Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Investigating the Temporal Dynamics of Auditory Processing and Selective Attention using Continuous and Ecologically Valid Stimuli: A Neural Engineering Approach.

A dissertation submitted to the University of Dublin

for the degree of Doctor of Philosophy

2011

Alan James Power

Department of Electronic and Electrical Engineering,

Trinity College Dublin

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Summary and Major Findings

Auditory Evoked Potentials (AEPs) are a measure of the neural activity in response to auditory stimulation and are used in many clinical and research applications. They are obtained from the electroencephalogram (EEG) by the repeated presentation of discrete auditory stimuli such as tones or noise bursts lasting from 5ms to 60ms. EEG activity measured in responses to each individual stimulus is averaged in order to eliminate neural activity not time-locked to the auditory events. The remaining activity is known as an AEP.

A standard AEP consists of a series of peaks and troughs, called components of the response, which researchers and clinicians study in order to investigate auditory processing in the brain. However, given that the auditory system has evolved to be able to process an infinitude of sounds ranging from discrete events to continuous stimuli with much higher order information, the use of discrete events alone may not be best suited to investigate auditory function. To circumvent some of these problems the Auditory Steady-State Response (ASSR) is employed. ASSRs are neural responses to periodic stimuli at the frequency of stimulation. This, unfortunately, results in the loss of the important components mentioned above. The only information obtainable is the power and phase of the responses. Thus the ability to obtain temporally detailed responses to complex and continuous stimuli would be advantageous. To this end the research outlined in this thesis introduces and describes the Auditory Evoked Spread Spectrum Analysis (AESPA) method.

The AESPA method utilises system identification method in order to characterise the auditory system using continuous stimuli. The AESPA method is

compared to the standard AEP method and its advantages and disadvantages are discussed. Furthermore the components of the novel AESPA response are analysed, defined and possible neural generators identified.

In addition, the AESPA method allows for the extraction of separate responses to simultaneously presented continuous stimuli. Exploiting this aspect of AESPA allows for the implementation of novel paradigms probing auditory attention effects. There is debate as to the source of attentional effects identified using the standard AEP method. In this regard the effects of endogenous auditory spatial attention on competing AESPA stimuli are investigated. These effects are localised to auditory cortex (AC) and are found to be due to a modulation of obligatory sensory activity in AC as opposed to a temporally overlapping influence of a separate attentional generator.

AESPA also allows for the extraction of responses to continuous natural speech taking into account the extended temporal nature of speech. Using this advantage data from an ecologically valid cocktail party like scenario are presented. These data show the modulation of activity in AC when two speech streams are presented simultaneously in an attention paradigm. In particular the likely suppression of auditory processing of the unattended stream in AC is highlighted. Thus the establishment of the AESPA method and the outcomes of its application to the investigation of phenomena that are of significant importance to the understanding of auditory physiology demonstrate AESPA to be an important and relevant tool in auditory neuroscience.

Acknowledgements

Firstly I'd like to thank my supervisor Prof. Richard Reilly without whose support, encouragement and good humour over the past several years this thesis would never have come to fruition.

I would also like to thank Dr. Ed Lalor whose enthusiasm and generosity with time, suggestions and advice were invaluable to me over the course of my PhD. His genuine passion for research and understanding the brain is palpable and I have undoubtedly learned a lot from him.

I would also like to thank Dr. Rob Whelan for always being willing to answer my boring stats questions. I'll master those stats yet! I'd also like to thank everybody in the Neural Engineering Group for their willingness to help with studies as well as the ever enjoyable lab nights out!

This work could not have been completed without the support of the Irish Research Council for Science, Engineering and Technology (IRCSET). This thesis is the result of a number of years of research in the Multimodal Signal Processing Lab in UCD and then the Neural Engineering Lab in TCD and I'd like to acknowledge the support of the staff of both universities who have helped me during the course of my PhD research.

I'd like to thank *tous les goys* and the UCD eng. crew for the general messing over the years which provided essential respite from the late nights in the lab. I'd also like to thank the *new amusement* boys for the mixed metaphors and, of course, the art of needle!.....an admission which will undoubtedly meet with some serious needle!

Last but most definitely not least I'd like to thank my parents, John and Clare, for their support and encouragement over the course of my PhD and my life. This would have been impossible without you. Thank you!

"Don't you wonder sometimes about sound and vision?"

David Bowie, *Sound and Vision*, 1977

Contents

Acronyms

AC	Auditory Cortex
AEP	Auditory Evoked Potential
AESPA	Auditory Evoked Spread Spectrum Analysis
AM	Amplitude Modulation
AN	Auditory Nerve
AP	Action Potential
ASA	Auditory Scene Analysis
ASSR	Auditory Steady-State Response
BBN	Broadband Noise
CN	Cochlear Nucleus
CNS	Central Nervous System
EEG	Electroencephalogram
ERP	Event Related Potential
EPSP	Excitatory Post-Synaptic Potential
FFR	Frequency Following Response
FM	Frequency Modulation
fMRI	functional Magnetic Resonance Imaging
FWE	Family-wise Error
GWN	Gaussian White Noise
GFP	Global Field Power
HG	Heschl's Gyrus
IC	Inferior Colliculus
IPSP	Inhibitory Post-Synaptic Potential

ILD	Inter-aural Level Difference
ITD	Inter-aural Time Difference
LL	Lateral Lemniscus
MEG	Magnetoencephalogram
MGN	Medial Geniculate Nucleus
MM	Mixed Modulation
NT	Neurotransmitter
PAC	Primary Auditory Cortex
PSP	Post-Synaptic Potential
RMS	Root Mean Square
SEP	Somato-sensory Evoked Potential
SOC	Superior Olivary Complex
SSA	Stimulus Specific Adaptation
SSR	Steady-State Response
SNR	Signal-to-Noise Ratio
SPM	Statistical Parametrical Mapping
TTG	Transverse Temporal Gyri
VEP	Visual Evoked Potential

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

"The brain is the organ of destiny. It holds within its humming mechanism secrets that will determine the future of the human race." Wilder Penfield, first director of the Montreal Neurological Institute.

1

Introduction

1.1 Introduction

The brain is an incredibly intricate system which enables us to function in, and make sense of, the world around us. Our senses are constantly bombarded with information from the surrounding environment. Often the information presented to just a single sense consists of a disparate array of congruent and incongruent sources. Highly complex processing is required to make sense of all this information and to separate the useful information from the unwanted information. Our brains continually and continuously carry out this task with ease. 100 billion neurons and innumerable interconnections make up the huge computing power of this spongy organ. The complexity and versatility of function carried out by the brain far exceeds any man made computer and it is due to this complexity that brain function is not fully understood.

A lot of progress has been made since the early theories of brain function espoused by Aristotle and others that the brain was a heat regulator and that sensation actually resided in the heart. Input from many different fields of research and expertise has contributed to this and psychologists, neuroscientists, engineers and others have sought to add to the understanding of brain function through investigation of all levels of neural activity from single neurons right up to sensory and cognitive networks of brain activity, using hemodynamic responses (fMRI) and electrophysiological (EEG) and magneto-physiological (MEG) responses.

When investigating auditory processing with EEG or MEG auditory evoked potentials (AEPs) are often employed. AEPs are obtained by averaging the EEG over many discrete stimulus presentations. AEPs are an important tool in understanding auditory function and are also used to investigate the clinical disorders such as depression (Gallinat et al. 2000), Alzheimer's disease (Cancelli et al. 2006), schizophrenia (Nagamato et al. 1991), multiple sclerosis (Chiappa et al. 1980), and anxiety disorder (Drake et al. 1991). Investigations into how AEPs change with disease can lead to a greater understanding of the disease. They also have the potential to be utilised as diagnostic indicators of disease and help to follow disease progression or improvement (Kiiski et al., 2011). Further to this AEPs have been used extensively in the investigation of the mechanisms and dynamics of auditory attention and have informed us a great deal as to the processes involved in self-directed attention (e.g. Hillyard et al., 1973) and attentional capture of salient stimuli (e.g. Escera et al., 2000). There some drawbacks to the standard AEP method,

however, such as the limited flexibility in stimulus and experimental design when restricted to using discrete stimuli as well as when trying to interrogate auditory function with more natural stimuli such as speech. AEPs are, however, essentially a method by which the properties of the auditory system are investigated. That said, many engineering methods also exist to investigate the properties of an unknown system. The research presented in this thesis describes a system identification method for the investigation of auditory processing resulting in temporally detailed responses to non-discrete stimuli. The method introduced is known as Auditory Evoked Spread Spectrum Analysis (AESPA). The AESPA method addresses some of the limitations of the standard AEP and allows for elicitation of responses to more natural and ecologically valid stimuli than are currently employed in auditory neuroscience. Furthermore, it allows for the extraction of responses to truly competing and simultaneous stimuli. These advantages allow for the design of more ecological experimental scenarios. Results of three studies are presented 1) the AESPA method is described and the responses obtained are characterised and compared and contrasted to standard Auditory Evoked Potentials (AEPs) 2) the effect of endogenous auditory spatial attention on auditory sensory activity employing continuous and truly competing stimuli. 3) the investigation of the attentional dynamics of the "cocktail party problem" when two natural speech streams are presented competitively.

A common theme throughout this thesis is the use of natural and ecologically valid stimuli in experimental protocols for the investigation of auditory function; in particular auditory attention and auditory scene analysis (ASA). The use of more

natural stimuli to investigate brain function is of great importance as to truly understand the brain we must be able to assess its functionality in controlled yet natural manner. The establishment of the AESPA method and the outcomes of its application to the investigation of auditory physiology in novel and more natural scenarios highlight the AESPA as a useful and complementary tool to standard AEP/EEG techniques for investigating the auditory system.

1.2 Thesis Outline

Chapter 2 seeks to outline the anatomy and function of the auditory system. Here the basic principles of the electroencephalogram (EEG) and the auditory evoked potential (AEP) technique are introduced. Auditory attention and the cocktail party problem are described and related effects on the AEP are outlined. The discrete nature of the standard AEP method and the methodological and ecological limitations associated with that approach calls for a new method which can take account of the extended temporal nature of real life stimuli.

Chapter 3 introduces the AESPA method and elucidates how it overcomes some of the limitations of the standard AEP technique. The AESPA and AEP responses are compared and contrasted and advantages and disadvantages of the AESPA method are outlined along with some possible clinical and research applications.

Chapter 4 highlights the debate in the literature regarding the generators involved in the N1 attention effect of the standard AEP. Specifically it examines the debate as to whether obligatory sensory activity is modulated by attention or not

(Hillyard et al., 1973 vs. Näätänen et al., 1987). By employing the AESPA in an endogenous auditory spatial attention paradigm effects on obligatory sensory processing in primary auditory cortex (PAC) are shown. Thus, endogenous attention does, at least in part, result in the modulation of sensory activity.

Chapter 5 seeks to investigate the cocktail party problem using two competing natural speech stimuli. The AESPA method was applied to extract responses to the simultaneously presented speech streams. A strong attentional effect on activity in AC is seen and indicates a suppression of sensory information related to the unattended speech stream in AC. This effect has a left bias and is likely related to the preferences of left auditory cortex for processing of fine temporal information. Furthermore this deployment of attention likely involves top-down control from higher order areas, including language areas, in the left hemisphere, and thus the effect that we see in AC likely reflects the effect of this contribution on early auditory processing.

Finally Chapter 6 summaries the thesis and the advances achieved therein. Further applications and possible future work employing the AESPA method are also discussed. The use of AESPA to probe differing effects due to spatial and non-spatial attention is outlined and preliminary results are outlined. The application of AESPA to the investigation of the of dyslexia/specific language impairment is also considered as are investigations into the AEP N1 deficit and impaired sensory gating in schizophrenia.

1.3 Contributions of this Thesis

- A novel method for interrogating auditory function using EEG – Auditory Evoked Spread Spectrum Analysis (AESPA). This method allows for the use of more natural stimuli than are commonly used in auditory neuroscience. The responses obtained using AESPA are characterised in terms of components and sources thereof.
- The application of the AESPA method to the investigation of endogenous attention. Results show that spatial attention does indeed modulate obligatory sensory processing in primary auditory cortex. There has been debate in the literature in this regard.
- The application of the AESPA method to the investigation of the Cocktail party problem using natural speech. The attention effects identified highlight the importance of the suppression of irrelevant information in primary auditory cortex.

1.4 Publications arising from this thesis

Journal:

Power AJ, Forde E-J, Reilly RB, Foxe JJ, Lalor EC. Attention to Speech Modulates Early Activity in Primary Auditory Cortex. S*ubmitted.*

Power AJ, Lalor EC, Reilly RB. Endogenous Auditory Spatial Attention Modulates Obligatory Sensory Activity in Primary Auditory Cortex., *Cereb Cortex*, 21:1223-1230, 2011.

Lalor EC, **Power AJ**, Reilly RB, Foxe JJ. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli., *J Neurophysiol*, 102:349-359, 2009.

International Peer-reviewed Conferences:

Power AJ, Lalor EC, Reilly RB, Extracting responses to Simultaneously presented Continuous Stimuli: An Auditory Attention Study. Proceeding of 4th Intl IEEE/EMBS Conference on Neural Engineering, Antalya, Turkey, 2009.

Power AJ, Lalor EC, Reilly RB. Eliciting Audio Evoked Potentials Using Continuous Stimuli. Proceeding of 29th Intl IEEE Engineering in Medicine and Biology Conference, Lyon, France, 2007.

2

The Brain, the Auditory System and the Electroencephalogram.

2.1 The Brain

The central nervous system (CNS) consists of the spinal cord and the brain. The brain is the part of the CNS that is in the skull. The brain can be broken down into three subdivisions: the hindbrain, the midbrain and the forebrain (see Fig. 2.1). The hindbrain is in turn made up of the cerebellum, the medulla oblongata and the pons. The hindbrain and the midbrain together make up the brainstem. The forebrain can be subdivided into two regions: the diencephalon and the telencephalon. The thalamus is located in the diencephalon and acts as a relay between various subcortical areas and the cortex. The telencephalon is made up of cerebral cortex and it is where the most complex sensory processing takes place as well as higher order tasks such as executive function and memory.

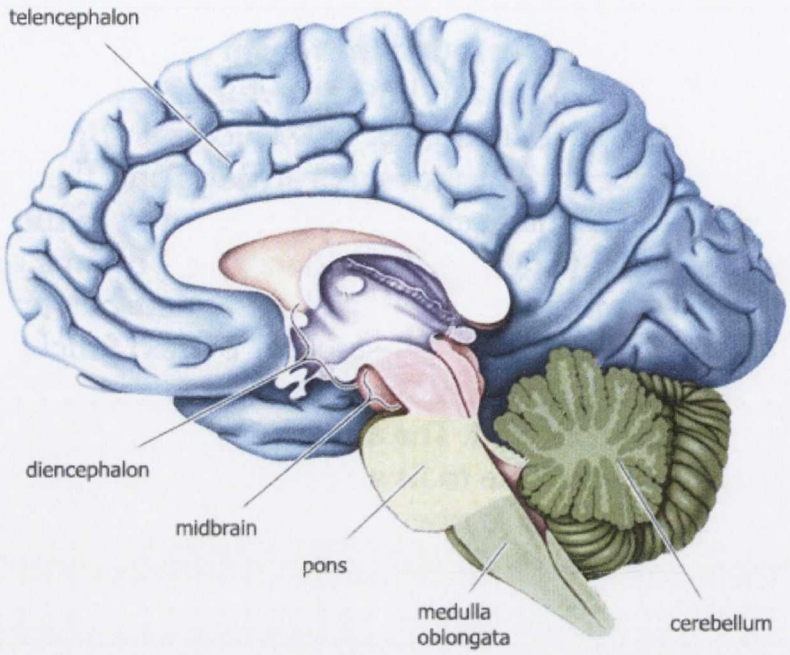

Figure 2.1 The internal structure of the brain (adapted from Bear et al., 2007).

The cerebral cortex is typically divided into four region based on anatomical landmarks. These landmarks are prominent fissures (known as sulci. Singular: sulcus) and protrusions (known as gyri. Singular: gyrus) in the cortical surface. Based on their location the regions are known as the frontal, parietal, occipital and temporal lobes (see Fig. 2.2). The frontal lobe is involved in higher functions such as attention, planning, language and movement. The parietal lobe processes lots of sensory information allowing us to perceive the world and our place in it. The occipital lobe deals primarily with vision; it is here that signals from the eyes become transformed into a useful representation. Finally, the temporal lobes focus on sound and language, and, by way of their connection with hippocampus in the diencephalon, memory formation and retrieval (Gibb, 2007).

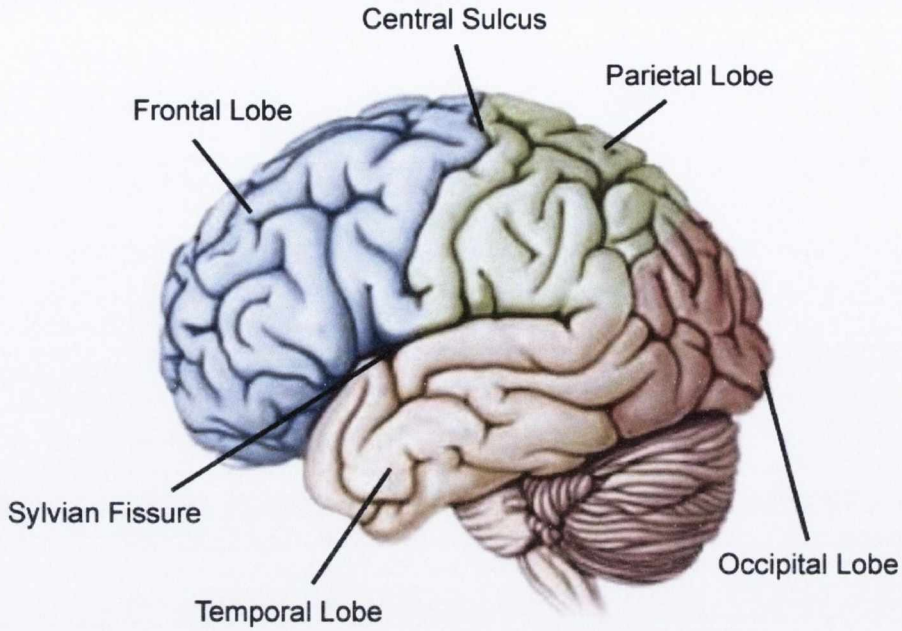

Figure 2.2 The lobes of the cerebral cortex (adapted from Bear et al., 2007).

2.2 The Auditory System

The external, middle and inner ear

The auditory system is made up of the external ear, the middle ear, the inner ear and the central auditory system. The external ear consists of the pinna, the ear canal and the external face of the tympanic membrane (ear drum) (see Fig. 2.3). The pinna acts to collect and focus the sound waves down the ear canal towards the middle ear which acts as an impedance matching network, i.e. allows for efficient transfer of energy, between the open air of the external ear and the fluid filled cochlea of the inner ear. The middle ear consists of the internal side of the tympanic membrane, three bones know as the ossciles and the oval window of the cochlea. The three ossciles bones are the malleus, the incus and the stapes. The ear drum is fused to the malleus which connects to the incus which in turn connects to the stapes. The foot plate of the stapes then connects to the inner ear via the oval window of the

cochlea (see Fig. 2.3). The inner ear contains the osseous labyrinth. The osseous labyrinth is made up of the semicircular canals of the vestibular system and the cochlear canal. The cochlea is a fluid-filled snail shell shaped structure which makes 2.5 turns from base to apex and contains the auditory sensory organ known as the organ of Corti (see Fig 2.4a). The organ of Corti sits on the basilar membrane and consists of inner and outer hair cells (see Fig 2.4a,b). The inner hair cells are responsible for hearing and sensitive to mechanical stimulation caused by the deformation of the basilar membrane. Bending of the hair cells in one direction causes depolarisation and bending in the other direction causes hyperpolarisation (see Fig 2.4b). Depolarisation is a change in a cell's membrane potential (i.e. the electrical potential difference between the interior and exterior walls of a cell) making it more positive. Hyperpolarisation is the opposite of this whereby the membrane potential is made more negative.

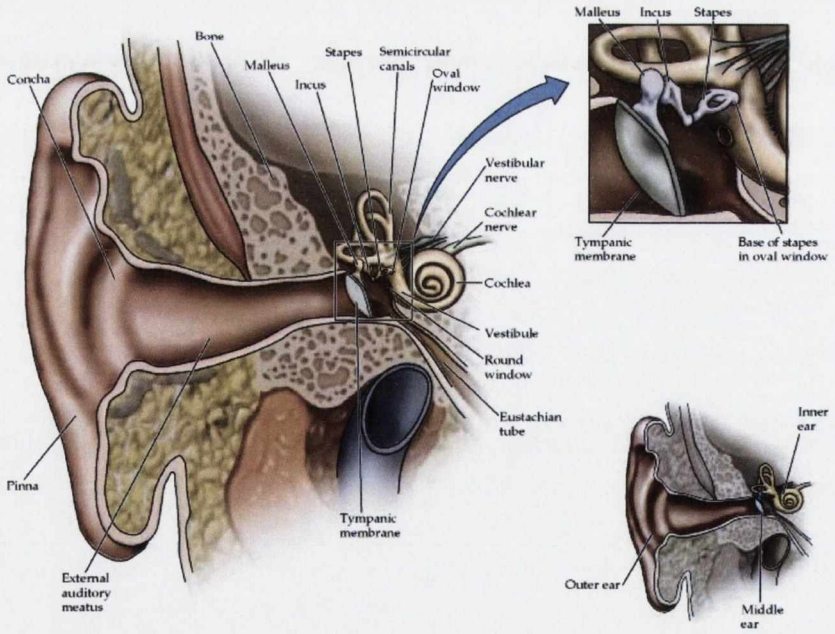

Figure 2.3 The outer, middle and inner ear (Purves et al., 2001).

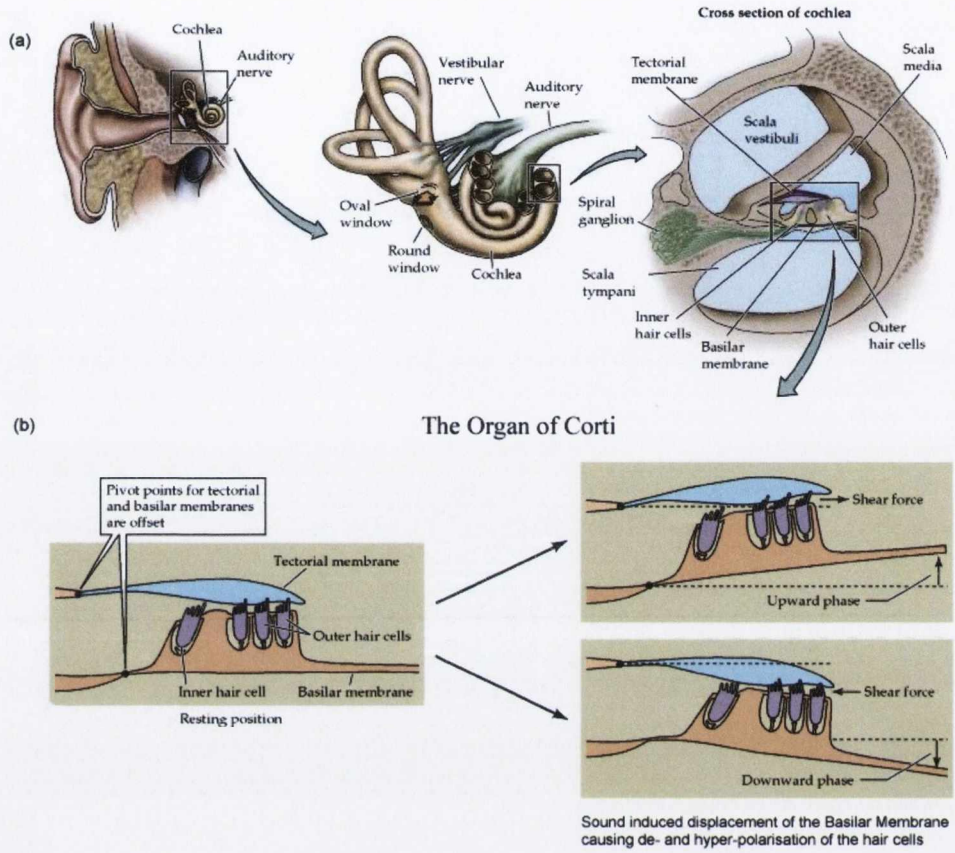

Figure 2.4 (a) The Cochlea, the organ of Corti and **(b)** inner hair cell displacement (adapted from Purves et al., 2001)

Sound waves travelling down the ear canal cause the tympanic membrane to vibrate which in turn causes the ossciles to vibrate. The vibration of the stapes against the oval window of the cochlea sets up a travelling wave in the cochlear fluid which propagates from the base towards the apex of the basilar membrane, growing in amplitude and slowing in velocity until a point of maximum displacement is reached. High frequencies maximally displace the base of the basilar membrane and low frequencies maximally displace the apex giving rise to a topographical mapping of frequency (see Fig2.5). This tonotopic organisation continues in most ascending

auditory structures between cochlea and cortex and indeed primary auditory cortex is tonotopically organised.

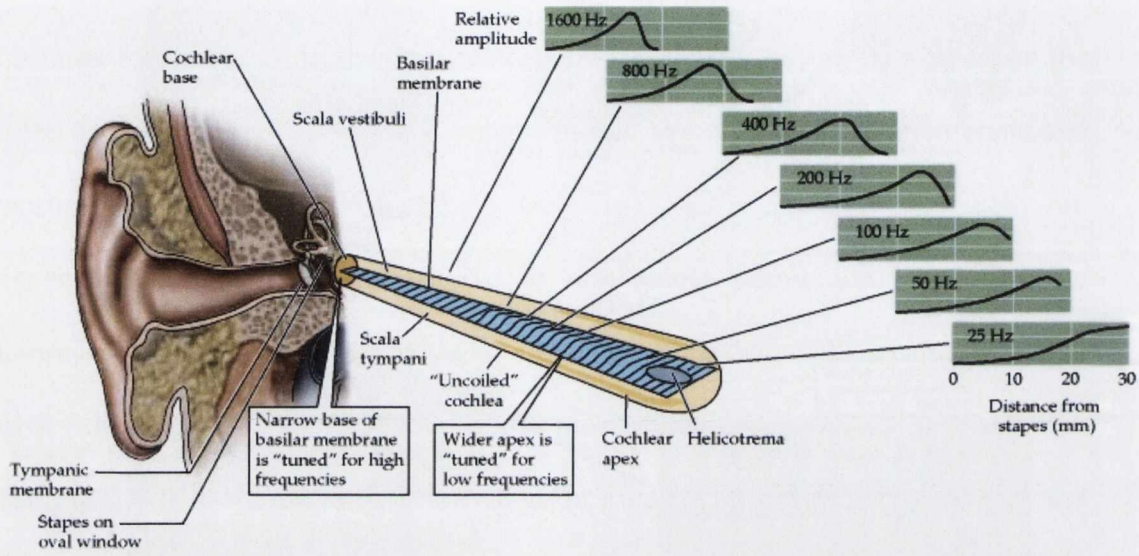

Figure 2.5 The uncoiled cochlea showing the frequency organisation from base to apex (Purves et al., 2001).

The Central Auditory System

The progression of the neural processing representation of sound through the auditory system continues from the basilar membrane via many different nuclei from brainstem to midbrain to auditory cortex to adjacent portions of the temporal and onto the frontal lobes. Firstly, spiral ganglion cells innervate the inner hair cells of the organ of Corti, which in turn send nerve projections to the central nervous system via the auditory nerve (AN). Depolarisation of the hair cells due to the displacement of the basilar membrane excites these cochlear nerve fibres. Axons of the spiral ganglion cells synapse in the cochlear nucleus (CN). The role of the CN is to receive the information delivered by the AN fibres, to process this information and distribute the modified signals to higher auditory nuclei. Projections from the anteroventral division of the CN essentially terminate in the principal nuclei of the superior olivary

complex (SOC). The various nuclei of the SOC receive their ascending input from the two CNs. Therefore they are the first stage of auditory pathways where a massive convergence of information from both ears takes place. With the exception of binaural processing the functional aspects of the SOC are unclear however. Carrying on from the CN and SOC neurons travel in a bundle called the lateral lemniscus (LL) towards the inferior colliculus (IC), which is in the in the midbrain. The main output projections from the LL are to the opposite LL via the commissure of Probst, and the IC on both sides as well as descending projections towards the SOC. The IC is an obligatory relay nucleus along ascending auditory pathways from lower brainstem to thalamus, since the vast majority of axons originating in the CN, SOC and LL terminate in the IC. It is thought all aspects of neuronal coding of sound are affected by some sort of transformation in the IC and representational maps of neuronal response features are formed, from which perceptional qualities of sound may be derived (Ehret & Romand, 1997). Following from IC the auditory thalamus receives its ascending afferents mainly from the IC. The principal auditory thalamic region is the Medial Geniculate Nucleus (MGN) and acts as a relay between IC and auditory cortex (AC). Thus axons from MGN project to AC on the upper surface of the superior temporal gyrus (Heschl's gyrus). AC is located inside the lateral sulcus of the brain which contains gyri known as the transverse temporal gyri (TTG) or Heschl's gyri. AC is generally divided into three subregions: the core, the belt and the parabelt. The core region is deep within the lateral sulcus and is made up of the medial 2/3 of the most anterior HG. This receives input from the MGN. The core is tonotopically organised. It has been suggested that the majority of sensory processing has been carried out by the time the signals reach AC (Nelken, 2004) and

that AC is (primarily) involved in sensory memory as well as some lower level sensory processing. The core is surrounded by a narrow band of cortex which receives input from the MGN as well as the core itself (see Fig.2.7). This is the belt region. The belt is thought to have 7 or 8 distinct fields each with distinct representation of the cochlea. Belt areas respond less vigorously to tone than the core but that said tonotopic gradients are present. Belt regions are strongly interconnected with one another but principally distribute to parabelt regions (a third level of cortical processing). Parabelt is located just lateral to the belt and is defined as that region of the temporal lobe where injections of tracers label large numbers of neurons in the belt but few or no neurons in the core (i.e. areas connected to the belt but not the core). The parabelt is interconnected with adjacent portions of the temporal and parietal lobes as well as several regions in the frontal lobe.

Temporal lobe connections:

1. Nearby cortex of the upper and lower banks of the superior temporal sulcus (STS).
2. Caudal regions close to middle temporal (MT) visual regions but there is no evidence for connections with these visual areas themselves.
3. Rostral cortex in superior temporal polysensory (STP) cortex where neurons respond to auditory, visual and even somatosenory stimulation.

Frontal connections:

1. Near or within the frontal eye fields (FEFs). This area is important for directing gaze and obviously sounds in space would be of visual interest.

2. Dorsolateral prefrontal cortex of the principal sulcus in monkey. This area is important for working memory and some locations in this prefrontal area respond to auditory stimuli.
3. More rostral and ventral frontal areas which may be involved in multimodal characterisation of objects and though also to subserve working memory for non-spatial tasks
4. Orbital-frontal cortex which is a multimodal region with role in assigning value to stimuli and is associated with reward and is considered emotive or motivational in function. Many neurons in this region respond to auditory stimuli.

It has also been suggested that there are two processing stream in auditory system: a ventral "what" stream and a dorsal "where" stream with connections from caudal parabelt travelling to areas primarily related to spatial processing and those from rostral travelling to areas primarily related to object recognition/non-spatial processing. Indeed some have suggested that this stream splitting may occur earlier - even as early as primary auditory areas (Romanski et al., 1999).

