The Acoustic Change Complex for Measuring Speech Perceptual Performance in Normal Hearing Listeners in Noise, Cochlear Implant Users, and Second Language Listeners

Emma Brint

A dissertation submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

of

UCL

Department of Speech, Hearing and Phonetic Sciences

Division of Psychology and Language Sciences

University College London

2017

Declaration

I, Emma Brint confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

Firstly, I would like to give my greatest thanks to my marvellous supervisor, Paul Iverson, for his support and guidance throughout this PhD. Paul's expertise and knowledge has been invaluable, and he has been there at all hours of the day, always ready to have a discussion. I have learnt so much from Paul both in the lab and on the softball pitch, and for that I am truly grateful. I would also like to thank my second supervisor, Stuart Rosen for his advice and support, as well as providing materials for this thesis. In addition, I would like to thank Jyrki Tuomainen for his advice on EEG, Bronwen Evans for sharing her expert ear, Yasuaki Shinohara and Yue Zhang for their statistical advice, Steve Nevard, Andrew Clark and Dave Cushings for their technical help, and Karen Wild for her help with recruiting cochlear implant participants. Without them all this PhD would have been a lot trickier. In addition, I would like to thank Andrew Faulker and Valerie Hazan for their support and dedication to the department (and of course for their musical talent). I would also like to thank the ESRC for funding my studies, and all the willing participants who took part in this research.

Special thanks go to Jieun Song for her help with phase coherence and her comments on this thesis, and to Kathleen McCarthy for her comments and much needed pep talks. The moral support I have received from these two has been invaluable. My colleagues and friends in the PhD room of course receive my sincerest thanks. Without you all, this PhD wouldn't have been half as fun. Among many, I thank José Joaquín Atria, Tim Schoof, Mauricio Figueroa Candia, Kurt Steinmetzger, Louise Stringer, Petra Hödl, Gisela Tome Lourido, Faith Chiu, Daniel

Kennedy-Higgins, Helen Willis, Giulia Borghini, Gwijde Maegherman, Hao Liu, Xiaoluan Liu, Sonia Granlund, Wafaa Al-Shangiti, Cristiane Hsu, Yasna Pereira, Albert Lee, Nada Al-Sari, Lucy Carroll, Mark Wibrow, Dong-Jin Shin, and Csaba Redey-Nagy. In addition, I would like to thank Georgina Roth, Katherine Mair, Laurianne Cabrera, and Melanie Pinet for always looking out for me.

My thanks also go to Nicholas Haywood for his invaluable comments on this thesis and his loving support through the final crunch. His imagination and vivacity are contagious, and he has been a great source of inspiration for me in this final year. My friends, who have kept my life full over the last four years, also receive my loving thanks. Amongst others, I thank Alessandra Carballal, Barnabas Chappell, Louise Chappell, Becky Moore, Llŷr Williams, the Rudkins family, the Jersey girls, the Knights, Louisa Finch and my magical partner in crime Emily Rolland. Furthermore, I am forever indebted to the one-of-a-kind Victoria Rudkins, who has been there for me through thick and thin, providing just the right balance of distraction and encouragement. Her unwavering support could not be matched. I would also like to thank my brother, Matthew Brint for his unquestioning friendship. Finally, I thank my wonderful parents, Alan and Cheryl Brint. Without their unconditional support, belief, and encouragement I would not have been able to achieve what I have done. They always have and always will inspire me.

For Shiela, Valar Morghulis.

Abstract

A series of experiments recorded the acoustic change complex (ACC) to a broad range of speech stimuli. The ACC is an electrophysiological response to an acoustic change within an on-going stimulus. Recordings were obtained from normal hearing (NH) listeners in noise, cochlear implanted (CI) listeners, and second language listeners. Chapter 2 presented a mixture of transitions between four vowels and four fricatives in quiet and at three noise levels to NH listeners. We investigated how the ACC was affected by noise, and how the ACC relates to individuals' behavioural speech-in-noise ability. Chapter 3 details a CI study using the same fricative and vowel stimuli but in quiet only. The ACC was measured and again compared to behavioural speech perception performance. Participants' neural entrainment to continuous speech was also measured, and compared to their speech perception performance and ACC responses. Chapter 4 details an ACC study using a mixture of transitions between eight voiceless fricatives presented to native English, Finnish and Polish speakers. The ACC magnitude was used to create similarity matrices that were analysed by non-metric MDS and an acoustic analysis of the fricative stimuli was performed. The ACC and its relationship to the spectra of the stimuli were used to investigate cross language differences between the groups. Overall, results suggest that the ACC is not merely a measure of general auditory detection as it is often described, but rather it is a measure at the border between auditory and linguistic processing in the auditory cortex. Furthermore, the results indicate that the ACC has potential for further use in research in a variety of listener populations, as well as potential clinical benefits.

Contents

List of Figures

List of Tables

1 General Introduction

Speech perception can be highly variable, due to differences in peripheral hearing function (e.g. hearing loss), differences in underlying perceptual and cognitive processing (e.g. second-language learners), or factors that are harder to identify (e.g. individual differences in speech-in-noise performance by normal-hearing individuals). Such individual differences are evident in people's daily experiences, but may also be apparent under clinical observation. In order to recognise and treat people with particular speech perception difficulties in a clinical setting, research focuses on both how listeners cope with challenging speech perception, and from where individual differences originate.

Speech processing can be measured and assessed at different levels of the human auditory system. Many behavioural measures of speech processing when used alone, such as tests of identification or sentence recognition, encompass the whole pathway from outer ear hearing, to central speech processing in the brain. This means that whilst performance on such measures may identify a problem, such tests may not necessarily locate the source of impairment. For example, when compared to a young adult normal-hearing listener, an elderly listener with normal audiometric thresholds may perform either equally as well or worse on a behavioural speech-in-noise task (e.g. Frisina & Frisina 1997; Schoof 2014). In such a scenario, worse performance is commonly attributed to age-related central auditory dysfunction, and/or age related cognitive decline (e.g. Füllgrabe et al. 2015). However, when compared to a young adult normal-hearing listener, an elderly *hearing impaired* listener will also likely perform worse on a behavioural speech recognition test. As such, performance may indicate either the effect of hearing loss, or the effect of aging plus hearing loss, thereby implying possible peripheral dysfunction, central dysfunction, or cognitive factors, or any combination of these factors (e.g. Frisina & Frisina 1997). For such a listener, it would not be possible to identify the source(s) of their reduced SIN ability from a conventional behavioural test of speech performance alone. This is also a practical issue in the clinic; clinicians currently rely on the audiogram and behavioural speech testing to treat a patient with speech processing difficulties, however, as outlined above, these measures alone may be insufficient to identify the source(s) of difficulty. Therefore, if an elderly listener with a hearing impairment is fitted with a hearing aid, but is not achieving high speech recognition performance, the clinician does not know if it is a peripheral issue (e.g. a fault with the hearing aid) or a central issue (e.g. an effect of ageing). As such, it may be difficult to tailor a rehabilitation program to the individual.

The growing recognition that global behavioural speech tests may be affected by multiple factors has led to renewed interest in developing further tests to better identify the cause(s) of impairment. For example, Füllgrabe et al. (2015) demonstrated that both composite measures of cognition and measures of temporal fine structure sensitivity are predictors of speech-in-noise performance in elderly listeners with normal audiograms. Furthermore, neurophysiological and neuro-imaging techniques provide an additional means to investigate the sources of speech processing difficulties, whilst testing with speech stimuli directly. Such methods can provide a direct measurement of speech processing at different stages of the auditory pathway (see Section 1.1). Furthermore, neurophysiological measures are objective, and they can often be recorded passively from subjects (i.e. many measures do not require the subject to perform or even attend to the acoustic stimuli). As such, certain objective measures are commonly used in hearing screening for new-born babies and infants, and can also be used for non-responsive or non-compliant subjects. The reduced participant effort associated with objective measures is also desirable when testing clinical or ageing populations. However, there is not currently a clinically-used objective measure of speech processing.

One group of objective measures that can examine the neural encoding of incoming speech are auditory evoked potentials (AEPs). AEPs are commonly measured using electroencephalography (EEG), where scalp-placed electrodes are used to record electrical activity in the brain. When a population of neurons fire in synchrony, changes in the voltage of the extracellular fluid are detectable from the scalp recording sites. Evoked voltage changes are, however, relatively small in comparison to those from noise sources, and so neural EEG signals are typically observable only after repeated measurement and averaging, or after further processing. Neural responses to sound have been successfully observed from passive activation of the auditory nerve (Wave I of the ABR: Pratt et al. 1999) to high-level cortical semantic processing of complex stimuli (N400: Kutas & Hillyard 1983; P600: Osterhout & Holcomb 1992; Hagoort et al. 1993). One benefit of using EEG to study auditory processing is the high temporal resolution the technique affords. This allows for accurate measurement of a system that characteristically displays rapid and precise temporal encoding. Furthermore, good temporal resolution allows for the identification of the area of processing from the latency of the response from the stimulus presentation. That is, the further along the pathway that the response occurs, the longer the observed latency. However, the spread of electrical current results in relatively poor spatial resolution. Nevertheless, modern EEG systems are capable of recording from a large number of electrodes, and post-recording signal processing allows for some degree of source localisation, to further identify sites of activation.

The experiments presented in this thesis investigated the effect of peripheral and central factors of speech perception on the acoustic change complex (ACC). The remainder of this chapter provides a background into EEG measures of speech perception, including the ACC. The following three chapters detail experiments that use a combination of the ACC and behavioural measures to explore speech perception in normal-hearing adults, cochlear implanted adults, and second language learners. Chapters 2 and 3 address peripheral manipulations (i.e. background noise and cochlear implants) that may limit speech perception performance. Specifically, Chapter 2 explores the ACC in normal-hearing adults listening to speech in noise, and Chapter 3 investigates the ACC in cochlear implanted adults listening to speech in quiet. Chapter 4 controls these periphery factors (i.e. normal-hearing in quiet) and addresses a central factor of speech perception, by measuring responses from listeners that differ in native language experience. Chapter 5 provides a discussion of the implications of the main findings from this thesis, and the potential use of the ACC in both research and clinical settings.

1.1 EEG Measures of Auditory Speech Processing

Commonly used AEP measures of auditory processing include the auditory brainstem response (ABR), the brainstem frequency following response (FFR) and the cortical mismatch negativity (MMN). The ABR is a series of seven peaks that occurs within the first 10 ms after stimulus onset (Pratt et al. 1999; Trainor

2008). These peaks represent subcortical activity along the auditory pathway from the cochlear nucleus to the thalamus, and are currently used clinically as a screening tool to assess hearing loss in babies (Trainor 2008). Owing to their short latency, and maximal simulations firing to transient sounds, ABRs are typically elicited using brief tone pips, tone sweeps/chirps, or clicks (Dau et al. 2000). Although ABRs can be evoked from the onset of speech sounds also, the measure reflects a transient response to sound onset, and so it alone cannot be used to characterize the sustained responding relevant for speech encoding. Encoding at the level of the ABR is primarily thought to preserve or enhance the spectral and temporal signal transmitted by the auditory nerve, with little high-level processing occurring at this stage in the pathway. Therefore, ABRs are not used to evaluate higher auditory functions, such as speech discrimination (Whiting 1998).

