stroeme in koerperlichen Leitern mit Anwendung auf die tierisch-elektrischen Versuche." Ann Phys Chem **89**: 211-233, 353-377.
- Herdman, A. T., T. Fujioka, W. Chau, B. Ross, C. Pantev and T. W. Picton (2004). "Cortical oscillations modulated by congruent and incongruent audiovisual stimuli." Neurol Clin Neurophysiol **2004**: 15.
- Hershenson, M. (1962). "Reaction time as a measure of intersensory facilitation." J.Exp.Psychol. **63**: 289-293.
- Hikosaka, K., E. Iwai, H. Saito and K. Tanaka (1988). "Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey." J.Neurophysiol. **60**(5): 1615-1637.

- Hinke, R. M., X. Hu, A. E. Stillman, S. G. Kim, H. Merkle, R. Salmi and K. Ugurbil (1993). "Functional magnetic resonance imaging of Broca's area during internal speech." Neuroreport **4**(6): 675-8.
- Horwitz, B. and D. Poeppel (2002). "How can EEG/MEG and fMRI/PET data be combined?" Hum Brain Mapp **17**(1): 1-3.
- Howard, M. A., I. O. Volkov, R. Mirsky, P. C. Garell, M. D. Noh, M. Granner, H. Damasio, M. Steinschneider, R. A. Reale, J. E. Hind and J. F. Brugge (2000). "Auditory cortex on the human posterior superior temporal gyrus." Journal of comparative neurology **416**(1): 79-92.
- Hubel, D. H. and T. N. Wiesel (1968). "Receptive fields and functional architecture of monkey striate cortex." J Physiol **195**(1): 215-43.
- Hubel, D. H. and T. N. Wiesel (1977). "Ferrier lecture. Functional architecture of macaque monkey visual cortex." Proc R Soc Lond B Biol Sci **198**(1130): 1-59.
- Hudspeth, A. J. (2000). Hearing. Principles of neural science. E. Kandel, J. H. Schwartz and T. M. Jessel. New York, McGraw-Hill: 590-612.
- Isomura, Y. and M. Takada (2004). "Neural mechanisms of versatile functions in primate anterior cingulate cortex." Rev Neurosci **15**(4): 279-91.
- Jaencke, L. and N. J. Shah (2004). "'Hearing' syllables by 'seeing' visual stimuli." European Journal of Neuroscience **19**(9): 2603-2608.
- Jahnsen, H. and R. Llinas (1984). "Electrophysiological properties of guinea-pig thalamic neurones: an in vitro study." J Physiol **349**: 205-26.
- Jahnsen, H. and R. Llinas (1984). "Ionic basis for the electro-responsiveness and oscillatory properties of guinea-pig thalamic neurones in vitro." J Physiol **349**: 227-47.
- Jancke, L., T. Wustenberg, H. Scheich and H. J. Heinze (2002). "Phonetic perception and the temporal cortex." Neuroimage **15**(4): 733-46.
- Jenkinson, M., P. Bannister, M. Brady and S. Smith (2002). "Improved optimization for the robust and accurate linear registration and motion correction of brain images." Neuroimage **17**(2): 825-41.
- Jezzard, P. and S. Clare (2001). Principles of nuclear magnetic resonance and MRI. Functional MRI: An introduction to methods. P. Jezzard, P. M. Matthews and S. M. Smith. Oxford, Oxford University Press: 3-34.

- Jiang, W. and B. E. Stein (2003). "Cortex controls multisensory depression in superior colliculus." J Neurophysiol **90**(4): 2123-35.
- Johnson, M. K., S. F. Nolde, M. Mather, J. Kounios, D. L. Schacter and T. Curran (1997). "Test format can affect the similarity of brain activity associated with true and false recognition memory." Psychological Science **8**: 250-257.
- Jones, J. A. and D. E. Callan (2003). "Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect." Neuroreport **14**(8): 1129-33.
- Jones, J. A. and K. Munhall (1997). The effects of separating auditory and visual sources on audio-visual integration of speech. Canadian Acoustics. **2**: 13-19.
- Kaas, J. H. and T. A. Hackett (1998). "Subdivisions of auditory cortex and levels of processing in primates." Audiol Neurotol **3**(2-3): 73-85.
- Kaas, J. H. and T. A. Hackett (1999). "'What' and 'where' processing in auditory cortex." Nat. Neurosci. **2**(12): 1045-1047.
- Kaas, J. H. and T. A. Hackett (2000). "Subdivisions of auditory cortex and processing streams in primates." Proc Natl Acad Sci U S A **97**(22): 11793-9.
- Kadunce, D. C., J. W. Vaughan, M. T. Wallace, G. Benedek and B. E. Stein (1997). "Mechanisms of within- and cross-modality suppression in the superior colliculus." J Neurophysiol. **78**(6): 2834-2847.
- Kastner, S., M. A. Pinsk, W. P. De, R. Desimone and L. G. Ungerleider (1999). "Increased activity in human visual cortex during directed attention in the absence of visual stimulation." Neuron **22**(4): 751-761.
- Keysers, C., E. Kohler, M. A. Umiltà, L. Nanetti, L. Fogassi and V. Gallese (2003). "Audiovisual mirror neurons and action recognition." Exp Brain Res **153**(4): 628-36.
- Klimesch, W., M. Doppelmayr, D. Rohm, D. Pollhuber and W. Stadler (2000). "Simultaneous desynchronization and synchronization of different alpha responses in the human electroencephalograph: a neglected paradox?" Neurosci Lett **284**(1-2): 97-100.
- Klimesch, W., M. Doppelmayr, H. Russegger, T. Pachinger and J. Schwaiger (1998). "Induced alpha band power changes in the human EEG and attention." Neurosci Lett **244**(2): 73-6.
- Klimesch, W., M. Doppelmayr, H. Wimmer, W. Gruber, D. Rohm, J. Schwaiger and F. Hutzler (2001). "Alpha and beta band power changes in normal and