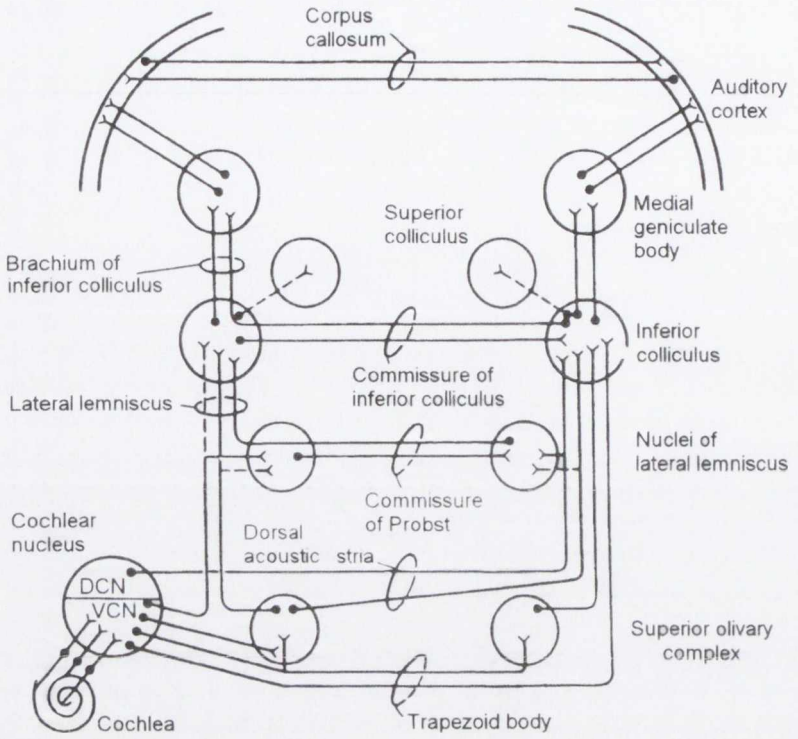

Figure 2.6 The auditory pathways from auditory nerve to auditory cortex. Symmetrical pathways from right ear not shown (Ehret and Romand., 1997).

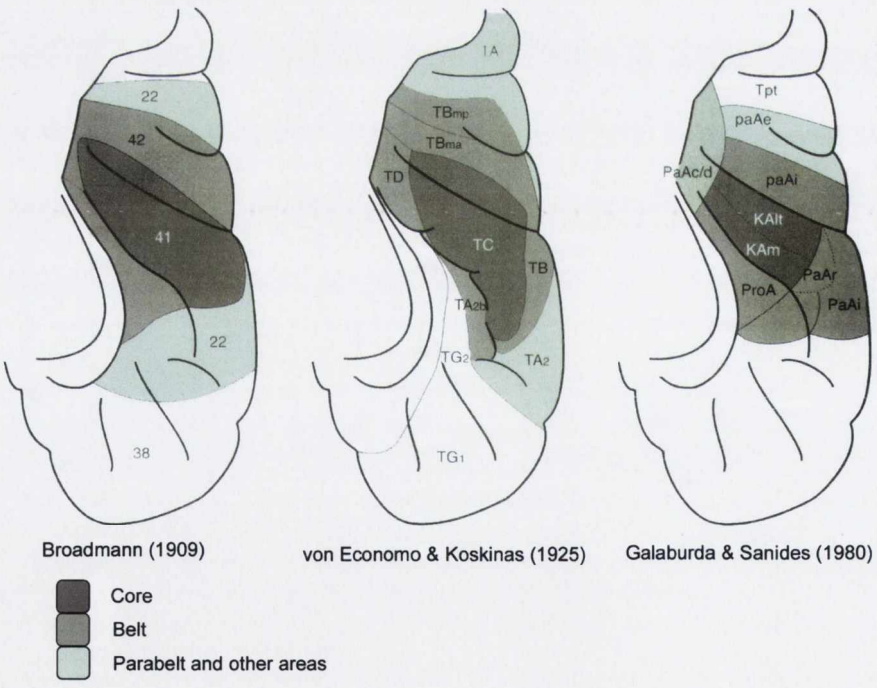

Figure 2.7 The results of three studies identifying auditory core, belt and parabelt regions of auditory cortex (adapted from Burkard et al., 2007).

2.3 The Electroencephalogram and Auditory Evoked Potentials

2.3.1 From Neuron to EEG to ERP

Neurons carry out the most important and unique function of the brain. They sense changes in the environment, communicate these changes to other neurons and command the body's responses to these sensations. Input into the neuron is through the dendrites (Fig 2.8). The number of inputs varies from 1 to about 100,000. This range reflects the fundamental purpose of nerve cells, namely to integrate information from other neurons. A neuron's output is via the axon. The axon is highly specialised for the transfer of information over distance in the nervous system. Axons can range from 1mm to 1m in length. The point of contact between an axon and a dendrite is called a synapse (see Fig 2.8). Synapses have two sides: the pre-synaptic side from which the information is flowing and the post-synaptic side to which the information is flowing. Information is carried along axons by electrical activity known as an action potential (AP). APs are electrical impulses that are initiated at the beginning of an axon (called the axon hillock) when the membrane potential of the neuron rises above a threshold. The membrane potential is the difference in voltage between the interior and exterior of a cell (i.e. the voltage across the membrane). The AP travels along the neuron until it reaches a synapse with another neuron. The post-synaptic neuron integrates this input with all the others it is receiving. If the integrated activity here exceeds threshold then this neuron will in turn initiate an AP.

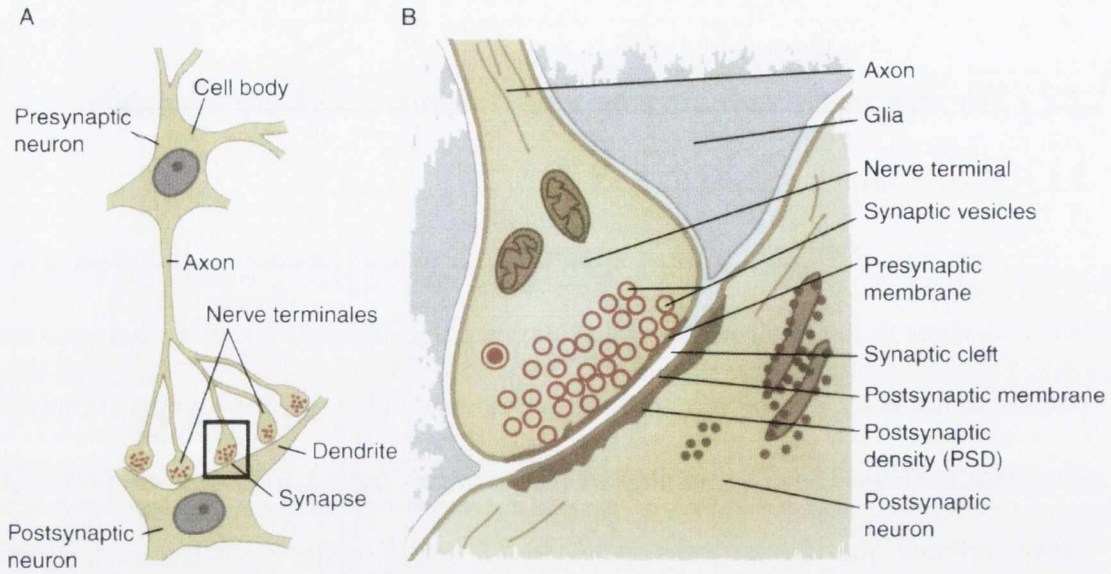

Figure 2.8 The Neuron and Synaptic organisation (adapted from Purves et al., 2001).

There is, however, another type of electrical activity in the brain called post-synaptic potentials (PSP). These PSPs are the basis of the electroencephalogram (EEG). For the majority of synapses there is no physical continuity between the pre- and post-synaptic cells and the signals must traverse the gap between them which is known as the synaptic cleft (see Fig 2.8). When an AP reaches a synapse it triggers the release of chemicals, known as neurotransmitters (NTs), across the gap and these NTs attach to the post-synaptic neuron. This is how one neuron transfers information to another across the synaptic cleft. The neurotransmitter then diffuses along the post-synaptic neuron and back across the cell wall (see Fig 2.9). Depending on the neurotransmitter released the post-synaptic neuron is either excited or inhibited. Due to this neurotransmitter cycle synapses act like tiny batteries driving NT (i.e. electrical current) in small loops generating electric fields. These electric fields are PSPs. When a PSP results in a depolarisation of the post-synaptic membrane (i.e. the membrane potential is brought closer to threshold) they are termed excitatory PSPs (EPSPs). On the other hand if the PSP results in hyperpolarisation of the post-

synaptic membrane (i.e. the membrane potential is driven further from threshold) they are termed inhibitory (IPSPs).

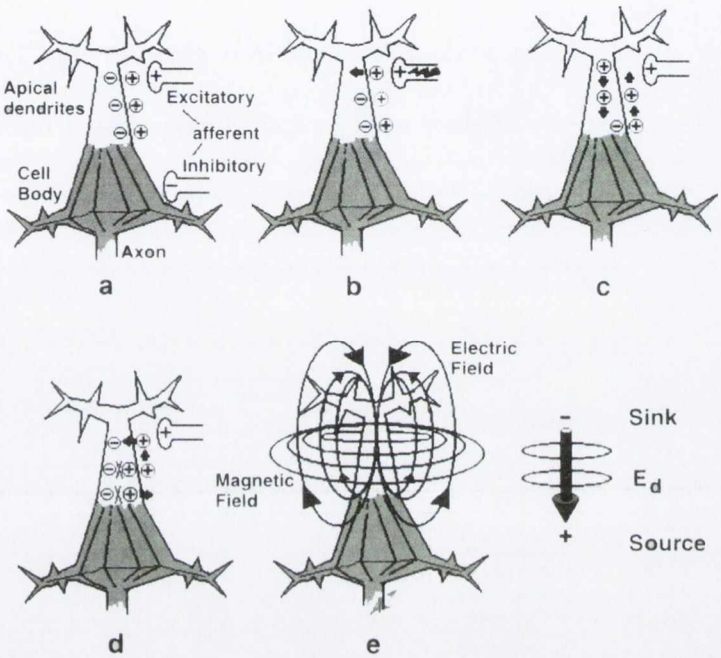

Figure 2.9 The generation of a post-synaptic potential due to neurotransmitter loops set up along the neuron (Zani and Proverbio., 2003).

The electroencephalogram (EEG) consists of excitatory and inhibatory post-synaptic electrical activity of large numbers of synchronously activated neurons. The field generated by a single (PSP) is not large enough to be detected at the scalp. Thus for a detectable potential to be established thousands of PSPs need to be simultaneously generated. Simultaneity is very important because if adjacent neurons are activated sequentially even though the total activity would be the same the net EEG activity at any given time would be negligible. Furthermore, in order for a far-field potential to be established the active neurons need to be oriented in the same direction. If this is not the case PSPs can cancel leaving no net field (i.e. closed-field see Fig 2.10a). EEG activity is measured using electrodes on the scalp to pick up the far-field potentials generated by the PSPs. As regards the topography of activity

measured at the scalp a radial source produces field activity which is maximal directly above the direction of neuronal orientation. Scalp distributions for variously oriented sources are shown in Fig 10b. Equal and opposite activity such as on opposing walls of a sulcus will cancel and not produce scalp activity despite synchronised neural activity. This is seen in scalp topography 3 in Fig 10 (b). Here the neurons orientated radially to the scalp contribute most to the topography whereas contributions of the tangentially orientated neurons on either side of the sulcus tend to cancel. That said if neurons on just one side of the sulcus were active they would alter the topography.

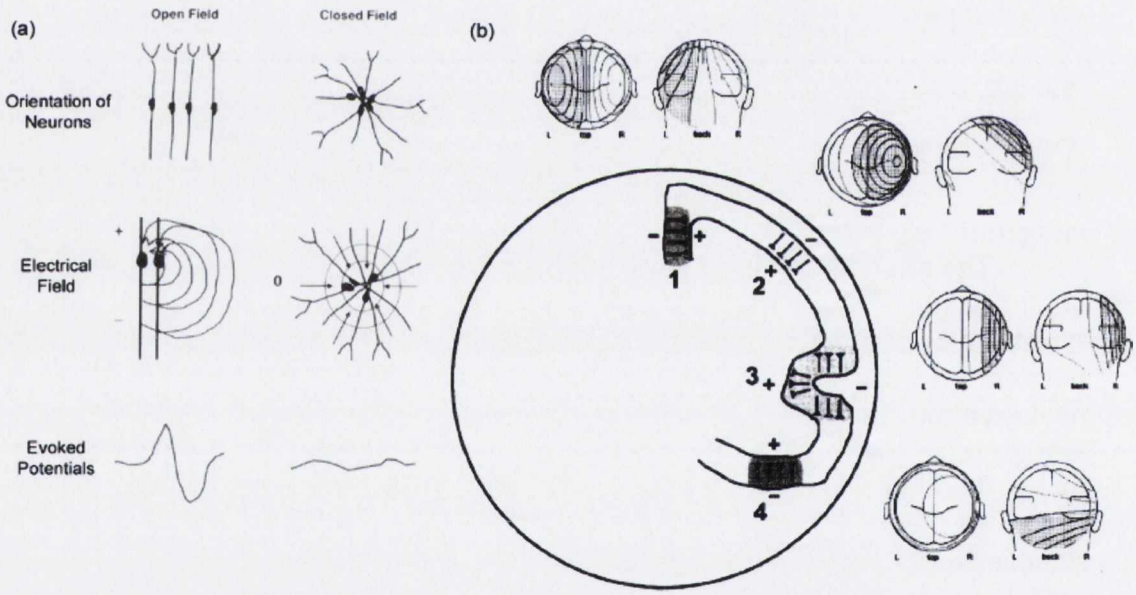

Figure 2.10 (a) The orientation of neurons leading to open and closed electric fields and the resulting scalp potentials (Burkard et al., 2007). **(b)** The orientation of open fields and the resulting scalp topographies (Olejniczak, 2006).

When using EEG to investigate sensory and cognitive activity many researchers employ the Event-Related Potential (ERP) method. Here a discrete

stimulus is repeatedly presented and the EEG response is averaged over all presentations of the stimulus. Thus activity not time-locked to the stimulus is averaged to zero leaving only EEG activity time-locked to the stimulus (see Fig 2.11). In the auditory modality these are known as auditory evoked potentials (AEPs).

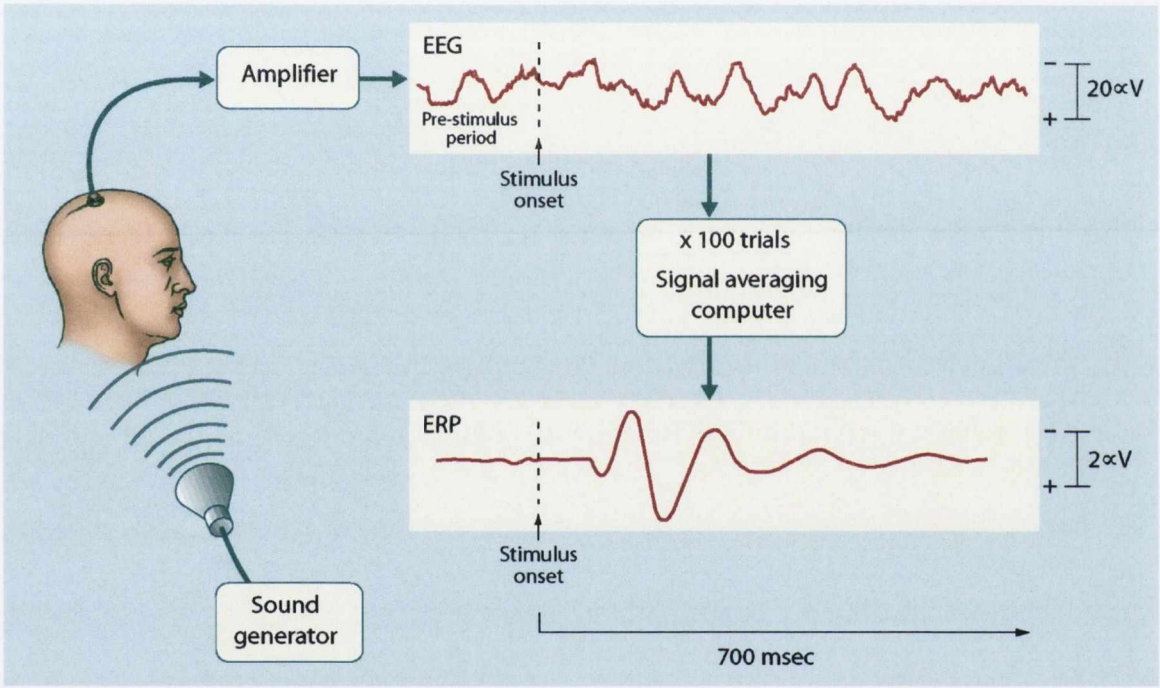

Figure 2.11. Averaging procedure from continuous EEG to ERP.

2.3.2 Auditory Evoked Potentials

The auditory-evoked potential (AEP) is an electrical response recorded from the nervous system following presentation of a sound stimulus (Davis 1939). The AEP obtained from the scalp using electroencephalography (EEG) can be subdivided into three sequences of waves reflecting activity of cell populations at various levels along the sensory processing hierarchy. The first sequence is known as the brain stem response, because of its purported origins, and consists of seven

positive peaks (labeled I–VII) occurring within the initial 8–12 ms. The following is an outline of the ABR components and their proposed generators. The anatomy of the auditory nerve and their contribution to ABR is relatively simple. That said, however, the relationship between the components of the ABR and the anatomy of more rostral structures of ascending auditory pathways (i.e. from CN to IC) is a less direct and more complicated.

Wave I:

This is due to the compound action potential at the distal end of the 8th cranial nerve (i.e. the auditory nerve). Wave I represents the afferent activity of the auditory nerve as it leaves the cochlea and enters the internal auditory canal (Burkard et al., 2007)

Wave II:

Wave II is also generated in the auditory nerve (Burkard et al., 2007). This time the generator lies at the proximal end near the brainstem. This is supported by the fact that the inter-peak interval between waves I and II is ~1.1ms. Given that the average length if the auditory nerve is ~2.6cm and the conduction velocity of the auditory nerve is ~22m/s this implies that conduction from cochlea to brainstem is ~1.18ms. The cranial nerve origin of wave II is also supported by the reliable measurement of wave II in brain death and from direct intraoperative recordings from auditory nerve as it enters the brainstem.

Wave III:

Wave III arises from neuronal activity in or near the cochlear nucleus (CN). Møller and colleagues (Møller et al., 1995) found an association between the latency of potentials recorded directly from the cochlear nucleus and the scalp recorded wave III. Wave III is most likely generated ipsilaterally in the caudal portion of the auditory pons. Dendrites in dorsal CN show a parallel arrangement which may allow for an evoked response dipole generator to be established and would be expected to generate a major ABR component such as wave III.

Wave IV:

Wave IV is not a consistent peak in the ABR and large individual variations even occur in individuals who have normal hearing. It often appears as a leading shoulder to wave V and these waves are thus sometimes referred to as the IV/V complex. Møller et al., 1995 suggest that wave IV is generated in structures close to the midline at the level of the Superior Olivary Complex (SOC). Anatomic features argue against a major contribution of the Lateral Lemniscus in ABR generation: Given the relatively small nuclei and also the fact that it receives innervations from different pathways the chances of simultaneous activity are limited (Moore et al., 1987a,b).

Wave V:

Wave V is the most analysed component of the ABR. Wave V is generated at the termination of lateral lemnicus fibres as they enter the inferior colliculus (IC) contralateral to the ear of stimulation. The ipsilateral IC contributes little to the

generation of wave V (Møller et al., 1995). Furthermore it Møller et al., 1995 suggest that wave V reflects activity in the LL that has not been interrupted in the SOC or the nucleus of the LL.

Waves VI and VII:

The origin of these peaks are open to question. Møller et al., 1995 suggest that they are due to continued synchronous activity of neurons on the IC.

ABR responses are followed by the middle latency sequence. These are thought to be caused by activity in thalamic nuclei and neurons in primary auditory cortex and consists of three negative (No, Na, and Nb) and two positive (Po and Pa) peaks that occur from around 8–12 ms to 40–50 ms. MLR responses are sensitive to stimulus intensity, with amplitudes of the MLR increasing with intensity up to about 70 dB above threshold. Latencies decrease with increasing stimulus intensity up to about 40 to 50 dB above threshold and then latencies stabilise. MLRs are also effected by subjects attention (Kadobayashi and Toyoahima 1984) and state of arousal (Deiber et al., 1989). A brief summary of the suggested component generators is outlined below.

No and Po:

These components are thought to be a slow response component belonging to ABR (Hashimoto, 1982)

Na:

Na is believed to originate subcortically from thalamus or thalamocortical radiations (Burkard et al., 2007). It occurs between 15 and 20ms post-stimulus.

Pa:

Pa is the most consistently recorded MLR component. It peaks around 25ms post-stimulus and typically has amplitude in the 0.5-1μV range. It has been suggested that Pa measured at the vertex has different generators than Pa measured over the temporal scalp. The vertex-recorded Pa is thought to be generated bilaterally from vertical dipoles within primary auditory cortex with contributions from the midbrain reticular formation (Kraus et al., 1992) and the MGB (McGee et al., 1992).

Nb:

The detectabililty and scalp distribution closely resemble those of Pa. The source of the Nb is unclear. Using MEG dipole localisation no difference between the location of Pa and Nb sources were found (Yoshiura et al., 1995). Thus either these two successive components reflect activities of the same neural population, the sources are too close to distinguish or the assumptions underlying the source localisation were invalid.

The final sequence is known as the long-latency or cortical response and reflects activity in higher-order auditory and association cortex. This sequence is made up of two positive (P1 and P2) and two negative (N1 and N2) peaks proceeding from 50 to 500 ms (Celesia and Peachey 1999; Picton et al. 1974). ALR

are elicited with a variety of stimuli including clicks, tone bursts, complex sounds and speech. ALR are effected by various stimulus factors. ALR amplitude increases with intensity saturating at about 70 dB above threshold and in general latencies decrease as intensity increases. As stimulus frequency increases the amplitude decreases. Amplitude increases as inter-stimulus interval decreases until ISI is about 10 seconds. ALR amplitudes increase as stimulus duration increases up to about 30 to 50 ms. ALR are also subject to participant attentional (particularly N1 and P2) and arousal effects (sleep effects responses in a complex manner)

P1:

P1 is thought to have generators primarily located in primary auditory cortex and specifically Heschl's gyrus. It is largest over midline central and lateral to central scalp. P1 is often described as part of the mid-latency response and is thus sometimes referred to as Pb. Recent work however has suggested that these are different components. P1 source configuration is likely more complicated than this though and various additional regions have been identified as contributing to the response (hippocampus, planum temporale and lateral temporal cortex (Burkard et al., 2007)). Neocortical areas may also be involved.

N1:

N1 is a negative peak that often occurs around 100ms after sound onset. N1 is known to have multiple generators in primary and secondary auditory cortex. It is thought to have three underlying components. The first of which is $N1_b$ and is generated in or near auditory cortex in the superior region of the temporal lobe. This is thought to

reflect attention to sound arrival, the reading out of sensory information from the auditory cortex or the encoding of a sensory memory in auditory cortex. The second component is the T complex. This is a positive wave occurring approximately 100ms after onset followed by a negative wave occurring approximately 150ms after onset. This is generated by radial source in secondary auditory cortex in secondary auditory cortex. The third component is a negativity occurring approximately 100ms after onset that is best recorded using long ISI. The generator of this component is unknown (Näätänen & Picton. 1987) and it may reflect a widespread transient arousal that acts to facilitate efficient sound processing. In general, the N1 is maximally recorded over midline central scalp locations.

P2:

The P2 is not as well understood as P1 and N1 components. It is thought to have generators in primary and secondary auditory cortex as well as the mesencephalic reticular activating system (Burkard et al., 2007). P2 is best recorded over central midline central scalp regions.

This detailed componentry, which can be measured with great precision due to the excellent temporal resolution of EEG, renders the AEP an extremely valuable and widely used tool in both research and clinical settings. For example, as well as being used to assess hearing function (for review see Celesia and Peachey 1999), specific components of the AEP have been used in the study of many neurological disorders including depression (Gallinat et al. 2000), Alzheimer's disease (Cancelli et al. 2006), schizophrenia (Nagamato et al. 1991), multiple sclerosis (Chiappa et al.

1980), and anxiety disorder (Drake Jr et al. 1991). In addition a number of cognitive processes such as selective attention (Bidet-Caulet et al. 2007; Picton and Hillyard 1974; Snyder et al. 2006) and auditory scene analysis (Bidet-Caulet et al. 2007; De Sanctis et al. 2008; Rahne et al. 2007; Snyder et al. 2006; Sonnadara et al. 2006) have been studied using AEPs. Due to the low magnitude of the AEP relative to the ongoing EEG it can be estimated only by averaging responses over a large number of discrete, repeated trials using stimuli of generally short duration (e.g., 10–100 ms). This method elicits responses with high signal-to-noise ratios (SNRs) and good reproducibility across trials and subjects.

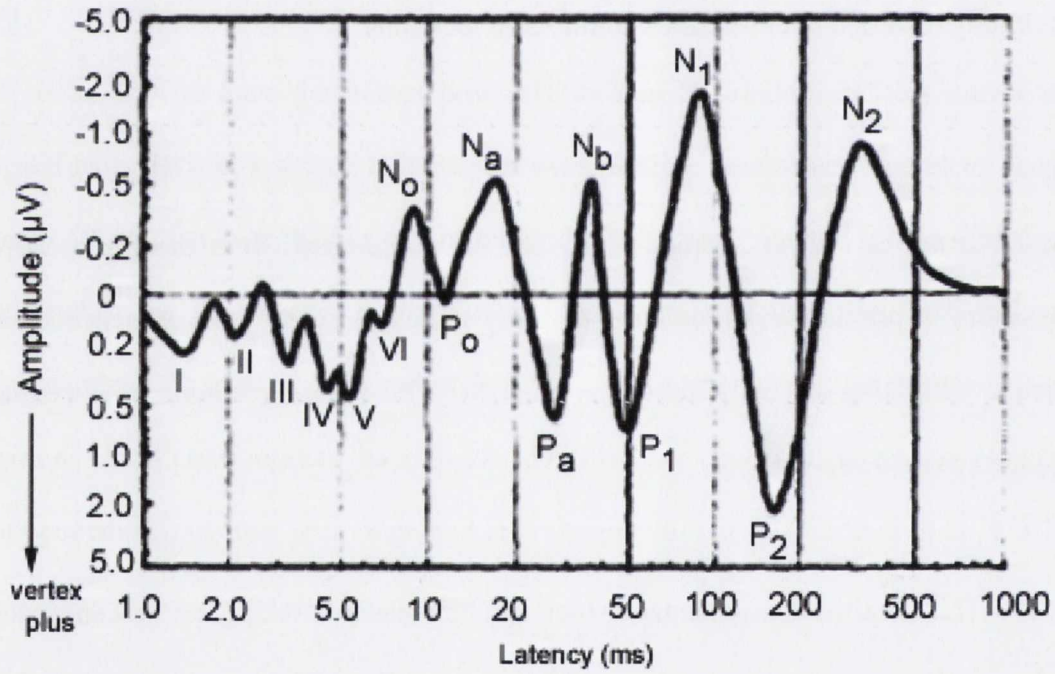

Figure 2.12 The Auditory Evoked Potential including brainstem, middle latency and long latency components. (adapted from Picton et al., 1974. Note log scales and also positive plotted downwards)

2.3.3 Others Modalities

Of course similar ERP methods are applied to other sensory modalities. In order to extract visual evoked potentials (VEPs) discrete flashes or checkerboard

phase reversals are employed and responses to several hundred flashes or reversals are averaged in an attempt to isolate activity time-locked to the stimulus. In the case of somato-sensory evoked potentials (SEPs) discrete mild electrical or pressure stimulation is applied and the responses to several hundred individual events are once again averaged in order to eliminate background EEG which is not time-locked to the stimulus.

2.3.4 Drawbacks of onset ERP technique

Obtaining responses to such repetitive stimulation has some inherent disadvantages, however. In reality, events are not isolated and in natural environments complex and continuous auditory stimulation is virtually ubiquitous. The human auditory system has evolved to efficiently process an infinitude of everyday sounds, which range from short, simple bursts of noise to signals with a much higher order of information such as speech. Investigation of temporal processing in this system using the event-related potential (ERP) technique has led to great advances in our knowledge. However, this method is restricted by the need to present simple, discrete, repeated stimuli to obtain a useful response. Furthermore, the need to present discrete stimuli generally precludes the resolution of responses to more than one stimulus at a time. This limitation can severely hamper the design of environmentally valid electrophysiological experiments looking at cognitive processing of multiple audio streams, such as those investigating auditory attention and auditory scene analysis. Some of the drawbacks of the onset averaging technique are overcome by employing the Auditory Steady-State Response (ASSR).

2.3.5 The Auditory Steady-State Response

Auditory event-related potentials can be obtained to non-discrete periodic stimuli. These are known as auditory steady-state responses (ASSRs). The ASSR was first identified by Galambos et al. in 1981 (Galambos et al., 1981). In this study the brain response to a continuous 400Hz sine wave amplitude modulated (AM) by a 40Hz sine wave was obtained. The EEG responses had significantly higher activity at 40Hz than other frequencies. Galambos termed this the "40Hz auditory potential". It is more commonly known as the ASSR or frequency following response (FFR) i.e. the response follows the modulating frequency of the stimulus. The ASSR is thought to have two dominant components: the brainstem and auditory cortices on the temporal lobe. Purcell et al., 2004 proposed that the brainstem follows modulation frequencies up to several hundred Hertz (and possibly even up to 1000 Hz) whereas the cortex can follow frequencies up to 70 Hz. Whatever the contributing neuronal centres ASSRs are not easily recorded at low intensities or at frequencies higher then 1000 Hz. Various stimuli have been employed to elicit ASSRs such as repeated click stimuli, AM noise, AM tones, beats (i.e. combining two tones of similar frequency), exponential AM (i.e. exponential of a standard modulating envelope can evoke larger responses than standard AM; John et al., 2002), frequency modulation (FM) and mixed modulation (MM; a combination of AM and FM).

As mentioned above employing ASSRs can circumvent the some of the drawback of onset AEPs e.g. separate responses at two concurrently presented stimuli with different modulation frequencies can be obtained (e.g. Linden et al., 1987). That said, however, although ASSRs are useful in certain clinical (Picton et

al., 2002) and research applications (Linden et al., 1987, Ross et al., 2004) a major disadvantage of the ASSR method is the fact that the only extractible features are a measure of response power and the phase delay relative to the stimulus. Consequently the useful information on the timing of response transmission through the auditory system, which can be gleaned from the onset AEP, is forfeited in the case of the ASSR.