A potentially more speech-relevant brainstem measure is the frequency following response (FFR: Moushegian et al. 1973). The FFR represents sustained neural activity that is phase locked to the waveform or the envelope of a periodic stimulus (Krishnan 2008), and is attributed to generator sites in the inferior colliculus (e.g. Greenberg et al. 1987). It is observed as a series of periodic peaks whose intervals correspond to the period of the stimulus frequency (Krishnan 2008). The FFR can be measured using speech sounds, such as /da/, and so is a step forward from the ABR towards successfully measuring speech perception (e.g. Cunningham et al. 2001; Skoe & Kraus 2010). The FFR can only be recorded using periodic stimuli within the range of phase locking at the brainstem (below ~ 1.5 - 2 kHz; Krishnan 2008), a range well encompassing that of the fundamental frequencies of speech (Traunmüller & Eriksson 1995). Furthermore, the FFR has shown some cross-language effects. When listeners are presented with Mandarin tones, native

Chinese listeners show a more robust FFR than English listeners (Krishnan et al. 2005, 2009a, 2009b). Additionally, short- and long-term training (such as musical experience) is associated with stronger FFR encoding (Wong et al. 2007; Carcagno & Plack 2011). Whilst the FFR may reflect temporal encoding beneficial for speech and/or pitch processing, recent studies suggest that the measure may not reflect pitch-specific encoding, and so the experimental and clinical applicability of the FFR remains debated (Dau 2003; Gockel et al. 2011; Gockel et al. 2015).

Cortical AEPs (CAEPs) are sensitive to a more complex range of speech materials and manipulations than brainstem measures. CAEPs are closely linked to auditory processes that underlie speech perception, such as auditory detection and discrimination (Picton 1990; Stapells 2002; Martin et al. 2008). A cortical response that is often used in speech processing research is the mismatch negativity (MMN: Näätänen et al. 1978). The MMN is a pre-attentive response that occurs as a negative peak 100-250 ms after a deviation is detected from an established auditory pattern (Szymanski et al. 1999; Näätänen 2001). The response is elicited by presenting standard tokens interspersed with deviant tokens; when a deviant is heard, the MMN response is elicited (Näätänen et al. 2001). The MMN can be elicited from a wide variety of infrequent changes including tone duration, spatial location, intensity, violations of a repeating tone pattern, and acoustic characteristics such as frequency (see Näätänen et al. 2001 for a review). Furthermore, evidence from neurophysiological studies using the MMN has shown language specific perceptual sensitivities (Dehaene-Lambertz 1997; Näätänen et al. 1997; Winkler et al. 1999).

Unlike the FFR, the MMN can be elicited using a wide range of both periodic and aperiodic stimuli spanning low and high frequencies. At a group level the MMN is fairly replicable; however, on an individual basis it is less reliable in terms of detection and reproducibility (Näätänen & Kreegipuu 2012). Furthermore, the MMN paradigm is not time-efficient, as a standard must be established before each response evoking deviant can be presented. This means that for every trial the target is preceded by two to six standard stimuli, drastically increasing the testing time. This limits the clinical applicability of the measure, and in experimental research, limits the number of conditions which can be reasonably tested within a single experiment. These disadvantages have led to interest in more time-efficient measures of cortical speech processing, such as the P1-N1-P2 complex or the acoustic change complex, which is the main focus of this thesis.

1.2 The P1-N1-P2 Complex

The auditory-evoked P1-N1-P2 complex takes the form of three distinct peaks: a positive peak at around 50 ms after stimulus onset, a negative peak at about 100 ms, and a second positive peak at about 180 ms (Martin et al. 2007). The complex is considered an early cortical response, and is generated in the thalamocortical segment of the central auditory system (Vaughan & Ritter 1970; Wolpaw & Penry 1975; Näätänen & Picton 1987; Hyde 1997). It can be elicited from sound onset or offset, or from the occurrence of a change within a sound. When elicited from a change within a sound, the response is known as the acoustic change complex (ACC; Ostroff et al. 1998; Martin & Boothroyd 1999, 2000). P1-N1-P2 responses to sound onset or offset share similar morphologies and they are not considered to be physiologically independent (as evidenced by temporal interactions; Hillyard & Picton 1978; and source localisation; Pantev et al. 1996a). Furthermore, onset responses have been studied more extensively than offset responses. As such, offset responses are not discussed further. In this section, P1-N1-P2 responses to onsets are first considered, and the ACC is discussed further in section 1.3. Because each of the three components of the response is considered to be differentially affected by factors such as stimulus parameters, age, and attentional state, each component is considered individually after a summary of the global response.

As the global P1-N1-P2 onset complex is elicited irrespective of stimulus type, it is often described as an obligatory response that reflects the neural encoding of the acoustic properties of an incoming signal, and so is necessary for behavioural detection (e.g. Martin et al. 2008). However, there is growing evidence that at least some aspects of the P1-N1-P2 complex represent more than pure detection only, and may be indicative of attentional and training effects (Tremblay & Kraus 2002; Crowley & Colrain 2004; Wagner et al. 2013). Furthermore, stimulus parameters can affect the amplitudes and latencies of the three components of the complex differentially (Crowley & Colrain 2004).

Early research on the P1-N1-P2 complex used the onset response to estimate hearing sensitivity (Perl et al. 1953; Appleby et al. 1963; McCandless & Best 1964; Cody & Bickford 1965; Davis 1966; Davis et al. 1967; Beagley & Kellogg 1969; Tyberghein & Forrez 1971; Mendel et al. 1975) and more recently, the response has been found to provide a reasonable estimate of behavioural thresholds (Stapells 2002; Lightfoot & Kennedy 2006). Despite this, the ABR is used for the clinical measurement of hearing thresholds rather than the P1-N1-P2 complex. This is because the ABR is not affected by sleep, sedation or attention (Stapells et al. 1994, 1995), whereas the P1-N1-P2 complex requires the subject to be awake. Although it is not as practical as the ABR in clinical audiometry, the P1-N1-P2 complex is well suited to speech processing research as it can be elicited using a broad range of speech material, and it has a larger amplitude than the ABR.

1.2.1 P1

For adults with normal hearing, the P1 occurs at approximately 50-80 ms after the onset of a sound, or after a change within a sound, and was first described by Geisler et al. (1958). It is thought to reflect the initial detection of sound by the auditory cortex, and the encoding of the sound's acoustic characteristics such as frequency, amplitude and timing at a pre-perceptual level (Shtyrov et al. 1998; Čeponienė et al. 2005; Anderson 2010). It has not been shown previously to be very sensitive to the magnitude of acoustic differences (Čeponienė et al. 2005). However, there is some evidence that the P1, when measured in the ACC, has a larger amplitude for changes between vowels than changes between fricatives (Iverson et al. 2016), indicating that it could be useful for more varied stimuli. Furthermore, it is possible that the P1 magnitude saturates quickly when a stimulus is detected, and so it may be more useful in experiments where the stimuli are near detection threshold (Iverson et al. 2016).

P1 is generated in the primary auditory cortex at Heschl's gyrus (Wood & Woolpaw 1982; Reite et al. 1988; Pool et al. 1989; Liégeois-Chauvel et al. 1994; Huotilainen et al. 1998; Ponton et al. 2002), however more recent work suggests that it has more complex generators including the hippocampus, the planum temporale, the lateral temporal cortex, and neocortical areas (Liégeois-Chauvel et al. 1994, 1999; Howard et al. 2000; Grunwald et al. 2003; Kisley et al. 2003).

The P1 is often described as an obligatory or gating response to the onset of a sound (Pratt 2012), and is not thought to be modulated by attention (Picton & Hillyard 1974). However, there is some evidence of its amplitude increasing with greater cognitive effort (Rao et al. 2010). The P1 is the least investigated of the three peaks, as it is the smallest in amplitude in adults and the least reliable to measure.

1.2.2 N1

The N1, first described by Davis (1939), is a negative peak that follows the P1, occurring at approximately 80-110 msec after the onset of a sound or the change within a sound. It is thought to reflect the encoding of time-varying aspects of a sound, such as amplitude, spectral, and voice onset time changes (Sharma & Dorman 1999; Martin & Boothroyd 2000; Sharma et al. 2000; Hoonhorst et al. 2009). The N1 can be used to assess hearing thresholds in infants and children and was the first evoked potential to be used for this purpose (Barnet & Lodge 1966; Rapin & Graziani 1967; Suzuki & Origuchi 1969; Taguchi et al. 1969), however there are reports that it does not fully mature until adolescence (Barnet 1975; Paetau et al. 1995; Bruneau et al. 1997; Pang & Taylor 2000; Ponton et al. 2000; Čeponienė et al. 2002; Kushnerenko et al 2002).

In general, the N1 response predominantly reflects the encoding of an audible stimulus onset or change, but unlike the P1, it is thought to vary with the magnitude of the acoustic change (Näätänen & Picton 1987; Picton 1990; Steinschneider & Dunn, 2002). However, the relationship of the N1 amplitude to acoustic characteristics of stimuli is not always monotonic. For example, at high intensity onset amplitudes, it has been suggested that the magnitude of the N1 starts to level off or even decrease (Buchsbaum 1976; Adler & Adler 1989; Paiva et al. 2016). Moreover, when noise is added to an auditory signal, at high signal-to-noise-ratios (SNR; i.e. > +30 dB SNR), the N1 peak can actually increase in amplitude, whereas the P1 and P2 peaks decrease in amplitude even at low noise levels (Kaplan-Neeman et al. 2006; Parbery-Clark et al. 2011; Papesh et al. 2015).

Näätänen & Picton (1987) describe three components of the N1. The first is fronto-centrally predominant (Vaughan and Ritter, 1970), generated bilaterally by vertically oriented sources in the supratemporal plane. The second component is biphasic, with a positive peak at around 100 ms and a negative peak at around 150 ms, generated in the lateral aspect of the superior temporal gyrus (secondary auditory cortices) and recorded over the temporal areas. This second component was first described by Wolpaw & Penry (1975) who termed it the 'T-complex,' and it has recently been found to be affected by the language background of the listener (Wagner et al. 2013). The third component is measured at the vertex and it is suggested that it induces a widespread arousal of the neural networks that facilitates stimulus detection, analysis, and response generation (Näätänen & Picton 1987; Crowley & Colrain 2004).

The amplitude of the N1 can be modulated with attentional focus and the predictability of the stimuli (Näätänen & Picton 1987; Hillyard et al., 1973; Woldorff & Hillyard 1991). The magnitude of the N1 is increased with selective attention to the stimuli (Näätänen & Picton 1987) and salience of the stimuli (Escera et al. 1998; Melara et al. 2002), and decreases rapidly with habituation of repeated stimuli (Fruhstorfer 1971; Rust 1977; Megela & Teyler 1979; Kenemans et al. 1989). Furthermore, there is emerging evidence that the N1 amplitude can be predictive of speech-in-noise ability (Billings et al. 2013).