- dyslexic children." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **112**(7): 1186-95.
- Klimesch, W., G. Pfurtscheller, W. Mohl and H. Schimke (1990). "Event-related desynchronization, ERD-mapping and hemispheric differences for words and numbers." Int J Psychophysiol **8**(3): 297-308.
- Klimesch, W., B. Schack and P. Sauseng (2005). "The functional significance of theta and upper alpha oscillations." Exp Psychol **52**(2): 99-108.
- Klimesch, W., H. Schimke and G. Pfurtscheller (1993). "Alpha frequency, cognitive load and memory performance." Brain topography **5**(3): 241-51.
- Klimesch, W., H. Schimke and J. Schwaiger (1994). "Episodic and semantic memory: an analysis in the EEG theta and alpha band." Electroencephalogr Clin Neurophysiol **91**(6): 428-41.
- Klucharev, V., R. Mottonen and M. Sams (2003). "Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception." Brain research. Cognitive brain research **18**(1): 65-75.
- Kohler, E., C. Keysers, M. A. Umiltà, L. Fogassi, V. Gallese and G. Rizzolatti (2002). "Hearing sounds, understanding actions: action representation in mirror neurons." Science **297**(5582): 846-8.
- Kosaki, H., T. Hashikawa, J. He and E. G. Jones (1997). "Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys." J Comp Neurol **386**(2): 304-16.
- Krause, C. M., A. H. Lang, M. Laine, M. Kuusisto and B. Porn (1996). "Event-related EEG desynchronization and synchronization during an auditory memory task." Electroencephalogr.Clin.Neurophysiol. **98**(4): 319-326.
- Krause, C. M., A. H. Lang, M. Laine, M. Kuusisto and B. Porn (1996). "Event-related EEG desynchronization and synchronization during an auditory memory task." Electroencephalogr Clin Neurophysiol **98**(4): 319-26.
- Krause, C. M., H. Lang, M. Laine, M. Kuusisto and B. Porn (1995). "Cortical processing of vowels and tones as measured by event-related desynchronization." Brain Topogr. **8**(1): 47-56.
- Krause, C. M., H. A. Lang, M. Laine, S. I. Helle, M. J. Kuusisto and B. Porn (1994). "Event-related desynchronization evoked by auditory stimuli." Brain topography **7**(2): 107-12.

- Krauss, G. L. and W. R. S. Webber (2005). Digital EEG. Electroencephalography: Basic principles, clinical applications, and related fields. E. Niedermeyer and F. H. Lopes da Silva. Philadelphia, LWW. 5.
- Kuriki, S., N. Isahai and A. Ohtsuka (2005). "Spatiotemporal characteristics of the neural activities processing consonant/dissonant tones in melody." Exp Brain Res **162**(1): 46-55.
- Kuriki, S., T. Nogai and Y. Hirata (1995). "Cortical sources of middle latency responses of auditory evoked magnetic field." Hear Res **92**(1-2): 47-51.
- Kurland, J., M. A. Naeser, E. H. Baker, K. Doron, P. I. Martin, H. E. Seekins, A. Bogdan, P. Renshaw and D. Yurgelun-Todd (2004). "Test-retest reliability of fMRI during nonverbal semantic decisions in moderate-severe nonfluent aphasia patients." Behav Neurol **15**(3-4): 87-97.
- Kwong, K. K., J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner and et al. (1992). "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation." Proceedings of the National Academy of Sciences of the United States of America **89**(12): 5675-9.
- Lachaux, J. P., D. Rudrauf and P. Kahane (2003). "Intracranial EEG and human brain mapping." Journal of Physiology, Paris **97**(4-6): 613-28.
- Lancaster, J. L., M. G. Woldorff, L. M. Parsons, M. Liotti, C. S. Freitas, L. Rainey, P. V. Kochunov, D. Nickerson, S. A. Mikiten and P. T. Fox (2000). "Automated Talairach atlas labels for functional brain mapping." Human Brain Mapping **10**(3): 120-31.
- Laurienti, P. J., T. J. Perrault, T. R. Stanford, M. T. Wallace and B. E. Stein (2005). "On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies." Exp Brain Res. **166**(3-4):289-97
- Laurienti, P. J., M. T. Wallace, J. A. Maldjian, C. M. Susi, B. E. Stein and J. H. Burdette (2003). "Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices." Hum Brain Mapp **19**(4): 213-23.
- Lauter, J. L., P. Herscovitch, C. Formby and M. E. Raichle (1985). "Tonotopic organization in human auditory cortex revealed by positron emission tomography." Hear Res **20**(3): 199-205.

- Lee, T. S. (2003). "Computations in the early visual cortex." J Physiol Paris **97**(2-3): 121-39.
- Leinonen, L., J. Hyvarinen and A. R. Sovijarvi (1980). "Functional properties of neurons in the temporo-parietal association cortex of awake monkey." Exp.Brain Res. **39**(2): 203-215.
- Lewis, J. W., M. S. Beauchamp and E. A. DeYoe (2000). "A comparison of visual and auditory motion processing in human cerebral cortex." Cereb.Cortex. **10**/9(873-888): 888.
- Lewis, J. W. and D. C. Van Essen (2000). "Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey." Journal of comparative neurology **428**(1): 112-37.
- Lewkowicz, D. J. (1992). "Infants' responsiveness to the auditory and visual attributes of a sounding/moving stimulus." Percept.Psychophys. **52**(5): 519-528.
- Lewkowicz, D. J. (2000). "The development of intersensory temporal perception: an epigenetic systems/limitations view." Psychol.Bull. **126**(2): 281-308.
- Lewkowicz, D. J. and G. Turkewitz (1981). "Intersensory interaction in newborns: modification of visual preferences following exposure to sound." Child Dev. **52**(3): 827-832.
- Liegeois-Chauvel, C., A. Musolino, J. M. Badier, P. Marquis and P. Chauvel (1994). "Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components." Electroencephalogr Clin Neurophysiol **92**(3): 204-14.
- Linden, J. F., A. Grunewald and R. A. Andersen (1999). "Responses to auditory stimuli in macaque lateral intraparietal area. II. Behavioral modulation." J Neurophysiol. **82**(1): 343-358.
- Liu, A. K., J. W. Belliveau and A. M. Dale (1998). "Spatiotemporal imaging of human brain activity using functional MRI constrained magnetoencephalography data: Monte Carlo simulations." Proc Natl Acad Sci U S A **95**(15): 8945-50.
- Logothetis, N. K. (2002). "The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal." Philosophical transactions of the Royal Society of London. Series B Biological sciences **357**(1424): 1003-37.