2.3.6 AEPs and Auditory Attention

As highlighted previously AEPs have been used to investigate many neurological phenomena. One of these is auditory attention. There are two broad classes of attentional mechanisms: exogenous attention and endogenous attention (e.g., Hopfinger and West 2006). Exogenous attention is the attraction of attentional focus by an environmentally salient stimulus whereas endogenous attention is the self-directed focus of attention to a particular region or feature of the environment. Extensive research investigating the intriguing phenomenon of endogenous auditory attention has been carried out using the electroencephalogram and the magnetoencephalogram (EEG; Hillyard et al. 1973, Näätänen et al. 1992. MEG; Woldorff et al 1993). These studies throw up an interesting debate as to the generators affected by endogenous attention. Hillyard and colleagues (Hillyard et al. 1973), in their seminal study, found the N1 wave (~100ms) of the auditory evoked potential (AEP) to be enhanced by auditory selective attention (see Fig. 2.13). In that study a series of 800 Hz tone bursts were presented to the left ear and 1500 Hz tone bursts to the right. Both had randomised inter-stimulus interval between 250 and 1250 ms. Subjects were asked to attended to one stream and ignore the other. To

monitor their performance they were required to response to infrequent target stimuli in the attended stream. These targets occurred about 10% of stimuli were target and they consisted to tone bursts of 840 Hz for the left ear and 1560 Hz for the right ear. They suggested that the enhanced N1 they observed was due to increased sensory processing of the attended stimulus. This view is supported by others (Woldorff et al 1993). However, the idea that obligatory sensory components (i.e. activity related to the basic sensory representation of the stimulus) of the AEP are affected by endogenous attention has been challenged by Näätänen and colleagues (Näätänen et al. 1987, Näätänen et al. 1992). They suggest that the N1 enhancement is likely due to the engagement or enhancement of separate components of the N1 wave that are not related to obligatory processes but to a matching process between the sensory input and an "attentional trace". Although they suggest that under some conditions obligatory components may possibly be affected by endogenous attention (Näätänen et al. 1987), they remain somewhat sceptical about obligatory sensory involvement in endogenous auditory attention (Näätänen et al. 1992, Alho et al. 1994). Research into this question is ongoing (Ross et al. 2010, Bidet-Caulet et al. 2007).

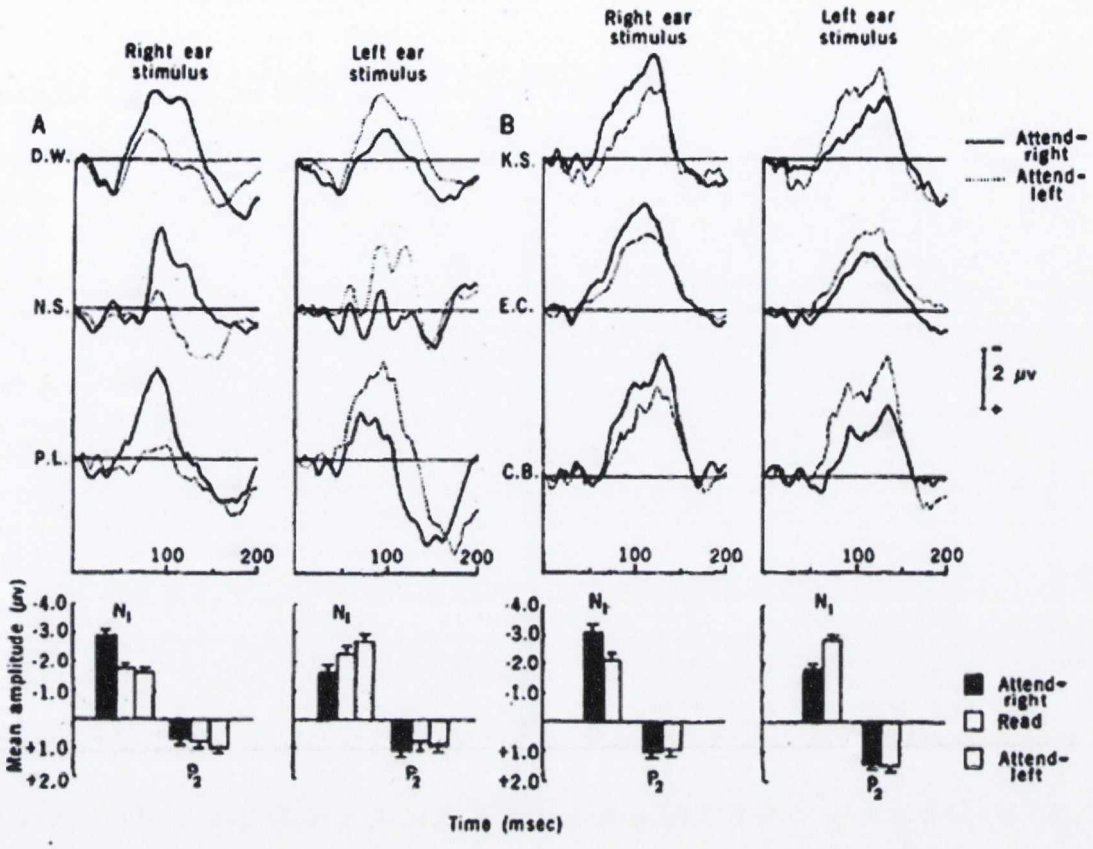

Figure 2.13 Panel A: Vertex responses from three subjects (D.W., N.S., P.L.) in Experiment 1 of Hillyard et al. (1973) separately for attended and unattended left-ear and right-ear stimuli. Bar graphs give the mean and standard error (10 subjects) of the baseline to peak amplitudes of Nl and P2 evoked via each ear under all three experimental conditions. Panel B: responses from three subjects (K.S., E.G., C.B.) in Experiment 2, with bar graphs giving mean amplitudes over all 10 subjects.

2.3.7 AEPs and the Cocktail Party Problem

A common and hugely important element of the rich mixture of sounds, with which the everyday environment presents us, is speech. Speech and language are essential for the everyday interaction between humans and our ability to single out and attend to one speaker in a crowded and noisy environment is integral to human communication. Since initial investigations in to the behavioural aspects of auditory attention by Cherry in 1953 there has been much interest in the mechanisms of auditory attention. Cherry, 1953 describes the ability to focus on and report from one speech stream and over another in a dichotic listening paradigm. Building on these behavioural studies the ERP method has been employed to interrogate the physiology of auditory attention such as those outlined above. Studies using more

natural stimuli have also been carried out. These studies have had to superimpose discrete probe stimuli on speech (Hink & Hillyard, 1976, Woods et al., 1984, Coch et al., 2005, Nager et al 2008) or average onset responses to a specific feature of speech (Teder et al., 1993). Interpretation of these results is likely hampered, however, by the fact that the choice of probe stimulus has a material effect on the responses and attentional effects found. Furthernore, neither of these methods takes account of the extended temporal nature of speech. That is to say that using these discrete events may result in confounding issues regarding imprecision in resolving the temporal locus of attention because of possible enhancement of ongoing speech prior to and following the onset event. Further to this, neuroimaging methodologies, such as functional magnetic resonance imaging (fMRI), have also been employed to expand our understanding of the cocktail party problem (Hill & Miller, 2009, Nakai et al., 2005). However, identifying the temporal dynamics of attention using fMRI is limited by the poor temporal resolution available with fMRI. Thus the precise temporal dynamics of attention to natural speech stimuli are unlikely to be accurately addressed in the above studies, whether employing ERP or fMRI, and studies which are more sympathetic to the dynamics of natural speech would be advantageous.

Thus, taking into account the various benefits and drawbacks of current AEP methods, the ability to obtain electrophysiological responses to complex, continuously presented stimuli with full temporal resolution would thus be very useful. To this end Chapter 3 of this thesis introduces and characterizes the Auditory Evoked Spread Spectrum Analysis (AESPA) method for obtaining temporally detailed response to extended continuous stimuli. The AESPA method also allows for extraction of two separate responses to two simultaneously presented stimuli.

This is exploited in Chapter 4 where the AESPA method is used to investigate the effects of truly endogenous auditory spatial attention on obligatory sensory processing in auditory cortex. In this chapter attentional effects on obligatory sensory processing in primary auditory cortex are identified. These results agree with the initial findings of Hillyard et al., 1973 that sensory processing in PAC is indeed affected by attention and counter the assertions of Näätänen that sensory activity is not affected by attention and that modulation of the N1 wave of the AEP is due to the engagement or enhancement of a separate attentional generators (Näätänen et al., 1987).

In chapter 5 of this thesis we use the AESPA method to investigate the temporal dynamics of the attention effects of the cocktail party problem by using the AESPA method to extract temporally detailed responses to simultaneously presented speech streams. These natural speech stimuli are uncontaminated by probing events and furthermore the extended temporal nature of the speech is accounted for in the AESPA responses. Results suggest that activity in response to the unattended stimulus is suppressed in PAC at ~209ms. The possible relationship between this suppression and the fact that working memory is highly engaged in the task is discussed as is the likely link between the left bias of the attention effect, the left bias of temporal processing and the left bias of language processing.

As a whole this thesis adds to the current progress made in investigating auditory processing and attentional mechanisms in more natural and realistic settings (Kerlin et al., 2010, , Teder et al., 1993). The ability of the AESPA to overcome

some of the limitations of standard AEP techniques shows it to be an important tool in auditory neuroscience. Furthermore the utility and flexibility of the AESPA method coupled with its ability to address basic neuroscientific questions such as the dynamics of attention to ecologically valid stimuli proves it to be of significant relevance to furthering our understanding of auditory processing.

3

Auditory Evoked Spread Spectrum Analysis

In natural environments complex and continuous auditory stimulation is virtually ubiquitous. The event-related potential (ERP) technique has led to great advances in our knowledge of the temporal processing properties of the auditory system. However, this method is restricted by the need to present simple, discrete, repeated stimuli to obtain a useful response. Alternatively the continuous auditory steady-state response is used. This method, however, reduces the evoked response to its fundamental frequency component at the expense of useful information on the timing of response transmission through the auditory system. In this chapter, we describe a method for eliciting a novel ERP, which circumvents these limitations, known as the AESPA (auditory-evoked spread spectrum analysis). This method uses

rapid amplitude modulation of audio carrier signals to estimate the impulse response of the auditory system. We show AESPA responses with high signal-to-noise ratios obtained using two types of carrier wave: a 1-kHz tone and broadband noise. To characterize these responses, they are compared with AEPs elicited using standard techniques. A number of similarities and differences between the responses are noted and these are discussed in light of the differing stimulation and analysis methods used. Data are also presented that demonstrate the generalizability of the AESPA method and a number of applications are proposed.

As outlined in chapter 2 AEPs can be elicited by the repeated presentation of a short discrete stimulus (Davis, 1939) and extracted from the EEG using an epoch averaging procedure. When evoked by such repetitive stimulation, it shows several distinct components, such as the N100 and P150 (see Fig. 2.12). However, in reality, events are not isolated and the human auditory system has evolved to efficiently process an infinitude of everyday sounds, which range from short, simple bursts of noise to signals with a much higher order of information such as speech. Thus given the drawbacks of the discrete onset AEP and ASSR methods outlined in chapter 2 the ability to obtain electrophysiological responses to complex, continuously presented stimuli with full temporal resolution would thus be very useful. The work outlined in this chapter seeks to contribute to this and a novel method for estimation of the impulse response of the auditory system using continuous and complex stimuli is introduced.

3.1 System Identification and AESPA

By considering the brain in simplified form as a linear system, with isolated events as input and EEG as output, the average event-related potentials (ERPs) can be said to approximate the system's time-domain impulse response function. However, white noise signals also are commonly used in characterization of physiological systems. Such a method was recently described in the visual domain (Lalor et al. 2006), wherein the impulse response of the visual system is obtained using a stimulus whose luminance or contrast is smoothly modulated by a stochastic signal with its power spread over a range of frequencies. This impulse response is known as the VESPA (visual-evoked spread spectrum analysis) response. The work presented in this chapter seeks to expand on this work by applying a similar approach to the auditory system. This is accomplished by smoothly, but stochastically modulating the amplitude of an auditory carrier stimulus and estimating the linear impulse response from the recorded EEG. The profile of the AESPA response and that of the AEP elicited using standard methods are compared and their similarities and possible cellular underpinnings of their differences are discussed. We further demonstrate the flexibility of the AESPA method using a stimulus that consists of continuous bursts of sounds separated by intervals of silence. The results suggest the utility of AESPA in a number of research and clinical experiments where the use of the standard AEP is not possible.

The AESPA method follows on from a broad variety of system identification methods that are used in auditory system analysis. Although the majority of these methods have been applied to neural spiking data (Eggermont 1993; Klein et al.

2000) some applications to EEG using m-sequenced presentation of binary stimuli have been reported (Eysholdt and Schreiner 1982; Shi and Hecox 1991). The AESPA can be considered as being most closely related to the Volterra–Wiener approach to system modeling (Marmarelis and Marmarelis 1978). Specifically, the estimation of the AESPA is based on the assumption that the output EEG, $y(t)$, consists of a convolution of the audio amplitude-modulated (AM) signal $x(t)$, with an unknown impulse response $w(\tau)$, plus noise,

i.e. $y(t) = w(\tau) * x(t) + noise$

Given the known audio AM signal and the measured EEG, we obtain $w(\tau)$ (i.e., the AESPA), by performing linear least-squares estimation. This continuous-time equation must be discretized and put in standard form for a least squares fit. To this end, we create a column vector consisting of the sampled points of the response function, $w=(w(F_0), w(F_0+\Delta t),\ldots, w(F_0+n_w\Delta t))$, and a column vector consisting of windows of sampled points of the modulating stimulus, $x_t=(x(t+F_1), x(t+F_1-\Delta t),\ldots, x(t+F_1-n_w\Delta t))$, where F_0 and F_1 are the limits of the region of support, $w(\tau)$ is allowed to be nonzero only for $F_0\leq\tau\leq F_1$. The number of samples in the window is nw+1 where $n_w=N(F1-F0)$, where N is the sampling rate and therefore $\Delta t=1/N$. For the case of the AM stimulis, the values of x(t) are simply the intensity values of the of the modulating waveform.

Rewriting the model in matrix notation in discrete time, and inserting the result into a formula for the mean squared error, gives:

$$y_t = \Delta t w^T x_t + noise$$

$$E = \left\langle \left| \Delta t w^T x_t - y_t \right|^2 \right\rangle$$

where $\langle \bullet \rangle$ indicates an average over t. Δt is included when the convolution integral is discretized i.e. $y(t) \approx \sum_{i=0}^{n_w} \Delta t w(F_0 + i\Delta t) x(t - F_0 - i\Delta t)$. This is to ensure that the magnitude of the estimated w(τ) is invariant to the sampling rate. We wish to fit the column vector w to a set of input column vectors x_t and corresponding output scalars y_t so as to minimize E. Expanding dE/dw=0 yields the linear system

$$\Delta t \left\langle x_t x_t^T \right\rangle w = \left\langle x_t y_t \right\rangle$$

Solving this involves inversion of the input signal's autocorrelation matrix $\left\langle x_t x_t^T \right\rangle$. In order to improve the quality of the estimate of w(τ), a regularization term is added. This reduces the variance of w(τ) at the expense of a bias, with a net decrease in the off-sample mean squared error. Standard ridge regression incorporates a quadratic penalty term, $\lambda \int_{F0}^{F1} w(\tau)^2 d\tau$ which in discrete time comes to $\lambda \sum_{i=0}^{n_w} \Delta t w_i^2 = \lambda w^T M w$ where the matrix M, introduced for generality, is here merely a scaled identity matrix, M=Δt I. This $\lambda w^T M w$ term is added to the squared error, resulting in a regularized objective,

$$E = \left\langle \left| \Delta t w^T x_t - y_t \right|^2 \right\rangle + \lambda w^T M w$$

Expanding dE/dw=0 yields a linear system

$$\left\langle \Delta t^2 x_t x_t^T + \lambda M \right\rangle w = \left\langle \Delta t x_t y_t \right\rangle$$

which can be solved for w as above.

The regularization parameter λ, used in the current study, was empirically chosen to be 4.4 x 10^{-3}. This resulted in good reduction of off-sample error without affecting the height or the latency of the response peaks.

Summarizing the AESPA method in qualitative terms, the response w(τ) is analogous to a filter which describes how the brain transforms the auditory input into the EEG output. Keeping this in mind the time axis for the AESPA carries a different meaning than the time axes in traditional ERP studies. Each point on the time axis can be interpreted as being the relative time between the continuous EEG and the continuous input intensity signal. Therefore the AESPA at -100 ms, for example, indexes the relationship between the input intensity signal and the EEG 100 ms earlier; obviously this should be zero. As another example, the AESPA at 100 ms indexes how the input intensity signal affects the EEG 100 ms later (Lalor et al., 2009).

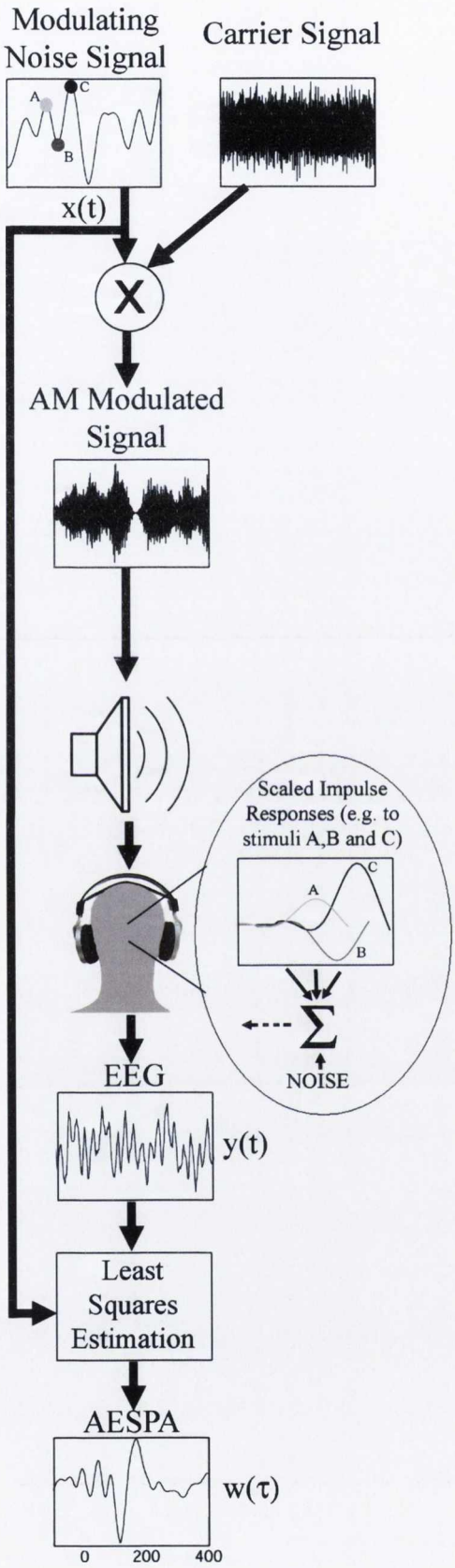

Figure. 3.1 Flow diagram of auditory-evoked spread spectrum analysis (AESPA) acquisition. A colored noise signal is used to modulate the amplitude of a carrier signal. The electroencephalogram (EEG) is modeled as a sum of overlapping scaled impulses in response to the amplitude modulations of the stimulus, plus noise. Three such scaled impulse responses are shown, corresponding to stimulus values *A*, *B*, and *C*. (from Lalor et al., 2009)

3.2 Initial Investigations

Two subjects aged 22-26 participated in the study. All had no history of hearing difficulties and EEG data were recorded from 128 electrode positions, filtered over the range 0 – 134 Hz and digitized at the rate of 512 Hz using a BioSemi Active Two system. Synchronization between the audio stimuli and the recorded EEG data was guaranteed by recording both stimuli and response simultaneously. EEG data were digitally filtered with a high-pass filter, where the passband was above 2 Hz and with a -60 dB response at 1 Hz and a low-pass filter with a 0-35Hz passband and a -50 dB response at 45 Hz. Channel location labels are given by the 10-5 nomenclature system outlined in (Oostenveld & Praamstra, 2001). The audio stimuli were generated with a SoundBlaster Extigy soundcard and presented to subjects using high fidelity Sennheiser HD650 headphones.

The Stimuli

Two audio signals formed the basis of the complex stimuli employed in this study. They were

1. a Gaussian broadband noise (BBN) waveform with energy limited to a bandwidth of 0-22.05kHz,
2. a 2kHz pure tone.

The modulating spread spectrum signals consisted of Gaussian white noise (GWN) with energy limited to 0-30 Hz. These signals were mapped to the intensity of the audio stimulus using an exponential relationship, in an attempt to exploit the logarithmic nature of auditory stimulus intensity perception.

$$x' = 10^{2x}$$

After the exponential mapping x' was normalized to the range 0 to 1. The modulating noise signal was then interpolated to give a smooth transition from one modulation amplitude to the next at the desired rate. Using the Nyquist sampling theorem and given that EEG power above 30 Hz is low compared to activity below 30 Hz, the modulation rate of each signal was set to be 60 Hz. Signals with the desired statistical properties were precomputed and stored. The basic BBN and tone stimuli were then modulated by these waveforms and also stored.

Thus, as an example, for a five minute long stimulus a GWN vector 18000 samples (i.e. 300 seconds x 60 samples/s) in length is created. This signal has energy from 0-30Hz at a sampling rate of 60Hz. This signal is then sent through the exponential mapping in an attempt to exploit non-linear intensity processing of the auditory system (Green, 1988). It is then resampled (interpolated) such that it has a sampling rate of 44100 Hz consistent with standard audio (the resulting length in samples with sampling frequency = 44.1 kHz is 13230000 samples). The auditory carrier is created using GWN or 1kHz tone 13230000 samples in length (300 seconds x 44100 samples/s). This is then modulated by the envelope resulting in the desired stimulus. This is presented to the participant while their EEG is recorded. The signal analysis procedure outlined in section 3.1 is then performed at every electrode to obtain the AESPA response (See Fig 3.1 for analysis flow diagram and Appendix A for an implementation of the estimation algorithm).

Experimental Procedure

Subjects were instructed to keep eye movements to a minimum for the duration of each session. While abstaining completely from eye-blinks is not possible for long periods, subjects were instructed to keep the number of eye-blinks to a minimum during each session. Subjects were also instructed to keep all other types of motor activity to a minimum during testing. Subject 1 undertook one 300 second session for the modulated noise stimulus, one 300 second session for the modulated tone stimulus and five 300 second sessions for the tone pip stimulus. Subject 2 undertook five 300 second sessions for each of the modulated noise, modulated tone and the tone pip stimuli. The tone pips occurred at a rate of one per second and the stimulus consisted of a 2 kHz tone 10ms in length including 2ms ramp-up and ramp-down

Quantification of Performance

Comparisons are made between the standard AEP and the AESPA by way of two methods. Firstly, SNRs were calculated for the AESPA and standard AEP. These SNRs were calculated by defining noise as the mean squared values in the 100ms interval immediately preceding the stimulus, and the signal was defined as the mean of the squared values in the interval 35-225ms post-stimulus. Secondly, correlation values between the AESPA and the standard AEP were also calculated. Again for the calculation

it is assumed that both the AEP and AESPA occur in the interval 35-225ms post-stimulus. The correlation index C was calculated as

$$C = \frac{COV(w,v)}{\sqrt{VAR(w)VAR((v))}}$$

Where *w(t)* is the AESPA and *v(t)* is the AEP. COV indicates covariance of the two waveforms and VAR the variance.

Results

Fig 3.2(a) and 3.2(b) respectively show the AESPA and standard AEP obtained from Subject 2 recorded from channel Cz. The plot of the AESPA is averaged over the 5 trials undertaken by the subject. The SNRs obtained for the AESPA for Subjects 1 and 2 at Cz are 4.52dB and 11.99dB respectively, whereas the SNR for the AEPs are much higher: 41.61dB and 36.9dB respectively. As can been seen from Fig 3.2(a) and (b), taken along side the SNR values, the proposed response (the AESPA) does indeed exist albeit with a smaller SNR than the standard AEP. The highest SNR for subject 2 in response to the noise modulated noise stimulus (33.46dB) was achieved at P9h. At this channel the AEP and AESPA also correlated strongly ($r=0.69$; $p=2.6\times10^{-15}$). In general the AESPA resulted in lower SNRs than the standard AEP but at some channel locations it resulted in higher SNRs.(e.g. at electrode T8h the AESPA resulted in a SNR that was 7.22 dB higher). The AESPA and AEP from Subject 1 at channel Cz correlate reasonably well ($r= 0.7$; $p=1.5 \times 10^{-15}$). Slightly stronger correlation is seen for Subject 2 ($r=0.77$; $p=2 \times 10^{-20}$). The strongest correlation between AEP and AESPA for subject 2 occurs at channel P7h ($r=0.91$; $p=7.1\times10^{10-39}$). A high AESPA SNR was also achieved at this location: 17.24dB. The AESPA average of Subject 1 and Subject 2 together resulted in an SNR of 12.85dB and also showed strong correlation to the standard AEP ($r=0.82$, $p=1.14\times10^{-24}$) at Cz. Fig. 3.2(c) and (d) respectively show the average over both subjects for the AESPA and AEP at Cz. The modulated 2kHz pure tone stimulus did

not seem to result in any appreciable responses. The highest SNR for Subject 2 was achieved at channel Cz (11.14 dB) but from the plot no definite response was apparent. Furthermore there was only a very low, albeit significant, correlation between the standard AEP and the response to this stimulus (r=0.29 p=0.0036). The highest correlation between the response to the modulated tone and the standard AEP was at FCC2h (r=0.78, $p=3.9x10^{-21}$). The SNR in this channel was 9.4 dB but again a plot of the response (not shown) does not show any specific waveform of note.

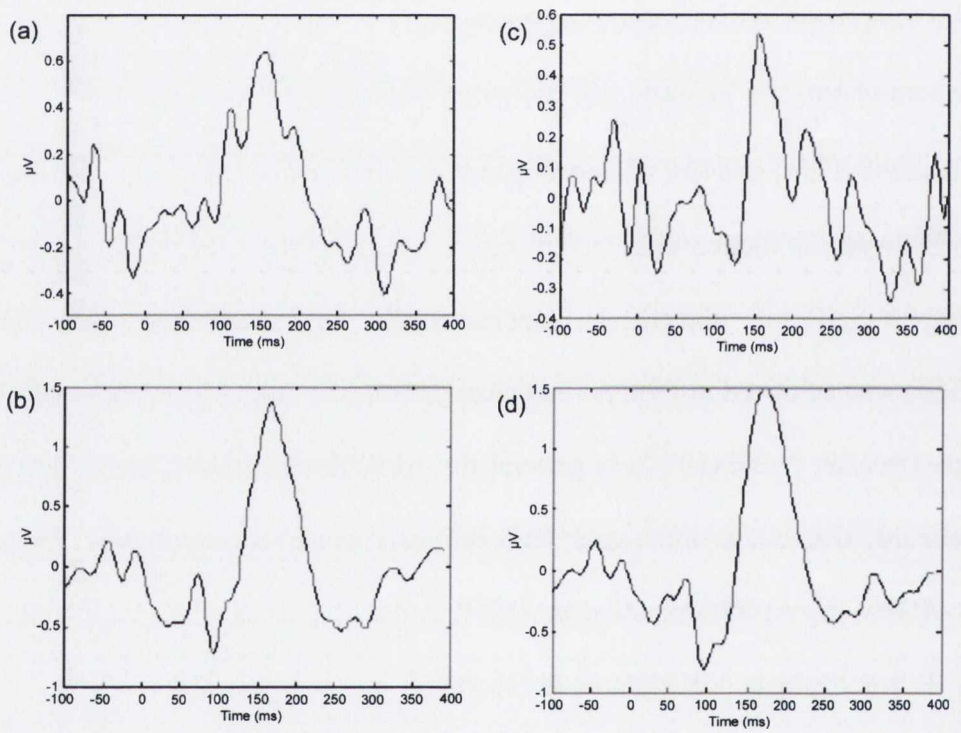

Figure 3.2 (a) The AESPA response to the modulated noise stimulus at Cz and (b) The standard AEP at Cz for subject 2. (c) The AESPA response to the modulated noise stimulus at Cz and (d) The standard AEP at Cz for subject average.

Discussion of Initial Investigations

It can be concluded from the results of the modulated noise stimulus that the proposed method of obtaining an AEP with broadband noise as the underlying sound

is valid. Although the SNR values are, in general, not as high as for the standard AEP, the fact that the response to the spread spectrum stimulus shows considerable correlation to the standard AEP suggests that the AESPA is a valid auditory evoked potential. It has also been shown that high SNR values can be achieved for the spread spectrum stimulus at specific channel locations. From these experimental results, it can also be concluded that that the modulated 2kHz tone is not as efficient in eliciting standard AEP-like responses compared to the modulated broadband noise. This may be due, in part, to the tonotopic nature of the auditory cortex i.e. the 2kHz modulated tone activates such a limited area of auditory cortex that a clear response at the scalp is not easily measured. The modulated broadband noise may exploit this tonotopic nature, stimulating less specific areas of auditory cortex thus resulting in larger responses.

These initial investigations have shown that impulse responses of the auditory cortex, known as AESPAs, can be extracted in response to stimuli consisting of broadband noise modulated by a noise signal whose power limited to 0-30Hz. The use of noise modulated tones has not resulted in any significant response. That said a more comprehensive EEG study with more subjects was warranted based on these results and is presented in the next section.

3.3 Follow-up investigation and in depth analysis of the AESPA response

3.3.1 Materials and Methods

Subjects

Participating in the study were 12 subjects (one female; aged 22–35 yr), all of whom had normal hearing. The experiment was undertaken in accordance with the Declaration of Helsinki. The Ethics Committee of St. Vincent's University Hospital in Dublin approved the experimental procedures and each subject provided written informed consent.

Stimuli

Once again two types of carrier signal were used in this study:

1. A Gaussian broadband noise (BBN) waveform, with energy limited to a bandwidth of 0–22.05 kHz.
2. A 1-kHz pure tone (TONE).

The amplitude of these carrier stimuli was modulated using Gaussian noise signals with uniform power in the range 0–30 Hz, i.e., at a rate of 60 modulations/s. This rate was chosen based on the fact that EEG power above 30 Hz is typically very low. Modulating signals with the desired statistical properties were precomputed and stored. Taking into account the logarithmic nature of auditory stimulus intensity perception, the values of these modulating signals (x) were then mapped to the amplitude of the audio stimulus x', using the same exponential relationship as before and normalized to between 0 and 1. The modulating noise signal were once again interpolated to give a smooth transition from one modulation amplitude to the next

and stored. An example of the amplitude-modulated BBN and TONE signals can be seen in Fig. 3.3. The audio stimuli were generated with a SoundBlaster Extigy soundcard and presented to subjects using high-fidelity Sennheiser HD650 headphones.

Experimental procedures

All subjects underwent ten presentations each of 120-s duration for both the amplitude-modulated BBN and amplitude-modulated TONE stimuli. Standard AEPs were also obtained from each subject during ten 120-s runs using repeated presentations of a discrete stimulus. These stimuli were presented at an average rate of 1 Hz with the interstimulus interval randomized between 750 and 1,250 ms and consisted of a 1-kHz tone, 50 ms in duration, which included 10-ms ramp-up and ramp-down. To demonstrate the flexibility of the AESPA method, a third stimulus was also tested, which represented something of a hybrid between the fully discrete AEP stimulus and the fully continuous stimuli otherwise used for the AESPA. It consisted of an amplitude modulated BBN carrier within each second of which a silent interval was randomly inserted. Each interval was also randomly chosen to be between 50 and 200 ms in duration. As with the AEP stimulus, these onsets and offsets consisted of 10-ms raised cosine ramps. Thus this stimulus contained bursts of continuous stimulation with multiple onsets and offsets and is referred to as a composite stimulus. Two subjects underwent 30 presentations each of 120-s duration for this stimulus. To directly compare the continuous AESPA data with those obtained using the composite stimulus, the same two subjects undertook a further 20 presentations each of 120 s duration using the original continuous BBN stimulus,

bringing their total number of continuous BBN runs to 30. Subjects were instructed to keep their eyes open and to keep eye movements and blinks to a minimum for the duration of each run. Subjects were also instructed to limit all other types of motor activity during each run.