1.2.3 P2

The P2 follows the N1, and in normal-hearing adults occurs at approximately 160-200 ms after the onset of or change within a sound. It matures as early as 2-3 years of age (Barnet 1975; Paetau et al. 1995; Bruneau et al. 1997; Pang & Taylor 2000; Ponton et al. 2000; Čeponienė et al. 2002; Kushnerenko et al 2002). The P2 is largest when measured at the vertex not only for auditory stimuli, but also for visual and somatosensory stimuli (Oades et al. 1995; Potts et al. 1998; Crowley & Colrain 2004), and it at least partly reflects the auditory driven output of the mesencephalitic reticular activating system (Knight et al. 1980; Näätänen & Picton 1987; Rif et al. 1991; Woods et al. 1993; Crowley & Colrain 2004). Using MEG, the sources of the P2 component have been identified to the anterior of the generators of the N1 (Sams et al. 1985; Hari et al. 1987; Pantev et al. 1991, 1996a, 1996b; Rif et al. 1991; Ross & Tremblay 2009).

Research suggests that the P2 magnitude does not decrease with habituation (Megela & Teyler 1979; Kenemans et al. 1989; Crowley & Colrain 2004), however, there is some evidence to suggest response adaption does occur, albeit at a slower rate than for the N1 (Ritter et al. 1968; Rust 1977). Furthermore, there is consistent evidence that P2 amplitude increases with psychoacoustic training and perceptual learning (Tremblay et al. 2001; Tremblay & Kraus 2002; Reinke et al. 2003; Ross & Tremblay 2009; Tremblay et al. 2014).

Some research suggests that the P2 may have a role in the classification or categorisation of auditory stimuli. In an oddball task, studies have found that the P2 magnitude is increased in response to the non-target stimuli, compared to the same stimuli in a task that does not require modulation of attention from the listener

(Garcia-Larrea et al. 1992; Novak et al. 1992). It has been suggested that this is an attention modulated effect (Novak et al. 1992), where the P2 contributes to identifying a stimulus as a target (Garcia-Larrea et al. 1992). Furthermore, this effect is largest at frontal electrode sites, which are situated on top of brain regions that involve inhibiting interference by irrelevant stimuli (Dempster 1991, 1992; van Zomeran & Bouwer 1994; Crowley & Colrain 2004).

The P2 decreases in amplitude when stimuli are attended. This contrasts with the N1, which increases in amplitude. However, this may reflect a combined effect of a broad negativity that occurs with attention (Hansen & Hillyard 1980; Näätänen & Picton 1987) making both the N1 and P2 appear more negative (i.e. resulting in a net increase in N1 magnitude, and decrease in P2 magnitude; Näätänen & Michie 1979; Näätänen & Picton 1987; Michie et al. 1990, 1993).

1.3 The Acoustic Change Complex

As previously outlined, when the P1-N1-P2 complex is measured in response to an acoustic change in an on-going sound, it is known as the acoustic change complex (ACC; Ostroff et al. 1998; Martin & Boothroyd 1999, 2000). A complex and constantly changing stimulus, such as continuous speech, produces multiple, often overlapping ACC responses. Although the ACC shares the morphology of a P1-N1-P2 complex, the extent to which the P1-N1-P2 complex and the ACC share neural processes is not fully understood. Whilst the onset-evoked P1-N1-P2 complex is itself a response to an acoustic change (i.e. from silence to sound), there is some evidence to suggest it does not represent identical processes to the ACC. For example, Jones et al. (1998) showed that the ACC to a change in pitch or timbre has a more posterior scalp distribution than the onset P1-N1-P2 response. Despite these distributional differences, as they are both a P1-N1-P2 complex to an acoustic change and share the same morphology they are often considered as variants of the same response. However, the extent of this variation is not fully understood.

The ACC is most commonly recorded from midline-central electrode sites which are dominated by contributions from the primary auditory cortex (e.g. Ponton et al. 2002; Tonnquist-Uhlen et al. 2003). It can be elicited using changes to basic stimulus properties such as frequency, intensity, and phase (Martin et al. 2007; Jerger & Jerger 1970; Näätänen & Picton 1987; Dimitrijevic et al. 2008; Ross et al. 2007), as well as more naturalistic changes, such as spectral and intensity changes in speech or speech like sounds (Ostroff et al. 1998; Martin & Boothroyd 1999, 2000).

The ACC, like the MMN, can be recorded without the listener actively listening or responding to the stimuli. However, it has three primary advantages over the MMN. Firstly, it is more time-efficient, as it can be elicited using stimuli that changes rapidly, at a rate of up to 2-3 Hz (Iverson et al. 2016) without having to create a standard. Secondly, it has a larger magnitude than the MMN, and so is more easily observed. Thirdly, it shows excellent test-retest reliability in individual adults (Martin & Boothroyd 1999; Tremblay et al. 2003). Despite such apparent benefits, the use of ACC as a tool to measure speech processing has not been extensively investigated.

Concerning speech stimuli, the ACC was first recorded in response to a change from a consonant to a vowel within a syllable (Kaukoranta et al. 1987; Ostroff et al. 1998). It has also been recorded to isolated acoustic characteristics of speech, such as amplitude and formant frequency changes within a vowel (Martin & Boothroyd 2000). As the ACC can be recorded using changes within speech stimuli, it may have important clinical implications, as the measure demonstrates cortical encoding of the acoustic features in speech that are needed for speech perception (Martin et al. 2007, 2008; Wagner 2013).

Usual measurement of the speech-evoked ACC typically involves recordings from isolated speech tokens with a phonetic change midway, such as /ui/ (Martin & Boothroyd 2000). This stimulus is repeated up to a few hundred times, so that the response can be observed easily after averaging, and a silent gap is presented between each token (typically about 2 s; Martin & Boothroyd 2000; Tremblay et al. 2003). Martin et al. (2010) improved the time-efficiency of the ACC paradigm by removing the inter-token silence, and so presenting a continuous chain of tokens (e.g. /uiuiuiuiuiui/). The ACC could be successfully measured at every change. However, when stimuli are repeated in this way, the magnitude of the N1 can be suppressed (Fruhstorfer 1971; Rust 1977; Megela & Teyler 1979; Kenemans et al. 1989).

Iverson et al. (2016) developed a method of recording the ACC that takes into account the multidimensional variability of natural speech, by allowing many more stimuli transitions to be measured in one experiment. These authors presented a long string of speech tokens concatenated in a random order, where token changes occurred at a rate of 2-3 times a second. Using this paradigm, the ACC was successfully recorded for over 50 different stimulus pairs in one experiment (i.e. a pair being a transition from one speech token to the next). Furthermore, by presenting the speech token stimuli in a random order, N1 suppression effects of stimulus repetition are minimised. The experiments described in this thesis used a paradigm similar to that of Iverson et al. (2016) in order to investigate the measurement of the ACC in different listening groups/conditions with a broad range of speech stimuli.

1.4 Aim

In summary, the ACC has great potential in the field of speech perception research, as it is characterised by good time efficiency, large response amplitude, and high test re-test reliability. The aim of this thesis was to investigate how the speech-evoked ACC is affected by manipulations of listening condition, and to explore the potential benefit of the measure for wider use in research and in the clinic. The work presented in this thesis describes three experiments that investigate the ACC in normal-hearing listeners in noise, cochlear implant users, and second language listeners. A combination of EEG and behavioural speech perception testing was used to examine how the ACC is informative of overall speech processing, and how it is affected by manipulations of speech perception. Chapter 2 details a study using a stimulus combination of four vowels and four fricatives, in quiet and in noise, to investigate how the ACC in normal-hearing listeners is affected by background noise, and how it relates to individual differences in speech-in-noise performance. Chapter 3 describes a study that presented cochlear implant (CI) users with the same fricative vowel stimuli as in Chapter 2, but presented in quiet only. Here, the ACC was measured in a group of CI users and was compared to their behavioural speech perception performance. The CI users' neural entrainment to continuous speech was also measured, and then compared to their speech perception performance and ACC responses. Finally, Chapter 4 details an ACC study using eight voiceless fricatives concatenated in a random order and played to native English, Finnish and Polish speakers. The ACC and its relationship to the spectra of the stimuli were used to investigate cross language differences between the groups.

2 Speech in Noise and the ACC

2.1 Introduction

In everyday life, many people struggle to understand speech in the presence of background noise. When in quiet, it is typically easy for a normal-hearing listener to follow a conversation with an individual speaker, whereas in a crowded room with multiple sound sources, the listener must selectively focus their attention on the desired speaker and tune out other voices and noise (Cocktail party effect: Cherry 1953). The attentional demand of selectively focusing on one speech source amongst competing sound sources remains a research topic of interest (e.g. Bronkhorst 2000). In addition to attentional demands, the presence of co-occurring acoustical signals can physically mask the speech signal of interest, making comprehension difficult or impossible.

Speech-in-noise (SIN) performance is often poor for the elderly, those with hearing loss, and those who are second-language listeners (e.g. Plomp 1978; Dubno et al. 1984; Rogers et al. 2006). Although age and peripheral hearing loss are known to increase SIN impairment, those with normal hearing often also experience difficulty. For example, a survey undertaken in the UK found that 26% of adults reported having difficulty understanding SIN, whereas only 16% had a hearing loss (≥ 25 dB HL, 0.5-4 kHz; Davis 1989). Recently, there has been growing interest in forms of hearing damage that may not be apparent from audiometric assessment alone. Animal studies have shown that noise exposure that results in temporary threshold elevations may also result in the permanent loss of peripheral (auditory nerve) fibers (Kujawa & Liberman 2009). This form of 'hidden' hearing loss, in which audiometric thresholds may appear normal, but elements of peripheral encoding are impaired, is currently being further explored in human listeners (Schaette & McAlpine 2011). It is also being investigated how this form of hearing damage may affect processes relevant to SIN encoding, such as the ability to encode precise temporal information (e.g. Plack et al. 2014).

There is great individual variability in speech recognition performance in noise. Better understanding of the variability seen in SIN performance provides an opportunity for improved prediction of SIN problems, improved diagnosis of conditions that effect SIN processes, and improved rehabilitation for individuals with problems with SIN. Such issues also concern our growing understanding of hidden hearing loss. If forms of hearing damage exist that are not diagnosed by standard audiometric tests, then there is scope to improve audiological assessment, by devising new measures of performance in conditions that reflect those challenging to hidden hearing loss, such as SIN. Furthermore, results from normal-hearing listeners can aid the identification of underlying mechanisms for speech processing in noise, which may then be relevant to understanding the causes of poor SIN performance associated with hearing impairment.

Both behavioural and physiological measures can be used to assess SIN ability. A wide range of SIN performance has been observed, even between those with normal hearing (Zhao & Stephens 2007). The reasons for these differences and the precise relationship between behavioural performance and physiological responses remain unclear. Nevertheless, physiological measures have been used to explore this variation by providing insight into how signals in noise are encoded in the auditory system (Billings et al. 2013). Observed cortical neural responses can provide information about the neural mechanisms used to process an incoming speech signal in noise.

This chapter details a study of the acoustic change complex (ACC) and the onset P1-N1-P2 complex to SIN in normal-hearing listeners. Four vowels and four fricatives were concatenated into sequences in a random order, and played to 30 native English normal-hearing listeners, in quiet, and in two levels of noise. The ACC response was recorded to each change of stimuli. Segments of silence were also occasionally present in the stimulus chain. This allowed for a comparison of the ACC and the onset P1-N1-P2 complex. Participants also completed a speech recognition task in noise, the results of which were compared to measures of the ACC and P1-N1-P2 complex responses. Variability across individuals was examined, as well as a comparison between measures.