- Logothetis, N. K., J. Pauls, M. Augath, T. Trinath and A. Oeltermann (2001). "Neurophysiological investigation of the basis of the fMRI signal." Nature **412**(6843): 150-7.
- Lopes da Silva, F. H. (2005). Event-related potentials: Methodology and Quantifications. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. E. Niedermeyer and F. H. Lopes da Silva. Philadelphia, LWW. **5**: 167-192.
- Loveless, N. E., J. Brebner and P. Hamilton (1970). "Bisensory presentation of information." Psychol.Bull. **73**(3): 161-199.
- Ludman, C. N., A. Q. Summerfield, D. Hall, M. Elliott, J. Foster, J. L. Hykin, R. Bowtell and P. G. Morris (2000). "Lip-reading ability and patterns of cortical activation studied using fMRI." Br J Audiol **34**(4): 225-30.
- Luethke, L. E., L. A. Krubitzer and J. H. Kaas (1989). "Connections of primary auditory cortex in the New World monkey, *Saguinus*." J Comp Neurol **285**(4): 487-513.
- Lutkenhoner, B., K. Krumbholz, C. Lammertmann, A. Seither-Preisler, O. Steinstrater and R. D. Patterson (2003). "Localization of primary auditory cortex in humans by magnetoencephalography." Neuroimage **18**(1): 58-66.
- Lutkenhoner, B., C. Lammertmann, C. Simoes and R. Hari (2002). "Magnetoencephalographic correlates of audiotactile interaction." Neuroimage **15**(3): 509-22.
- MacSweeney, M., E. Amaro, G. A. Calvert, R. Campbell, A. S. David, P. McGuire, S. C. Williams, B. Woll and M. J. Brammer (2000). "Silent speechreading in the absence of scanner noise: an event-related fMRI study." Neuroreport. **11**(8): 1729-1733.
- MacSweeney, M., R. Campbell, G. A. Calvert, P. K. McGuire, A. S. David, J. Suckling, C. Andrew, B. Woll and M. J. Brammer (2001). "Dispersed activation in the left temporal cortex for speech-reading in congenitally deaf people." Proc Biol Sci **268**(1466): 451-7.
- Makeig, S. (1993). "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones." Electroencephalogr Clin Neurophysiol **86**(4): 283-93.

- Makeig, S. and T. P. Jung (1996). "Tonic, phasic, and transient EEG correlates of auditory awareness in drowsiness." Brain research. Cognitive brain research **4**(1): 15-25.
- Makeig, S., M. Westerfield, T. P. Jung, S. Enghoff, J. Townsend, E. Courchesne and T. J. Sejnowski (2002). "Dynamic brain sources of visual evoked responses." Science **295**(5555): 690-4.
- Makinen, V., H. Tiitinen and P. May (2005). "Auditory event-related responses are generated independently of ongoing brain activity." Neuroimage **24**(4): 961-8.
- Maldonado, P. E. and G. L. Gerstein (1996). "Neuronal assembly dynamics in the rat auditory cortex during reorganization induced by intracortical microstimulation." Exp Brain Res **112**(3): 431-41.
- Mallat, S. (1989). "Multichannle frequency decompositions of images and wavelet models." IEEE Trans Acoust Speech Signa process **32**: 2091-2109.
- Mandelbrot, B. B. (1998). Multifractals and 1/f noise. New York, Springer-Verlag.
- Maris, E. and R. Oostenveld (in press). "Nonparametric Statistical Tests for Spatiotemporal Biological Data: Theory." Neuroscience.
- Massaro, D. W. (2004). From multisensory integration to talking heads and language learning. Handbook of Multisensory Processing. G. A. Calvert, C. Spence and B. E. Stein. Cambridge, MIT Press.
- Matsumoto, R., D. R. Nair, E. LaPresto, I. Najm, W. Bingaman, H. Shibasaki and H. O. Luders (2004). "Functional connectivity in the human language system: a cortico-cortical evoked potential study." Brain **127**(Pt 10): 2316-30.
- Matthews, P. M. (2001). An introduction to functional magnetic resonance imaging of the brain. Functional MRI: An introduction to methods. P. Jezzard, P. M. Matthews and S. M. Smith. Oxford, Oxford University Press: 3-34.
- McGurk, H. and J. MacDonald (1976). "Hearing lips and seeing voices." Nature **264**(5588): 746-748.
- McIntosh, A. R., R. E. Cabeza and N. J. Lobaugh (1998). "Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus." Journal of Neurophysiology **80**(5): 2790-6.
- McIntosh, A. R., N. J. Lobaugh, R. Cabeza, F. L. Bookstein and S. Houle (1998). "Convergence of neural systems processing stimulus associations and coordinating motor responses." Cereb.Cortex. **8**(7): 648-659.

- Menon, V., J. M. Ford, K. O. Lim, G. H. Glover and A. Pfefferbaum (1997). "Combined event-related fMRI and EEG evidence for temporal-parietal cortex activation during target detection." Neuroreport. **8**(14): 3029-3037.
- Meredith, M. A., M. T. Wallace and B. E. Stein (1992). "Visual, auditory and somatosensory convergence in output neurons of the cat superior colliculus: multisensory properties of the tecto-reticulo-spinal projection." Exp.Brain Res. **88**(1): 181-186.
- Merzenich, M. M. and J. F. Brugge (1973). "Representation of the cochlear partition of the superior temporal plane of the macaque monkey." Brain Res **50**(2): 275-96.
- Mesulam, M. and M. EJ (1982). "Insula of the old world monkey. III: Efferent cortical output and comments on function." J Comp Neurol **212**: 38-52.
- Miller, J. (1991). "Channel interaction and the redundant-targets effect in bimodal divided attention." J Exp.Psychol.Hum.Percept.Perform. **17**(1): 160-169.
- Miller, L. M. and M. D'Esposito (2005). "Perceptual fusion and stimulus coincidence in the cross-modal integration of speech." J Neurosci **25**(25): 5884-93.
- Mistlin, A. J. and D. I. Perrett (1990). "Visual and somatosensory processing in the macaque temporal cortex: the role of 'expectation'." Exp.Brain Res. **82**(2): 437-450.
- Molholm, S., W. Ritter, D. C. Javitt and J. J. Foxe (2004). "Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study." Cereb Cortex **14**(4): 452-65.
- Molholm, S., W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder and J. J. Foxe (2002). "Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study." Cognitive Brain Research **14**: 115-128.
- Morein-Zamir, S., S. Soto-Faraco and A. Kingstone (2003). "Auditory capture of vision: examining temporal ventriloquism." Brain research. Cognitive brain research **17**(1): 154-63.
- Morris, H. and H. Luders (1985). "Electrodes." Electroencephalogr Clin Neurophysiol **37**: 3-26.
- Mottronen, R., C. M. Krause, K. Tiippana and M. Sams (2002). "Processing of changes in visual speech in the human auditory cortex." Cognitive Brain Research **13**(3): 417-425.