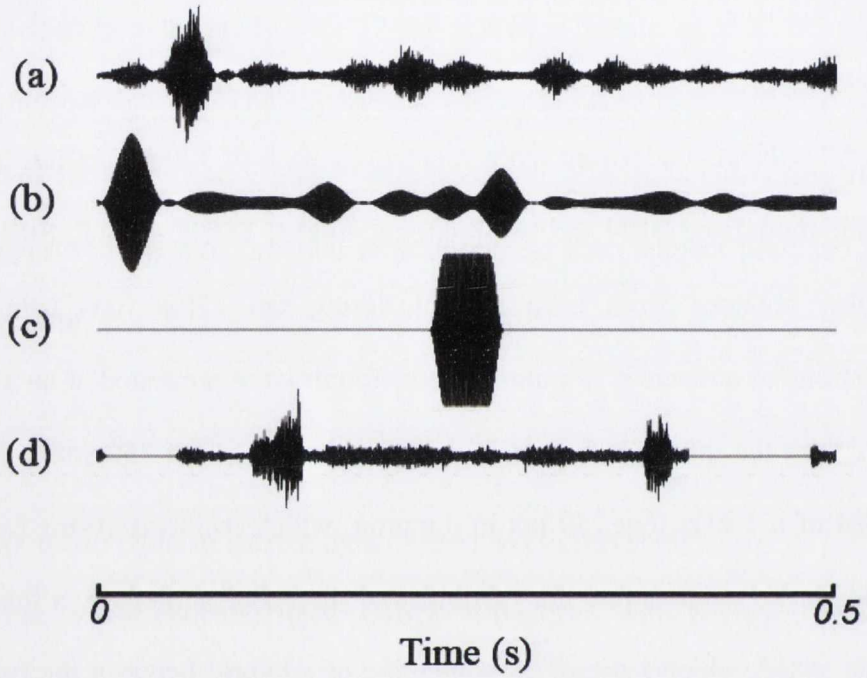

Figure. 3.3 A 0.5-s segment of (a) an amplitude modulated broadband noise (BBN) stimulus and (b) an amplitude-modulated 1,000-Hz TONE stimulus. c: a 50-ms 1,000-Hz tone pip with 10-ms rise and fall times. d: a composite of the BBN stimulus and the standard auditory-evoked potential (AEP) stimulus (i.e., a modulated BBN stimulus incorporating forced onsets and offsets). (from Lalor et al., 2009)

EEG acquisition

EEG data were recorded from 130 electrode positions, lowpass filtered to the range 0–134 Hz, and digitized at the rate of 512 Hz using a BioSemi Active Two system. Synchronization between the audio stimuli and the recorded EEG data was ensured by including the signal on the parallel port of the presentation computer, indicating the onset and offset of the stimuli, among the recorded signals. EEG data were digitally filtered with a high-pass filter, where the passband was 2 Hz and with

a 60-dB response at 1 Hz and a low-pass filter with passband 35 Hz and a 50-dB response at 45 Hz. The data at each channel were re-referenced to the average of the two mastoids. When recording EEG the signals are recorded as the potential between the electrodes and a reference. In order the change the reference the activity at the new reference is taken away from the activity on all other electrodes at each time point. The mastoids electrodes are positioned over the posterior part of the temporal bone behind the ear. These are *relatively* isolated from EEG activity as they reside over a thick region of bone and are thus often used as a reference point in EEG studies.

Once the EEG has been recorded the AESPA response is estimated (see analysis flow diagram Fig. 3.1). It should be noted that for the estimation the values of $x(t)$ were assumed to be constant across each 16.67-ms modulation period (i.e. 60Hz modulation implies that the amplitude changes every 16.67ms). These modulation values were determined from the original un-interpolated values. Furthermore the initial modulation values, i.e., the linear values obtained prior to the exponential mapping, were used. This seemed reasonable under the assumption that the exponential mapping would actually result in a more linear intensity perception. The AESPA was estimated using a sliding window from 200 ms pre-stimulus to 400 ms post-stimulus that was advanced sample by sample. This window was chosen to present the AESPA using an interval similar to that typically used for plotting the AEP. However, the meaning of the interval is slightly different because the AESPA, unlike the AEP, does not correspond to a specific discrete event occurring at *time 0*. Instead, each time point on the time axis can be interpreted as being the relative time

between the EEG and the input intensity signal. Therefore the AESPA at -100 ms, for example, indexes the relationship between the input intensity signal and the EEG 100 ms earlier; obviously this should be zero. As another example, the AESPA at 100 ms indexes how the input intensity signal affects the EEG 100 ms later. The steps involved in generating the stimuli and the estimation of the AESPA are illustrated in Fig. 3.1. AEPs were determined by averaging the EEG in response to a stimulus in epochs from 200 ms prestimulus to 400 ms poststimulus. SNRs for both methods were determined by considering data in the interval 0–300 ms as signal and -100 to 0 ms as noise. Grand average AEPs and AESPAs were baseline corrected by subtracting the mean of the prestimulus values (-100 to 0 ms) for each channel.

Determination of AESPA components

To ascertain the components of the AEPSA response we applied the topographic pattern analysis methods described in Murray et al. (2008). This analysis method splits responses into distinct segments called microstates (Lehmann 1987), which represent periods of topographic stability, and was implemented in the present study using CarTool (http://brainmapping.unige.ch/cartool.htm). We used a k-means clustering algorithm that was constrained such that clusters that were 75% correlated were merged and response segments less than or equal to three samples (5.9 ms) in duration were rejected. This is essentially a data-compression algorithm in which a number of template topographic maps that best account for the data are obtained. The distinct microstates are then identified by comparing the template maps to the responses by way of spatial correlation. This was carried out on a concatenated data set consisting of all three conditions (i.e., AEP, BBN AESPA, and TONE AESPA)

as recommended in Murray et al. (2008). The choice of 75% correlation for merging was based on the analysis of our AEP response into the standard P1, N1, and P2 components (Picton et al. 1974). For consistency, we then used this same threshold for the analysis of AESPA responses. The choice of four samples as the minimum number needed to constitute a response was somewhat ad hoc and was chosen to avoid division of the AESPA into an unjustifiably large number of candidate components.

3.3.2 Results

The group-average standard AEP and the group-average AESPAs obtained using the modulated BBN and TONE stimuli, along with the standard error (SE) across subjects, are shown in Fig. 3.4 for eight electrode locations distributed around the scalp, all referenced to the average of the mastoids. Clear responses can be seen for all three methods. The standard AEP exhibits clear P1, N1, and P2 components at around 70, 120, and 200 ms, respectively, and is larger in amplitude at frontal sites compared with those in occipital areas. For the BBN AESPA at frontal locations a succession of small peaks and troughs can be seen in the range 0–100 ms along with prominent negative and positive components at 145 and 200 ms, respectively. In more posterior locations, the earlier components remain visible, but the large, later peaks are absent. Similarly for the TONE AESPA, there is some notable activity in the 0 to 100-ms range at all electrodes shown with large negative and positive later components being much more evident at frontal sites.

Figure 3 (a)

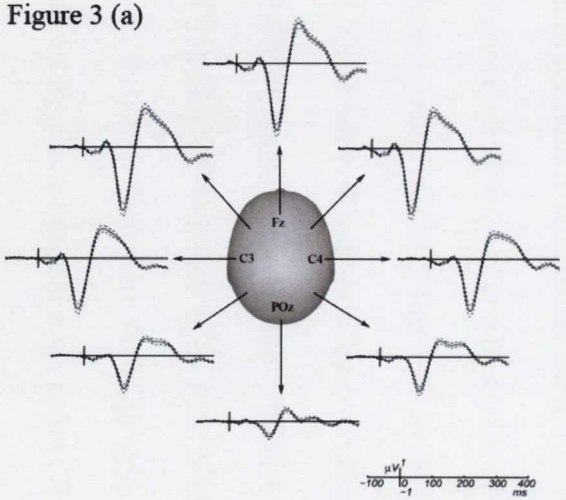

Figure 3. (b)

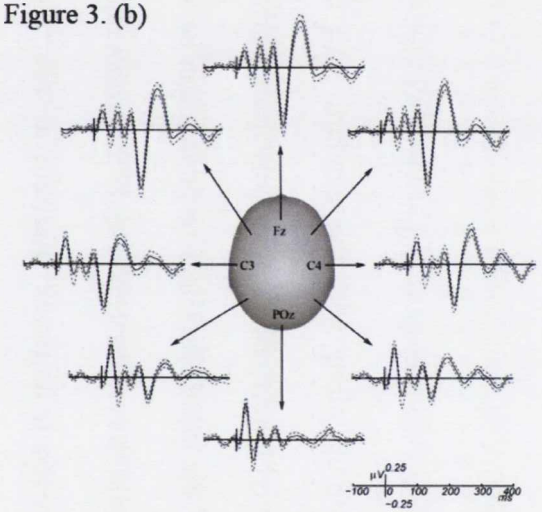

Figure 3. (c)

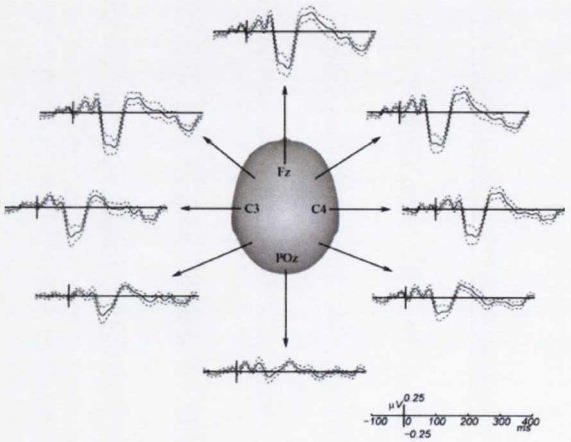

Figure. 3.4 Average (*A*) standard AEPs, (*B*) AESPAs using BBN carrier stimulus, and (*C*) AESPAs using ONE carrier stimulus at selected electrode locations, chosen to give a representation of the responses over the whole scalp. The solid lines represents the grand average, whereas the dashed lines represent the SE over the 12 subjects. (from Lalor et al., 2009)

AESPA components

Fig. 3.5 presents the group-average data in three ways. The *bottom panels* show the global field power (GFP; Lehmann and Skrandies 1980; Skrandies 1995) for each response, with the horizontal dashed line on this panel indicating twice the mean GFP in the interval -100 to 0 ms. The *middle panel* represents a butterfly plot showing all channels simultaneously. The microstates identified by the topographic analysis are indicated by vertical lines and their corresponding topographies are shown in the *top panel*. Not wanting to equate the componentry of the AESPA with that of the AEP, we tentatively designated seven BBN AESPA candidate components, based on the microstates identified by our analysis, as follows: Pa (21–45 ms), Na (45–61 ms), Pb (61–92 ms), Nb (92–104 ms), Pc (104–125 ms), Nc (125–170 ms), and Pd (170–238 ms). Although twice the mean GFP in the interval -100 to 0 ms was used as a component threshold, the response in the fifth interval (Pc) was also considered to be a candidate component, despite not exceeding this threshold, given that it was identified as a distinct microstate in the previous analysis. The same seven components are evident to a degree in the TONE AESPA, although the Nb, Pc, and Nc components appear to be less easily distinguished and are, in fact, identified as a single microstate by the topographic analysis. The Pd component is also present in the TONE response, but seems a little less well defined. Both of these results may be considered in light of the much lower SNR of the TONE AESPA response, which is discussed in the following text. The butterfly and GFP plots in Fig. 3.5 and the topographic analysis strongly suggest the presence of these seven components, although we sought to submit this hypothesis to a further statistical test. Figure 3.6 consists of a pair of statistical cluster plots marking time points at which

the BBN and TONE AESPAs differ significantly from zero for each channel. These plots were calculated using running *t*-tests across subjects at an alpha level of 0.05. In the case of the BBN AESPA, three highly distinct clusters are seen representing Pa, Nc, and Pd, whereas there appears to be some activity in central areas around 380 ms. The TONE AESPA also exhibits significant clusters for Pa, Nc, and Pd, although the Nc activity is much more widely distributed in time and the Pa activity is less evident over posterior and occipital scalp. The late activity around 380 ms, however, is much more evident. Although the four proposed components between Pa and Nc did not show up as significant for either response, the fact that the response has clearly begun by 20 ms -combined with evidence from the butterfly and GFP plots of Fig. 3.5 and the microstates analysis - suggests that perhaps the *t*-tests were simply underpowered. It is also worth noting that the small local maximum evident at around 10 ms in the BBN GFP plot, which is suggestive of an even earlier component, does not show up as statistically significant.

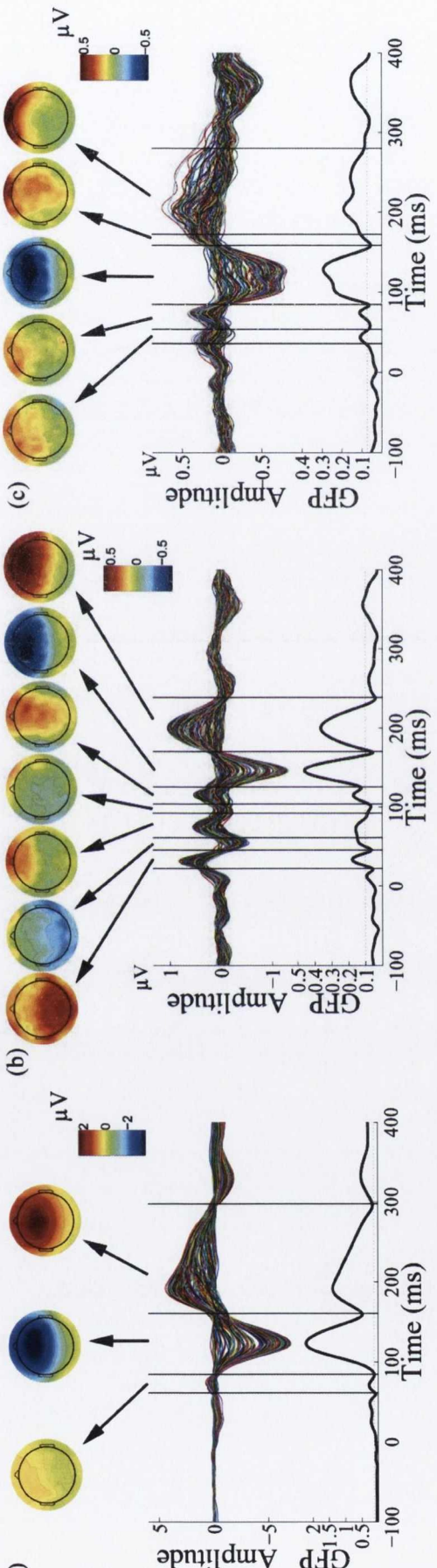

Figure 3.5 Global field power (GFP) responses, with the horizontal dashed line indicating twice the mean GFP in the interval -100 to 0 ms and the vertical lines indicating the candidate component intervals as established from the BBN AESPA (*bottom plot*), butterfly plots showing all channels plotted simultaneously (*middle*), and the average scalp topographies over the candidate component intervals for the (a) AEP, (b) BBN AESPA, and (c) TONE AESPA. (from Lalor et al., 2009)

Comparison of AEP and AESPA

Clearly the AEP exhibits a very different morphology with only three components being evident in Fig. 3.5(a). The prominence of the N1 and P2 components of the AEP and the Nc and Pd components of the AESPAs suggest that they may be related, although the timing is notably different. This is backed up somewhat by the topographical maps in the *top panels* of Fig. 3.5, which show that the AEP's N1 exhibits as a strong frontal negativity. The Nc of the BBN AESPA in the sixth interval appears to be similarly distributed, as does the negativity of the TONE AESPA in its third interval. Furthermore, these three components were all represented by the same template map in the topographic analysis and show high spatial correlation (AEP/BBN: $r = 0.96$; AEP/TONE: $r = 0.98$; BBN/TONE: $r = 0.94$; all with $P < 0.01$). All methods also display a strong late positivity in frontal areas—i.e., in the third interval for the AEP, the fourth and fifth intervals for the TONE AESPA, and in the seventh interval for the BBN AESPA. Again, all were significantly spatially correlated (AEP/BBN: $r = 0.93$; AEP/TONE: $r = 0.55$; BBN/TONE: $r = 0.77$; all with $P < 0.01$). This suggests that these components may derive from similar generators. Further comparison between responses can be seen in Fig. 3.7, which illustrates the values of the correlation coefficients at each electrode for each pair of methods. This analysis was aimed at providing information on how the latencies of the different responses differed at every electrode location. Somewhat surprisingly, the TONE AESPA and the AEP are seen to be the most strongly correlated pair, with $r = 0.6$ ($P < 0.01$) for all frontal and almost all central electrode locations. Both AESPA responses are seen to be widely correlated with each other, although not as strongly as the TONE AESPA and AEP. The BBN

AESPA and AEP exhibit significant, if not very strong, correlations in central and frontal locations only. As is evident from the SE traces in Fig. 3.4, there is a discrepancy in the SNR between the AEP and the AESPA responses. SNRs of the GFP of the group-average responses for the AEP, BBN AESPA, and TONE AESPA were 25.18, 11.42, and 11.93 dB, respectively.

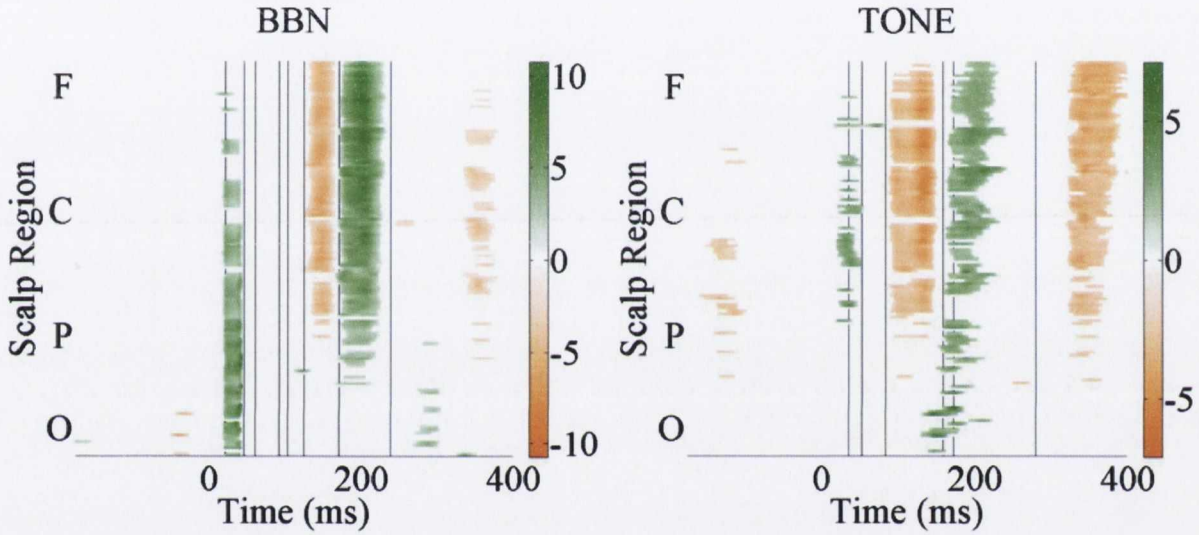

Figure 3.6 Statistical cluster plots marking the time points for all electrodes, at which the AESPA response differed significantly from zero on the basis of 2-tailed *t*-tests at a p-value of 0.05. White denotes nonsignificance, whereas positive *t* values (AESPA > 0) are marked on a green scale and negative *t*-values (AESPA < 0) are marked in gold. Electrodes are ordered from the *bottom*, occipital (O), parietal (P), central (C), and frontal (F) proceeding in the anterior direction in rows from *left* to *right*. In the case of the BBN AESPA, 3 highly distinct clusters are seen representing Pa, Nc, and Pd, whereas there appears to be some activity in central areas around 380 ms. The TONE AESPA also exhibits significant clusters for Pa, Nc, and Pd, although the Nc activity is much more widely distributed in time and the Pa activity is less evident over posterior and occipital scalp. The late activity around 380 ms, however, is much more evident. (from Lalor et al., 2009)

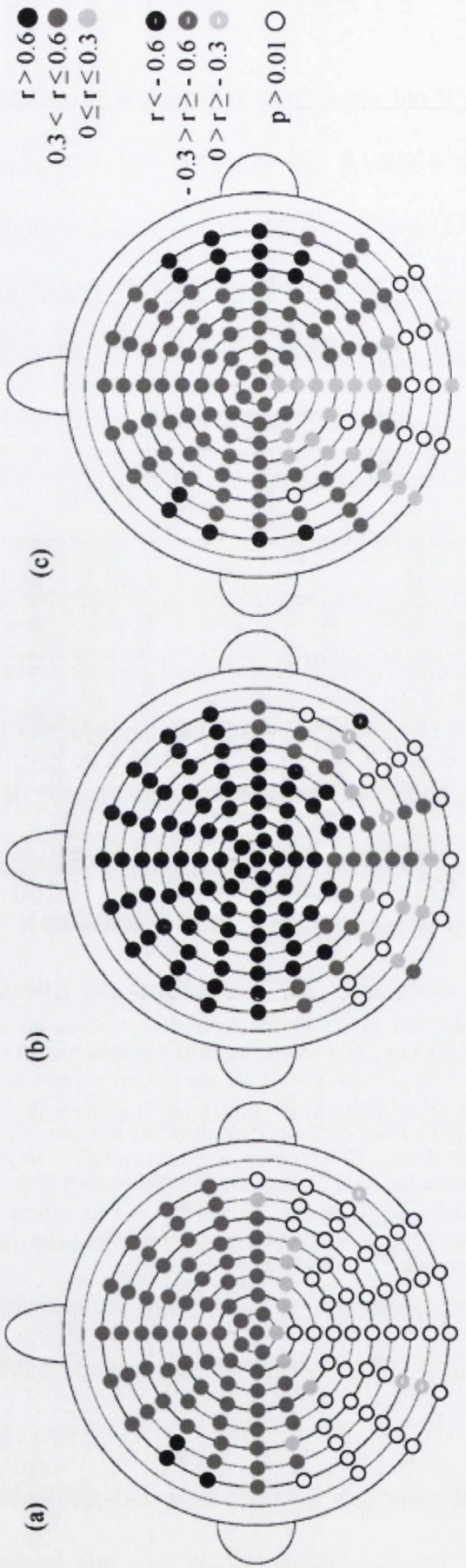

Figure. 3.7. Correlation coefficients for the standard AEP and the BBN AESPA (a), the standard AEP and the TONE AESPA (b), and the BBN AESPA and TONE AESPA (c). The black, dark gray, and light gray circles indicate a significant correlation with $r > 0.6$, $0.3 \leq r \leq 0.6$ and $0 \leq r \leq 0.3$, respectively. The white circles indicate electrodes at which correlations were insignificant (i.e., $P > 0.01$). Circles with "-" in the center denote negative correlations with strengths indicated using the same shading scheme as above. (from Lalor et al., 2009)

AESPA as a generalization of the AEP

Although the AESPA responses and the AEP exhibit clear differences, there are clearly some similarities. This is not surprising given that, as discussed in the study describing the visual analog to the AESPA (the VESPA; Lalor et al. 2006), the standard AEP can be considered a special case of the broader AESPA method. That is, although the AEP method is restricted to the use of discrete stimuli that are isolated in time, the AESPA stimuli can include such discrete events along with periodic bursts, temporal rate changes, and fully continuous stimulation. To demonstrate this, we analyzed the continuous (AESPA), composite, and discrete (AEP) stimuli using both the AESPA analysis method and simple signal averaging for the two subjects who undertook the composite stimulus resentations. Fig. 3.8 plots the corresponding responses at frontocentral electrode location Fz, referenced to the average mastoids. In the continuous stimulus case, signal averaging was carried out time-locked to positive intensity changes that exceeded 0.5, whereas in the composite case, we time-locked to all onsets. Although responses are evident when using signal averaging on the continuous and composite data, they exhibit much lower SNRs (3.92 dB for the continuous and 4.98 dB for the composite stimulus on the GFP) and the componentry is much less sharply defined than that in the case where the same data are analyzed using the AESPA method (SNRs: 10.95 dB for the continuous and 13.16 dB for the composite on the GFP). Furthermore, the responses produced by the AESPA method for both the continuous and composite responses are highly correlated ($r = 0.62$, $P < 0.01$ at Fz), whereas those produced by the averaging method are significantly *negatively* correlated ($r = 0.33$, $P < 0.01$ at Fz). This serves to illustrate that, for such nontrivial stimuli, the effect of the noise

on the averaging procedure has been sufficient to render the AEP an inconsistent measure of neural activity, whereas the AESPA method continues to deliver a stable response. Finally, because they are mathematically equivalent, using a train of impulses as input to the AESPA analysis (i.e. a series of delta functions at the time points where the discrete stimuli begin) unsurprisingly results in a response to the discrete stimulus that is identical to that obtained using the signal averaging method, reinforcing the notion that the AESPA method can be considered a generalization of the traditional AEP technique.

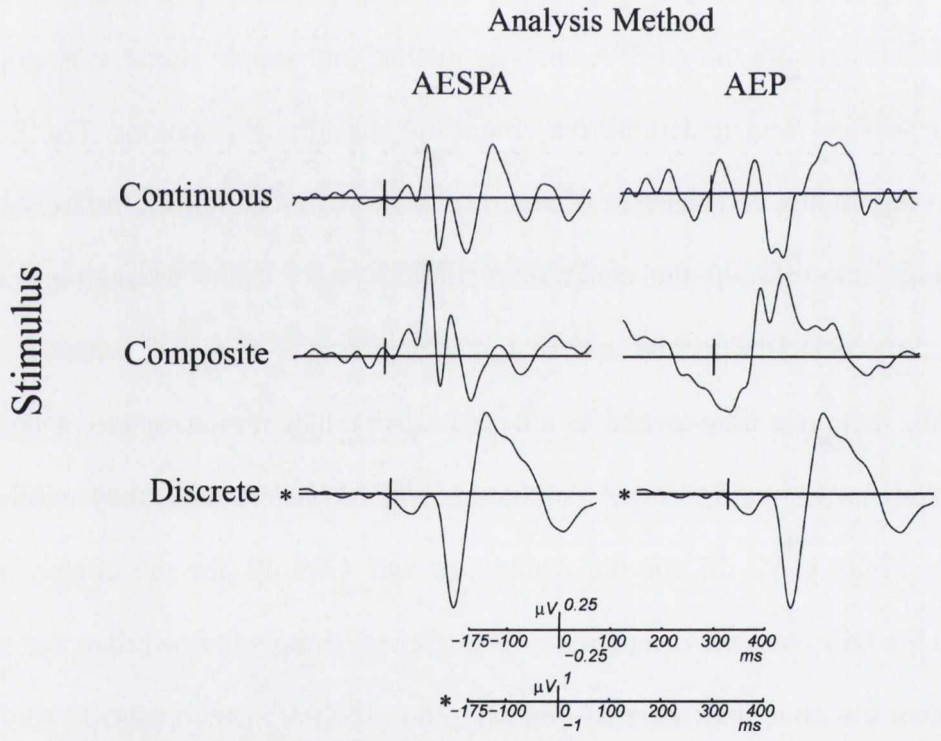

Figure. 3.8 Responses to the 3 different stimuli used analyzed by both the AESPA method and the time-locked averaging method used in standard AEP extraction. These responses are the grand average of 2 subjects who each underwent 30 trials of each of the continuous and composite stimulus type and 10 trials of the discrete stimulus. Note the plots for the discrete stimulus are on a different scale from that of the other plots. There seems to be some activity evident for the continuous and composite stimuli analyzed using the AEP technique, but the AESPA method results in a much higher signal-to-noise ratio, highlighting its ability (from Lalor et al., 2009)

Source analysis of the AESPA

To further investigate the origins of the AESPA componentry, we performed source analysis using the BESA software suite (http://www.besa.de). Because of the greater SNR of the BBN AESPA, we began our source analysis with the group average BBN response. We used the following strategy to arrive at a solution:

1) We first fit a window from 138 to 152 around the Nc peak with a pair of symmetrically constrained dipoles. The fit is exceptional, explaining 98.9% of the variance in this window. Looking at the model residual showed no remaining signal around the Nc period. These two dipoles, which are shown in red in Fig. 3.9, were localized very close to Heschl's gyrus (X,Y, Z -39, -15, 16). In fact, fixing these dipoles in the centroid of the auditory core (about -46, -24, 12), about 1 cm from where they landed when allowed to freely fit, and allowed their orientation to vary, the model still explained 98.6% of the variance in this window. As such, it is with good confidence that we say that the Nc is generated in and around the auditory core, consistent with the localization of the AEP's N1 component in many previous studies (see Leavitt et al. 2007).

2) The residual waveforms indicated that a number of obvious components preceding the Nc timeframe remain unexplained by primary auditory generators. For instance, the Pa component at 30 ms is still fully evident and mapping the residual shows a bilateral posterior map. Fitting a pair of symmetrically constrained dipoles to this period (23–41 ms) resulted in

localization to the inferior occipitotemporal surface. This seemed unlikely to be correct. Looking at the butterfly plot in Fig. 3.5, we noted a rather unipolar distribution for this peak, which fits the pattern expected of a deep source. This was confirmed by a reasonably good fit (explained variance 89.6%) using the addition of a single dipole in the brain stem (8, -39, -26). Although identification of brain stem sources using EEG is unlikely to be completely reliable, the fact that this component is significantly positive at almost every scalp electrode when referenced to the average of the mastoids and the short latency of the component strongly suggest a subcortical source. This source may actually be the thalamus, given previous work linking AEP components at this latency with thalamic sources (Picton et al. 1974a).

3) The distribution of the residual over the Pc component indicated that bilateral cortical sources were likely involved. Fitting bilateral symmetric dipoles from 110 to 130 ms resulted in a pair of locations in the temporal lobe in the region of Brodmann Areas 21 and 22, between the superior and middle temporal gyri (-48, 1, -31), with a different orientation to the primary Nc generators. Explained variance in this timeframe was 94.6% and there was no obvious signal left in this timeframe in the residual waveforms.

4) Working backward in time, another relatively distinct component was evident with a peak at about 70 ms. The topography of the residual during this timeframe was multifocal and complex and was fit well by a pair of dipoles in the middle frontal gyrus (-37, -2, 40). Between 21 and 238 ms, this

simple model, shown in Fig. 3.8, accounted for 97.3% of the variance. Allowing this seven dipole fit to go free (with the exception of the primary auditory generators) across the 21- to 238-ms range showed that the solution was very stable, with no improvement in explained variance and very little movement in dipole location. Using this solution to fit the average BBN AESPAs on a subject-by-subject basis gave us an average explained variance of 84.1%with SD of 8.4%. Given the difficulty in arriving at a stable solution for the TONE response based on its lower SNR, and the similarities in componentry (Fig. 3.4) and topographies between the responses, we examined how well the BBN source solution did in explaining the variance of the TONE response. The TONE response was very well accounted for by the solution with 94.2% of the variance explained. The average explained variance across subjects was 74.4% and SD of 7.3%. Interestingly, the same solution accounted for 98.7% of the variance of the group average AEP over the interval 21–238 ms.