2.1.1 CAEPs and Listening in Noise

CAEP studies concerning listening in noise have adopted a variety of measures, including the MMN (Muller-Gass et al. 2001), the N100m (Hiraumi et al. 2008; Alain et al. 2009), and the P1-N1-P2 onset response (e.g. Martin & Stapells 2005; Parbery-Clark et al. 2011; Billings et al. 2013). Both non-speech and speech stimuli have been investigated, and in general, cortical responses decrease in amplitude and increase in latency when stimuli are presented in noise, with some exceptions to this pattern, as discussed below (e.g., Martin & Stapells 2005; Parbery-Clark et al. 2011; Billings et al. 2013).

Using the P1-N1-P2 complex, Martin et al. (1997, 1999, 2005) and Whiting et al. (1998) explored how the N1 is affected by high pass, low pass, and broadband noise. Normal-hearing listeners were presented with the speech stimuli /ba/ and /da/. For all noise types, the N1 was present even when stimuli were not behaviourally discriminable from each other, and its amplitude decreased and latency increased with increasing noise level (or increasing/decreasing noise cut-off frequency; Martin et al. 1997; Whiting et al. 1998; Martin et al. 1999; Martin & Stapells 2005). Furthermore, they found that the N1 latency was more susceptible to degradation (increased latency) in noise than the N1 amplitude (Whiting et al. 1998). Several other studies have looked at adult P1-N1-P2 onset response to speech sounds, such as /ba/, /da/ or /ga/, or to non-speech sounds such as 1 kHz pure tones, in either white noise (Kaplan-Neeman et al. 2006; McCullagh et al. 2012; Kim et al. 2012; Sharma et al. 2014), speech-shaped noise (Papesh et al. 2015), broadband noise (Michalewski et al. 2009), or multi-talker babble (Parbery-Clark et al. 2011). It has been observed consistently that for high signal to noise ratios (SNR; > +30 dB SNR) the N1 peak increases in amplitude, whereas the P1 and P2 peaks decrease in amplitude and increase in latency even at high SNRs (Kaplan-Neeman et al. 2006; Parbery-Clark et al. 2011; Papesh et al. 2015). However, as noise increases (< +10 dB SNR) all three peaks decrease in amplitude and increase in latency (McCullagh et al. 2012; Kim et al. 2012; Papesh et al. 2015). Most previous studies have only addressed the N1 and the P2, as the P1 is often smaller in amplitude and more variable and so harder to measure reliably (Sharma et al. 2014). As such, several studies have excluded this peak from analysis (Kaplan-Neeman et al. 2006; Parbery-Clark et al. 2011; McCullagh et al. 2012), or have reported that it is affected differently by noise than the N1 and P2 peaks (Sharma et al. 2014). One aim of the current study was to obtain a reliable P1 response for analysis (i.e. from collecting from a large sample and by presenting a high number of stimulus presentations).

The majority of studies that have investigated the P1-N1-P2 response to speech presented in noise have used the onset response to /ba/, /da/, or /ga/ speech stimuli (e.g. Martin et al. 1997; Whiting et al. 1998; Kaplan-Neeman et al. 2006; Sharma et al. 2014). Fewer studies have measured responses to a broader range of speech stimuli (Zendel et al. 2015; Kuruvilla-Mathew et al. 2015). Zendel et al. (2015) recorded the onset response to French CVC tokens in a multi-talker babble from musicians and non-musicians. Consistent with previous studies, peak amplitudes decreased and peak latencies increased with decreasing SNR (e.g. Martin & Stapells 2005; Parbery-Clark et al. 2011; Billings et al. 2013). Musicians typically display larger P1 responses. Kuruvilla-Mathew et al. (2015) presented /di/, /ti/, /gi/, /mi/, /pi/, /si/ and /ʃi/ speech stimuli in a multi-talker babble at +10 dB SNR to normal-hearing listeners who were either wearing hearing aids or not. The unaided conditions saw an increase in latencies and a decrease in amplitudes with noise, whereas amplification produced complex effects on latencies and amplitudes across speech stimuli and peaks.

A related line of research has investigated the cortical responses to SIN of children with and without learning impairment or autism (Cunningham et al. 2001; Wible et al. 2002; Warrier et al. 2004; Russo et al. 2009; Anderson et al. 2010; Hassaan 2015). In general, children have a more defined P1 than adults (Sharma et al. 2002; Čeponienė et al. 2002) and a less defined N1 (Sussman et al. 2008) and P2. As such, child studies typically focus on P1-N1 amplitude and latency. As with the adult studies, /ba/, /da/, /ga/ or pure-tones are often used with white, broadband, speech shaped, or multi-talker babble noise. Again, the peaks consistently decrease in amplitude and increase in latency when the signal to noise ratio is decreased. Anderson et al. (2010) tested 32 children with a /da/ stimulus, and compared HINT

(Hearing in Noise Test: Nilsson et al. 1994) scores to P1-N1-P2-N2 response characteristics. Compared to speech presented in quiet, the P1 decreased in amplitude when the noise was added (multi-talker babble, +10 dB SNR). When they split the data into the top and bottom SIN performers, they found that those who performed worse on the HINT task had a larger N2 response, which they theorised showed a link between cortical inhibition and SIN perception. In contrast, Cunningham et al. (2001) found that children who performed worse in a just noticeable difference task between /ada/ and /aga/ also showed a smaller difference between their P2 and N2 peaks when listening to /da/ in broadband noise.

To my knowledge, there have been two studies that directly compared the onset P1-N1-P2 response to speech in noise to behavioural measures of SIN performance in adult listeners (Billings et al. 2013; Parbery-Clark et al. 2011). Billings et al. (2013) tested 15 normal-hearing adults on their SIN ability using IEEE sentences (Rothauser et al. 1969) presented in speech-shaped noise. They also measured P1-N1-P2 responses to /ba/ presented in speech-shaped noise at varying SNRs and varying signal levels. They concluded that N1 amplitude and latency measures best predicted SIN performance (compared to P1 and P2). Those performing worse in the IEEE task showed smaller-amplitude and longer-latency N1 responses. The authors also found significant correlations at certain SNRs for P1 amplitude and latency and P2 latency, but these correlations were overall less predictive of performance. Parbery-Clark et al. (2011) tested 22 normal-hearing adults on their SIN ability using the HINT test, and measured subcortical (ABRs) and cortical (N1-P2) responses to the speech token /da/ in a multi-talker babble of varying level. They found that at high SNRs the N1 amplitude and latency increased in comparison to in quiet. Even at high SNRs, however, the P2 amplitude decreased and latency increased in comparison to quiet. Looking at all their combined results, the authors concluded that those with better ABRs or better subcortical resilience to noise and better HINT scores showed larger N1 responses.

These two studies suggest that the onset P1-N1-P2 response, particularly the N1 component, shows some relationship with listeners' SIN abilities. To date however, there have been no studies investigating the relationship between listeners' SIN ability and ACC response. Although the ACC and the onset P1-N1-P2 complex share a similar morphology, it is not clear if they share the same neural generators (Martin et al. 2008). The ACC provides insight into the listener's ability to distinguish between acoustic features of speech, whereas the P1-N1-P2 complex represents the transition from silence to sound only. In other words, the former reflects the ability to discriminate speech-relevant acoustic features, whereas the latter may primarily reflect the obligatory detection/encoding of the onset of a new sound. For this reason, the ACC may be a more meaningful index of SIN performance than the P1-N1-P2 onset complex. However, it is not known whether the morphology of the ACC and the P1-N1-P2 complex are each similarly affected by the presence of noise. To this end, the aim of the following study was to investigate the effect of noise on the ACC compared to the P1-N1-P2 complex, and to compare the relationship of these responses to SIN performance. Furthermore, the speech stimuli used in previous studies have been limited, mainly consisting of /ba/, /ga/, and /da/. This study aimed to investigate the effect of noise on the ACC and P1-N1-P2 complex using a broad range of speech stimuli. We used a variety of speech tokens known to elicit a clear ACC in order to give a better representation of the variety of speech that a listener hears in daily life, and to minimize the adaptation effects associated with the repeated presentation of a single transition. By testing using a time-efficient presentation paradigm, each subject was presented with a large number of transitions between speech tokens. Additionally, a relatively large sample was tested (n = 29). As such, the experiment was designed with the aim of observing robust responses, allowing for a good representation of all early cortical components (especially the P1 peak).

In Experiment 1, SIN ability was measured using a sentence recognition task in noise. Participants listened to sentences in speech-shaped noise at three different SNRs, and were asked to repeat what they heard. For this behavioural task, sentence recognition was chosen in order to provide a broad measure of participants' SIN abilities. Experiment 2 measured the ACC and P1-N1-P2 onset responses to vowels and fricatives concatenated into chains in a random order, presented in speech-shaped noise at three different SNRs. The effect of noise on the ACC and the P1-N1-P2 complex was explored, and response characteristics were compared to the results of the sentence recognition task. Based on previous CAEP in noise studies, with the addition of noise, peak amplitudes of both the ACC and P1-N1-P2 complex were expected to be decreased, and component latencies were expected to be increased. Additionally, the N1 of the ACC and the P1-N1-P2 complex was expected to show a relationship with the participants' behavioural speech recognition in noise scores.

2.2 Experiment 1: Sentence Recognition in Noise

A sentence recognition task was used to measure the participants' SIN abilities. The results from this experiment accompany those from the following ACC and P1-N1-P2 onset complex in noise experiment (section 2.3).

2.2.1 Participants

Thirty native English-speaking normal-hearing adults contributed to the experiment. However, one participant was removed from analysis owing to excessively noisy data in the EEG recording, leaving a group of 29 participants (age range: 19-38). Native English was defined as those whose first language was English, and with no other language being spoken in their household before they started school. All participants passed a hearing screening test by demonstrating thresholds of at least 20 dB HL for pure-tones presented in quiet at the octave frequencies between 250 and 8000 Hz. All participants were right handed with no known neurological impairments.

The experiment and recruitment for this study was approved by the University College London ethics committee. All subjects provided their informed consent before beginning the experiment and were paid an honorarium for their time.

2.2.2 Stimuli

The Basic English Lexicon (BEL) sentence set was used to assess speech recognition (Calandruccio & Smiljanic 2012). This set comprises a large lexicon of 500 sentences created from non-native English conversation. All BEL sentences were recorded from a female Southern British English Speaker (500 sentences). The recordings were taken in a sound-proof booth with a 44.1 kHz sampling rate at 16 bits per sample. Speech-shaped noise was created based on the smoothed long-term average spectrum of the Bel sentence recordings. The sentences were embedded in this noise with SNRs of -9, -6, and -3 dB; with the noise level fixed and the sentence level adjusted.

2.2.3 Apparatus

A Fireface UC external audio interface (44.1 kHz sampling rate, 16 bits per sample) connected to Etymotic Research ER-1 insert earphones was used for stimulus presentation. Stimuli were presented binaurally at 65 dB SPL, and the sound level was verified with a Brüel & Kjaer 4157 artificial ear for insert earphones. Stimuli were presented using Praat (Boersma & Weenink 2013).

2.2.4 Procedure

Each participant listened to 20 sentences at each of the three SNR conditions (i.e. 60 sentences in total). Each sentence was selected randomly from the lexicon of 500 BEL sentences. For each participant, sentences were not repeated within or across SNR conditions. The stimuli were presented in a single block, for which the presentation order of SNR conditions was randomised. The participant and the experimenter were seated in a sound-proof booth. After the end of each stimulus presentation, the participant was asked to repeat back what they had heard, and the experimenter recorded the number of key words in the sentence they identified correctly (each Bel sentence contains four key words; e.g. The RABBIT and MOUSE EXPLORED the FIELD).