- Mottronen, R., C. M. Krause, K. Tiippana and M. Sams (2002). "Processing of changes in visual speech in the human auditory cortex." Brain research. Cognitive brain research **13**(3): 417-25.
- Mottronen, R., M. Schurmann and M. Sams (2004). "Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study." Neurosci Lett **363**(2): 112-5.
- Mufson, E. J. and M. M. Mesulam (1984). "Thalamic connections of the insula in the rhesus monkey and comments on the paralimbic connectivity of the medial pulvinar nucleus." J.Comp.Neurol. **227**(1): 109-120.
- Mukamel, R., H. Gelbard, A. Arieli, U. Hasson, I. Fried and R. Malach (2005). "Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex." Science **309**(5736): 951-4.
- Murakami, S., A. Hirose and Y. C. Okada (2003). "Contribution of ionic currents to magnetoencephalography (MEG) and electroencephalography (EEG) signals generated by guinea-pig CA3 slices." J Physiol **553**(Pt 3): 975-85.
- Murakami, S., T. Zhang, A. Hirose and Y. C. Okada (2002). "Physiological origins of evoked magnetic fields and extracellular field potentials produced by guinea-pig CA3 hippocampal slices." J Physiol **544**(Pt 1): 237-51.
- Musso, M., A. Moro, V. Glauche, M. Rijntjes, J. Reichenbach, C. Buchel and C. Weiller (2003). "Broca's area and the language instinct." Nature neuroscience **6**(7): 774-81.
- Naatanen, R. (2001). "The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm)." Psychophysiology **38**(1): 1-21.
- Nagata, S. I., K. Uchimura, W. Hirakawa and J. I. Kuratsu (2001). "Method for quantitatively evaluating the lateralization of linguistic function using functional MR imaging." AJNR Am J Neuroradiol **22**(5): 985-91.
- Newsome, W. T., K. H. Britten, C. D. Salzman and J. A. Movshon (1990). "Neuronal mechanisms of motion perception." Cold Spring Harb Symp Quant Biol **55**: 697-705.
- Niedermeyer, E. (2005). The normal EEG of the waking adult.
Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. E. Niedermeyer and F. H. Lopes da Silva. Philadelphia, LWW. **5**: 167-192.

- Nishitani, N. and R. Hari (2002). "Viewing lip forms: cortical dynamics." Neuron **36**(6): 1211-20.
- Nishitani, N., M. Schurmann, K. Amunts and R. Hari (2005). "Broca's Region: From Action to Language." **20**: 60-9.
- Nunez, P. L. (1990). "Localization of brain activity with electroencephalography." Advances in neurology **54**: 39-65.
- Nunez, P. L. and R. B. Silberstein (2000). "On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense?" Brain topography **13**(2): 79-96.
- Oertel, D. (1991). "The role of intrinsic neuronal properties in the encoding of auditory information in the cochlear nuclei." Curr Opin Neurobiol **1**(2): 221-8.
- Ogawa, S., R. S. Menon, D. W. Tank, S. G. Kim, H. Merkle, J. M. Ellermann and K. Ugurbil (1993). "Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model." Biophys.J. **64**(3): 803-812.
- Ojanen, V., R. Mottonen, J. Pekkola, I. P. Jaaskelainen, R. Joensuu, T. Autti and M. Sams (2005). "Processing of audiovisual speech in Broca's area." Neuroimage **25**(2): 333-8.
- Olson, I. R., J. C. Gatenby and J. C. Gore (2002). "A comparison of bound and unbound audio-visual information processing in the human cerebral cortex." Brain research. Cognitive brain research **14**(1): 129-38.
- Opitz, B., A. Mecklinger, A. D. Friederici and D. Y. von Cramon (1999). "The functional neuroanatomy of novelty processing: integrating ERP and fMRI results." Cereb Cortex **9**(4): 379-91.
- Oram, M. W. and D. I. Perrett (1994). "Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli." Journal.of.Cognitive.Neuroscience Vol **6**(2): 99-116.
- Pandya, D. N. and D. L. Rosene (1993). "Laminar termination patterns of thalamic, callosal, and association afferents in the primary auditory area of the rhesus monkey." Exp Neurol **119**(2): 220-34.
- Pandya, D. N. and F. Sanides (1973). "Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern." Z Anat Entwicklungsgesch **139**(2): 127-61.

- Pantev, C. (1995). "Evoked and induced gamma-band activity of the human cortex." Brain Topogr. 7(4): 321-330.
- Pantev, C., M. Hoke, K. Lehnertz, B. Lutkenhoner, G. Anogianakis and W. Wittkowski (1988). "Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields." Electroencephalogr Clin Neurophysiol 69(2): 160-70.
- Pantev, C., S. Makeig, M. Hoke, R. Galambos, S. Hampson and C. Gallen (1991). "Human auditory evoked gamma-band magnetic fields." Proc.Natl.Acad.Sci.U.S.A. 88(20): 8996-9000.
- Paulesu, E., C. D. Frith and R. S. Frackowiak (1993). "The neural correlates of the verbal component of working memory." Nature 362(6418): 342.
- Paulesu, E., D. Perani, V. Blasi, G. Silani, N. A. Borghese, U. De Giovanni, S. Sensolo and F. Fazio (2003). "A functional-anatomical model for lipreading." J Neurophysiol 90(3): 2005-13.
- Pearson, R., B. P, G. KC and P. TP (1982). "The organization of the connections between the cortex and the claustrum in the monkey." Brain Res 234: 435 441.
- Pekkola, J., V. Ojanen, T. Autti, I. P. Jaaskelainen, R. Mottonen, A. Tarkiainen and M. Sams (2005). "Primary auditory cortex activation by visual speech: an fMRI study at 3 T." Neuroreport 16(2): 125-8.
- Pelizzzone, M., R. Hari, J. P. Makela, J. Huttunen, S. Ahlfors and M. Hamalainen (1987). "Cortical origin of middle-latency auditory evoked responses in man." Neuroscience Letters 82(3): 303.
- Penfield, W. and T. Rasmussen (1950). The cerebral cortex of man: A clinical study of localization of function. New York, Macmillan.
- Penfield, W. and T. Rasmussen (1952). The cerebal cortex of man. New York, Macmillan.
- Pfurtscheller, B. and F. H. Lopes da Silva (2005). EEG Event-Related Desynchronization and Event-Related Synchronization. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. E. Niedermeyer and F. H. Lopes da Silva. Philadelphia, LWW. 5: 1003-1016.
- Pfurtscheller, G. (2001). "Functional brain imaging based on ERD/ERS." Vision research 41(10-11): 1257-60.