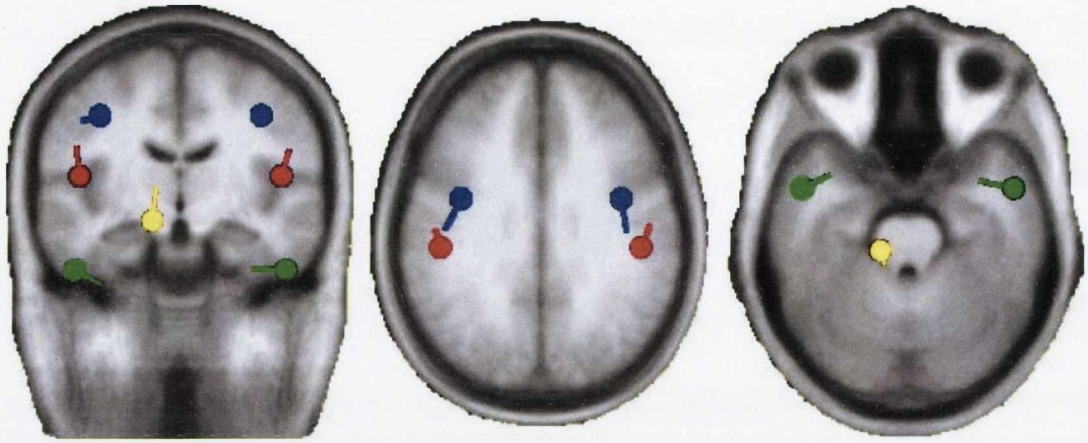

Figure. 3.9 Source dipoles for the group average BBN AESPA response obtained using BESA software. Between 22 and 237 ms, this simple model accounted for 97.3% of the variance. On a subject-by-subject basis the average explained variance was 84.1% with SD of 8.4%. The TONE response was also very well accounted for by the same solution, with 94.2% of the variance explained, as was the AEP, with 98.7% of the variance of the group average accounted for. Source locations with Talairach coordinates in parentheses: red: centroid of the auditory core (about -46, -24, 12); yellow: the brain stem (8, -39, -26); green: between the superior and middle temporal gyri (-48, 1, -31); blue: middle frontal gyrus (-37, -2, 40). (from Lalor et al., 2009)

3.3.3 Discussion

We have described a method that allows for the estimation of a novel auditory potential, known as the AESPA, using continuous, amplitude-modulated stimuli. These responses display temporally detailed componentry with high SNR and can be elicited using stimuli with differing characteristics. This method may facilitate research that cannot be conducted using the traditional ERP technique of discrete stimulation and averaging. Although the goal of this study is to present the AESPA response as a novel tool for carrying out research on the auditory system and *not* to equate it with the traditional AEP, comparison between the two methods can help in characterizing this new response. In fact, although the AESPA responses show significant correlation with the standard AEP, they are undoubtedly distinct, particularly in posterior regions. One obvious reason for this is that the AESPA signal processing assumes a linear relationship between the input stimulus intensity and the output EEG, whereas no such assumption is made in calculating the AEP. As

such, the AESPA likely reflects activity of a limited subset of auditory cells for which this assumption holds true. That said, it should be noted that the analysis outlined in this study can easily be extended to include higher-order processing effects (Lalor et al. 2008). The notion that activity of a limited subset of cells is reflected in the AESPA also speaks to the issue of the differing morphologies of the TONE and BBN AESPA responses. Specifically, given the tonotopic nature of auditory cortex (Howard et al. 1996; Morel et al. 1993; Pantev et al. 1988), it is likely that the response to the amplitude-modulated TONE stimulus reflects activity of an even more specific subset of cells than that of the BBN stimulus. Tonotopy of auditory cortex may also explain why the TONE AESPA is so highly correlated with the standard AEP (Fig. 3.7), given that both were elicited by 1-kHz tones in the current study. The dissimilar nature of the stimulation methods used to elicit the AEP and AESPA is another likely reason for the differing responses. It has been reported that only a limited number of auditory cortical neurons display stimulus-synchronized responses (Lu et al. 2001). As a result cortical representations of rapidly occurring stimuli are likely to be quite distinct from those of isolated, discrete stimuli. Specifically, it is thought that two processes in auditory cortex account for the representation of sound modulation. Slower modulation frequencies (30 Hz) are thought to be represented by neuronal discharges that are temporally synchronized to the stimulus, whereas higher modulation frequencies are represented by nonsynchronized rate-based discharges (Lu et al. 2001). Further evidence of this dichotomy was shown in a study using amplitude-modulated sound in squirrel monkeys (Bieser and Müller-Preuss, 1996). There it was reported that 78.1% of all acoustically driven neurons encoded the envelope of the AM sound with most

neurons using a combination of two distinct modes. One mode involved spikes that followed the AM envelopes in a phase-locked manner, whereas the other involved significant changes in spike rate with the changing stimulation. Based on this mixed encoding of AM modulation in nonhuman primates, it is unclear what percentage of human auditory cortical cells would respond to an AM signal in such a way as to be reflected in the AESPA. Bieser and Müller-Preuss (1996) also note that the 21.9% of neurons not encoding AM displayed simple on, on–off, or off responses at the beginning or the end of a stimulus sound. The activity of cells such as these would be reflected in a standard AEP and not in an AESPA obtained to a continuous stimulus. The fact that the SNR values for the AESPA responses obtained to the continuous and composite stimuli in this study are so similar and that the two responses are so highly correlated, however, suggests that further work is required to investigate this line of reasoning. The AEP and the AESPA responses are also notably different in terms of their amplitude. There are a number of possible reasons for this. First the aforementioned likelihood that the AESPA reflects activity from a much more specific subset of cells is probably a factor. Second, as can be seen in Fig. 3.3, the intensity of the AESPA stimuli in this study rarely reached that used in generating the AEP which, given the dependence of response amplitude on stimulus intensity (Beagley and Knight 1967), likely plays a role. Third, and perhaps most important, acquiring the AEP by signal averaging involves the assumption that a stimulus of nonzero duration (in our case 50 ms) is essentially an ideal unit impulse. Thus the only stimulus attribute that is taken into consideration when calculating an AEP is the stimulus onset time, whereas other features such as duration of ramp-up/down and offset time are typically ignored. The AESPA needs to make no such assumption

because the flexibility of the method allows the experimenter to take into account the fact that the auditory stimulus has its energy spread in time when calculating the response. Thus because the AESPA represents a *function* that produces an estimate of the EEG output based on a broader window of input stimulus values, it is not surprising that it differs in amplitude from that of the AEP, which represents a *response* to an assumed ideal impulse.

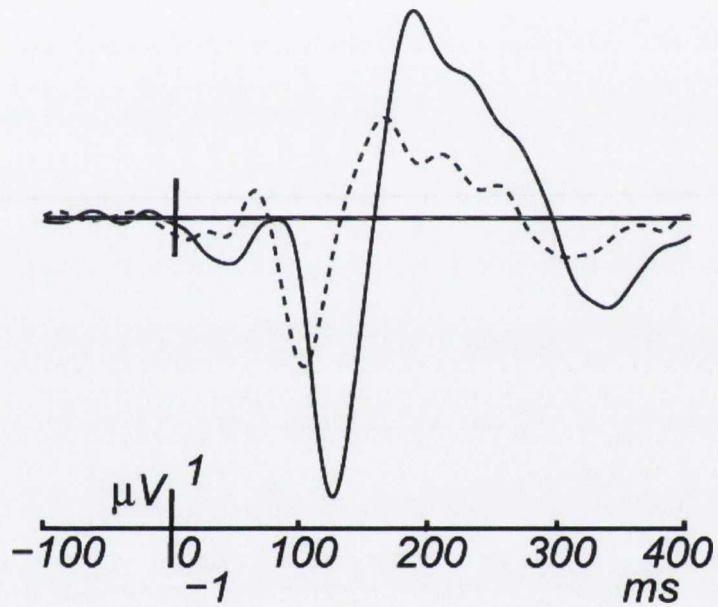

Figure. 3.10 AESPA responses due to the discrete stimulus analyzed with an impulse (solid trace) and a 50-ms window (dashed trace). The flexibility of the AESPA method allows the experimenter to take into account the fact that the auditory stimulus is not an ideal impulse but has its energy spread in time when calculating the response. As expected, stimulus power spread across 50 ms results in a smaller response function with a shorter latency. (from Lalor et al., 2009)

Fig. 3.10 plots the results of the AESPA analysis on the discrete data in the case where we assume the stimuli to be ideal impulses and the case where we take into account the full 50-ms profile of the stimulus. As expected, taking account of the fact that the stimulus power is spread across 50 ms results in a smaller response function with a shorter latency. Interestingly, some subcomponents of the P2 wave also become more evident when using the full stimulus profile, suggesting that the

assumption of a 50-ms stimulus as an impulse might result in some blurring of neighboring componentry. The effect of taking into account the detailed nature of the stimulus profile is also evident in Fig. 3.8, where the AESPA method clearly outperforms the signal-averaging technique for both the continuous and composite stimuli. This demonstrates the flexibility of the AESPA method in terms of producing high-quality responses to stimuli with diverse acoustic properties. Given the purported variation in the response of auditory cells to stimulus modulations with different temporal statistics and the prevalence of complex stimulation methods in the auditory processing literature as well as in the natural world, this flexibility suggests the AESPA as a useful addition to the palette of research tools currently used. Future work will aim to examine the characteristics of AESPA responses at more points along the "continuous– discrete" stimulus spectrum using a variety of carrier signals. The topographic analysis carried out in this study provides another means for comparing the AESPA responses and the AEP. Again, it is not the aim of the current study to equate the AESPA with the AEP, but to use the standard AEP to assist in characterizing the AESPA response. The high spatial correlations of the large late negativity and the large late positivity between each of the three methods and the fact that the same template map was assigned to the negativity for each of the three methods lends significant weight to the notion that they may be generated in similar cortical areas. The source solution generated for the BBN AESPA in this study provides additional support for some commonality in the cortical generators. The fact that the same simple solution accounted for 97.3, 94.2, and 98.7% of the variance for the BBN AESPA, TONE AESPA, and AEP, respectively, strongly suggests that similar neural pathways are involved in generating the responses—a

result that does not come as a surprise. Unlike their visual analogs, the VESPA and visually evoked potential (VEP; Lalor et al. 2006), the AESPAs shown in this study display lower SNR than that of the traditional AEP given a comparable amount of testing time. This is particularly the case for the TONE AESPA. The likely limited number of cells for which the AESPA's assumption of linearity hold true and perhaps a less than optimal choice of carrier signals are possible reasons for this. It should also be noted that no artifact rejection was performed on the AESPA data in this study, which would assuredly influence the SNR. The reason for this is that the continuous nature of the stimulus makes it difficult to know which data to reject when an eyeblink or movement artifact is identified. Because such artifacts are independent of the stimulus signal against which we are regressing the EEG, they do not have any systematic effect on the AESPA response and would influence the response only through a slight reduction in SNR. This can be somewhat addressed using independent component analysis as was previously carried out on VESPA data (Lalor et al. 2007) or by using shorter trials and averaging responses across them.

A number of research applications can be proposed for which the AESPA may be better suited than the traditional AEP. As mentioned earlier, the temporally constrained and simplistic nature of the stimuli required by the AEP method surely makes the method suboptimal for a system that has evolved to analyze an infinitely broad range of sounds. Much of today's research on auditory processing focuses on using complex, species-specific stimulation (see Suga 1995) for which the traditional ERP method is not entirely suited. The flexibility afforded by the complex AESPA stimuli may make it a very useful tool for analysis of the temporal processing of

speech or speech-like sounds in humans. Another potential advantage of the AESPA method is in simultaneously obtaining responses to multiple stimuli. Lalor et al. (2006) demonstrated the ability to obtain VESPA responses to two concurrently presented stimuli in different parts of the visual field. If we assume an analogy between tonotopy in the auditory domain and retinotopy in the visual domain, it is highly likely that AM of multiple carrier waves that do not overlap in frequency using independent modulating signals would enable the estimation of separate AESPA responses to each. This would render the method useful in various applications such as research on auditory scene analysis and on the effects of attention on the temporal processing of audition. The VESPA method has already been used to demonstrate the timing of endogenous visual spatial attention, something that is very difficult to do using suddenly onsetting VEP stimuli (Lalor et al. 2007). Similarly, questions that are difficult to address using isolated suddenly onsetting AEP stimuli may be investigated using the AESPA.

In summary, evoked responses with detailed temporal profiles, known as AESPAs, have been obtained using complex, continuous auditory stimuli. To begin characterizing these responses, comparisons have been made with the standard AEP and similarities and differences between the two responses have been discussed. The demonstrated flexibility of the AESPA method gives it a number of advantages over the traditional AEP technique and suggests its applicability to fundamental research on auditory processing as well as studies of attention, auditory scene analysis, and perhaps in a number of neurological disorders where auditory processing is affected.

Thus having characterized the AESPA and outlined some of its advantages I wish to exploit a particularly interesting trait of the AESPA method i.e. the ability to obtain responses to two (or more) simultaneous presented and truly competing stimuli. To this end the following chapter describes investigations into the dynamics of endogenous auditory attention using two concurrent AESPA stimuli.

"Choice of attention – to pay attention to this and ignore that – is to the inner life what choice of action is to the outer. In both cases, a man is responsible for his choice and must accept the consequences, whatever they may be." W.H. Auden (1907 - 1973). Anglo-American poet.

4

AESPA and endogenous auditory attention

Endogenous attention is the self-directed focus of attention to a region or feature of the environment. The research outlined in this chapter assesses the effects of endogenous attention on temporally detailed responses to continuous and competing auditory stimuli obtained using the AESPA method. As described in Chapter 2 there is some debate as to whether an enhancement of sensory processing is involved in endogenous attention. It has been suggested that attentional affects are not due to increased sensory activity but are due to engagement of separate temporally overlapping non-sensory attention-related activity. Due to the nature of the AESPA method the obtained responses represent activity directly related to the

stimulus envelope and thus predominantly characterize obligatory sensory processing (Lalor et al., 2009, Chapter 3). In this chapter attentional modulations at ~136ms (during the Nc component of the AESPA response) are identified and localised to auditory cortex. Although the involvement of separate non-sensory attentional centres cannot be ruled out, these findings clearly demonstrate that endogenous attention does modulate obligatory sensory activity in auditory cortex.

The majority of EEG studies investigating endogenous auditory attention (e.g. Hillyard et al., 1973, Näätänen et al., 1987) have employed the standard event-related potential (ERP) technique. These studies, although informative, are somewhat hampered by the fact that they use discrete stimuli to investigate endogenous attention. Specifically, the assertion that attentional effects to discrete onset stimuli are entirely due to endogenous top-down attention is complicated by the exogenous attention grabbing effects of onset stimuli. It has been suggested, in the visual domain, that endogenous and exogenous attention may be separate systems (Hopfinger and West 2006). Whether these proposed separate systems affect sensory processing in the same manner, however, is still unknown, as is how they interact when both are engaged. That said, despite the fact that the attention grabbing nature of discrete stimuli is well established in both visual (Jonides 1981) and auditory (Escera et al. 1998) modalities there have been very few studies which attempt to eliminate the possible confounding exogenous attentional influence on the investigation of endogenous attention.

The continuous nature of steady-state response (SSR) stimuli likely minimises the influence of exogenous attention on the SSRs. Studies employing SSR to investigate sustained auditory attention have had mixed results, however: One study using simultaneously presented click trains presented at a rate of either 37Hz or 40Hz found a transient N1 effect to stimulus onset but no steady-state effect (Linden et al. 1987), which suggests that sustained responses are not affected by endogenous attention. A more recent study using AM tone bursts of 800ms duration, however, found a steady-state effect but no transient N1 effect (Ross et al. 2004), which suggests that sustained responses are affected by endogenous attention. The latter study did not employ a stimulus competition design however and used a visual task in the unattended condition. Thus we are unable to discern, from this experimental protocol, whether one competing stimulus would result in increased activity relative to another. The somewhat contradictory results suggest that the effects of sustained endogenous auditory attention on competitive, continuous stimuli have yet to be adequately elucidated.

A further issue relates to the simultaneity of stimuli in most ERP based studies. In such studies the discrete stimuli are not presented truly simultaneously in the attended and unattended channels (Hillyard et al. 1973, Näätänen et al. 1992). Thus the argument that the effects found are due to competing stimuli is somewhat weakened. In an attempt to overcome this drawback, paradigms where simultaneous stimulation is employed have been carried out. A recent study assessed the transient onset responses of subjects who were asked to detect an occasional change in modulation frequency of amplitude modulated sounds presented to one ear while

ignoring concurrent sounds in the other ear (Ross et al. 2010). Interpretation of these results, however, is complicated by the fact that any increase in the transient response to the attended stimulus would be superimposed on the unaffected (or even possibly inhibited) response to the unattended stimulus perhaps diluting any attention affect. Indeed researchers investigating auditory scene analysis have been forced to go to great lengths to create the perception of auditory streaming using discrete stimuli (Sussman et al. 1999, Ritter et al. 2006, DeSanctis et al. 2008). Recently, however, there has been a move towards employing more natural stimulus paradigms to assess auditory function (Lalor et al. 2009, Lalor and Foxe. 2010, Kerlin et al. 2010). Lalor and Foxe (2010) obtained temporally detailed responses to natural speech timuli and Kerlin et al. (2010) found attentional modulation of activity in AC when using competing speech stimuli and a template matching analysis method.

Exploiting the AESPA method we employed a cocktail-party-like experimental approach to investigate the effects of endogenous attention on sensory processing in the auditory system. The continuous nature of the AESPA method minimises the confounding influence of the attention grabbing nature of discrete stimuli. We present two continuous auditory streams concurrently, one to the left and one to the right ear, and subjects are instructed to attend either to the left or right.

4.1 Materials and Methods

Subjects and Data Acquisition

17 subjects aged 21-33 (15 male) participated in the study. The experiment was undertaken in accordance with the Declaration of Helsinki. The Ethics

Committee of the School of Psychology at Trinity College Dublin approved the experimental procedures and each subject provided written informed consent. Subjects reported no history of hearing impairment or neurological disorder. EEG data were recorded from 130 electrode positions, lowpass filtered to the range 0 – 134 Hz and digitized at the rate of 512 Hz using a BioSemi Active Two system. EEG data were then digitally filtered off-line with a high-pass filter, where the passband was above 2 Hz and with a -60 dB response at 1 Hz and a low-pass filter with passband below 35Hz and a -50 dB response at 45 Hz. The data at each channel were re-referenced to the average of the responses at the left and right mastoids.

Stimuli

The AESPA stimulus consists of a carrier stimulus amplitude modulated by a spread spectrum signal. In this case two root mean square (RMS) normalised band-pass noise carriers of bandwidth 1kHz centred at 1kHz (LOW stream) and 5kHz (HIGH stream) respectively were employed (Fig. 4.1 (c) and (d)). These centre frequencies were chosen on the basis that 1kHz and 5kHz tones are perceived with approximately the same loudness (ISO:226) and also because they are far enough apart in frequency that they are perceived separately. These carriers were then modulated by independent spread spectrum signals and the low and high frequency streams were concurrently presented to the left and right ears respectively. The reason that the stimuli were separated in location as well as carrier stimulus centre frequency is because 1) it is spatial attention that is under investigation and 2) if both the left and right streams had identical carriers the stimuli may have been perceived

as one auditory object varying in inter-aural intensity difference as opposed to two spatially separate objects.

The spread spectrum signals consisted of Gaussian noise with energy uniformly distributed between 0-30 Hz. Taking into account the logarithmic nature of auditory stimulus intensity perception, the values of these modulating signals, x, were then mapped to the amplitude of the audio stimulus, x′, using the following exponential relationship:

$$x' = 10^{2x},$$

and normalized to between 0 and 1. It was expected that this would result in a more linear perception of audio intensity modulation. Transitions between levels were smoothed by using a 5ms ramp consisting of half a period of a 100Hz sine wave. Using the Nyquist sampling theorem and given that EEG power above 30 Hz is low, the modulation rate of each signal was set to be 60 Hz.

Experimental Procedure and Tasks

While abstaining completely from eye-blinks is not possible for long periods, subjects were instructed to keep the number of eye-blinks to a minimum during each session. Subjects were also instructed to keep all other types of motor activity to a minimum during testing. Testing was carried out in a dark room and in order to minimize eye movements subjects were asked to keep their eyes open and to fixate on a small cross presented in the centre of their visual field.

Each subject undertook ten trials where they were asked to attend to the HIGH stream in the right ear (attend-HIGH-RIGHT condition) and ten trials where they

were asked to attend to the LOW stream in the left ear (attend-LOW-LEFT condition). Each trial was 120 seconds in duration. The sequence of conditions was randomized for each subject. Stimuli were presented at an intensity level deemed comfortable by the subject before beginning the experiment.

In order to monitor each subject's progress, targets and distracters were inserted randomly in each stream. These events consisted of a specific pattern of amplitude modulation imposed on the random process. Targets consisted of a modulation level of -2.5dBfs for 25.5ms followed by -12dBfs for 16ms followed by -2.5dBfs for 25.5ms, giving a total length of 67ms, whereas distracters consisted of a flat modulation of -6dBfs for 67ms. dBfs refers to decibels full scale and represents a dB value relative to the maximum modulation level for each subject (see Fig. 4.1). Although the events are embedded in the stimulus they are still distinguishable from the on-going amplitude modulations. This is due to the fact that the events are generally somewhat louder than the on-going modulation. The reason for this is because of the exponential mapping outlined above which restricts the modulating waveform to spend ~90% of its time below -12 dBfs. Furthermore, due to the random nature of the modulating waveform non-event related amplitudes exceeding -12 dBfs generally have a short duration compared to the 67ms duration of events.

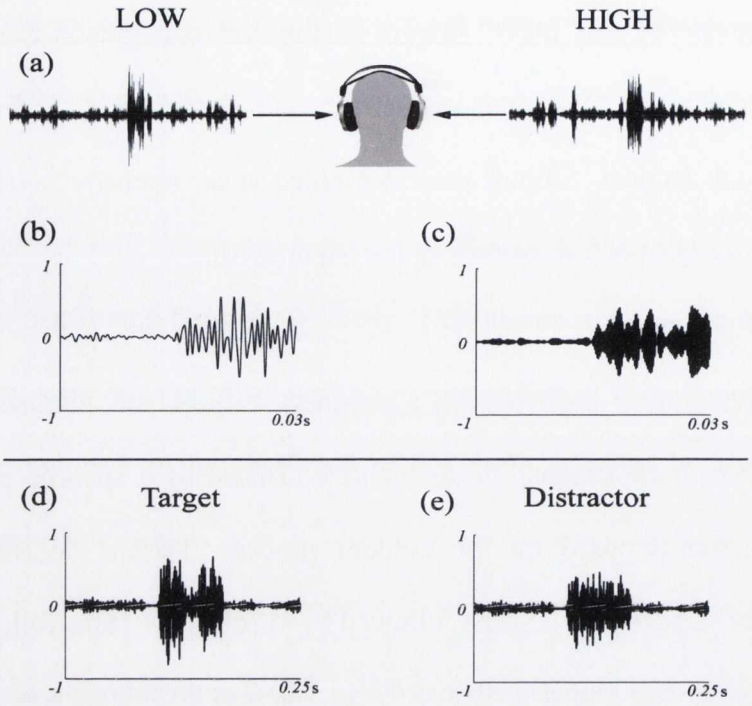

Figure 4.1. The stimulation setup. (a) LOW and HIGH streams were presented dichotically. A segment of the (b) LOW and (c) HIGH frequency carrier stimuli, respectively. Examples of (d) target and (e) distracter events.

Subjects were instructed to respond, by way of a button press, only when a target in the attended stream was heard. Each trial contained a total of 24 events (i.e., both targets and distracters). The proportion of targets and distracters in each trial was randomly assigned ranging from 8 targets (and therefore 16 distracters) to 16 targets (and therefore 8 distracters). On average 48.75% of events were targets and 51.25% were distracters. No event, either within or between streams, could occur within one second of another and the maximum separation between events within streams could not be more than 9 seconds. EEG was recorded for later analysis where both the responses to the HIGH-RIGHT and LOW-LEFT streams for each condition were extracted. The stimulation paradigm and the stimuli are outlined in Fig. 4.1.

Signal Processing

As outlined in Chapter 3 we obtain the AESPA by performing a linear least squares fit of the response model

$$y(t) = w(\tau) * x(t) + noise$$

Where y(t) is the measured EEG response, x(t) is the amplitude envelope of the stimulus, the symbol * indicates convolution, w(τ) is the impulse-response function to the amplitude of the stimulus, and the noise is assumed to be Gaussian (Lalor et al., 2009).

Quantification of Results

When calculating task performance any response occurring within a one second period after an event was considered to be a response to that event. We calculated the percentage of correct responses, percentage of responses to distracters in the attended stream and also percentage of responses to events in the to-be ignored stream, (see Table 1). To test the behavioural results statistically they were submitted to a 2 x 2 repeated measures ANOVA using factors of stimulus (LOW-LEFT vs. HIGH-RIGHT) and event in the attended stream (target vs. distracter).

The Global Field Power (GFP, Lehmann and Skrandies 1980) was obtained for the grand average responses to each stimulus condition. The GFP is a reference free measure of the response power over the whole scalp. The GFP was used for preliminary visualization of the data and to indicate possible periods of attentional

modulation of the responses. The periods of interest were obtained by submitting the GFPs to running t-tests. A component was considered to be of interest if responses were significantly different ($p<0.01$) for a period of at least 11 consecutive samples (~20ms). In order to investigate how different areas are affected by attention we also partitioned the scalp into 4 regions (Frontal, Central, Left Temporal and Right Temporal) and averaged over responses at electrodes within each region. Later AESPA components (i.e. Nc and Pd) have been shown to have very low signal-to-noise ratio in parietal-occipital regions (Lalor et al. 2009) and thus this region was not included in analysis. Fig 4.2 shows the responses in the included regions as well as the GFPs for the HIGH-RIGHT and LOW-LEFT responses when attended and unattended. In the GFP plots the areas identified as being of significant interest by the running t-tests are shaded grey. Topographic maps of the Nc component when attended and unattended are also shown, as are difference topographies.

Statistical differences in components identified as being possibly effected by attention were tested using a repeated measures ANOVA. The root-mean-squared (RMS) values in a ~10ms window around the component peak were used in the statistical analysis. Topographic maps of affected components were also obtained using the average amplitude in the relevant windows. Component topographies were compared using the TANOVA method (Murray et al. 2008). This method assesses whether two topographies are statistically different using a non-parametric randomisation procedure. Given that topographic maps reflect source configuration, this was done in order to assess whether the attentional modulation of any component was due to an engagement of additional generators or whether it was

merely due to enhanced activity of generators already engaged in the unattended condition (i.e., if topographies are not statistically different then we can assert that they are likely due to the same generators). A further more detailed investigation of the component generators was also carried out using the BESA (5.2) source analysis package.

4.2 Results

Behavioural results

In order to keep subjects highly engaged in the attentional task, discriminating between targets and distracters in the attended stream was deliberately made difficult. The difficulty of the task was evidenced by the relatively high number of distracters to which subjects responded (see Table 4.1). However, the fact that subjects responded to a significantly greater percentage of targets than distracters (main effect of event, $F(1,16) = 28.088$, $p < 0.001$) shows that subjects were capable of performing the task. Furthermore, the very low percentage of events responded to in the ignored stream indicates that attention to the intended stream was taking place. Despite the fact that we found a significant main effect of stimulus ($F(1,16) = 14.203$, $p = 0.002$), due to higher number of event responses when the right ear was attended than when the left ear was attended, there was no stimulus x event interaction ($p>0.05$). This suggests the ability to carry out the task (i.e., to distinguish between targets and distracters) did not differ depending on which ear was attended.

	LOW-LEFT	HIGH-RIGHT
p(T) ± std dev %	82.89 ± 11.59	86.75 ± 9.35
p(D) ± std dev %	57.83 ± 26.81	65.41 ± 26.84
P(TBI) ± std dev %	1.79 ± 1.29	1.41 ± 1.07

Table 4.1. Percentage correct responses (p(T)), percentage of responses to distracters in the attended stream (p(D)) and percentage of events responded to in the to-be-ignored stream (p(TBI)) for each condition.

Electrophysiology Results

Based on the running t-tests carried out on the GFPs we identified the Nc and Pd components as being of particular interest for further investigation. Nc and Pd were defined as the RMS amplitude in a ~10 ms interval around the mean latency of each component of the grand averages of the LOW-LEFT and HIGH-RIGHT responses (i.e. Nc: 136 ms and Pd: 208 ms, see Fig. 4.2). In order to test the effects of attention for each stimulation condition, we performed a 4-way 2 x 2 x 4 x 2 repeated-measures ANOVA using factors of stimulation (LOW-LEFT stream vs. HIGH-RIGHT stream), attention (Attended vs. Unattended), electrode region (Left, Right, Central and Frontal), and component (Nc vs. Pd). Greenhouse-Geisser corrections were applied to the repeated measures factors where the sphericity assumption was violated with the corrected degrees of freedom reported. Sphericity relates to the equality of the variances of the difference between levels of a repeated measure factor. Sphericity is an assumption of an ANOVA with a repeated measures factor and Greenhouse-Geisser is a correction method employed when sphericity is violated.

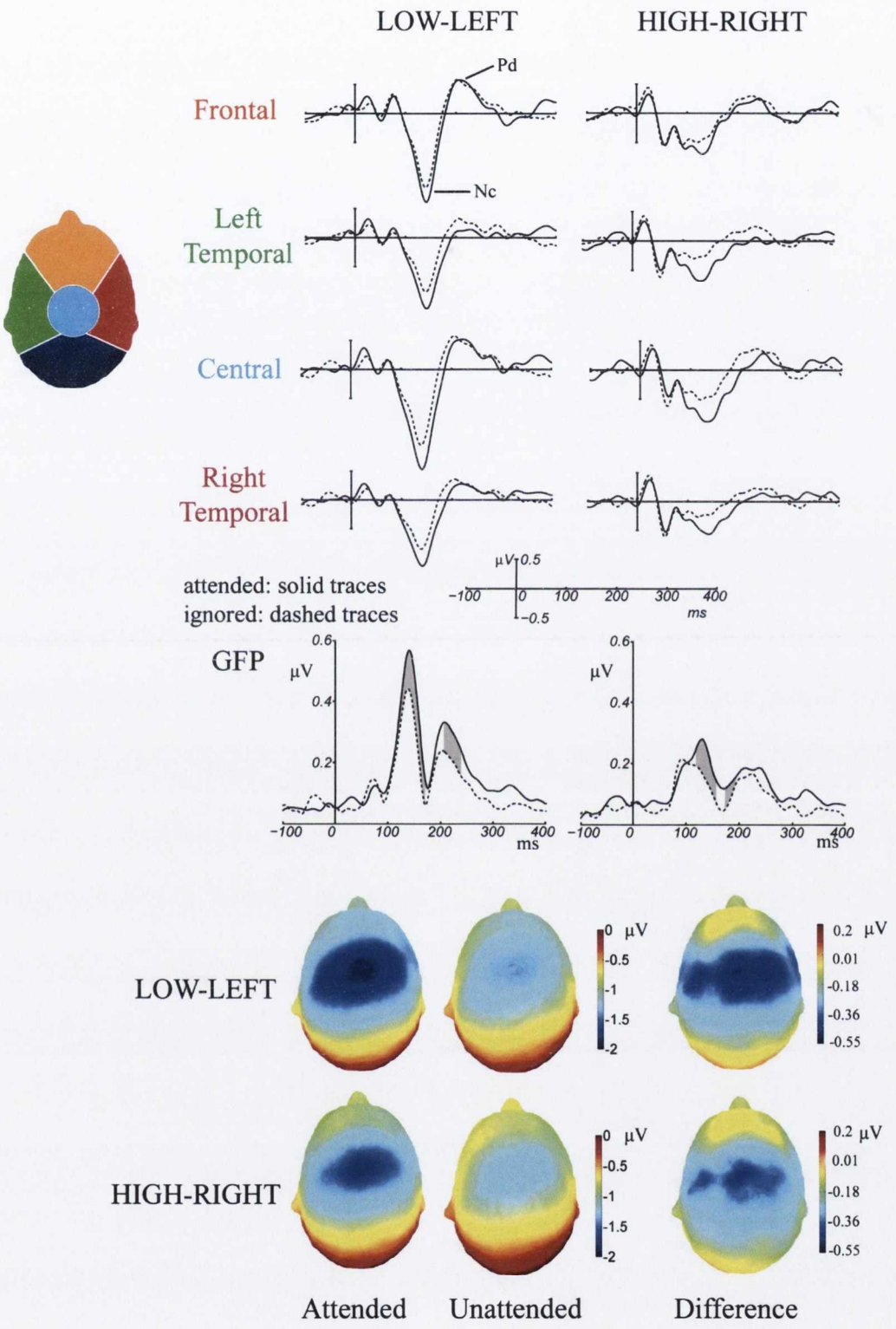

Figure 4.2. Responses for indicated regions and GFPs for the stimuli presented at the left and right ears when attended and unattended (upper panels). Time intervals of interest identified by running t-tests are shaded gray in the GFPs. Topographic maps of the Nc component when attended and unattended as well as difference topographies are shown in lower panel.