2.2.5 Results

For each SNR condition, the percentage of key words correct was calculated for each participant. This data is presented in Figure 2.1. Overall, participants performed worse as SNR was decreased (-9 dB SNR: *mean* = 37.97%, *s.d.* = 10.89, -

6 dB SNR: *mean* = 78.92%, *s.d* = 8.17, -3 dB SNR: *mean* = 94.01%, *s.d* = 4.10). As the variances were not equal between the results at each SNR (variability in performance increased as the SNR decreased, as evident from the increased ranges and interquartile ranges observed in Figure 2.1), a logistic mixed model was used. Sentence and subject were random intercepts, SNR was the independent variable and the sentence recognition response (i.e. coded as 1 for all key words correct within the sentence, and 0 for less than all key words correct, as a binary response is needed for a logistic regression with mixed factors) as the dependent variable. The results showed that the effect of SNR on the scores was significant ($\chi^2(2) = 189.65$, $p < .001$).

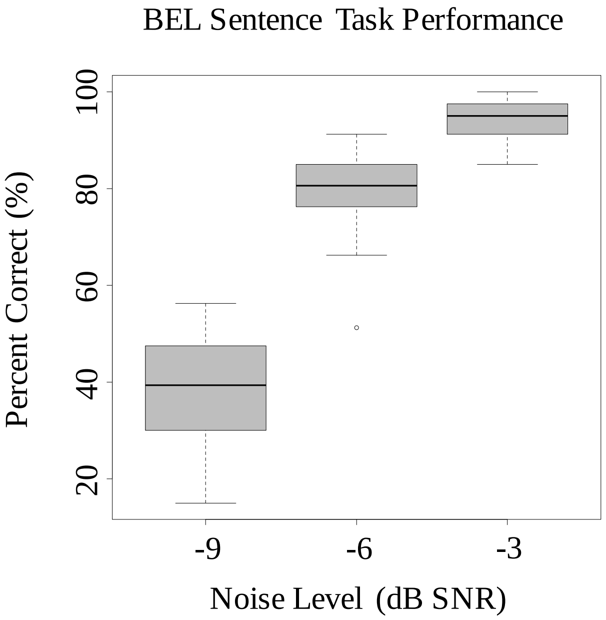

Figure 2.1 Boxplots showing the median (38.75, 80.00, 95.00), range (41.25, 25.00, 15.00) and interquartile ranges (16.25, 8.75, 6.25) for the BEL sentence task performance at – 9 dB SNR, -6 dB SNR, and -3 dB SNR respectively.

The sentence recognition results at -3 dB SNR were close to 100% and showed a ceiling effect. Furthermore, at -9 dB SNR some listeners performed very poorly, and so individual sentence scores at -6 dB SNR were used as the measure of SIN ability for Experiment 2. This intermediate condition reflected an SNR where speech recognition remained moderately high, but where the noise caused some impairment to performance.

2.3 Experiment 2: The Acoustic Change Complex in Noise

Experiment 1 showed that participants performed worse at the sentence recognition task as the SNR decreased. Furthermore, as SNR decreased, performance also became more variable between individuals. In Experiment 2, the ACC and onset P1-N1-P2 responses were recorded for a variety of speech sounds. The effect of noise on these responses was assessed, and then characteristics of the ACC and onset responses were compared to SIN performance.

2.3.1 Participants

The same 29 participants that were tested in Experiment 1 were tested in Experiment 2.

2.3.2 Stimuli

The stimuli comprised of four vowels (/ɑ/, /i/, /ɔ/, /u/), as heard in the words *Bart*, *beat*, *bought* and *boot*, and four fricatives (/ʃ/, /s/, /v/, /z/) as heard at the beginning of the words *short*, *sing*, *vine* and *zip*, and silence. Sustained vowels and fricatives were recorded from a female Southern British English speaker. The recordings were made in a sound-proof booth with a 44.1 kHz sampling rate at 16 bits per sample. Speech stimuli manipulations were then conducted in Praat (Boersma & Weenink 2013). For each recorded sound, a one second segment that was deemed to be reasonably static (minimal spectral change) was extracted from the recording. The amplitude envelope of each segment was then flattened by calculating the envelope of the original recording (full rectification and a 50-Hz low-pass filter, and then normalized). The original recording was then divided by this envelope, and the overall amplitudes of all eight phonemes were rescaled to be equal. For the vowels and voiced fricatives, the pitch was flattened to the mean pitch of the recordings (200 Hz) using an overlap-add method (Boersma & Weenink 2013).

The four vowels, four fricatives, and portions of silence were then concatenated into a random order without replacement (the same stimulus could not appear twice in a row) to create two-minute sequences that included all possible pair transitions (i.e. 72 possible transitions including each sound to silence and vice versa). The phonemes were concatenated by splicing each transition with 50 ms raised-cosine ramps at each overlapping transition (i.e. the phoneme was ramped down at the same time as the new phoneme was ramped up). The duration of each stimulus was randomly jittered from 350-450 ms, with the segment randomly selected from the longer, one-second recording. This minimised the possibility of any spectral discontinuity within the recorded stimuli from being consistently time-aligned to the spectral changes. After the phonemes were concatenated into a two-minute sequence, the entire amplitude envelope was flattened to correct for amplitude fluctuations during the brief overlapping transitions. These concatenations were made into two-minute audio files either in quiet or with speech-shaped noise added at +4 dB SNR or -3 dB SNR. For CAEPs from signals presented in noise, the effect of SNR has been shown to obscure any effect of signal level (Billings et al. 2009; Billings et al. 2012; Billings et al. 2013; Sharma et al. 2014). As such, the noise level was held constant across conditions, and the signal level was adjusted to create the desired SNRs. Each two-minute sequence was only used once.

2.3.3 Apparatus

All stimuli were presented binaurally via Praat (Boersma & Weenink 2013). A Fireface UC external audio interface (44.1 kHz sampling rate, 16 bits per sample) connected to Etymotic Research ER-1 insert earphones was used for stimulus presentation via a custom-built headphone amplifier. Stimuli were presented at 65 dB SPL, and the sound level was verified with a Brüel & Kjaer 4157 artificial ear for insert earphones.

An additional audio channel was used to provide stimuli timing information to the EEG system. A custom converter was used to convert the audio signal into TTL pulses, which was recorded as a time-aligned trigger by the EEG system. Testing occurred in an electromagnetically shielded sound-proof booth that was kept at 19^{o}C to minimize sweat artefacts. Within the booth, mains electricity was turned off during testing to reduce electrical artefacts as far as possible.

Evoked responses were recorded with a BioSemi ActiveTwo EEG system. Sixty-four active surface electrodes were placed in accordance with the international 10-20 system. Seven additional electrodes were placed on the left and right mastoid, left and right canthus, above and below the left eye, and on the tip of the nose. Responses were recorded at a sampling rate of 2048 Hz.

2.3.4 Procedure

Each participant was presented with 20 minutes (i.e. 10 two-minute sequences) of each condition (quiet, +4 dB SNR, and -3 dB SNR). This means that in total, each participant heard approximately 42 repetitions of each possible pair at each noise level (i.e. each possible phoneme transition pair). The entire stimulus presentation time was around one hour. During testing, participants watched a silent animated film, and were given the opportunity for a short break every 10 minutes.

2.3.5 Analysis

Data were processed offline in MATLAB (MathWorks 2013a), using the toolboxes Fieldtrip (Oostenveld et al. 2011), EEGlab (Delorme & Makeig 2004), ERPlab (Lopez-Calderon & Luck 2014), and NoiseTools (de Cheveigné & Simon 2007, 2008a, 2008b; de Cheveigné 2010, 2012, 2016; de Cheveigné & Parra 2014; de Cheveigné & Arzounian 2015). For each electrode channel, data were first high pass filtered at 0.1 Hz (Butterworth filter, as implemented by the ERPlab plugins within EEGlab), and then referenced to the mastoid electrodes. Using Fieldtrip, the recording was then segmented into epochs spanning from 100 ms before to 350 ms after each stimulus onset. Epochs were then baseline corrected to a 100 ms window before the stimulus change. Data were then low pass filtered at 30 Hz (Butterworth filter, as implemented by Fieldtrip) and downsampled to a 512 Hz sampling rate.

Using NoiseTools, denoising source separation (DSS[1]) was used to extract the ACC/P1-N1-P2 components for each subject, using a method that increased the signal-to-noise ratio of the neural signals, by selecting linear combinations of electrodes that maximised the repeatability of the ACC/P1-N1-P2 activity. After visual inspection of the DSS components, the first three were selected and the response was taken as the projection of these three components back into sensor space at electrode FCz. Epochs in which the signal amplitude exceeded 150 µV were rejected due to their high noise level, and the remaining epochs were averaged for each pair transition at each noise level.

Statistical analysis was run using R (R Core Team 2013). For linear mixed-model analyses the R 'lme4' package (Bates et al. 2015) was used, and type II analysis-of-variance tables calculated were using the 'CAR' package (Fox

---

[1] Denoising source separation (DSS) is a component analysis, which uses principle components analysis (PCA) alongside normalisation and a set of bias filters to isolate artefactual components and correct for them across the whole sensor space. It is assumed that components can be removed without affecting the rest of the data, or that reliable components of interest can be selected. DSS is thought to be more effective than other linear methods, such as ICA (de Cheveigné, A., Parra, L. 2014). Furthermore, it can be used to isolate the desired response, which can then be projected back into the sensor space. For the current study, DSS was applied using the Matlab toolbox NoiseTools (de Cheveigné & Simon 2007, 2008a, 2008b; de Cheveigné 2010, 2012, 2016; de Cheveigné & Parra 2014; de Cheveigné & Arzounian 2015).

&Weisberg 2011). The ‘lsmeans’ package (Lenth 2016) was used for post-hoc analysis.

2.3.6 Results

The effects of noise on the ACC and on P1-N1-P2 onset responses

The data was subdivided into two groups: ‘ACC’ trials (those which comprised of transitions from speech to speech; i.e. all trials with silence involved were excluded leaving 56 pairs), and ‘onset’ trials (those containing a transition from silence to speech only, leaving 8 pairs; i.e. providing a P1-N1-P2 onset complex). Trials containing a transition from sound to silence were excluded from analysis. For both of these groups, the scalp distributions showed that the averaged response across all pairs was most prominent in fronto-central locations, with the largest response being observed at FCz (Figure 2.2). Response magnitude was highest for the onset trials in quiet. Figure 2.2 also shows the averaged response at FCz across all subjects for the three noise conditions, in quiet (blue), at +4 dB SNR (black), and at -3 dB SNR (red), for both groups of trials. In both trial groups, all three peaks (P1, N1, and P2) were largest in the quiet condition and smallest in the -3 dB SNR condition. The latency of all three peaks also appeared to increase with increasing noise level.

To estimate the peak amplitude for each participant at each noise level, the maximum (P1 and P2) and minimum (N1) points within windows taken from the relative peaks in the grand mean response (Figure 2.2) were identified. The peak amplitude was estimated as the average amplitude of the waveform from a 7 ms window around that maximum/minimum point (i.e. 3.5 ms either side of the peak).