- Pfurtscheller, G. and C. Andrew (1999). "Event-Related changes of band power and coherence: methodology and interpretation." Journal of clinical neurophysiology official publication of the American Electroencephalographic Society **16**(6): 512-9.
- Pfurtscheller, G. and R. Cooper (1975). "Frequency dependence of the transmission of the EEG from cortex to scalp." Electroencephalography and clinical neurophysiology **38**(1): 93-6.
- Pfurtscheller, G. and W. Klimesch (1992). "Functional topography during a visuo-verbal judgment task studied with event-related desynchronization mapping." Journal of clinical neurophysiology official publication of the American Electroencephalographic Society **9**(1): 120-31.
- Pfurtscheller, G. and F. H. Lopes da Silva (1999). "Event-related EEG/MEG synchronization and desynchronization: basic principles." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **110**(11): 1842-57.
- Pfurtscheller, G., H. Maresch and S. Schuy (1977). "Inter- and intrahemispheric differences in the peak frequency of rhythmic activity within the alpha band." Electroencephalography and clinical neurophysiology **42**(1): 77-83.
- Pfurtscheller, G. and C. Neuper (1992). "Simultaneous EEG 10 Hz desynchronization and 40 Hz synchronization during finger movements." Neuroreport **3**(12): 1057-60.
- Pfurtscheller, G. and C. Neuper (1994). "Event-related synchronization of mu rhythm in the EEG over the cortical hand area in man." Neuroscience Letters **174**(1): 93-6.
- Pfurtscheller, G. and C. Neuper (1997). "Motor imagery activates primary sensorimotor area in humans." Neuroscience Letters **239**(2-3): 65-8.
- Polich, J. (1986). "P300 development from auditory stimuli." Psychophysiology **23**(5): 590-7.
- Polich, J., P. C. Ellerson and J. Cohen (1996). "P300, stimulus intensity, modality, and probability." Int.J Psychophysiol. **23**(1-2): 55-62.
- Ponton, C. W., J. J. Eggermont, B. Kwong and M. Don (2000). "Maturation of human central auditory system activity: evidence from multi-channel evoked potentials." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology **111**(2): 220-36.

- Price, C., G. Thierry and T. Griffiths (2005). "Speech-specific auditory processing: where is it?" Trends Cogn Sci 9(6): 271-6.
- Price, C. J., R. J. Wise, E. A. Warburton, C. J. Moore, D. Howard, K. Patterson, R. S. Frackowiak and K. J. Friston (1996). "Hearing and saying. The functional neuro-anatomy of auditory word processing." Brain 119(Pt 3): 919-931.
- Puce, A., T. Allison, M. Asgari, J. C. Gore and G. McCarthy (1996). "Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study." J.Neurosci. 16(16): 5205-5215.
- Puce, A., T. Allison, S. Bentin, J. C. Gore and G. McCarthy (1998). "Temporal cortex activation in humans viewing eye and mouth movements." Journal.of.Neuroscience Vol 18(6): 2188-2199.
- Puce, A., T. Allison, J. C. Gore and G. McCarthy (1995). "Face-sensitive regions in human extrastriate cortex studied by functional MRI." J.Neurophysiol. 74(3): 1192-1199.
- Pulvermuller, F., W. Lutzenberger, H. Preissl and N. Birbaumer (1995). "Spectral responses in the gamma-band: physiological signs of higher cognitive processes?" Neuroreport. 6(15): 2059-2064.
- Radeau, M. (1994). Auditory-visual spatial interaction and modularity. Cahiers.de.Psychologie.Cognitive./Current.Psychology.of Cognition. 13: 3-51.
- Rademacher, J., V. S. Caviness, Jr., H. Steinmetz and A. M. Galaburda (1993). "Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology." Cereb Cortex 3(4): 313-29.
- Rademacher, J., P. Morosan, T. Schormann, A. Schleicher, C. Werner, H. J. Freund and K. Zilles (2001). "Probabilistic mapping and volume measurement of human primary auditory cortex." Neuroimage 13(4): 669-83.
- Rahn, E. and E. Basar (1993). "Enhancement of visual evoked potentials by stimulation during low prestimulus EEG stages." Int J Neurosci 72(1-2): 123-36.
- Rahn, E. and E. Basar (1993). "Prestimulus EEG-activity strongly influences the auditory evoked vertex response: a new method for selective averaging." Int J Neurosci 69(1-4): 207-20.
- Raichle, M. E. (1998). "Behind the scenes of functional brain imaging: a historical and physiological perspective." Proc Natl Acad Sci U S A 95(3): 765-72.

- Raichle, M. E. (2001). "Cognitive neuroscience. Bold insights." Nature **412**(6843): 128-30.
- Raij, T., K. Uutela and R. Hari (2000). "Audiovisual Integration of Letters in the Human Brain." Neuron **28**(2): 617-625.
- Rauschecker, J. P. (1998). "Cortical processing of complex sounds." Curr Opin Neurobiol **8**(4): 516-21.
- Rauschecker, J. P. and B. Tian (2000). "Mechanisms and streams for processing of "what" and "where" in auditory cortex." Proc Natl Acad Sci U S A **97**(22): 11800-6.
- Rauschecker, J. P., B. Tian and M. Hauser (1995). "Processing of complex sounds in the macaque nonprimary auditory cortex." Science **268**(5207): 111-4.
- Rauschecker, J. P., B. Tian, T. Pons and M. Mishkin (1997). "Serial and parallel processing in rhesus monkey auditory cortex." J Comp Neurol **382**(1): 89-103.
- Recanzone, G. H., C. E. Schreiner, M. L. Sutter, R. E. Beitel and M. M. Merzenich (1999). "Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey." J Comp Neurol **415**(4): 460-81.
- Regan, D. (1989). Human Brain Electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine. New York, Elsevier.
- Reite, M., J. E. Zimmerman, J. Edrich and J. Zimmerman (1976). "The human magnetoencephalogram: some EEG and related correlations." Electroencephalogr Clin Neurophysiol **40**(1): 59-66.
- Ribary, U., A. A. Ioannides, K. D. Singh, R. Hasson, J. P. Bolton, F. Lado, A. Mogilner and R. Llinas (1991). "Magnetic field tomography of coherent thalamocortical 40-Hz oscillations in humans." Proc Natl Acad Sci U S A **88**(24): 11037-41.
- Rizzolatti, G., L. Fadiga, L. Fogassi and V. Gallese (1999). "Resonance behaviors and mirror neurons." Arch Ital Biol **137**(2-3): 85-100.
- Rizzolatti, G., L. Fadiga, V. Gallese and L. Fogassi (1996). "Premotor cortex and the recognition of motor actions." Brain research. Cognitive brain research **3**(2): 131-41.
- Rizzolatti, G., L. Fadiga, V. Gallese and L. Fogassi (1996). "Premotor cortex and the recognition of motor actions." Brain Res.Cogn.Brain Res. **3**(2): 131-141.