Firstly and most importantly there was a main affect of attention ($F(1,16) = 27.75$, $p < 0.001$) indicating that components of the responses to the attended stream were enhanced. There was also a significant effect of stimulus ($F(1,16) = 54.004$, $p < 0.001$). This is due to the fact that the later cortical responses (i.e. Nc and Pd) to the LOW-LEFT stream are greater in amplitude than responses to the HIGH-RIGHT steam. This is likely due to the logarithmic nature of frequency representation in auditory cortex i.e. the higher the frequency the smaller the amount of cortex devoted to it (Romani et al. 1982). Using a wider carrier stimulus bandwidth for the higher frequency stream may result in more similar HIGH and LOW responses. There was no stimulus x attention interaction ($p>0.05$) suggesting that both the HIGH-RIGHT and LOW-LEFT streams were similarly affected by attention. A significant attention x region x component interaction ($F(1.98,31.71) = 7.636$, $p=0.002$) was found. To further interrogate the components driving this interaction we performed separate 2-way 2 x 4 repeated measures ANOVAs on Nc and Pd with factors of attention (attended vs. unattended) and region (left, right, frontal and central). In the case of Nc we found a significant effect of attention ($F(1,16) = 20.28$, $p< 0.001$) as well as a significant attention x region interaction ($F(3,48) = 3.44$, $p = 0.024$). In the case of Pd the effect of attention was not significant ($p>0.05$) although there was an attention x region interaction ($F(3,48) = 7.41$, $p< 0.001$). In order to ascertain the regions driving these interactions post-hoc t-test were carried out and identified the Nc effect to be driven by attention in left ($t(16) = 5.21$, $p < 0.001$), right ($t(16) = 4.1$, $p = 0.001$), central ($t(16) = 4.6$, $p< 0.001$) and frontal ($t(16) = 2.4$, $p = 0.027$) regions. The Pd interaction was driven by an effect in the frontal region ($t(16) = 2.4$, $p = 0.027$). Employing bonferonni correction, however, resulted in a significant Nc

effect only in left, right and central regions but no Pd effect. This suggests that the Pd effect indicated by the running t-test performed on the GFP is marginal whereas the Nc effect is robust. This is further indicated by an attention x component interaction in the initial 4-way ANOVA that approached significance ($F(1,16) = 3.138$, $p = 0.096$).

Just because both left and right regions are similarly affected by attention does not rule out the possibility that responses may be biased to one hemisphere over the other. Since it has been suggested that spatial processing may by lateralized (Zatorre and Penhune. 2001, Spierer et al. 2009) we sought investigate whether this was the case here. The initial 4-way ANOVA resulted in a stimulus x region interaction ($F(3,48) = 10.02$, $p<0.001$). This allowed us to perform post-hoc t-tests on the LEFT-LOW and HIGH-RIGHT responses to inspect the regional differences driving this interaction. We found no difference between the right and left regions for either stimulus type ($p>0.05$ for both). This tells us that there was no hemispheric bias due to the spatial nature of the stimuli.

The possibility that increased activity in the Nc component window may be due to an engagement of additional non-obligatory generators and not increased activity of the sensory activity was investigated using the TANOVA method (Murray et al. 2008). Topographies in attended and unattended conditions for the LOW-LEFT response were not found to be statistically different (LOW-LEFT attended vs. unattended: $p = 0.5855$). This suggested that the same generators were involved in both attended and unattended conditions for the responses to the LOW-LEFT

stimulus. In the case of the HIGH-RIGHT responses, however, the TANOVA did suggest that the topographies were statistically different between conditions ($p = 0.0147$). That said, the TANOVAs did not indicate statistically dissimilar topographies between the LOW-LEFT unattended and HIGH-RIGHT unattended responses ($p = 0.3365$) or between the LOW-LEFT attended and HIGH-RIGHT attended responses ($p = 0.0874$) suggesting that regardless of stimulus similar generators were engaged.

To further investigate the location and strength of the Nc generators and the possible attention affects a dipole analysis was carried out on the Nc component. Starting with the Nc component of the unattended LOW-LEFT responses we attempted to fit two symmetrical regional sources (a regional source in BESA consists of three orthogonal dipoles at the same location). Fitting the sources to a window encompassing the whole Nc component (111 – 154ms: identified from the GFPs) placed the sources at Talairach coordinates $x = -32.6$, $y = -16.8$, $z = 13.6$. This configuration accounted for 98.88% of the variance in the fitted window and the sources are located within 1cm of Heschl's Gyrus (HG). In fact fixing the source locations to the centre of the auditory core (at talairach $x = -46$, $y = -24$, $z = 12$) only slightly reduced the variance explained to 98.12%. Applying this same model to the attended LOW-LEFT responses accounted for 96.76% of the variance. Thus bilateral sources in AC provided an excellent model of the Nc component in both the attended and unattended conditions. This coupled with the insignificant dissimilarity in topographies shown by the TANOVA indicates that the same sources were involved when the stimulus was attended and unattended. Following the same procedure for

the HIGH-RIGHT responses resulted in an initial localisation of the unattended Nc to talairach coordinates x = -37.7, y = -24.9, z = 18.5 with 97.73% of the variance explained. Again this is within 1 cm of HG and fixing the sources to the centre of the auditory core as before only slightly reduced the variance explained to 97.24%. Applying this model to the attended HIGH-RIGHT responses accounted for 93.36% of the variance. Although the model accounted for slightly lower variance than in the case of the LOW-LEFT stimuli bilateral sources in AC again provided an excellent model of the Nc component. Employing the same model for the LOW-LEFT and HIGH-RIGHT responses is backed up by the TANOVA results which suggest that both the LOW-LEFT and HIGH-RIGHT stimuli employ similar generators when unattended as well as when attended. Indeed, it is possible that the lower signal power of the HIGH-RIGHT responses (indicated by the main effect of stimulus mentioned above) may have played a part in TANOVA results which suggested topographical differences between HIGH-RIGHT conditions. These noisier responses may also account for the slightly lower explained variance in the BESA model. Thus the fact that the higher SNR responses (i.e. the responses to the LOW-LEFT stimulus) did not result in statistically significant topographical differences combined with the fact that the Nc component in all stimulus conditions is well explained by sources located in AC suggests that the Nc modulation results from an enhancement of the obligatory sensory activity in AC and not the engagement of supplementary non-obligatory activity.

4.3 Discussion

In this study we used continuous and simultaneous competing stimuli to assess endogenous auditory attention in a cocktail-party like environment. The use of continuous stimuli has eliminated possible confounding effects associated with exogenous attention and discrete stimulation. We found a strong attentional effect of the Nc component in left and right hemispheres as well as central areas. Since the AESPA response primarily represents obligatory sensory processing and since Nc generators have been localised to AC we have shown that sensory processing in AC is modulated by endogenous top-down attention.

Our results are at odds with Näätänen's attentional trace hypothesis (Näätänen 1982), which suggests that obligatory sensory components are not affected by attention. He proposes that attention acts by way of an endogenous processing negativity (PN) which is due to a comparison process between the neural representation of the stimulus and the relevant attentional trace. This PN overlaps and is superimposed on true sensory processes giving the impression of modulation of sensory activity in many cases. Näätänen et al. (1992) do concede that in some instances simultaneous effects on sensory activity cannot be ruled out entirely due to the inability of current methods to distinguish between obligatory sensory activity and overlapping voluntary activity. They remain sceptical, however, of the involvement of sensory processes in attention effects (Näätänen et al. 1992). Although we have isolated an endogenous attention effect on obligatory sensory processes, due to the nature of AESPA responses we are precluded from investigating the undeniably important voluntary components, such as the PN, which

are not well synchronised to stimulus fluctuations. Our results do agree with Hillyard et al. 1973, however, who suggests that sensory processes are affected by attention. The results also agree with the findings of Woldorff and colleagues (Woldorff et al. 1993) who found attentional modulation of activity localised to AC in the ranges 20-50ms and 80-130ms and argue that these effects are due to attentional modulation of sensory processes.

Recently evidence has been emerging that the majority of sound feature processing is achieved subcortically and that AC represents sounds in terms of auditory objects (Nelken 2004). Furthermore it has been suggested that AC is involved in sensory memory (Näätänen and Winkler 1999, Näätänen et al. 2001, Ulanovsky et al. 2003). A suggested mechanism for sensory memory is stimulus-specific adaptation (SSA; Ulanovsky et al. 2003), which has been posited as a possible neuronal correlate both of the decreased N1 to repeated stimuli and of the mismatch negativity (MMN). SSA is the process by which neurons decrease their responses to sequences of identical stimuli, i.e., the activity of neurons is affected by stimulus history. Furthermore, activity of sustained responses has been shown to be significantly affected by SSA (Ulanovsky et al. 2003) and thus, due to the continuous nature of the AESPA stimulus, it is likely that our responses primarily represent activity of the subset of neurons that are least susceptible to SSA. That is to say those cells most susceptible to SSA would contribute minimal activity in response to a continuous stimulus whereas those least susceptible to SSA, i.e., the neurons least involved in sensory memory and most involved in feature processing, would contribute most to the response. This would suggest that the attention effect found

here is due to enhanced feature processing and not related to sensory memory representation. Also the fact that SSA has been shown to be more prominent for sustained responses than for onset responses of neurons in primary AC (Ulanovsky et al. 2003) may account for the smaller size of the Nc components relative to the AEP-N1 component (Lalor et al., 2009).

Although we only found attentional effects around 136ms it is possible that earlier AESPA components, especially the Nb and Pc components, may be affected by attention but these components are somewhat ill-defined. This may be due to the frequency content of the carrier stimuli: previously these components were seen to be ill-defined when a 1kHz tone was used as the carrier stimulus as opposed to broadband noise (Lalor et al. 2009). A wider carrier stimulus bandwidth may allow for a more detailed investigation of these components. The lack of an early effect on well defined components such as Pa maybe due to the nature of the task employed here. It has been shown that the locus of attention is flexible and varies depending on the processing stage most heavily loaded by the task in question (Vogel et al. 2005, Kelly et al. 2008). We employed a difficult event discrimination task (i.e. discrimination between targets and distracters) which is likely to load later processes as opposed to simpler frequency deviant identification tasks (e.g., Woldorff et al.,1993).

Recent efforts at assessing attention to simultaneously presented stimuli, have looked at transient responses to the onset of amplitude modulated sounds (Ross et al. 2010). Attentional modulation was found as early as 143 ms and was localised to

AC. However, whether this is due to increased sensory activity or a separate endogenous effect is not clear. Furthermore whether this effect has been diluted by unaffected or inhibited responses to the simultaneous unattended stimulus is unclear. Thus not only has employing the AESPA method allowed for the investigation of endogenous attention effects on obligatory sensory processes which are unaffected by simultaneous voluntary components, it has also allowed for the isolation of truly separate responses to concurrent stimuli. That said, however, we can only assert that sensory processes are affected by endogenous attention and cannot investigate the certain effects of endogenous attention on non-sensory cognitive processes.

A recent attempt was made to examine attention in a competing stimulus environment using an intricate paradigm (Bidet-Caulet et al. 2007). The authors of that study assessed both transient responses and SSRs from depth electrode recordings in patients with epilepsy. While the results from this study were encouraging, they exhibited a number of inconsistencies: responses to certain stimuli were enhanced when attended in some conditions and reduced when attended in other conditions, and no attention effects were found in a significant number of subjects. Furthermore, the use of onsets and SSRs precluded the assessment of the timing of attentional enhancement during sustained attention. This study also posited a dominant role for left hemisphere in attentional selection with right hemisphere being inhibited as a function of attentional load. This is at odds with our results which show voluntary attentional enhancement effects over both left and right hemispheres and no lateralisation of responses. However, further inspection of their results reveals attentional enhancement in right hemisphere for a number of

attentional conditions, suggesting that it may be premature to make strong conclusions about the hemispheric specialization of auditory attention in sensory areas.

There is much debate relating to lateralisation of attentional and spatial processing with some studies favouring a right hemisphere dominance (Tanaka et al. 1999, Spierer et al. 2009), others a left lateralisation (Pinek at al 1989), others a bias to the hemisphere contralateral to the stimulus (Zatorre et al 1995), others whole field neglect following unilateral lesions i.e. no difference between left and right lesion subjects (Zatorre and Penhune. 2001). Indeed, in their study that encountered mixed lateralisation results when studying patients with unilateral temporal lobe excisions either encroaching on or sparing Heschl's gyrus Zatorre and Penhune summed up the observed variation succinctly: "the existence of individual differences likely illustrates differential patterns of functional lateralisation" (Zatorre and Penhune. 2001). Thus the possibility of a predominant lateralisation of spatial processing in auditory cortex is still an open question. Our results suggest that intensity processing is not affected by the spatial properties of a stimulus. This does not rule out the possibility of the involvement of spatially sensitive centres, not represented in the intensity processing characterised by the AESPA responses, which may be lateralised: Frequency-specific inter-aural time differences (ITD), inter-aural level differences (ILD) as well as monaural amplitude spectrum cues are thought to integrate in a non-linear fashion to create a map of space represented by location specific neurons (Cohen and Knudsen. 1999). Thus since the activity of these location specific neurons is not linearly related to the intensity of the stimulus then it

is unlikely to be accounted for in the AESPA response. This would further back up our assertion that the AESPA response is due to the activity of intensity processing neurons only (Lalor et al. 2009). In light of this the current results, although enlightening as to the effect of endogenous attention on intensity processing, are not directly comparable with previous studies on the lateralisation of spatial processing. Furthermore there has been much interest in the purported separation of 'what' and 'where' streams in the auditory system (Rauschecker and Tian. 2000, Tian et al. 2001, Ahveninen et al. 2006). Given that amplitude modulation is the driving feature of the stimulus employed here (i.e. a likely 'what' feature) it may be that the 'what' stream is preferentially driven by the current implementation of the AESPA method and thus activity of the location relevant 'where' stream is not represented. Implementation of a paradigm whereby the intensity of a stimulus is kept constant but the location of the stimulus is modulated would shed further light on this possibility.

In this chapter we have used the AESPA method to investigate endogenous auditory attention in a novel paradigm employing two simultaneously presented, continuous stimuli. Having identified that obligatory sensory activity is affected by endogenous attention using this paradigm and given that one of the main goals of this research is the investigation of the auditory system in ecologically valid paradigms the next chapter seeks to probe the dynamics of auditory attention in a natural cocktail party scenario. Lalor and Foxe, 2010 have previously shown that it is possible to extract responses to natural speech using the AESPA method and in

Chapter 5 I interrogate the dynamics of attention in a cocktail party scenario where two natural speech streams are presented simultaneously.

"Cocktail Party: A gathering held to enable forty people to talk about themselves at the same time. The man who remains after all the liquor is gone is the host" Fred Allen (1894 - 1956) American Comedian/Humorist.

5

AESPA and the Cocktail Party Problem

The cocktail party problem refers to the scenario when subjects are asked to follow one of two or more competing speech streams resulting in a highly accurate recall rate for the information in the followed stream along with a poor (close to zero) recall of information in the ignored stream. Not all information in the unattended stream is discarded however; the ignored stream can capture attentional focus when a subject's own name or words semantically related to the attended stream are presented in the ignored stream (Moray, 1959, Lewis, 1970). Since these pioneering studies many have sought to elucidate the physiology and mechanisms of the stimulus selection in a multi-stimulus environment. That said, however, conflicting results concerning attentional modulations in primary auditory cortex

(PAC) during cocktail party scenarios have come from studies using functional magnetic resonance imaging (fMRI) (Hill & Miller 2010) and studies using dipole modelling of EEG activity in particular frequency bands (Kerlin et al., 2010). Furthermore, neither of these methods permits the precise temporal localization of these attention effects. Efforts to circumvent this shortcoming and to exploit the temporal resolution of EEG have utilized the ERP method by averaging responses to either probe stimuli superimposed on speech (Hink & Hillyard, 1976, Woods et al., 1984, Coch et al., 2005, Nager et al 2008) or to specific features of the speech stream itself (Teder et al., 1993). Interpretation of these results is complicated, however, by possible exogenously alerting effects relating to the probe stimuli. Furthermore, the averaging procedure assumes probe stimuli and speech features to be delta functions, which is rarely the case and which can lead to temporal smearing of the response (Lalor et al., 2009 & Chapter 3), and, accordingly, the effect. This difficulty is compounded by the extended temporal nature of the speech itself, which can result in the enhancement of auditory responses both before and after the time-point used to select epochs for averaging. The aim of the current work is to utilize the AESPA method outlined in Chapters 3 and 4 (see also Lalor et al., 2009, Lalor & Foxe, 2010, Power et al., 2011) to extract temporally detailed responses to the envelopes of each of two concurrent speech streams, allowing us to precisely determine the spatial and temporal locus of naturally deployed attention to speech.

5.1 Materials and Methods

Subjects and Data Acquisition

40 subjects took part (mean age: 27.3 ± 3.2 yrs; 32 Male; 7 left handed). 20 subjects (age: 27.5 ± 3.44 yrs; 18 Male; 2 left handed) attended to the left ear and 20 subjects attended to the right ear (age: 27.24 ± 3.03 yrs; 14 Male; 5 left handed). The experiment was undertaken in accordance with the Declaration of Helsinki. The Ethics Committee of School of Psychology at Trinity College Dublin approved the experimental procedures and each subject provided written informed consent. Subjects reported no history of hearing impairment or neurological disorder. EEG data were recorded for 34 of the subjects using 130 electrode positions (17 of these subjects attended to the left and the remaining 17 to the right). Data for the remaining 6 participants were collected using 162 electrode positions (3 of these subjects attended to the left and the remaining 3 to the right). The data were filtered over the range 0 – 134 Hz and digitized at the rate of 512 Hz using a BioSemi Active Two system. EEG data were then digitally filtered off-line with a high-pass filter, where the passband was above 2 Hz and with a -60 dB response at 1 Hz and a low-pass filter with passband below 35Hz and a -50 dB response at 45 Hz. The data at each channel were re-referenced to the average of the responses at the left and right mastoids. After extracting responses from the data acquired using the 162 electrode positions they were mapped down to the 130 electrode positions used for all other subjects resulting in a coherent dataset with identical channel configuration. This mapping consisted of using splining information for the 162 channel montage (obtained from the splining algorithm in EEGLAB; http://sccn.ucsd.edu/eeglab/) to map the data to "splined-space". The data were then mapped back to electrode space

using the splining information for the 130 electrode configuration. Over all this resulted in a down-sampling in electrode space from 162 electrode to 130 electrodes.

Stimuli, Experimental Procedure and Tasks

Two classic works of fiction are presented one to the right ear and the other to the left ear. These were Twenty Thousand Leagues Under the Sea (**TLS**) and Journey to the Centre of the Earth (**JCE**). These works were segmented into 30 passages each approximately 60s in length. Further to this, and in order to minimise the possibility of the unattended stream capturing the subjects' attention during silent periods in the attended stream, silent gaps exceeding 0.5 s were truncated to 0.5 s in duration. No subject reported abnormal speech patterns as a result of this. Subjects were instructed to attend to either the story in left or right ear for all 30 passages (i.e. ~1800s of data per subject). After each passage subjects answered multiple choice questions on *both* stories (i.e. on the attended story and the unattended story). There were between 4 and 6 questions on each story and subjects were given four possible answers for each question. Each passage took up from where the previous passage left off in the story and stimulus amplitudes in each stream within each run were normalised to have the same root mean squared (RMS) intensity. We used a between subject design as we wanted subject to follow the stories as naturally as possible (i.e. only hear each passage once) and thus repeating any individual trial and switching attended stream would require presenting a passage already presented and would likely result in biased behavioural results (as the questions for the unattended stream in this case would already have been answered in a previous condition where it was attended).

Signal Processing and the AESPA method

Once again the AESPA responses are obtained by performing a linear least squares fit of the response model

$$y(t) = w(\tau) * x(t) + noise$$

Where y(t) is the measured EEG response, x(t) is the amplitude envelope of the stimulus, the symbol * indicates convolution, w(τ) is the impulse-response function to the amplitude of the stimulus, and the noise is assumed to be Gaussian (Lalor et al., 2009).

In order to estimate the AESPA it was necessary to determine the amplitude envelope of the speech signals. Given that the sampling rate of the audio signal was 44,100 Hz and that of the EEG was 512 Hz, the root mean square (RMS) of an average of 83.133 audio signal samples as calculated for each sample of EEG. This represents a simple resampling of the audio signal. Further to this since envelope frequencies between 2 Hz and 16 Hz contribute most to speech intelligibility (Drullman et al., 1994ab, van der Horst et al., 1999) the envelope was then low pass filtered with a corner frequency of 20Hz. Since intensity processing varies by the log of stimulus intensity each envelope was transformed by taking the log to the base ten of the intensity and normalizing between zero and the maximum before mapping (i.e. the relative intensities of the maxima were unaffected by the log mapping).

Since our model assumes a linear relationship between the EEG and stimulus envelope the EEG was also low pass filtered with corner frequency of 20 Hz. In a linear system the output can only have energy at frequencies which are in the input. Thus low pass filtering at 20Hz was done in order to minimise the contribution of higher frequency activity, whether unrelated to the envelope or non-linearly related to the envelope (i.e. exclude responses at frequencies not contained in the envelope input), which our linear model and estimation method would consider as noise. The contribution of these higher order non-linear responses is of obvious importance to speech processing and by extending our analysis, as has been done previously in the (Lalor et al 2008), these non-linear responses can be investigated. This is, however, beyond scope of this paper.

Exploratory Analysis

Initially we used global field power (GFP) to identify time periods/components of interest. In order to increase the signal-to-noise ratio (SNR) of the responses and also since GFPs take no account of topographical distribution of responses and only represent a reference-free quantification of response power over the whole scalp we collapsed across stimulus for this initial exploratory stage of analysis i.e. any topographical differences based on ear attended or other left/right hemisphere biases in processing is not accounted for in the GFP. This resulted in ATTENDED and UNATTENDED responses. Differences between these two GFPs were assessed using running unpaired t-tests. A period was considered of interest if the GFPs were significantly different ($p<0.05$) for 11 continuous samples.

Statistical Parametric Mapping - SPM

The periods of interest identified by the GFPs t-tests (i.e. 140-230ms) were then subjected to statistical parametric mapping (SPM; Kiebel and Friston, 2004a, 2004b). When using SPM for EEG the EEG signals are mapped from the electrode domain to a flattened 2 dimensional scalp space, consisting of a 64 x 64 pixel grid, by way of a linear interpolation. This mapping is done at every time point of the ERP and results in a three dimensional representation of the data (2 dimensions in space and one in time). This representation consists of what is called a voxel-based representation of the EEG but the representations are still of the scalp activity and no underlying source configuration has been derived. Thus a voxel in this case refers to a position on the scalp (in 2-D space) at a particular time. These spatio-temporal representations are used to test for temporally and regionally specific effects. Here we employed separate factorial ANOVAs for each stimulus type. Each ANOVA had 2 levels (attended vs unattended). We further employed conjunction analysis to establish what effects are present in responses to both stimuli. The statistical threshold for the SPM analysis was set to the stringent level of $p_{uncorrected} < 0.001$ and 500 contiguous voxels. Peak effects are also reported and these have been subject to family-wise error (FWE) correction for multiple comparisons.

5.2 Results

Questions on **TLS** were answered correctly 82.26% (SD ± 5.25%) of the time when attended and 26.18% (SD ± 5.83%) of the time when unattended, which was not statistically different from chance performance ($p>0.05$), in line with previous reports (Cherry, 1953). Questions on **JCE** were answered correctly 82.04% (SD ±

5.9%) of the time when attended and 30.27 (SD ± 5.11%) of the time when unattended. Although extremely low, the performance on **JCE** when unattended was greater than chance ($p<0.05$). This may have been due to the ability of subjects to occasionally infer the most likely answer from the surrounding questions and or to occasional attentional capture by the non-cued story. A 2x2 ANOVA with levels of stimulus (**TLS** vs. **JCE**) and attention (attended vs. unattended) returned a significant main effect of attention ($p<0.001$), no main effect for stimulus ($p = 0.154$) and no stimulus x attention interaction ($p = 0.113$).

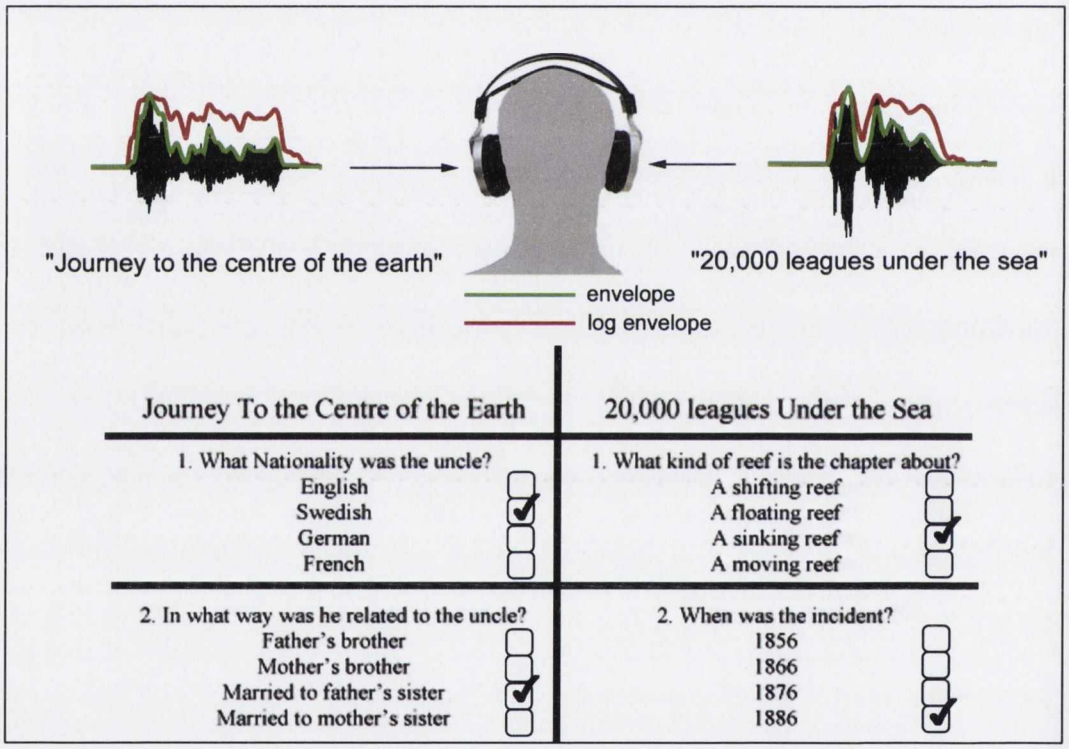

Figure 5.1 The experimental procedure. In each trial, subjects listened to ~60s of two stories presented concurrently and dichotically. After each 60s trial, subjects were presented with between 4 and 6 multiple choice questions on each story with each question having four possible answers. Subjects were asked to preferentially attend to one story, but to attempt to answer all questions on both stories. Answers were indicated on screen using a mouse click. Each 60s trial continued from the point in the stories where the previous trial ended, with no repetition of stimuli. EEG data were not collected while subjects answered the questions. Responses were extracted using the log envelope of the speech stimuli.

EEG data were recorded from 135 scalp electrodes while subjects listened to the stories. Temporally detailed responses to the attended and unattended speech

streams were obtained for the data from each electrode using the AESPA method and the speech envelope as described above in section 5.1 under the heading *Stimuli, Experimantal Procedures and Tasks* (see also Lalor et al., 2009, Lalor et al., 2010, Power et al., 2011) (Fig. 5.2a and Fig. 5.2b). In an exploratory analysis of these responses, running t-tests on the global field power (GFP) (Lehmann & Skrandies, 1980) identified the negative component around 170 ms (known as the Nc) and the positive component around 220 ms (known as the Pd) as possibly being affected by attention (Fig. 5.2c, also see supplemental methods). Accordingly, these components were subjected to statistical parametrical mapping for ERPs as described above. We applied a temporal mask (140 to 230 ms) in order to analyse the EEG responses around the Nc and Pd components. Separate family-wise error (FWE) corrected SPM analyses were carried out comparing the effects of attention on the two different stories. A significant contralateral attention effect over the left hemisphere was seen during the Pd timeframe for TLS (Fig. 5.2d). The peak effect occurred at 207 ms ($p_{FWE} = 0.017$). A small region in the Nc window survived the stringent criterion of $p_{uncorrected} = 0.001$ and 500 contiguous voxels. This effect did not survive FWE correction ($p_{FWE} = 0.079$ at 143ms) when using a temporal mask containing both Nc and Pd, however it did survive FWE correction when the mask was narrowed around the Nc component ($p_{FWE} = 0.014$ at 141ms; see Fig 5.3). With regard to **JCE**, the Pd component was once again strongly affected by attention over the contralateral hemisphere (Fig. 5.2e; $p_{FWE} = 0.005$; peak at 213ms). However, this response was even more strongly affected over the ipsilateral left hemisphere ($p_{FWE} = 0.001$; peak at 213ms). Comparing the average GFP in the timeframe of the Pd to the average baseline GFP in the window -100 to 0ms with paired t-tests did not find any

statistical differences between Pd activity when unattended and baseline noise power (p_{TLS} = 0.1881 and p_{JCE} = 0.3658). Since the Pd component has previously been shown to be a stable and robust component in passive listening[9,10], this suggests that processing of irrelevant speech may suppressed at around 210 ms. It should be noted that, despite the results for the **TLS** responses, the Nc was not found to be significantly affected by attention for the **JCE** responses.