The peak latencies were estimated by taking the time from the stimulus change onset to the maximum/minimum points used to calculate the peak amplitudes.

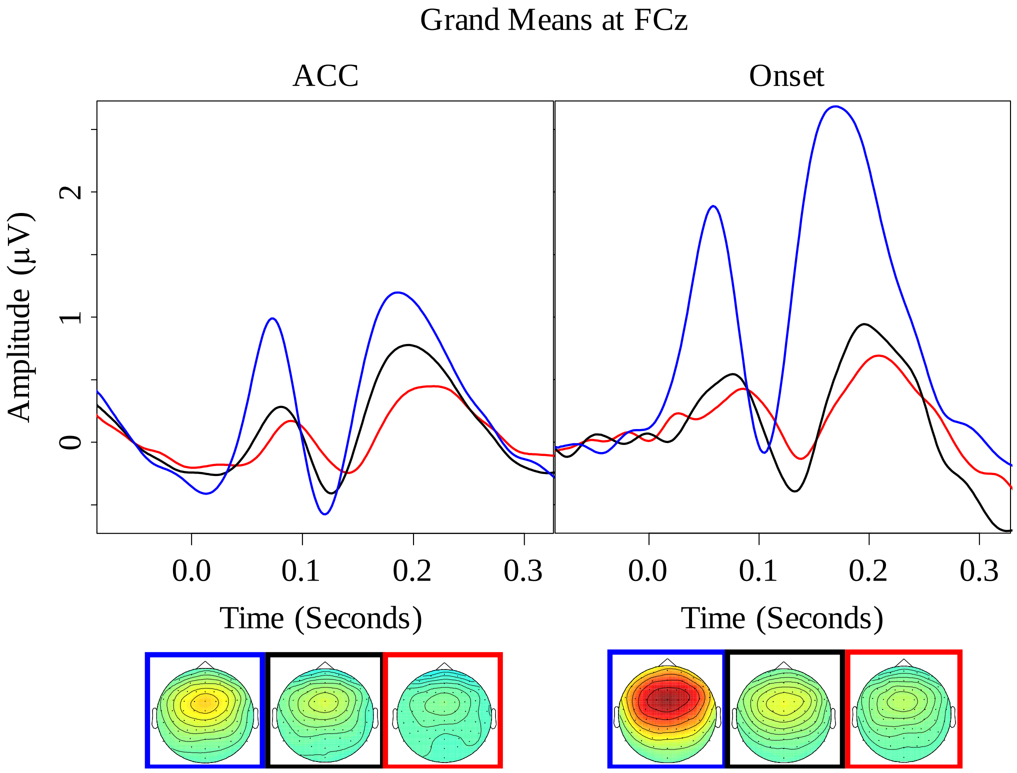

Figure 2.2 Top panels: Grand mean responses at FCz, for ACC (56 pairs) and onset (8 pairs) in quiet (blue), +4 dB SNR (black) and -3 dB SNR (red). The stimulus onset is at 0 s. Bottom panels: Scalp distributions were calculated for P2 (150-250 ms), and the colour range used in all plots ranges from -1.4 to +1.8 µV. Plot outlines denote the noise condition (as for top panels).

To investigate the effect of noise on the ACC and P1-N1-P2 onset responses, a linear mixed models analysis was run. By-subject and by-pair were entered as random intercepts, with the peak amplitude as the dependent measure. P1, N1, and P2 components were each analysed in separate models. For each model, SNR (quiet, +4, and -3 dB SNR), trial type (ACC or onset), and their interaction were fixed factors. For the P1, there were main effects for both SNR and trial type ($\chi^2(2)$ =

506.68, *p* < .001, and $\chi^2(1) = 14.81$, *p* < .001, respectively), and a significant interaction between SNR and trial type ($\chi^2(2) = 24.99$, *p* < .001). For the N1, there was a main effect of SNR ($\chi^2(2) = 47.88$, *p* < .001), and a significant interaction between SNR and trial type ($\chi^2(2) = 60.15$, *p* < .001). For the P2, there was also a main effect of SNR ($\chi^2(2) = 371.96$, *p* < .001), and a significant interaction between SNR and trial type ($\chi^2(2) = 54.85$, *p* < .001). These main effects demonstrate that the response amplitude of all three peaks decreased with decreasing SNR. The main effect of trial type for the P1 only suggests that only this peak showed a distinction between onset and ACC trials, but the significant interaction between SNR and trial type for all three peaks suggests that the SNR affected responses differentially between trial groups.

A Tukey HSD-adjusted pair wise comparison showed that for all three peaks in the ACC trials, the effect of noise was graded. Peak amplitudes were largest in quiet, smallest at -3 dB SNR, and intermediate at +4 dB SNR (*p* < .01 for all comparisons). For the onset trials, peak amplitudes (P1, N1, and P2) were larger in quiet than in +4 or -3 dB SNR (*p* < .01 for all comparisons). However, there was no difference in the peak amplitudes between the SNRs of +4 or -3 dB (P1: *p* = .050; N1: *p* = .061; P2: *p* = .14). The effect of noise on peak amplitudes in both ACC and onset trials is visualised in the top row of Figure 2.3. The figure shows that the reduction in ACC magnitude was graded by noise condition. In contrast, for the onset trials, peak amplitudes were reduced in the presence of noise (barring the N1 component), but were less differentiated by SNR (i.e. +4 or -3 dB SNR).

In quiet, the P1 and P2 amplitudes were larger in the onset trials than in the ACC trials (*p* < .01 for both comparisons). In contrast, the N1 amplitude was smaller in the onset trials than in the ACC trials (*p* < .01). From figures 2.2 and 2.3, one may speculate that the smaller N1 in the onset trials likely reflects the larger P1 and P2 amplitudes in those trials having a global effect on the entire response complex, making it more “positive” overall. At +4 dB SNR, only the P1 remained larger in the onset trials compared to the ACC trials ($p < .05$), as the N1 and P2 amplitudes were not significantly different between trial groups (N1: $p = .78$; P2: $p = .70$). At -3 dB SNR there was no difference in any of the peak amplitudes between the ACC trials and the onset trials (P1: $p = .12$; N1: $p = .78$; P2: $p = .70$).

A linear mixed model analysis was then run to compare peak latencies. As for the amplitude analysis, by-subject and by-pair were random intercepts, and the peak latency was the dependent measure. The P1, N1, and P2 components were analysed in separate models. The SNR (quiet, +4 and -3 dB SNR), trial type (ACC or onset), and their interaction were fixed factors. In all three models there was a main effect of SNR (P1: $\chi^2(2) = 165.08$, $p < .001$; N1: $\chi^2(2) = 699.34$, $p < .001$; P2: $\chi^2(2) = 5220.80$, $p < .001$), but there were no main effects of trial type, nor any interactions between SNR and trial type ($p > .05$ in all cases). These findings indicate that the noise had a significant influence on the response latency (i.e. latency increased with noise), and that this effect was not significantly different between the ACC and the onset trials. Response latencies are plotted in the bottom panels of Figure 2.3.

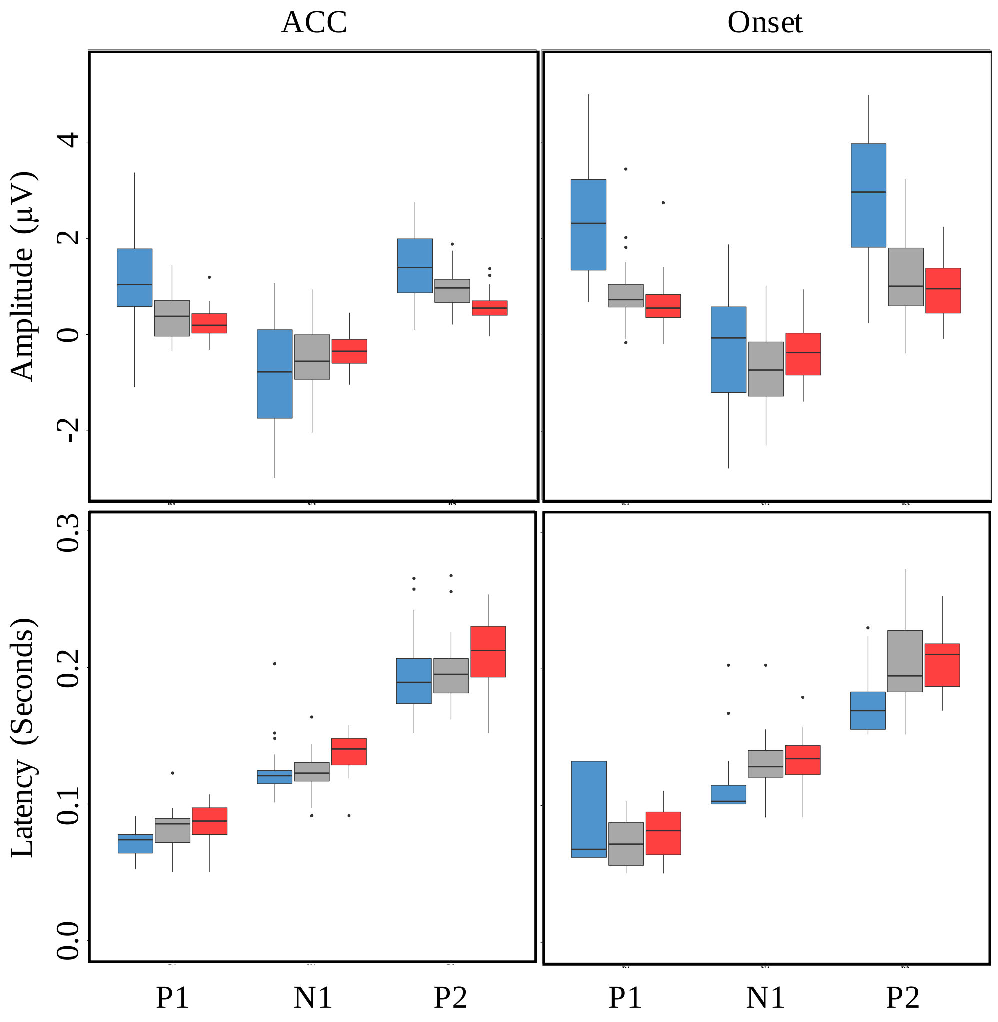

Figure 2.3 Boxplots showing the P1, N1, and P2 peak amplitudes in the top row, and peak latencies in the bottom row. The different SNR conditions are plotted in different colour bars: in quiet (blue), at +4 dB SNR (grey) and at -3 dB SNR (red). All data are from FCz, and averaged across either ACC trials (left) or onset trials (right).

Comparison of the ACC and onset to SIN performance

For the following analysis, behavioural SIN ability for each participant was compared to their CAEP responses. In Experiment 1, each participant performed a sentence recognition task at three different noise levels (-3, -6, and -9 dB SNR). The sentence recognition results at -3 dB SNR were close to 100% and showed a ceiling effect. The results at -9 dB SNR were low, and overall reflected poor recognition.

Therefore, performance at -6 dB SNR was used as the measure of SIN ability for the following analysis.

The latencies of the peaks were difficult to measure reliably for individual stimulus pairs with low-amplitude responses, particularly at low SNRs. As such, only peak amplitudes are considered in this section. A linear mixed models analysis was used to investigate the effect of SIN performance on the peak amplitudes of the ACC and onset responses. Participants and stimulus-pair were random intercepts, and the peak amplitude was the dependent measure (P1, N1, and P2 in separate models). SNR (quiet, +4, and -3 dB SNR), SIN score, trial type (ACC or onset), and their interactions were fixed factors. The main effects of trial type and interactions between trial type and SNR were reported in the previous analysis and so are not reported again here.