- Robinson, S. E. and J. Vrba, Eds. (1999). Functional neuroimaging by synthetic aperture magnetometry (SAM). Recent Advances in Biomagnetism. Sendai, Tohoku University Press.
- Rockland, K. S. and H. Ojima (2002). "Multimodal convergence in calcarine visual areas in macaque monkey." International Multisensory Research Forum, Geneva, Switzerland.
- Rosahl, S. K. and R. T. Knight (1995). "Role of prefrontal cortex in generation of the contingent negative variation." Cerebral cortex New York, N.Y. 1991 5(2): 123-34.
- Rosen, B. R., R. L. Buckner and A. M. Dale (1998). "Event-related functional MRI: past, present, and future." Proc Natl Acad Sci U S A 95(3): 773-80.
- Roskies, A. L. (1999). "The binding problem." Neuron 24(1): 7-9, 111-25.
- Rossi, S., F. Tecchio, P. Pasqualetti, M. Ulivelli, V. Pizzella, G. L. Romani, S. Passero, N. Battistini and P. M. Rossini (2002). "Somatosensory processing during movement observation in humans." Clinical neurophysiology official journal of the International Federation of Clinical Neurophysiology 113(1): 16-24.
- Rossini, P. M. and F. Pauri (2000). "Neuromagnetic integrated methods tracking human brain mechanisms of sensorimotor areas 'plastic' reorganisation." Brain research. Brain research reviews 33(2-3): 131-54.
- Rupp, A., S. Hack, A. Gutschalk, P. Schneider, T. W. Picton, C. Stippich and M. Scherg (2000). "Fast temporal interactions in human auditory cortex." Neuroreport 11(17): 3731-6.
- Salenius, S., M. Kajola, W. L. Thompson, S. Kosslyn and R. Hari (1995). "Reactivity of magnetic parieto-occipital alpha rhythm during visual imagery." Electroencephalogr Clin Neurophysiol 95(6): 453-62.
- Salenius, S., A. Schnitzler, R. Salmelin, V. Jousmaki and R. Hari (1997). "Modulation of human cortical rolandic rhythms during natural sensorimotor tasks." Neuroimage 5(3): 221-8.
- Sams, M. (1991). "Cortical responses to changes in auditory stimuli. Magnetoencephalographic studies." Acta Otolaryngol Suppl 491: 124-30; discussion 131.
- Sams, M., R. Aulanko, M. Hamalainen, R. Hari, O. V. Lounasmaa, S. T. Lu and J. Simola (1991). "Seeing speech: visual information from lip movements

- modifies activity in the human auditory cortex." Neuroscience Letters **127**(1): 141-145.
- Scherg, M., R. Hari and H. Hamalainen (1989). Frequency-specific sources of the auditory N19-P30-P50 response detected by a multiple source analysis of the evoked magnetic fields and potentials. New York, Plenum.
- Scherg, M. and D. von Cramon (1985). "A new interpretation of the generators of BAEP waves I-V: results of a spatio-temporal dipole model." Electroencephalogr Clin Neurophysiol **62**(4): 290-9.
- Scherg, M. and D. Von Cramon (1986). "Evoked dipole source potentials of the human auditory cortex." Electroencephalogr Clin Neurophysiol **65**(5): 344-60.
- Schmidt, D., B. J. Krause, F. M. Mottaghy, U. Halsband, H. Herzog, L. Tellmann and H. W. Muller-Gartner (2002). "Brain systems engaged in encoding and retrieval of word-pair associates independent of their imagery content or presentation modalities." Neuropsychologia **40**(4): 457-70.
- Schoenfeld, M. A., C. Tempelmann, A. Martinez, J. M. Hopf, C. Sattler, H. J. Heinze and S. A. Hillyard (2003). "Dynamics of feature binding during object-selective attention." Proc Natl Acad Sci U S A **100**(20): 11806-11.
- Schoffelen, J. M., R. Oostenveld and P. Fries (2005). "Neuronal coherence as a mechanism of effective corticospinal interaction." Science **308**(5718): 1111-3.
- Schroeder, C. E. and J. J. Foxe (2002). "The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex." Cognitive Brain Research **14**(1): 187-198.
- Schroger, E. and A. Widmann (1998). "Speeded responses to audiovisual signal changes result from bimodal integration." Psychophysiology **35**(6): 755-759.
- Schurmann, M., C. Basar-Eroglu and E. Basar (1997). "A possible role of evoked alpha in primary sensory processing: common properties of cat intracranial recordings and human EEG and MEG." Int J Psychophysiol **26**(1-3): 149-70.
- Scott, S. K., C. C. Blank, S. Rosen and R. J. Wise (2000). "Identification of a pathway for intelligible speech in the left temporal lobe." Brain **123 Pt 12**: 2400-6.
- Scott, S. K. and I. S. Johnsrude (2003). "The neuroanatomical and functional organization of speech perception." Trends Neurosci **26**(2): 100-7.
- Sekuler, R., A. B. Sekuler and R. Lau (1997). "Sound alters visual motion perception." Nature **385**(6614): 308.