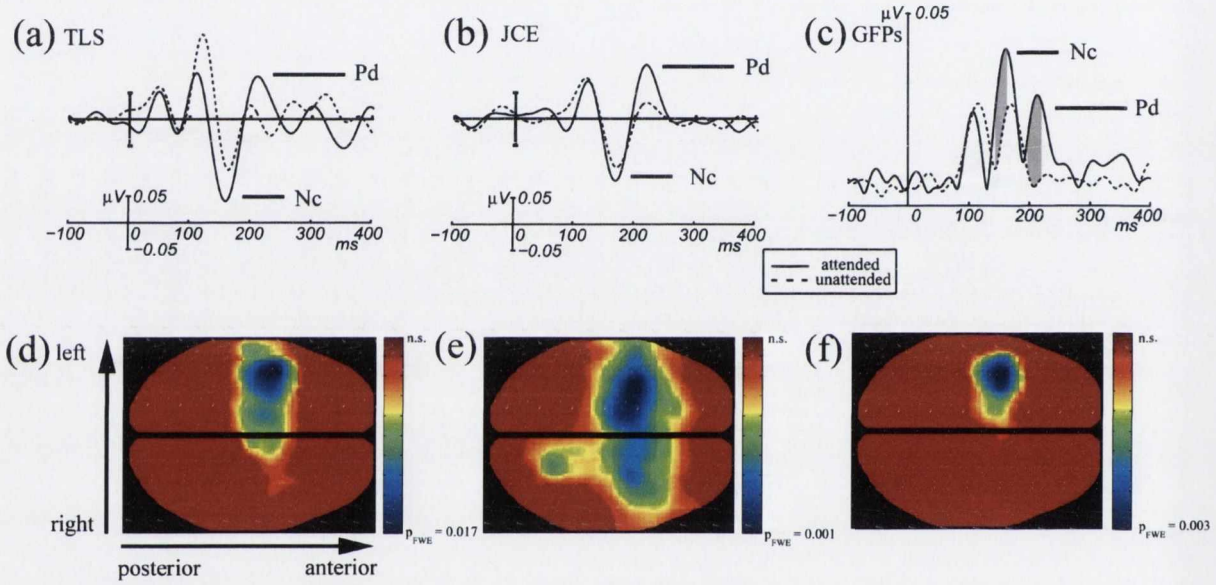

Figure 5.2. The effect of attention on the scalp recordings. **(a)** AESPA responses over frontal midline (Fz) to TLS when attended and unattended showing enhancement of the Nc and Pd components. **(b)** AESPA responses over frontal midline (Fz) to JCE when attended and unattended showing enhancement of the Pd component. **(c)** Average global field power across both stories when attended and unattended. The gray shading indicates temporal periods where the GFP to the attended story exceeded that of the unattended story according to a one-tailed pointwise t-test. **(d)** and **(e)** show the spatial distribution of the Pd effect identified by the SPM analysis for TLS (right ear) and JCE (left ear) respectively. **(f)** shows the left biased spatial distribution of effects common to both TLS and JCE. Note: (d), (e) and (f) represent scalp effects that have been projected onto a brain template and do not show inverse source solutions.

Source Analysis

BESA source analysis was then carried out on the responses in order to localise the sources of the components affected by attention. Firstly we investigated the simplest response; the unattended responses averaged across stimulus. Using two symmetrical regional sources we localised activity in the Nc window (140-185ms) to

Talairach co-ordinates 44.2, -17.9, 9.8. This is consistent with activity in Broadmann areas 41 and coincides with primary auditory cortex. This accounted for 96.8% of the variance in this window. Given the suppressed nature of the Pd component in the unattended responses we then looked at the average of the attended responses across stimulus to localise Pd activity. The bilateral regional sources in PAC accounted for 98.34% of the variance in the Nc window for the attended response. Extending the localisation window to include the Pd component (i.e. 140-230ms) resulted in the bilateral sources in PAC accounting for 97.36% of the variance in the combined Nc/Pd window. Fixing the bilateral regional sources in their current position and localising an additional source only increased the variance explained by 1.47% and was localised to the limbic lobe. This was considered unlikely to be truly representative of the responses, which have been previously linked to low-level sensory activity (Lalor et al., 2009, Power et al., 2010). Further to this, since the bilateral regional sources in PAC provide an excellent model of the Nc and Pd activity the additional source was discarded. This grand average response across stimulus and condition accounted for 97.83% of the variance in the combined Nc/Pd window (i.e. 140-230ms).

In line with previous AESPA source analysis reports (Lalor et al.,2009, Power et al.,2011) the dipole model localised Nc and Pd to Broadmann's area 41 which coincides with Primary Auditory Cortex (PAC; see supplementary methods). The model accounts for 97.83% of the variance in the combined window of 140ms to 230ms around Nc and Pd for the grand average. This simple PAC source model accounts for 96.8% of the variance for the JCE responses when attended and 94.74%

of the variance when unattended. 96.12% of the variance is explained for the TLS when attended and 91.32% when unattended. Thus bilateral sources in PAC provide an excellent model of Nc and Pd activity for all responses. The model also accords well with other reports of attention effects in a cocktail party scenario (Kerlin et al., 2010).

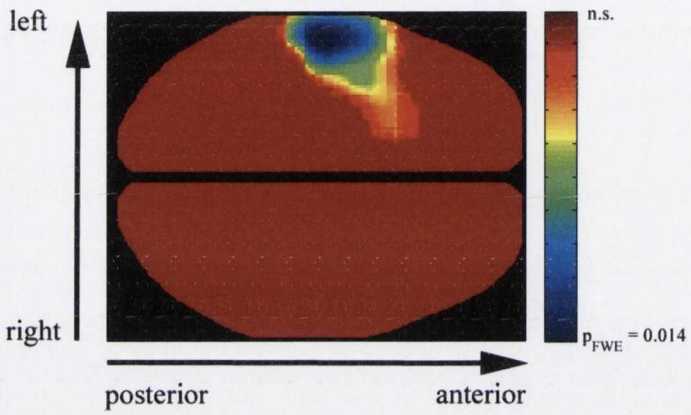

Figure 5.3. The effect of attention on the scalp recordings for the Nc window for **TLS**. The cluster shown survives $p<0.001$ and 500 contiguous voxels with a peak effect significance of $p_{FWE} = 0.014$.

5.3 Discussion

We have shown robust attentional effects on the processing of spatially separate competing natural speech stimuli. These effects occur in PAC at around 200 ms. The timing of these effects is somewhat later than effects found in studies employing probe stimuli superimposed on the speech streams (Hink & Hillyard, 1976, Woods et al., 1984, Coch et al., 2005, Nager et al 2008). These studies suggest attentional effects by 100 ms. Similarly, a study that obtained responses by averaging to the onset of a particular speech feature found an effect as early as 40ms, peaking at 80-110 ms and lasting until ~400 ms (Teder et al., 1993). Possible reasons for this discrepancy include exogenous attention effects to the superimposed probes (Escera

et al., 2000), difficulty determining the precise onset of the epoch for averaging (particularly when the onset is a feature of the speech itself (Teder et al., 1993)), and imprecision in resolving the temporal locus of attention because of possible enhancement of ongoing speech prior to and following the probe stimulus. Using the AESPA method, we have found a temporally precise effect during the timeframe of the Pd component. The fact that this component was almost completely absent in the case of responses to the unattended speech streams (Fig. 5.2c) suggests that unimportant information is heavily suppressed at this stage of processing.

This almost complete suppression of the Pd component contrasts with attentional effects on the AESPA when obtained to non-speech stimuli, where the AESPA in response to ignored stimuli displayed a robust Pd component (Power et al., 2011). This may be due to the fact that speech discrimination is a highly familiar task and that, as such, suppression of irrelevant speech may be more easily accomplished than when using artificial stimuli. Another possible explanation for this result involves the high demands placed on working memory as a result of the task employed in this study. In order for subjects to adequately perform the task they are required to hold the content of ~60s long story segments in memory until the questions are presented at the end of the segment. It has been shown that neural suppression of irrelevant information underlies optimal working memory performance (Zanto & Gazzaley, 2009). No such working memory demands were involved when assessing attentional effects on the AESPA using artificial stimuli (Power et al., 2011).

We noticed a strong left hemisphere bias in our attention results (Fig. 5.2f). Because the AESPA is obtained to the (log) amplitude envelope of the speech signal, i.e., a low-level feature of the signal, we suggest that this bias is likely due to the pre-linguistic advantage of left primary auditory cortex to process temporal information, which has been linked to the left hemisphere bias for language in higher areas (Zatorre & Belin, 2001). However, when employing non-speech stimuli with similar envelope statistics to derive the AESPA, no significant left bias was observed (although there was a trend; Power et al., 2011). Because the AESPA is likely to be insensitive to contributions from higher order language areas as a result of the assumption of linearity between amplitude envelope and EEG, we do not directly attribute these differing levels of left hemisphere bias to greater levels of processing activity within higher order language areas in the case of speech versus non-speech stimuli. Rather, we speculate that, because of the highly familiar nature of the speech task employed here, the bias stems from a more efficient deployment of attentional resources than when using less realistic stimuli. This deployment of attention likely involves top-down control from higher order areas, including language areas, in the left hemisphere, however, and thus the effect that we see in AC likely reflects the effect of this contribution on early auditory processing.

The results presented in this chapter represent the achievement of one of the main goals of the research outlined in this thesis i.e. the investigation of endogenous attention in a natural and ecologically valid scenario. Of course further work is necessary to elucidate the contributions of working memory or otherwise to the sensory effects found here but the fact that sensory information in PAC is effected

while a subject is engaged in a cocktail party scenario is a significant finding especially given the temporal detail provided by the AESPA method. To my knowledge this is the first study to obtain temporally detailed responses to two simultaneously presented speech stimuli in a cocktail party scenario in which the extended temporal nature of the stimuli is taken into account. Furthermore the speech is uncontaminated by potentially confounding probe stimuli. This represents a significant step towards investigation of the auditory system in ecologically valid scenarios.

6

Future work and Summary.

The flexibility of AESPA has proven of great use in addressing important questions of auditory physiology. This useful flexibility comes from the ability to tailor the stimuli to have desired statistics depending on which aspect of auditory processing is under investigation. The ability of the method to extract responses to continuous and complex stimuli (including natural speech) is of huge relevance to advancing our knowledge of auditory function in natural environments. Further to this the ability to achieve experimental paradigms where true simultaneity of stimuli is possible without the loss of temporally detailed responses highlights AESPA's significance in the study of the dynamics of auditory attention, as has been expanded on in this thesis. The AESPA method is by no means presented as a replacement of

standard methods, or indeed as a be-all and end-all solution to the methodological obstacles to assessing neurological phenomena. It is, however, an undoubtedly useful and flexible tool that can be used to complement our understanding of auditory physiology and a number of further applications are outlined below.

6.1 Future Work

AESPA and non-spatial attention – Endogenous Attention and the purported Auditory "What" and "Where" Processing Streams.

In order to investigate the temporal dynamics of non-spatial attention we employed the same paradigm as in Chapter 4 except that stimuli were presented binaurally. To date no attention effects on the AESPA responses have been identified (see Fig 6.1). Interestingly, however, when presented binaurally (i.e. the stimuli have no spatial element and left and right ears both receive the same stimulation) subjects seemed unable to distinguish between targets and distracters in the attended stream. That said they are clearly able to distinguish between the two stimuli as very few responses to events in the to-be-ignored stimulus are seen. There are two possible explanations for this:

1) The subjects are highly engaged and attending to the desired stream but subjects are just unable to distinguish targets and distracters within the stream. If this is the case we could conclude that non-spatial attention does not significantly effect sensory processing in auditory cortex. Thus

highlighting an important difference between spatial and non-spatial attention.

2) The discrimination task was too hard but the event recognition task was too easy such that subjects gave up on trying to discriminate between targets and distracters in the attended stream but found the identification of an event in the relevant stream trivially easy and thus were not truly engaged in the task. If this is the case the lack of engagement could account for the lack of effect in the AESPA responses.

Further investigation into the development of a suitably engaging non-spatial task is obviously necessary. The preliminary results do, however, raise some interesting questions regarding the nature of spatial attentional effects highlighted in Chapter 4 and non-spatial attention. Further work investigating how endogenous attentional mechanisms might interplay with the purported separation of "what" and "where" processing streams in the auditory system is on-going.

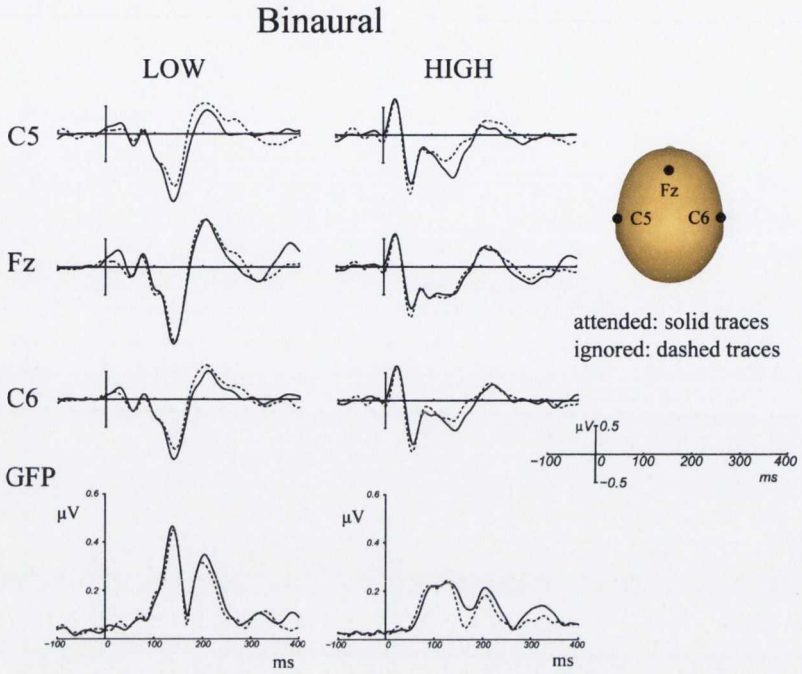

Figure 6.1 AESPA responses and GFPs in a non-spatial attention task.

AESPA and Specific Language Impairment (SLI)/Dyslexia

Dyslexia is traditionally defined as impairment in reading ability in spite of adequate intelligence and tuition. A prominent theory that attempts to account for the deficits associated with dyslexia is that of an impairment in the ability to adequately process brief or rapidly varying sounds. This was first suggested as a possible cause of specific language impairment by Tallal and Piercy in 1973 (Tallal & Piercy, 1973) and was later extended to specific reading impairment (Tallal, 1980).

In an attempt to investigate this theory further McAnally and Stein (1997) showed a significantly reduced ASSR amplitude for dyslexics compared to controls. This deficit was shown for a modulation index of 100% at modulation frequencies

20, 40, 60, 80 Hz. Furthermore Mennel et al. (1999) showed that dyslexics are less sensitive to AM at modulation frequencies between 10 and 160Hz than controls. This was done by establishing the modulation depth threshold for dyslexics and controls at 10, 20, 40, 80, and 160 Hz. They also showed that at these modulation frequencies the amplitude of the ASSR, similarly to McAnally and Stein (1997), is much reduced in dyslexic subjects. Furthermore, Witton et al. (2002) have shown, by varying modulation depth, that ability to detect 2 Hz modulation of a 1 kHz tone is not impaired in dyslexics whereas at 20 Hz modulation performance is degraded significantly compared to controls suggesting that, at lower frequencies dyslexics are as adept as controls and their deficit stems from an inability to process faster modulations adequately. This agrees with the rapid auditory processing (RAP) deficit hypothesis (Tallal, 1980).

AESPA stimuli consist of a broadband noise carrier stimulus which is amplitude modulated by a randomly varying envelope. The statistics of the envelope waveform can be chosen at will. This makes the AESPA stimulus particularly suited to the investigation of the affect of AM on sensory activity. The majority of studies investigating this theory have either been psychophysical or have employed the ASSR, which as outlined previously lacks temporal detail and is limited to measures of responses power and phase relative to the stimulus. The temporal detail afforded by the AESPA method will hopefully shed further light on the origins of the deficits described above. The AESPA stimuli employed in Chapter 3 and 4 of this thesis employed envelopes with uniform energy in the range 0-30Hz. Manipulation of the envelope such that it only has energy either in the range 0-2Hz or 2-30Hz would be useful in further interrogating the nature and neural locus of the deficits in auditory

processing associated with dyslexia. Furthermore, since dyslexia is a language disorder extracting the AESPA response to natural speech stimuli may also be informative.

A competing theory, however, suggests that the core phonological deficit found in dyslexia is not due to impaired RAP but that it is sound envelope and rise time processing that is deficient and that this is at frequencies commensurate with envelope modulations inherent in natural speech, especially frequencies in the delta (0.5-4Hz) and theta (4-8Hz) bands (Goswami, 2010). Justification for this comes from fact that empirical work has provided only modest support for RAP assumptions. Firstly, studies showed that children with dyslexia could perform either as well as controls in RAP tasks, or showed deficits, but deficits that were significant at long as well as rapid temporal rates (e.g., Studdert-Kennedy & Mody, 1995; Nittrouer, 1999; Marshall, Snowling, & Bailey, 2001; Rosen & Manganari, 2001; Waber et al., 2001). Second, it was demonstrated that elongating or stretching the formant transitions in syllables (or stretching the syllables in time i.e. slowing inherent transitions) did not improve syllable perception in dyslexia, which is a necessary corollary of RAP (McAnally et al., 1997). Findings such as these led to a growing view that deficits in rapid temporal processing were not central to the phonological difficulties found in children with dyslexia (e.g., McArthur & Bishop, 2001; Rosen, 2003).

Obviously further research is needed to assess the primary deficit leading to dyslexia. AESPA would a useful tool in investigating the above theories. Given the

possible use of natural speech stimuli the processing of natural rhythms in speech can be assessed. Furthermore given the flexibility of stimulus design using stimuli to specifically probe fast temporal changes would also be useful. Of course a better understanding of the root causes of dyslexia would give important insights into remediation. For example, if the latter theory is true (i.e. deficit in processing low frequency envelope cues) it may be possible to use rhythmic interventions based on music and singing to improve phonological awareness via enhancing sensitivity to language rhythms. Certainly, anecdotally preschool teachers believe that musical experiences (nursery rhymes, songs) improve language development. The proposed theoretical framework might help to explain why. On this hypothesis, the perception and production of structured rhythmic and temporal patterns is a crucial part of language acquisition and phonological representation. That said if the former theory were to be supported it would point towards a different approach to remediation such as training to help discriminate stimuli containing rapid spectro-temporal changes (Tallal, 2004).

The AESPA and Schizophrenia – N1 deficit and P50 Sensory Gating.

The N1 component of the standard AEP is known to be reduced in schizophrenic patients compared to normal control subjects (see Fig. 6.2, from Shelley et al, 1999). The N100 reflects a complex multigenerator process and there is much debate as to whether the reduction in patients is due to all N1 generating cells being affected or just a functional subset (Rosburg et al., 2008). We have previously argued that the N1 and Nc have similar, but unlikely identical, generators due to the

nature of the stimulus and analysis method (Chapter 3, Lalor at al., 2009). Furthermore the flexibility of the AESPA method allows for greater control of the stimuli to target specific cell groups (as was shown for the visual analogue of AESPA, VESPA, Lalor & Foxe, 2009). Thus AESPA's application to the investigation of the N1 deficit is apparent given that it may be helpful in establishing whether the deficit is due to decreased activity of the whole N1 complex or just a functional subset.

Previous research has suggested that both exogenous and endogenous activity is affected in schizophrenia. This has come from studies showing reduced or missing N1 attention effects as well as reduced N1 in studies not requiring allocation of attention. Although N1 deficit cannot be used as a simple test of the presence or absence of schizophrenia studying the N1 can be used to help us to understand abnormalities in basic elemental mechanisms of brain function and structure (Rosburg et al., 2008). Thus employing AESPA stimuli in both passive and attention paradigms would shed further light as to the origin of the N1 deficit. Whether exogenous processing is effected and further to this whether attentional mechanisms that enhance exogenous processing are affected. Also topographical and/or source analysis of responses would further help establish whether general exogenous processing is effected or if a subset of generators are affected.

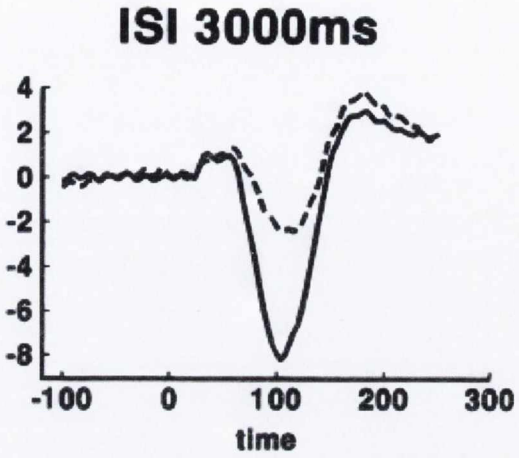

Figure 6.2. N1 waveform for control (solid lines) and schizophrenic (dashed lines) schizophrenic subjects at ISI of 3 seconds at the Fz electrode. Waveforms reflect responses to repetitive standard stimuli. (Figure from Shelley et al, 1999)

Additionally pulse paired sensory gating stimuli have also been used in the investigation of auditory gating of the P50 component of the standard AEP in schizophrenia (Patterson et al., 2008). In this paradigm a series of pairs of short pulse stimuli are presented. The first pulse in a pair is known as the conditioning stimulus and the second pulse is known as the test stimulus. Successive conditioning stimuli are separated by 10s while conditioning and test stimuli within a pair are separated by 0.5 s. In normal controls it has been shown that the response to the test stimulus (known as S2) is much lower than to that of the conditioning stimulus (known as S1). The level of suppression in schizophrenics has been shown to be much lower than in controls (see Fig. 6.3, from Jin et al., 1997). Indeed this gating ratio has been suggested as a candidate endophenotype i.e. a possible marker for those at risk of developing the condition (Patterson et al., 2008). The reduced gating ratio is thought to be due to a deficit in sub-cortical and cortical neuronal inhibitory pathways. Thus those with schizophrenia are unable to inhibit irrelevant sensory information resulting in an overload of information reaching consciousness. How this reduced gating of temporally close discrete stimuli relates to the responses obtained using continuous stimuli could be investigated using the AESPA method. While previous

studies eliciting ASSRs and continuous stimuli have shown deficits in gamma-band (40Hz) synchronization and power and have also shown that phase-locking factor was positively correlated with auditory hallucination symptoms (Spencer et al., 2008, Spencer et al., 2009) the temporal detail afforded by the AESPA response may be useful in further elucidating possible gating deficits in schizophrenia adding to the understanding and characterization of this possible endophenotype.

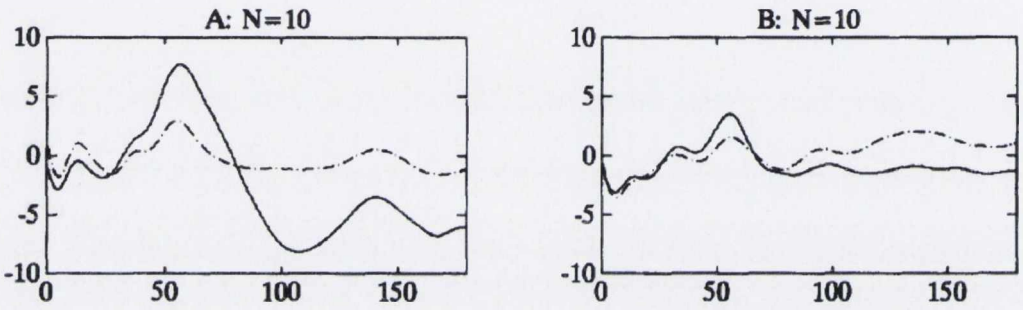

Figure 6.3. Grand Average ERPs in normals (A: N=10) and schizophrenics (B: N=10). Compared with schizophrenic patients, normal subjects had higher S1 P50 amplitude (solid lines) and lower S2/S1 ratio, while S2 responses (dotted lines) remained the same between groups. Figure taken from Jin et al, 1997

6.2 Summary and Discussion

In this thesis we have discussed the limitations of the standard AEP method. We have sought to overcome some of these limitations by establishing the AESPA method as outlined in Chapter 3. The flexibility of this method allows for the application of more natural stimuli to investigate auditory function. To this end we employed the AESPA to investigate the auditory processing centres affected by endogenous auditory attention using continuous artificial stimuli (Chapter 4). We

then went on to probe the temporal dynamics of auditory attention to natural speech passages in a cocktail party environment (Chapter 5).

In chapter 3 the AESPA method is introduced and the responses obtained using it are characterised and compared and contrasted to standard AEP responses. The AESPA was found to have four prominent positive components which we labelled Pa, Pb, Pc and Pd and three prominent negative components (Na, Nb and Nc). AESPA responses were found to be highly correlated with standard AEP responses. The AESPA Nc component was further identified to have similar generator as the AEP N1 component in primary auditory cortex. The AESPA response is shown to be a generalisation of the standard AEP averaging procedure. Furthermore the flexibility of the AESPA method allows for the response to the full temporal profile of discrete stimuli (which have non-zero duration) to be extracted as opposed to merely the onset response extracted using of the averaging method. Pa, Na, Pb and Nb were found to likely result from subcortical centres where as Pc, Nc and Pd were found to result from cortical structures specifically auditory cortex. The smaller amplitude of the AESPA responses as compared to the AEP responses is thought to result from a number of possible reasons such as subset of linear cells and possibly stimulus specific adaptation (SSA; Power et al., 2011).

Having established and characterised the AESPA we then sought to apply it to the study of endogenous auditory attention in Chapter 4. The Nc component of the AESPA responses was found to be significantly effected by endogenous attention. Nc was once again localised to auditory cortex and the identified effect was shown to

be due to increased activity of the same generators as when the stimulus was ignored. Since the AESPA likely represents activity of obligatory sensory processing and no additional generator we recruited when the stimulus was attended we concluded that obligatory sensory processing in auditory cortex was effected by endogenous attention. As outlined in the discussion in that chapter this agrees with the initials assertions of Hillyard et al in their pioneering paper in 1973. That said the additional and/or complementary "attentional trace theory" (Näätänen et al., 1987) is by no means ruled out by these findings. Earlier components may be susceptible to attentional modulation and it is possible that the lack of modulation of these earlier components may be due to the nature of the task employed i.e. a within stream event discrimination task was employed whereas employing a simple difficult-to-identify event may bias earlier processing.

To fully understand auditory function, and brain function in general, it is important to study it in response to natural stimuli in ecologically valid paradigms. Given the versatility of the AESPA method we were able to extract responses to two competing natural speech streams. Again the results support the filter theory of attention. Here we found that activity was highly significantly affected at ~209ms (i.e. the Pd component) furthermore the activity was marginally effect at ~143ms (i.e. the Nc component). Both of these effects were localised to auditory cortex. The results suggest that the Pd effect is primarily due to the suppression of processing to unattended/irrelevant information.

In its current form, the AESPA somewhat simplistically assumes a linear relationship between the stimulus envelope and the EEG response. Due to the network-based nature of information processing in the brain and its undoubted root in feedforward, feedback and recurrent activity, it is clear that a linear feedforward assumption would render the AESPA an incomplete measure of auditory processing. For example, the modulation of early afferent activity by efferent activity from higher order areas is likely to be a highly non-linear process that would not be well characterized by the AESPA method. Indeed, recent work has unsurprisingly shown that expanding the AESPA method to higher dimensions, does increase the ability to predict EEG by passing a novel input through the AESPA model/transfer function (Power et al., 2011b). This suggests that there are meaningful interactions between stimuli at different time points that affect the subsequent EEG activity. The fact that evidence for nonlinear processing exists is not at all surprising given the necessarily nonlinear nature of real world objects such as the brain. In fact, the fact that the auditory system responds in a highly nonlinear way to auditory intensity has been well documented (Green, 1988). We attempted to correct for some of this nonlinearity using our logarithmic intensity mapping (see page 46), however many other types of nonlinearity exist in the brain, such as saturation and burst firing. That said, a large body of evidence indicates the usefulness of describing EEG using linear models (eg Franaszczuk and Blinowska 1985; Blinowska and Franaszczuk 1989, Wright 1990, Blinowska and Malinowski, 1991). Indeed the studies contained in this thesis highlight the application of the linear AESPA as useful and relevant in many scenarios especially those in which more natural experimental protocol is

desired. In sum, despite the absence of non-linear contributions effects on the linearly related activity are necessarily and evidently important.

The work outlined in the thesis emphasizes that AESPA response primarily represent exogenous (obligatory sensory) activity. Given that the AESPA response results from activity that is directly related to a driving input of rapid intensity fluctuations it is unlikely that endogenous activity, such as the PN (Näätänen, 1982), contingent negative variation (CNV) (van Boxel & Böcker, 2004) and Bereitschaftspotential (or readiness) Potenital (Kornhuber & Deecke, 1965), which are not well synchronised to stimulus fluctuations would be present in the responses obtained. The activity of the above mentioned phenomena are the result of stimulus expectation (CNV), response preparation (Bereitschaftspotential Potenital) and a voluntarily maintained endogenous cognitive representation of a stimulus (PN) as opposed to lower level direct sensory processing of stimulus fluctuations and thus would not be represented in AESPA responses.

The results outlined in Chapter 4 highlight some debates related to the basic neural processing of stimuli and attentional processing. Much debate has revolved around the issue of whether attention operates at an early stage of processing based on the physical characteristics of the stimulus or at a later stage based on a cognitive representation (Broadbent, 1958; Moray, 1959; Deutsch and Deutsch, 1963; Johnston and Wilson, 1980). This debate, which was raised on the back of behavioural data, is echoed in electrophysioloical studies which generally argue that attention acts either on sensory processing or voluntary endogenous activity (e.g.

Hillyard et al., 1973 vs. Naatanen et al., 1982). Here the fact that exogenous sensory activity has been shown to be effected supports the idea that attention does operate on the sensory processing of auditory stimuli. Recently, however, there has also been much interest in the possibility of a flexible locus of attention that acts at the stage of processing most overloaded by the task (Vogel, 2002). Indeed, for many auditory attention tasks previously employed (e.g., deviant frequency detection) it is likely that the early endogenous processing negativity, PN, (which has been linked to a flexible mechanism of rough selection between sounds based on overall discriminability, Hansen and Hillyard, 1980) and the flexible task-related effects underpinned by sensory gain mechanisms (loaded at a stage of processing inherent to sensory processing e.g. frequency processing/discrimination) would act at a similar stage of processing and would thus overlap. That is to say if overall discriminability *and* task requirements are reliant on the same information effects are likely to occur around the same time. Alternatively, if competing stimuli were discriminable based on frequency difference but the task required responses based on some other aspect of the stimulus that was reliant on information preceeding or succeeding frequency processing the effects resulting from the separate mechanisms may not overlap. Thus, this inconsistent overlap of activity based on specific task demands may be the primary contributor to the conflicting data showing either support for selection at stage of obligatory sensory processing or voluntary endogenous activity. Indeed studies using very similar paradigms but slightly different task demands have shown support for one theory or the other (e.g. Hillyard et al., 1973 vs. Naatanen et al., 1982). Taking into account the multitude of studies supporting either theory it seems reasonable to deduce that both mechanisms would be active in any given paradigm in

order to achieve optimal attentional deployment and selection. Thus the ability of the AESPA method to disentangle sensory driven activity from endogenous voluntary activity has proved to be hugely important to the establishing a truer idea of auditory attentional processing mechanisms. Further investigation of the flexible locus and how this relates to attentional deployment on obligatory-sensory and voluntary endogenous activity is obviously required and is discussed further taking the results of Chapters 4 and 5 into account.