For P1, there was a main effect of SNR ($\chi^2(2) = 512.41$, $p < .001$), but no main effect of SIN score ($\chi^2(1) = 3.12$, $p = .077$). There was a significant interaction between SNR and SIN score ($\chi^2(2) = 59.70$, $p < .001$), but there was no significant interaction between SIN score and trial type ($\chi^2(1) = 1.17$, $p = .28$), or between SIN score, SNR, and trial type ($\chi^2(2) = 1.03$, $p = .60$).

For N1, there was a main effect of SNR ($\chi^2(2) = 48.06$, $p < .001$), but no main effect of SIN score ($\chi^2(1) = 0.63$, $p = .43$). There was a significant interaction between SNR and SIN score ($\chi^2(2) = 19.12$, $p < .001$), but there was no significant interaction between SIN score and trial type ($\chi^2(1) = 1.15$, $p = .28$), or between SIN score, SNR, and trial type ($\chi^2(2) = 0.067$, $p = .97$).

For P2, there was a main effect of SNR ($\chi^2(2) = 372.16$, $p < .001$), but no main effect of SIN score ($\chi^2(1) = 0.88$, $p = .35$), and there was no significant interaction between SNR and SIN score ($\chi^2(2) = 2.82$, $p = .24$). Furthermore, there was no significant interaction between SIN score and trial type ($\chi^2(1) = 0.15$, $p = .70$), or between SIN score, SNR, and trial type ($\chi^2(2) = 0.012$, $p = .99$).

The lack of any significant interactions between trial type and SIN score suggests that the relationships between the peak amplitudes and SIN score were not statistically different between the ACC and the onset trials. The interaction between SNR and SIN score for the P1 and N1 peaks is explored in Figure 2.4. The figure shows that participants with lower SIN scores showed a larger P1 response, particularly in quiet (left panel). When compared to quiet, all participants showed a reduced P1 in the presence of noise, nevertheless worse performers continued to exhibit larger amplitudes than better performers. For the N1, an opposite pattern is observed (Figure 2.4, right panel). Participants with lower SIN scores showed smaller N1 responses, particularly in quiet. For the quieter noise condition (+4 dB SNR), N1 magnitude also increased with better behavioural performance, however this effect was not as pronounced as in quiet. For the louder noise condition (-3 dB SNR), N1 magnitude varied little with SIN performance in the behavioural task.

To explore the interaction further, Pearson correlations were used to characterise the relationship between SIN scores and the P1 and N1 peak amplitudes for each of the SNRs. Significance was tested using linear mixed-models with by-subject and by-pair random intercepts, the peak amplitude as the dependent measure, and SIN score as the independent measure. A different model was calculated for each SNR. P1 in quiet was significantly correlated with SIN score ($r = -0.37$, $\chi^2(1) = 4.69$, p = .030), but not at +4 dB SNR ($r = -0.13$, $\chi^2(1) = 2.00$, $p = .16$), nor at -3 dB SNR ($r = -0.25$, $\chi^2(1) = 0.50$, $p = .48$). N1 amplitude was not significantly correlated with SIN score at any SNR, however, the relationship in quiet was stronger than in noise (quiet: r = -0.21, $\chi^2(1)$ = 1.35, p = .25; +4 dB SNR: r = -0.0047, $\chi^2(1)$ = 0.39, p = .53; -3 dB SNR: r = -0.12, $\chi^2(1)$ = 0.0006, p = .98).

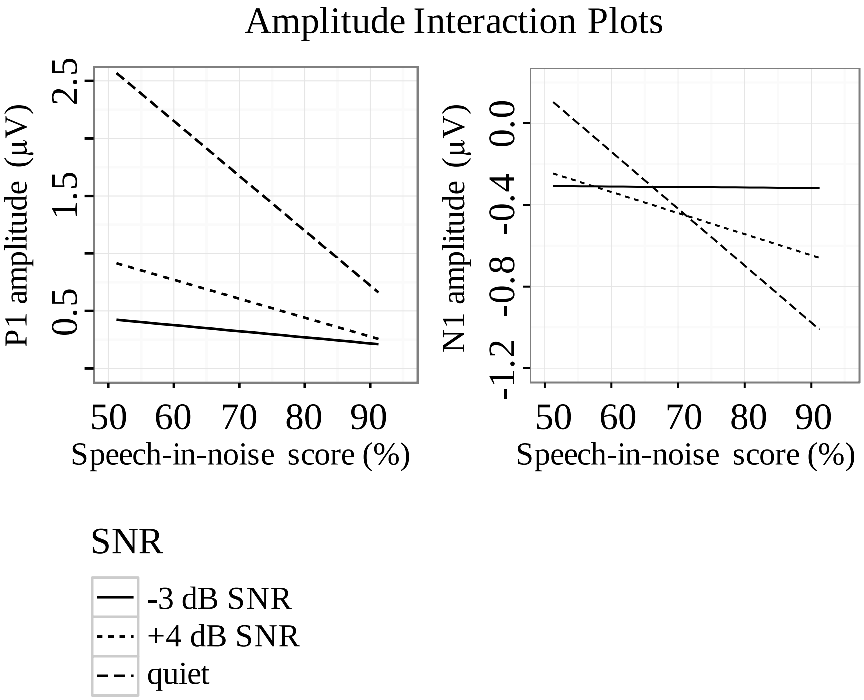

Figure 2.4 Plots showing the interactions between SNR and SIN score for the P1 (left), and the N1 (right), by showing linear regression lines for SIN score versus N1/P1 peak amplitude at each SNR.

2.4 Discussion

The aim of this study was to investigate the effect of noise on the ACC and the P1-N1-P2 onset complex to a broad range of speech stimuli in normal-hearing listeners. Additionally, the properties of these responses were compared to speech-in-noise (SIN) performance in a sentence recognition task. There is good evidence that SIN performance is challenging for certain listeners, such as the hearing impaired and the elderly (e.g. Plomp, 1978; Dubno et al. 1984; Rogers et al. 2006).

Additionally, there is growing interest in variable performance in people with seemingly normal audiograms. Whilst the reasons for such variability remain unclear, recent evidence suggests certain peripheral hearing damage may affect hearing performance in certain situations but not audiometric thresholds (hidden hearing loss; Schaette & McAlpine 2011; Plack et al. 2014). On the basis of such considerations, there is renewed interest in characterising and understanding SIN performance, especially for clinical benefit. Whilst previous studies have characterised how early cortical responses to speech are affected by noise (e.g. Martin & Stapells 2005; Parbery-Clark et al. 2011; Billings et al. 2013), there has been no direct comparison between the ACC to speech and behavioural SIN performance. The ACC to changing speech phonemes may be a good indicator of SIN performance, as it relies on the discrimination of speech-relevant spectral fluctuations. This was assessed in the current experiment. The paradigm used in this experiment represented a time-efficient measure of the ACC, and allowed for a robust ACC response to be observed in all individuals.

In summary, as the SNR decreased, for both the ACC and the onset trials the amplitude of the response decreased and the latencies increased. However, the ACC was more sensitive to the change in the level of presented noise, as its amplitude reduced in a graded manner as the SNR decreased. The amplitude of the onset complex decreased in the presence of noise, but did not decrease further with further SNR reduction. A second aim of the study was to explore how onset and ACC responses predict behavioural SIN performance. When comparing performance on a sentence recognition task in noise to the CAEP responses, poorer performers showed smaller N1 responses and larger P1 responses. This pattern was observed most strongly for CAEPs to stimuli presented in quiet. This observed relationship between peak amplitudes and behavioural performance was the same for both the ACC and the P1-N1-P2 complex responses.

The ACC shows a graded response to the addition of noise

When noise was added to the signal, the amplitudes of both the ACC and the onset responses were reduced. This is consistent with previous studies of cortical responses to SIN (Kaplan-Neeman et al. 2006; Kim et al. 2012; Papesh et al. 2015). The current results suggest that the ACC amplitude was more sensitive to changes in SNR, whereas the onset response was only sensitive only to the addition of noise, irrespective of SNR. This comparison has not been made in previous literature, but some studies have shown that the P1-N1-P2 complex does not show a graded decrease in amplitude with decreasing SNR (e.g. McCullagh et al. 2012; Papesh et al. 2015). This could be an indication of the onset response reflecting primarily the obligatory detection of an incoming sound, and so may be relatively invariant for spectral differences in detected stimuli. In contrast, the ACC to changing speech tokens, as measured here, relies on the detection of spectral changes in a speech stimulus. It is likely that the salience of the spectral change is reduced with decreasing SNR; for example, we observed a graded deterioration in behavioural performance as SNR is decreased. To this end, the ACC measure tested here may provide a more meaningful index of SIN performance than the P1-N1-P2 complex; as the former relies on the *discrimination* of speech-relevant spectral changes whereas the latter is indicative of stimuli detection only (i.e. the onset of an unmasked signal). Alongside some distributional differences between the ACC and the onset P1-N1-P2 complex (Jones et al. 1998), these behavioural differences seen in the current study suggest that there are some distinct characteristics between the two, despite them sharing the same morphology. However, whether they are distinct responses or variants of the same response remains uncertain.

Furthermore, the graded ACC response in noise observed currently also included the P1 amplitude of the ACC. The P1 has previously been shown to be insensitive to the magnitude of acoustic differences (Čeponienė et al. 2005). However, the current results suggest that the peak is sensitive to changes in the SNR of the stimulus.

P1 is larger for poorer SIN performers

Typically, the P1 is small in adults and so is often excluded from analyses (Kaplan-Neeman et al. 2006; Parbery-Clark et al. 2011; McCullagh et al. 2012). A benefit of the current design is that responses to multiple acoustic changes can be recorded in a short space of time. This factor combined with the relatively large sample size of the current study meant that the P1 could be reliably measured. As such, the P1 observed in the current study can be considered robust, and can be analysed with confidence. The results show that poorer performers in the sentence task showed larger P1 responses (particularly in quiet). Previous findings have suggested that the N1 is the most reliable predictor of SIN ability (Parbery-Clark et al. 2011; Billings et al. 2013). The current finding, that worse SIN performance is related to a *larger* P1, would initially appear counter intuitive to these previous observations. However, the current P1 findings are partially supported by a study by Billings et al. (2013). These authors tested 15 normal-hearing adults on their SIN ability using IEEE sentences (Rothauser et al. 1969) in speech-shaped noise. They also measured P1-N1-P2 responses to a /ba/ stimulus presented in speech-shaped noise at varying SNRs and varying signal levels. They found that for behavioural performance with a 60 dB SPL signal, the best predictor of the individual score was the N1 amplitude at +5 dB SNR (r = 0.77). However, the next best predictor was the P1 amplitude at +35 dB SNR. Here, those that performed worse on the behavioural task showed a *larger* P1 amplitude (r = 0.74). This is consistent with the current results.