- Seltzer, B. and D. N. Pandya (1978). "Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey." Brain Res. **149**(1): 1-24.
- Seltzer, B. and D. N. Pandya (1994). "Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: a retrograde tracer study." J.Comp.Neurol. **343**(3): 445-463.
- Senkowski, D., M. Gomez-Ramirez, P. Lakatos, G. R. Wylie, S. Molholm, C. E. Schroeder and J. J. Foxe (2005). A new method for analyzing multisensory effects in induced oscillatory responses in humans and simians. Cognitive Neuroscience Society, New York.
- Shadlen, M. N. and J. A. Movshon (1999). "Synchrony unbound: a critical evaluation of the temporal binding hypothesis." Neuron **24**(1): 67-77, 111-25.
- Shadlen, M. N. and W. T. Newsome (1998). "The variable discharge of cortical neurons: implications for connectivity, computation, and information coding." J Neurosci **18**(10): 3870-96.
- Shams, L., S. Iwaki, A. Chawla and J. Bhattacharya (2005). "Early modulation of visual cortex by sound: an MEG study." Neurosci Lett **378**(2): 76-81.
- Shannon, B. J. and R. L. Buckner (2004). "Functional-anatomic correlates of memory retrieval that suggest nontraditional processing roles for multiple distinct regions within posterior parietal cortex." J Neurosci **24**(45): 10084-92.
- Shapleske, J., S. L. Rossell, P. W. Woodruff and A. S. David (1999). "The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance." Brain research. Brain research reviews **29**(1): 26-49.
- Shofner, W. P. and E. D. Young (1985). "Excitatory/inhibitory response types in the cochlear nucleus: relationships to discharge patterns and responses to electrical stimulation of the auditory nerve." J Neurophysiol **54**(4): 917-39.
- Singer, W. (1993). "Synchronization of cortical activity and its putative role in information processing and learning." Annu.Rev.Physiol. **55**: 349-374.
- Singer, W. (1994). "Coherence as an organizing principle of cortical functions." Int.Rev.Neurobiol. **37**: 153-183.
- Singer, W. and C. M. Gray (1995). "Visual feature integration and the temporal correlation hypothesis." Annual Reviews in Neuroscience **18**: 555-586.

- Singh, K., G. Barnes, A. Hillebrand, E. Forde and A. Williams (2002). "Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response." Neuroimage **16(1):103-14(1)**: 102-14.
- Singh, K. D., G. R. Barnes and A. Hillebrand (2003). "Group imaging of task-related changes in cortical synchronisation using nonparametric permutation testing." Neuroimage **19(4)**: 1589-601.
- Specht, K. and J. Reul (2003). "Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task." Neuroimage **20(4)**: 1944-54.
- Stein, B. E. (1978). "Development and organization of multimodal representation in cat superior colliculus." Fed.Proc. **37(9)**: 2240-2245.
- Stein, B. E. (1981). "Organization of the rodent superior colliculus: some comparisons with other mammals." Behav.Brain Res. **3(2)**: 175-188.
- Stein, B. E., W. S. Huneycutt and M. A. Meredith (1988). "Neurons and behavior: The same rules of multisensory integration apply." Brain Research. **Vol 448(2)**: 355-358.
- Stein, B. E., W. Jiang and T. R. Stanford (2004). Multisensory integration in single neurons of the midbrain. Handbook of multisensory processing. G. A. Calvert, C. Spence and B. E. Stein. Cambridge, MIT Press: 243-262.
- Stein, B. E. and M. A. Meredith (1993). Merging of the Senses. Cambridge, MIT Press.
- Stein, B. E., M. A. Meredith, W. S. Huneycutt and L. McDade (1989). "Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli." Journal.of Cognitive.Neuroscience **Vol 1(1)**: 12-24.
- Stein, B. E., M. A. Meredith and M. T. Wallace (1993). "The visually responsive neuron and beyond: multisensory integration in cat and monkey." Prog.Brain Res. **95**: 79-90.
- Stein, B. E., M. W. Wallace, T. R. Stanford and W. Jiang (2002). "Cortex governs multisensory integration in the midbrain." Neuroscientist a review journal bringing neurobiology, neurology and psychiatry **8(4)**: 306-14.
- Stok, C. J., J. W. Meijis, M. J. Peters and W. L. Rutten (1986). "Model evaluation using electroencephalography and magnetoencephalography." Acta Otolaryngol Suppl **432**: 5-10.

- Sumby, W. H. and I. Pollack (1954). "Visual contribution to speech intelligibility in noise." Journal of the Acoustical Society of America **26**: 212-215.
- Summerfield, Q. (1979). "Use of visual information for phonetic perception." Phonetica. **36**(4-5): 314-331.
- Summerfield, Q. (1991). "Visual perception of phonetic gestures."
- Summerfield, Q. (1992). "Lipreading and audio-visual speech perception." Philos.Trans.R.Soc.Lond.B.Biol.Sci. **335**(1273): 71-78.
- Summerfield, Q. and P. Assmann (1987). "Auditory enhancement in speech perception."
- Summerfield, Q. and M. McGrath (1984). "Detection and resolution of audio-visual incompatibility in the perception of vowels." Quarterly.Journal.of.Experimental.Psychology.: Human Experimental.Psychology. Vol **36**(1-A): 51-74.
- Talairach, J. and P. Tournoux (1988). Co-planar stereotactic atlas of the human brain. Stuttgart: Thieme.
- Tallon-Baudry, C., O. Bertrand, C. Delpuech and J. Permier (1997). "Oscillatory gamma-band (30-70 Hz) activity induced by a visual search task in humans." J Neurosci **17**(2): 722-34.
- Tanji, J. (1996). "New concepts of the supplementary motor area." Curr Opin Neurobiol **6**(6): 782-7.
- Teder-Salejarvi, W. A., J. J. McDonald, F. Di Russo and S. A. Hillyard (2002). "An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings." Cognitive Brain Research **14**(1): 106-14.
- Tenke, C. E., C. E. Schroeder, J. C. Arezzo and H. G. Vaughan, Jr. (1993). "Interpretation of high-resolution current source density profiles: a simulation of sublaminar contributions to the visual evoked potential." Exp Brain Res **94**(2): 183-92.
- Tettamanti, M., G. Buccino, M. C. Saccuman, V. Gallese, M. Danna, P. Scifo, F. Fazio, G. Rizzolatti, S. F. Cappa and D. Perani (2005). "Listening to action-related sentences activates fronto-parietal motor circuits." J Cogn Neurosci **17**(2): 273-81.
- Thesen, T., J. F. Vibell, G. A. Calvert and R. A. Osterbauer (2004). "Neuroimaging of multisensory processing in vision, audition, touch and olfaction." Cognitive Processing **5**(2): 84-93.