The Nc component did not show a consistent attention effects in the cocktail-party experiment using natural speech (Chapter 5) as was found for in Chapter 4. This may speak to the issue of a task-related flexible temporal locus of selective attention (Lavie et al., 2004, Vogel et al., 2005). Given the semantic requirements of the task attentional recourses would likely be deployed in order to maximize the ability to semantically process attended speech. Given previous reports of the semantic processing of unattended speech without the ability to consciously identify that speech (Lewis, 1970; Holender, 1986), the left hemispheric Pd suppression effect observed in our data may relate to the prevention of memory trace formation for the semantic information in the unattended speech stream. This accords reasonably well with EEG/MEG data suggesting that semantic memory use during language comprehension can be indexed by the so-called N400 component (Kutas and Hillyard, 1980; Kutas and Federmeier, 2000), which has been purported to onset as early as 200-250 ms in the posterior half of the left superior temporal gyrus (Kutas and Federmeier, 2011). Furthermore, it has been shown in a dichotic listening word list task that, while both attended and unattended words are semantically processed

and activate semantic representations, the N400 elicited by unattended words is insensitive to semantic manipulation (Bentin et al., 1995). The attentional suppression of our Pd component at 200-220 ms, combined with the general lack (or at least severe reduction – see Kutas and Federmeier, 2011) of any semantic manipulation effects on the N400 to unattended stimuli, may suggest a specific temporal locus of attentional suppression before which semantic processing occurs and after which semantic information would otherwise be encoded into working memory. Thus considering the studies outlined in Chapters 4 and 5 identified differing dominant loci of attention based on differing task demands it seems likely that a flexible locus of attention is at play and is evident at the stage of processing most relevant to the task demands as suggested by Vogel and colleagues (Vogel., 2005).

Furthermore, the impact of using AESPA in this and future attention studies using natural speech is apparent given that previous investigations using ERPs and continuous speech have had to superimpose short probe stimuli onto the natural speech streams to enable ERPs to be extracting by averaging to the probes (Hink and Hillyard, 1976; Woods et al., 1984; Coch et al., 2005; Nager et al., 2008). This introduces the possibility of unwanted exogenous effects attention-grabbing effect (Hopfinger and West, 2006) and indeed whether the resultant ERP is actually due to speech processing or processing related to the non-speech probe stimuli has been questioned (Woods, 1984). In addition, the averaging procedure could also result in temporal smearing of the ERP and any corresponding attention effects as a result of incorporating speech input immediately surrounding the probes. That said, the

averaged ERP procedure has revealed a great deal about the electrophysiology of auditory attentional selection with evidence of attentional effects having been shown in a large number of dichotic speech studies (e.g., Hink and Hillyard, 1976; Woods et al., 1984; Teder et al., 1993; Coch et al., 2005; Nager et al., 2008). However, as mentioned previously, using the ERP to disentangle the contributions of endogenous attentional processes (such as the PN component; Näätänen, 1982; Woods, 1990) from those that act on exogenously driven processing is difficult given their temporal overlap. As such, it may be the case that the averaged ERP technique may not be the best suited tool for determining the potentially highly precise temporal locus of attention for suppressing the memory encoding of unattended but semantically processed speech. This is made clear when considering the markedly early onset and broad temporal extent of the PN component. Because of the fact that it has been shown to onset as early as ~50 ms and because of its fronto-central distribution, it is highly unlikely that the early phase of the PN component (the PN_e) relates to semantic processing/working memory. Indeed, it is thought to index a flexible mechanism of rough selection between sounds based on their feature-based discriminability (Hansen and Hillyard, 1980). The later phase of the PN wave (PN_l) is less well understood than its earlier counterpart (Woods et al., 1990), but is thought to be related to the maintenance of an attentional trace (Näätänen, 1982). This response can be particularly broad in temporal duration leading to the notion that it relates to how relevant the superimposed probe stimuli are to the speech stream and not to the processing of the speech itself (Woods et al., 1984). Thus AESPA adds specificity in identification of effects on sensory processing as well as

various methodological and paradigm design advantages over standard ERP methods.

The research contained in this thesis highlights the utility of the AESPA for research on questions relevant to basic neuroscience and suggests possible applications to research on neurological deficits such as dyslexia and schizophrenia. This thesis also adds to the drive towards investigating brain function using more natural paradigms. This is of great importance as investigating the brain with more natural stimuli in more natural scenarios will lead to a better insight into the everyday function and dysfunction (in the case of neurological disorders) of the brain.

Bibliography

Ahveninen J, Jääskeläinen IP, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin F-H, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW. 2006. Task-modulated ''what'' and ''where'' pathways in human auditory cortex. *Proc Natl Aca Sci*, 103: 14608–14613.

Alho K, Teder W, Lavikainen J, Näätänen R. 1994. Strongly focused attention and auditory event-related potentials. *Biol Psych* 38: 73 – 90.

Beagley HA, Knight JJ. 1967. Changes in auditory evoked response with intensity. *J Laryngol Otol* 81:861-73.

Bear AF, Connors BW, Paradiso MA (2007). *Neuroscience: Exploring the Brain.* Philadelphia: Lippincott Williams and Wilkins.

Bentin S, Kutas M, Hillyard SA. 1995. Semantic processing and memory for attended and unattended words in dichotic listening: behavioral and electrophysiological evidence. J Exp Psychol Hum Percept Perform 21:54-67.

Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O. 2007. Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J Neurosci* 27:9252-9261.

Bieser A, Müller-Preuss P. 1996. Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. *Exp Brain Res* 108:273–284.

Blinowska KJ, Franaszczuk PJ. 1989. A model of the generation of electrocortical rhythms. Springer Series in Brain Dynamics, Vol. 2. Springer, Berlin Heidelberg New York, pp 192-201.

Blinowska KJ, Malinowski M. 1991. Non-linear and linear forecasting of the EEG time series *Biol. Cybern.* 66, 159-165.

Broadbent D. 1958. *Perception and communication.* London, UK: Pergamon Press.

Burkard RF, Don M, Eggermont JJ. (Eds.). 2007. *Auditory Evoked Potentials: Basic Principles and Clinical Applications.* Philadelphia: Lippincott Williams & Wilkins.

Cancelli I, Cadore IP, Merlino G, Valentinis L, Moratti U, Bergonzi P, Gigli GL, Valente M. 2006. Sensory gating deficit assessed by P50/Pb middle latency event related potential in Alzheimer's disease. *J Clin Neurophysiol* 23:421-425.

Celesia GG, Peachey NS. 1999. Visual evoked potentials and electroretinograms. In: *Electroencephalography: Basic Principles, Clinical Applications and Related Fields, fourth edition,* edited by Niedermeyer E, Lopes da Silva F. Baltimore: Lippincott Williams & Wilkins.

Cherry, E.C. 1953. Some experiments on the recognition of speech, with one and with two ears. *J. Acous. Soc. Am.* **25**, 975-979.

Chiappa KH, Harrison JL, Brooks EB, Young RR. 1980. Brainstem auditory evoked responses in 200 patients with multiple sclerosis. *Ann Neurol* 7:135-143.

Cohen YE and Knudsen EI. 1999. Maps versus clusters: different representations of auditory space in the midbrain and forebrain. *Trends Neurosci.* 22: 128-135.

Coch, D., Snaders, L.D. & Neville, H.J. 2005. An Event-related Potential Study of Selective Auditory Attention in Children and Adults. *J. Cog. Neuro.* **17**, 605-622.

Cornelissen P., Richardson A., Mason A., Fowler S., Stein J. 1995. Contrast sensitivity and coherent motion detection measured at photopic luminance levels in dyslexics and controls. *Vision Research.* 35: 1484-1494.

Davis PA. 1939. Effects of Acoustic Stimulation on the Waking Human Brain. *J Neurophysiol* 2:494-499.

De Sanctis P, Ritter W, Molholm S, Kelly SP, Foxe JJ. 2008. Auditory scene analysis: the interaction of stimulation rate and frequency separation on pre-attentive grouping. *Eur J Neurosci* 27:1271-1276.

Deiber MP, Ibanez V, Bastuji H, Fischer C, Mauguière F. 1989. Changes of middle latency auditory evoked potentials during natural sleep in humans. *Neurology*. 39:806-813.

Deutsch RP, Deutsch D. 1963. Attention: some theoretical considerations. *Psychological Review*. 70:80-90.

Drake Jr ME, Pakalnis A, Phillips B, Padamadan H, Hietter SA. 1991. Auditory evoked potentials in anxiety disorder. *Clin Electroencephalogr* 22:97-101.

Drullman R, Festen JM, Plomp R. 1994a. Effects of temporal envelope smearing on speech reception. *J Acous Soc Am.* 95:1053-1064

Drullman R, Festen JM, Plomp R. 1994b. Effects of reducing slow temporal modulation on speech reception. *J Acous Soc Am.* 95:2670-2680

Eggermont JJ. 1993. Wiener and Volterra analysis applied to the auditory system. *Hear Res* 10:167-190.

Ehret G & Romand R. 1997. *The Central Auditory System.* New York: Oxford University Press.

Escera C, Alho K, Schröger E, Winkler I. 2000. Involuntary attention and distractability as evaluated with event-related potentials. *Audiol Neurootol.* 5:151 – 166.

Eysholdt U, Schreiner C. 1982. Maximum length sequences – a fast method for measure brain-stem evoked responses. *Int J Aud* 21:242-250.

Franaszczuk PJ, Blinowska KJ. 1985. Linear model of brain electrical activity, EEG as a superposition of damped oscillatory modes. *Biol Cybern* 53:19-25.

Galambos R, Makeig S, Talmachoff PJ. 1981. A 40-Hz auditory potential recorded from the Human Scalp. *Proc Nat Acad Sci USA* 78: 2643 – 2647.

Gallinat J, Bottlender R, Juckel G, Munke-Puchner A, Stotz G, Kuss H-J, Mavrogiorgou P, Hegerl U. 2000. The loudness dependency of the auditory evoked N1/P2-component as a predictor of the acute SSRI response in depression. *Psychopharmacology* 148:404-411.

Gibb BJ. 2007. A Rough Guide to the Brain, Penguin Books.

Goswami U. 2010. A temporal sampling framework for developmental dyslexia. *Trends Cog Sci.* 15(1):3-10.

Green DM. 1988. Profile Analysis: Auditory Intensity Discrimination, Oxford: Oxford University Pres.

Hall III JW. 2007. The New Handbook of Auditory Evoked Responses. Boston, MA: Allyn and Bacon.

Hansen JC, Hillyard SA. 1980. Endogenous brain potentials associated with selective auditory attention. *Electroencephalogr Clin Neurophysiol.* 49: 277-290.

Hashimoto I. 1982. Auditory evoked potentials from the human midbrain: slow brain stem responses. Electoenceph and Clin Neurophysiol. 53:652-657.

Hill, K.T. & Miller, L.M. 2010. Auditory Attentional Control and Selection in Cocktail Party Listening. *Cereb. Cortex* **20**, 583-590.

Hillyard SA, Hink RF, Schwent VL, Picton TW. 1973. Electrial signs of selective attention in the human brain. *Science*. 182: 177 – 180.

Hink, R.F. & Hillyard, S.A. 1976. Auditory Evoked Potentials during Selective Attention to Dichotic Speech Maeeages. *Percep. & Psychophys*. **20**, 236-242.

Holender D. 1986. Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behav Brain Sci* 9:1-66.

Hopfinger JB and West VM. 2006. Interactions between endogenous and exogenous attention on cortical visual processing. *NeuroImage*. 31: 774 – 789.

Howard III MA, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA. 1996. A chronic microelectrode investigation of the tonotopic organization of the human auditory cortex. *Brain Res* 724:260–264.

John MS, Dimitreijevic A, Picton TW. 2002. Auditory Steadt-State Responses to Exponential Modulation Envelopes. *Ear & Hearing* 23:106-117.

Johnston WA, Wilson J. 1980. Perceptual processing of nontargets in an attention task. *Mem Cognit*. 8:372-377.

Jin Y., Potkin S.G., Patterson J.V., Sandman C.A., Hetrick W.P., Bunney W.E. Jnr. 1997. Effects of P50 temporal variability on sensory gating in schizophrenia, *Psychiatry Research*, 70:71 – 81.

Jonides J. 1981. Voluntary versus automatic control over the mind's eye movement. *Attention Performance*. 9: 187 – 203.

Kadobayashi I, Toyoshima A. 1984. Effects of attention on middle latency potentials. *Folia Psychiatr Neurol Jpn*. 38:425-463

Kelly SP, Gomez-Ramirez M, Foxe JJ. 2008. Spatial attention modulates initial afferent activity in human primary visual cortex. *Cerebral Cortex*. 18:2629-2636.

Kerlin JR, Shahin AJ, Miller LM. 2010. Attentional gain control of ongoing cortical speech respresentations in a "Cocktail Party". *J Neurosci* 30: 620 – 628.

Kiebel SJ, Friston KJ. 2004a. Statistical parametric mapping for event-related potentials I: generic considerations. *NeuroImage* 22:492-502.

Kiebel SJ, Friston KJ. 2004b. Statistical parametric mapping for event-related potentials II: a hierarchical temporal model. *NeuroImage* 22:503-520.

Kiiski H, Reilly RB, Lonergan R, Kelly S, O'Brien M, Kinsella K, Bramham J, Burke T, O'Donnchadha S, Nolan H, Hutchinson M, Tubridy N, Whelan R. 2011. Change in PASAT performance correlates with the change in P3 ERP amplitude over a 12-month period in multiple sclerosis patients. *J Neurol Sci*. 305: 45-52.

Klein DJ, Depireux DA, Simon JZ, Shamma SA. 2000. Robust spectrotemporal reverse correlation for the auditory system: Optimizing stimulus design. *J Comp Neuro* 9:85-111.

Kornhuber HH, Deecke L. 1965. Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale [Brain potential changes during arbitrary and passive movements in humans: Bereitschaftspotential and reafferent potentials]. *Pflügers Archiv*, 284, 1–17.

Kraus N, McGee T, Littman T, Nicol T. 1992. Reticular formation influences on primary and non-primary auditory pathways as reflected by the middle latency response. *Brain Res.* 587: 186-194.

Kutas M, Hillyard SA. 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*. 207:203-205.

Kutas M, Federmeier KD. 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol.* 62:621-647.

Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ. 2006. The VESPA: A method for the rapid estimation of a visual evoked potential. *Neuroimage* 32:1549-1561.

Lalor EC, Kelly SP, Pearlmutter BA, Reilly RB, Foxe JJ. 2007. Isolating endogenous visuo-spatial attentional effects using the novel Visual Evoked Spread Spectrum Analysis (VESPA) technique. *Eur J Neurosci* 26:3536-3542.

Lalor EC, Yeap S, Reilly RB, Pearlmutter BA, Foxe JJ. 2008. Dissecting the cellular contributions to early visual sensory processing deficits in schizophrenia using the VESPA evoked response. *Schizophr Res* 98:256-264.

Lalor E.C., Foxe J.J. 2009. Visual evoked spread spectrum analysis (VESPA) responses to stimuli biased towards magnocellular and parvocellular pathways, *Vision Research*, 49: 127 – 133.

Lalor EC, Power AJ, Reilly RB, Foxe JJ. 2009. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. *J Neurophysiol* 102:349-359.

Lalor EC, Foxe JJ. 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur J of Neurosci.* 31:189-193.

Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT. 2000. Automated Talairach Atlas labels for functional brain mapping. *Hum Brain Mapp* 10:120-131.

Lavie N, Hirst A, de Fockert J, Viding E. 2004. Load theory of selective attention and cognitive control. *J Exp Psych.* 133:339-354.

Leavitt VM, Molholm S, Ritter W, Shpaner M, Foxe JJ. 2007. Auditory processing in schizophrenia during the middle latency period (10-50 ms): high-density electrical mapping and source analysis reveal subcortical antecedents to early cortical deficits. *J Psychiatry Neurosci* 32:339-53.

Lehmann D. 1987. Principles of spatial analysis. In: *Handbook of electroencephalography and clinical neurophysiolog, vol. 1: methods of analysis of brain electrical and magnetic signals*, edited by Gevins AS, Redmond A. Amsterdam: Elsevier, 1987.

Lehmann D, Skrandies W. 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol* 48, 609– 621.

Lewis JL. 1970. Semantic processing of unattended messages using dichotic listening. *J Exp Psychol.* 85: 225-228.

Linden RD, Picton TW, Hamel G, Campbell KB. 1987. Human auditory steady-state evoked potentials during selective attention. *Electroencephalogr Clin Neurophysiol* 66:145-59.

Lovegrove W.J., Bowling A., Badcock B., Blackwood M. 1980. Specific reading disability: differences in contrast sensitivity as a function of spatial frequency. *Science.* 210: 439-440.

Lu T, Liang L, Wang X. 2001. Temporal and rate representations of time-varying signals in auditory cortex of awake primates. *Nature Neurosci* 4:1131-1138.

Marmarelis PZ, Marmarelis VZ. 1978. Analysis of Physiological Systems: The White Noise Approach. New York, NY: Plenum Press.

Marshall CM, Snowling MJ, Bailey PJ. 2001. Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. *J Speech Lang Hear Res, 44*(4): 925-940.

McAnally KI, Hansen PC, Cornelisson PL, Stein JF. 1997. Effect of time and frequency manipulation on syllable perception in developmental dyslexics. *Journal of Speech, Language and Hearing Research, 40,* 912-24.

McAnally K.I., Stein J.F. 1997. Scalp Potentials Evoked by Amplitude-Modulated Tones in Dyslexia, *Journal of Speech, Language and Hearing Research*. 40: 939-945.

McArthur GM, Bishop DVM. 2001. Auditory perceptual processing in people with reading and oral language impairments: current issues and recommendations. *Dyslexia, 7*, 150-170

McGee T, Kraus N, Littman T, Nicol T. Contribuations of medial geniculate body subdivisions to the middle latency response. *Hear Res.* 61: 147-154.

Mennel P., McAnally K.I., Stein J. 1999. Psychophysical Sensitivity and Physiological Response to Amplitude Modulation in Adult Dyslexic Listeners., *Journal of Speech, Language and Hearing Research* 1999; 42: 797-803.

Merigan W.H., Eskin T.A. 1986. Spatio-temporal vision of macaques with severe loss of Pb retinal ganglion cells. *Vision Research*. 26: 1751-1761.

Moray NP. 1959. Attention in dichotic listening: Affective cues and the influence of instructions. *Q J Exp Psychol*. 9:56-60.

Morel A, Garraghty PE, Kaas JH. 1993. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol* 335:437–459.

Murray MM, Brunet D, Michel CM. 2008. Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr* 20:249-64.

Näätänen R. 1982. Processing negativity: An evoked-potential reflection of selective attention. *Psychol Bull*, 92: 605 – 640.

Näätänen R, Picton T. 1987. The N1 wave of human electric and magnetic response to sound: A review and analysis of component structure. *Psychophysiology*. 24: 375 – 425.

Näätänen R, Teder W, Alho K, Lavikainen J. 1992. Auditory attention and selective input modulation: a topographical ERP study. *NeuroReport* 3: 493 – 496.

Näätänen R, Winkler I. 1999. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull*. 125: 826-859.

Näätänen R. 2001. 'Primitive Intelligence' in the auditory cortex. *Trends in Neurosci.* 24: 283-288.

Nagamoto HT, Adler LE, Waldo MC, Griffith J, Freedman R. 1991. Gating of auditory response in schizophrenics and normal controls. Effects of recording site and stimulation interval on the P50 wave. *Schizophr Res* 4:31-40.

Nager, W., Dethlefsen, C. & Munte, T.F. 2008. Attention to human spearkers in a virtual Auditory Environment: Brain Potential Evidence. *Brain Research*. **1220**, 164-170.

Nakai T, Kato C & Matsuo K. 2005. An fMRI Study to Investigate Auditory Attention: A Model of the Cocktail Party Phenomenon. *Magn Reson Med Sci.* 4:75-82.

Nelken I. 2004. Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol*. 14:474-480.

Nittrouer S. 1999. Do temporal processing deficits cause phonological processing problems? *J Speech Lang Hear Res, 42*(4), 925-942.

Olejniczak P. 2006. Neurophysiologic Basis of EEG. *J Clin Neurophys*. 23: 186-189.

Oostenveld R, Praamstra P. 2001. The five percent system for high-resolution EEG and ERP measurements. *Clin Neurophys*. 112:713-719.

Patterson P.V., Hetrick W.P., Boutros N.N., Jin Y., Sandman C., Stern H., Potkin S., Bunney W.E. Jnr. 2008. P50 sensory gating ratios in schizophrenics and controls: A review and analysis, *Psychiatry Research*, 158: 226 – 247.

Pantev C, Hoke M, Lehnertz K, Lütkenhöner B, Anogianakis G, and Wittkowski W. 1988. Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalogr Clin Neurophysiol* 69:160–170.

Picton TW, Hillyard SA, Krausz HI, Galambos R. 1974a. Human auditory evoked potentials. I: evaluation of components. *Electroenceph Clin Neurophys* 36:179-190.

Picton TW, Hillyard SA. 1974b. Human auditory evoked potentials. II. Effects of attention. *Electroencephalogr Clin Neurophysiol* 36:191-199.

Picton TW, John MS, Dimitrijevic A. 2002. Possible roles for the auditory steady-state responses in identification, evaluation and management of hearing loss in infancy. *Audiol Today*. 14:29-34.

Pinek B, Duhamel JR, Cavé C, Brouchon M. 1989. Audio-spatial deficits in human: differential effects associated with left versus right hemisphere parietal damage. Cortex 25:175–186.

Power A.J., Lalor E.C., Reilly R.B. 2007. Eliciting Audio Evoked Potentials Using Continuous Stimuli, *Engineering in Medicine and Biology Society, 2007, 29th Annual International Conference of the IEEE.*

Power, A.J., Lalor, E.C. & Reilly, R.B. 2011. Endogenous Auditory Attention Modulates Obligatory Sensory Activity in Primary Auditory Cortex. *Cereb. Cortex* 21:1223-1230.

Power AJ, Reilly RB, Lalor EC. 2011b. Comparing linear and quadratic models of the human auditory system using EEG. Proceedings of 33rd Intl IEEE Engineering in Medicine and Biology Conference, Boston.

Purcell DW, John MS, Schneider BA, Picton TW. 2004. Human Temporal Auditory Acuity as assessed by envelope following responses. *J Acoust Soc Am.* 82:165-178.

Purves D, Augustine GJ, Fitzpatrick D, Hall WC, LaMantia A-S, McNamara JO, White LE. 2001. *Neuroscience*. Sunderland, MA, USA: Sinauer Associates, Inc.

Rahne T, Böckmann M, von Specht H, Sussman ES. 2007. Visual cues can modulate integration and segregation of objects in auditory scene analysis. *Brain Res* 1144:127-135.

Ramus F., Rosen S., Dakin S.C., Day B.L., Castellote, J.M., White S. Frith U. 2000. Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*. 126, 841-865.

Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci USA*. 97: 11800 – 11806.

Ritter W, De Sanctis P, Molholm S, Javitt DC, Foxe JJ. 2006. Preattentively grouped tones do not elicit MMN with respect to each other. *Psychophysiology*, 43, 423–430.

Romani GL, Williamson SJ, Kaufman L. 1982. Tonotopic organization of the human auditory cortex, *Science*, 216: 1339-1340.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci.* December; 2(12): 1131–1136.

Rosburg T., Boutros N.N., Ford J.M., 2008. Reduced auditory evoked potential component N100 in schizophrenia – A critical review, *Psychiatry Research*, 161: 259 – 274.

Rosen S. 2003. Auditory processing in dyslexia and specific language impairment: is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics, 31*, 509-27.

Rosen S, Manganari E. 2001. Is there a relationship between speech and nonspeech auditory processing in children with dyslexia? *J Speech Lang Hear Res, 44*(4), 720-736.

Ross B, Picton TW, Herdman AT, Hillyard SA, Pantev C. 2004. The effect of attention on the auditory steady-state response. *Neurol Clin Neurophysiol*. 22: 1-4.

Ross B, Hillyard SA, Picton TW. 2010. Temporal dynamics of selective attention during dichotic listening. *Cereb Cor*. 20:1360-371

Skrandies W. 1995. Visual information processing: topography of brain electrical activity. *Biol Psychol* 20:1-15.

Shapley R., 1990. Visual sensitivity and parallel retinocortical channels. *Annual Review of Psychology*. 41: 635-658.

Shelley A-M., Silipo G., Javitt D.C. 1999. Diminished responsiveness of ERPs in schizophrenic subjects to changes in auditory stimulation parameters: implication for theories of cortical dysfunction, *Schizophrenia Research*, 37: 65 – 79.

Shi Y, Hecox KE. 1991. Nonlinear system identification by m-pulse sequences: application to brainstem auditory evoked responses. *IEEE Trans Biomed Eng* 38:834-845.

Snyder JS, Alain C, Picton TW. 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci* 18:1-13.

Sonnadara RR, Alain C, Trainor LJ. 2006. Effects of spatial separation and stimulus probability on the event-related potentials elicited by occasional changes in sound location. *Brain Res*. 1071:175-185.

Spencer KM, Niznikiewicz MA, Nestor PG, Shenton ME, McCarley RW. 2009. Left auditory cortex gamma synchronization and auditory hallucination symptoms in schizophrenia. *BMC Neurosci*, 10:85.

Spencer KM, Salisbury DF, Shenton ME, McCarley RW. 2008. γ-band Auditory Steady-State Responses are impaired in First Episode Psychosis. *Boil Psych*, 64:5, 369-375.

Spierer L, Bellmann-Thiran A, Maeder P, Murray MM, Clarke S. 2009. Hemispheric competence for auditory spatial representation. *Brain*. 132: 1953-1966.

Suga N. 1995. Processing of auditory information carried by species-specific complex sounds. In: *The Cognitive Neurosciences*, edited by Gazzaniga M. Cambridge, MA: MIT.

Sussman E, Ritter W, Vaughan HG Jr. 1999. An investigation of the auditory streaming effect using event-related brain potentials. *Psychophysiology*. 36: 22–34.

Studdert-Kennedy M, Mody M. 1995. Auditory temporal perception deficits in the reading impaired: A critical review of the evidence. Psychonomic Bulletin & Review, 2: 508-14.

Tallal P. & Piercy M. 1973. Defects of non-verbal auditory Perception in Children with Developmental aphasia. *Nature*. 241: 468-469

Tallal P. 1980. On the 'specifics' of specific reading disability and specific language impairment. *J. Child Psychol. Psychiatry*. 41: 869-847.

Tallal P. 2004. Improving language and literacy is a matter of time. *Nat Rev Neurosci*, 5(9): 721-728.

Tanaka H, Hachisuka K, Ogata H. 1999. Sound lateralisation in patients with left or right cerebral hemispheric lesions: relation with unilateral visuospatial neglect. *J Neurol Neurosurg Psychiatry*. 67:481–486.

Teder, W., Kujala, T. & Näätänen, R. Selection of Speech Messages in free-field listening. 1993. *NeuroReport*. **5**, 307-309.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. 2001. Functional Specialization in Rhesus Monkey Auditory Cortex. *Science*. 292: 290-293.

Ulanovsky N, Las L., Nelken I. 2003. Processing of low-probability sounds by cortical neurons. *Nat Neurosci.* 6:391-398.

Van Boxtel GJM, Böcker KBE. 2004. Cortical measures of anticipation. *Journal of Psychophysiology*, 18, 61-76.

Van der Horst R, Leeuw AR, Dreschler WA. 1999. Importance of temporal-envelope cues in consonant recognition. *J Acoust Soc Am*. 105:1801-1809.

Vogel EK, Woodman GF, Luck SJ. 2005. Pushing around the locus of selection: evidence for the flexible-selection hypothesis. *J Cogn Neurosci.* 17:1907--1922.

Waber DP, Weiler, MD, Wolff PH, Bellinger D, Marcus DJ, Ariel R, Forbes P, Wypij D. 2001. Processing of rapid auditory stimuli in school-age children referred for evaluation of learning disorders. *Child Development, 72*(1): 37-49.

Witton C., Stein J.F., Stoodley C.J., Rosner B.S., Talcott J.B. 2002. Separate Influences of Acoustic AM and FM Senisitivity on the Phonological Decoding Skills of Impaired and Normal Readers. *Journal of Cognitive Neuroscience*. 14:6: 866-874.

Woldorff M, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE. 1993. Modulation of early sensory processing in human auditory cortex during selective attention. *Proc Natl Acad Sci USA*. 90: 8722 – 8726.

Woldorff M. G., 1993. Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction., *Psychophysiology*, 30, 98-119.

Woods, D.L., Hillyard, S.A. & Hansen, J.C. 1984. Event-related Potentials reveal similar Attentional Mechanisms during Selective Listening and Shadowing. *J. Exp. Psych.* **10**, 761-777.

Woods DL. 1990. The physiological basis of selective attention: Implications of event-related potential studies. In: Event-related brain potentials: basic issues and applications (Rohrbaugh JW, Parasuraman R, Johnson Jr. R, ed), pp178-209. New York: Oxford UP.

Wright JJ. 1990. Reticular activation and the dynamics of neuronal networks. *Biol Cybern* 62:289-298.

Yoshiura T, Ueno S, Iramina K, Masuda K. 1995. Source localization of middle latency auditory evoked magnetic fields. *Brain Res.* 703:139-144.

Zani A & Proverbio AM. 2003. *The Cognitive Electrophysiology of Mind and Brain.* Burlington, MA: Academic Press.

Zanto, T.P. & Gazzaley A. 2009. Neural Suppression of Irrelevant Information underlies Optimal Working Memory Performance *J. Neurosci.* **29**, 3059-3066.

Zatorre RJ, Ptito A, Villemure JG. 1995. Preserved auditory spatial localization following cerebral hemispherectomy. *Brain.* 118:879–889

Zatorre RJ, Penhune VB. 2001. Spatial localization after excision of human auditory cortex. *J Neurosci*. 21:6321–6328

Zatorre, R.J. & Belin, P. 2001. Spectral and Temporal Processing in Human Auditory Cortex. *Cereb. Cortex* **11**, 946-953.

Appendix A: A MATLAB implementation of basic AESPA estimation

```
%% load EEG data and the stimulus envelope

load EEG                          % dimensions are num_electrode x
num_sample_points
load stimulus_envelope        % dimension is 1 x num_sample_points
num_electrodes = size(EEG,1);
window_size = 320;   % This is the number of sample points of the AESPA response
                                  % being estimated at an EEG sampling rate of 512Hz
this is                           % 625ms
EEG = [EEG  zeros(num_electrodes,window_size)];
stimulus_envelope = [stimulus_envelope'; zeros(window_size,1)];

%% Solve for the AESPA response for each channel

mxxt = zeros(window_size,window_size);

for i = window_size:length(stim_envelope)

   x = [stim_envelope(i:-1:i-window_size+1)];
   xt = x';
   xxt = x*xt;
   mxxt = mxxt + xxt;

end

mxxt = mxxt/(i-window_size+1);

for chan = 1: num_electrodes          % Estimate one electrode at a time

   mxy = zeros(window_size,1);

   for i = window_size:length(stim_envelope)

      x = [stim_envelope(i:-1:i-window_size+1)];
      y = EEG(chan,i-61);      % -61 is the number of pre-stimulus samples that we
                               % wish to calculate. At 512 sampling it corresponds to
                               % roughly -120ms.
      xy = x*y;
      mxy = mxy + xy;
   end

   mxy = mxy/(i-window_size+1);
   w(chan,:) = mxxt\mxy;
end
```