- Tiitinen, H., J. Sinkkonen, P. May and R. Naatanen (1994). "The auditory transient 40-Hz response is insensitive to changes in stimulus features." Neuroreport 6(1): 190-2.
- Todd, J. W. (1912). "Reaction to multiple stimuli." Archives of Psychology, N.Y. 3(25).
- Treisman, A. (1996). "The binding problem." Curr.Opin.Neurobiol. 6(2): 171-178.
- Treisman, A. (1998). "Feature binding, attention and object perception." Philos.Trans.R Soc.Lond.B.Biol.Sci. 353(1373): 1295-1306.
- Turner, B., M. M and K. M (1980). "Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey." J Comp Neurol 191(515 543).
- Turner, R. (1995). "Functional mapping of the human brain with magnetic resonance imaging." Seminars in the Neurosciences 7: 179-194.
- Ulbert, I., G. Karmos, G. Heit and E. Halgren (2001). "Early discrimination of coherent versus incoherent motion by multiunit and synaptic activity in human putative MT+." Hum Brain Mapp 13(4): 226-38.
- Ungerleider, L. G. and J. V. Haxby (1994). "'What' and 'where' in the human brain." Curr.Opin.Neurobiol. 4(2): 157-165.
- Uttal, W.R. (2001). *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, MA : MIT Press
- van Atteveldt, N., E. Formisano, R. Goebel and L. Blomert (2004). "Integration of letters and speech sounds in the human brain." Neuron 43(2): 271-82.
- Van Veen, B. D. and K. M. Buckley (1988). "Beamforming:a versatile approach to spatial filtering." IEEE ASSP Magazine 5(4): 4-24.
- Van Veen, B. D., W. van Drongelen, M. Yuchtman and A. Suzuki (1997). "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering." IEEE transactions on bio medical engineering 44(9): 867-80.
- van Wassenhove, V., K. W. Grant and D. Poeppel (2005). "Visual speech speeds up the neural processing of auditory speech." Proc Natl Acad Sci U S A 102(4): 1181-6.
- Varela, F., J. P. Lachaux, E. Rodriguez and J. Martinerie (2001). "The brainweb: phase synchronization and large-scale integration." Nat Rev Neurosci 2(4): 229-39.

- Vitacco, D., D. Brandeis, R. Pascual-Marqui and E. Martin (2002). "Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing." Hum Brain Mapp 17(1): 4-12.
- Wagner, K., L. Frings, A. Quiske, J. Unterrainer, R. Schwarzwald, J. Spreer, U. Halsband and A. Schulze-Bonhage (2005). "The reliability of fMRI activations in the medial temporal lobes in a verbal episodic memory task." Neuroimage.
- Wallace, M. T., M. A. Meredith and B. E. Stein (1992). "Integration of multiple sensory modalities in cat cortex." Exp.Brain Res. 91(3): 484-488.
- Wallace, M. T., M. A. Meredith and B. E. Stein (1993). "Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus." J Neurophysiol. 69(6): 1797-1809.
- Watanabe, J. and E. Iwai (1991). "Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task." Brain Res.Bull. 26(4): 583-592.
- Watanabe, M. (1992). "Frontal units of the monkey coding the associative significance of visual and auditory stimuli." Exp.Brain Res. 89(2): 233-247.
- Watkins, K. E., A. P. Strafella and T. Paus (2003). "Seeing and hearing speech excites the motor system involved in speech production." Neuropsychologia 41(8): 989-94.
- Watson, J. D., R. Myers, R. S. Frackowiak, J. V. Hajnal, R. P. Woods, J. C. Mazziotta, S. Shipp and S. Zeki (1993). "Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging." Cereb Cortex 3(2): 79-94.
- Welch, R. B. and D. H. Warren (1986). Intersensory interactions. In Handbook of Perception and Human Performance (K R Kaufman & J P Thomas (eds), John Wiley & Sons. Volume 1: Sensory Processes and Perception: 1-36.
- Westbury, C. F., R. J. Zatorre and A. C. Evans (1999). "Quantifying variability in the planum temporale: a probability map." Cereb Cortex 9(4): 392-405.
- Wilkinson, L. K., M. A. Meredith and B. E. Stein (1996). "The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior." Exp.Brain Res. 112(1): 1-10.

- Wilson, F. A., S. P. O'Scalaidhe and P. S. Goldman-Rakic (1994). "Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex." Proc Natl Acad Sci U S A **91**(9): 4009-13.
- Wise, R. J., S. K. Scott, S. C. Blank, C. J. Mummery, K. Murphy and E. A. Warburton (2001). "Separate neural subsystems within 'Wernicke's area'." Brain **124**(Pt 1): 83-95.
- Woldorff, M. G., C. C. Gallen, S. A. Hampson, S. A. Hillyard, C. Pantev, D. Sobel and F. E. Bloom (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.
- Woldorff, M. G., M. Liotti, M. Seabolt, L. Busse, J. L. Lancaster and P. T. Fox (2002). "The temporal dynamics of the effects in occipital cortex of visual-spatial selective attention." Brain research. Cognitive brain research **15**(1): 1-15.
- Woolrich, M. W., T. E. Behrens, C. F. Beckmann, M. Jenkinson and S. M. Smith (2004). "Multilevel linear modelling for fMRI group analysis using Bayesian inference." Neuroimage **21**(4): 1732-47.
- Woolrich, M. W., T. E. Behrens and S. M. Smith (2004). "Constrained linear basis sets for HRF modelling using Variational Bayes." Neuroimage **21**(4): 1748-61.
- Woolrich, M. W., B. D. Ripley, M. Brady and S. M. Smith (2001). "Temporal autocorrelation in univariate linear modeling of fMRI data." Neuroimage **14**(6): 1370-86.
- Worsley, K. J., A. C. Evans, S. Marrett and P. Neelin (1992). "A three-dimensional statistical analysis for CBF activation studies in human brain." Journal of cerebral blood flow and metabolism official journal of the International Society of Cerebral Blood Flow and Metabolism **12**(6): 900-18.
- Wright, T. M., K. A. Pelphrey, T. Allison, M. J. McKeown and G. McCarthy (2003). "Polysensory interactions along lateral temporal regions evoked by audiovisual speech." Cereb Cortex **13**(10): 1034-43.

- Young, E. D. and M. B. Sachs (1979). "Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers." J Acoust Soc Am **66**(5): 1381-1403.
- Yvert, B., A. Crouzeix, O. Bertrand, A. Seither-Preisler and C. Pantev (2001). "Multiple supratemporal sources of magnetic and electric auditory evoked middle latency components in humans." Cereb Cortex **11**(5): 411-23.
- Zatorre, R. J. and P. Belin (2001). "Spectral and temporal processing in human auditory cortex." Cereb Cortex **11**(10): 946-53.
- Zatorre, R. J., A. C. Evans, E. Meyer and A. Gjedde (1992). "Lateralization of phonetic and pitch discrimination in speech processing." Science **256**(5058): 846-9.
- Zatorre, R. J., E. Meyer, A. Gjedde and A. C. Evans (1996). "PET studies of phonetic processing of speech: review, replication, and reanalysis." Cereb.Cortex. **6**(1): 21-30.

Appendix