The enhancement of certain aspects of cortical responses has been seen with musicians (Lee et al. 2009; Strait et al. 2009; Zendel et al. 2015). Zendel et al. (2015) found that the P1 was enhanced in musicians compared to non-musicians when actively listening to speech in a multi-talker babble. Additionally, the P1 peak was also delayed in the musically experienced group. The musicians also performed better in a behavioural SIN task than the non-musicians. The authors theorised that musicians rely more heavily on acoustic cues when listening to SIN than non-musicians. This P1 enhancement and delay during active listening could be related to the current results. More specifically, poorer SIN listeners may require more attention to perform compared to those with better SIN processing abilities. It could be that pre-conscious attentional processes affect the top-down modulation of responses, even though participants were not instructed to actively attend the stimuli (e.g. Tervaniemi et al. 2009; Anderson et al. 2010). Whereas musicians outperform non-musicians and have an enhanced and delayed P1, perhaps non-musicians who weigh more on acoustic cues do worse in SIN tasks. Their strategy for listening in quiet does not withstand the presence of noise.

A related possibility is that increased P1 amplitude with poorer SIN performance could be indicative of listening effort. Rao et al. (2010) demonstrated that the P1 was enhanced in a task that involved greater listening effort. The authors used Garner’s interference paradigm (Garner 1974), where participants classified the pitch of either a pure tone or a filtered noise band, presented alongside compound sounds containing both a pure tone and a filtered noise band. The frequency of the pure tone was either ‘high’ or ‘low’, and the frequency band of the filtered noise was also either ‘high’ or ‘low’, yielding four compound sound combinations. Participants were asked to classify either the pitch of the pure tone or of the filtered noise band as high or low in blocks whilst the pitch of the non-target was either kept constant or varied (i.e. constant vs. varied background condition). Furthermore, the SNR was kept at +15 dB in favour of the pure tone. The SNR meant that the classification of the pitch of the filtered noise band was a more difficult task (and required more cognitive effort) than the classification of the pitch of the pure tone, and the varied background condition was more difficult (and required more cognitive effort) than the constant background condition. In this paradigm, the cognitive effort needed to do the tasks was different, whilst the compound auditory stimuli were physically identical. The results showed increased P1 amplitudes for the varying background condition and for the pitch classification of the filtered noise bands, therefore the larger P1 was attributed to increased cognitive effort. In the current study, it is plausible that the lower SIN performers required greater cognitive effort than the higher performers because they find listening in noise challenging. However, as participants were not instructed to attend to the sounds, this interpretation would assume that participants were either paying some level of attention to the sound stimuli, or that cognitive effort was obligatorily allocated to the speech stimuli. Additionally, P1 variations were observed when the speech stimuli were presented in quiet, so one could speculate that variation in cognitive listening effort is extended to listening in quiet also. Of the three peaks, the N1 is most often associated with an increase in amplitude with selective attention (Hillyard et al. 1973; Näätänen & Picton 1987; Woldorf & Hillyard 1991), however the results of this study suggest that the P1 may also be modulated by certain aspects of subconscious attention or effort. It is possible that these effects are not observed in studies where attention is modulated explicitly because they are masked by the broad processing negativity that occurs with attention (Hansen & Hillyard 1980; Näätänen & Picton 1987; Näätänen 1990; Alho et al. 1994). Further study would be required to directly assess the cognitive effort associated with the current task.

It is also possible that the greater amplitude of the P1 in quiet in poorer SIN performers is indicative of reduced neural efficiency. The better SIN performers could require fewer neural resources in order to achieve accurate speech perception than the poor performers. This theory has been used in the past to explain correlations between cortical responses and IQ scores (Robaey et al. 1995; Zhang et al. 2006), and to explain correlations between N2 amplitude and SIN performance (Anderson et al. 2010). Anderson et al. (2010) tested 32 children, and compared HINT (Hearing in Noise Test: Nilsson et al. 1994) scores to P1-N1-P2-N2 response characteristics to /da/. Compared to speech presented in quiet, the P1 decreased in amplitude when the noise was added (multi-talker babble, +10 dB SNR), but it wasn’t effected by HINT performance. However, when they split the data into to the ‘good’ and ‘poor’ SIN performers, they found that those who performed worse on the HINT task had a *larger N2* response. They suggest that the top SIN performers may have been recruiting fewer neural resources due to greater neural efficiency and that this is shown as a smaller N2 peak. In the current study, the smaller P1 in the higher SIN performers could reflect greater neural efficiency than the lower SIN performers. However, in the current study, the N1 peak was reduced in poorer performers, suggesting that if this interpretation has merit, any differences in neural efficiency would affect response components in a non-linear manner. Again, further study is required to confirm any such effects.

N1 is smaller for poorer SIN performers

The N1 is thought to reflect the encoding of time-varying aspects of a sound, such as amplitude, spectral, and voice onset time changes (Sharma & Dorman 1999; Martin & Boothroyd 2000; Sharma et al. 2000; Hoonhorst et al. 2009). The amplitude of the N1 varies with the magnitude of the acoustic change (Näätänen & Picton 1987; Picton 1990; Steinschneider & Dunn 2002). Accurate cortical representations of the acoustic information in a speech signal are important for speech perception in noise (Anderson et al 2010; Parbery-Clark et al. 2011). Therefore, it appears that the better listeners were more sensitive to the acoustic change, and so showed a larger N1. In this way the larger N1 is indicative of better cortical representation of the acoustic changes in the better performers (Parbery-Clark et al. 2011).

Parbery-Clark et al. (2011) tested 22 normal-hearing adults on their SIN ability using the HINT test, and measured subcortical (ABRs) and cortical (N1-P2) responses to the speech token /da/ in a multi-talker babble of varying level. They found that those with better ABRs in noise and better HINT scores had a larger N1s. They suggest that those with a better subcortical resilience to noise have a larger N1. In the current study, it is possible that the individuals who were better at the SIN task, also have a greater resilience to noise, which is indexed by a larger N1 amplitude.

The N1 is thought to increase in amplitude with selective attention (Hillyard et al. 1973; Näätänen & Picton 1987; Woldorf & Hillyard 1991). As such, the better SIN performers may have shown a larger N1 because they paid more attention subconsciously to the speech (that is, despite not being instructed to actively attend the stimuli). Whilst this may contradict the prior argument that the P1 was larger in poorer SIN performers because they subconsciously paid greater attention to, or used more cognitive effort when listening to speech, early cortical peaks do not exist in isolation. Therefore, it is possible that any attentional effects influencing the N1 may also modulate the amplitude of the P1. For example, whilst the better performers exhibit a smaller P1, this may reflect a combined effect of a broad negativity occurring with attention (Hansen & Hillyard 1980; Näätänen & Picton 1987), making both the N1 and P1 more negative. However, this overlapping broad processing negativity usually effects the P2 (Näätänen & Michie 1979; Näätänen & Picton 1987; Michie et al. 1990, 1993) and there was no effect of SIN ability on P2 amplitude in this study. Therefore, in this case, if there was an attentional effect on the N1, it is perhaps more likely to be isolated to this component only, as opposed to being a broad processing negativity effecting the whole response.

2.4.1 Conclusion

Whilst the relationship between the ACC and P1-N1-P2 onset complex and SIN ability is not fully understood, it has been consistently observed that SIN ability correlates with certain components of these evoked responses (Cunningham et al. 2001; Anderson et al. 2010; Parbery-Clark et al. 2011; Billings et al. 2013). The current results are therefore consistent with these previous findings. This gives promise that CAEPS have utility as tools to predict SIN ability for normal hearing listeners. It would be of further interest to develop CAEP measures able to predict SIN performance for those showing SIN impairment (e.g. individuals with hearing or learning impairment, or elderly listeners). Better understanding of the variability seen in SIN ability provides opportunity for improved prediction of SIN problems, improved diagnosis of conditions that effect SIN processes, and improved rehabilitation for individuals with SIN problems. An objective test for SIN ability would be clinically valuable for diagnosis and to aid measurement of rehabilitation outcomes. Further research into the use of the ACC and the P1-N1-P2 complex in noise in clinical populations (e.g. hearing aid users) and the elderly is needed to further test the potential of these responses as tools to predict SIN performance.

This study suggests that the underlying mechanisms of the ACC provide a more detailed representation of SIN than the P1-N1-P2 onset response. Furthermore, notably, this study supports previous findings that noise affects the various cortical peaks in a differential manner. This supports the growing acceptance that early cortical responses are not purely obligatory. Results suggest that the P1 may reflect aspects of greater pre-conscious attention and cognitive effort, and lower neural efficiency in poorer SIN performers. Whereas, the N1 reflects a greater resilience to noise and a better cortical representation of the acoustic information in the speech signal by good SIN performers. There is growing evidence for a relationship between higher level processing and obligatory cortical activity, suggesting that there may be top-down influences on these early cortical auditory processes.

3 Cortical Measures of Speech Perception in Cochlear Implant Users

3.1 Introduction

Chapter 2 investigated the effect of noise on CAEPs in normal hearing (NH) listeners and how this related to their speech-in-noise (SIN) ability. The speech stimuli were made harder to perceive by adding background speech-shaped noise. However, all listeners displayed normal hearing, and so individual differences in their SIN performance cannot be explained by peripheral factors (at least in terms of audiometric thresholds). This next chapter focuses on CAEPs in cochlear implant (CI) users. Here, both peripheral and central processes are likely to be much more variable amongst subjects than for a NH population. For a group of CI participants, Experiment 3 measured behavioural speech performance, and Experiment 4 measured the ACC response using a paradigm similar to that presented in Chapter 2. Finally, Experiment 5 measured entrainment of low frequency neural oscillations to continuous speech in the same CI participants. This initial introduction considers both the factors that affect CI performance, as well as research into CAEP measures of CI performance. After Experiments 3 and 4 are presented, neural entrainment (or phase coherence) measures are introduced, and then Experiment 5 is presented. Finally, a general discussion summarises the findings from all three experiments.

A CI is a surgically implanted auditory prosthesis, designed to restore hearing to the profoundly deaf. A common cause of hearing loss is the loss of cochlea hair cells; from aging, noise exposure, certain medical or genetic conditions, or exposure to ototoxic substances. CIs bypass absent or damaged hair cells by innovating the auditory nerve directly through electrical stimulation. Modern implanted devices comprise of a number of electrically stimulating electrodes spaced along an array. The array is surgically inserted into the cochlear, placed typically in the scala tympani, and optimally spanning 18-26 mm of the length of cochlea (depending of the length of the implanted array, and the success of the implantation). Acoustic information is received by an external microphone, and is converted by a behind-the-ear speech processor into a corresponding sequence of electrode activation. These stimulation parameters are transmitted to a subcutaneous receiver/stimulator via electromagnetic induction. The receiver/stimulator then activates the electrode array. The delivered electrical pulses from the electrodes provide the input to the auditory nerve that is broadly analogous to acoustic hearing (see Zeng 2004 for a general review). For example, the electrodes spaced along the array are designed to replicate the place-of-excitation based encoding of frequency (Wilson et al. 2011). However, the pattern of auditory nerve activity elicited from CI stimulation is markedly different from the normally functioning cochlear. The broad spread of electrical current from electrical stimulation results in a much more restricted frequency selectivity. Even though a device may have up to 22 electrode channels, the effective number of independent frequency channels may be markedly less, due to current spread (e.g. Friesen et al. 2001). Additionally, rapid electrical stimulation typically saturates auditory nerve firing, and so whilst implants are able to provide rapid pulse trains, users are typically insensitive to temporal cues above approximately 300-500 Hz (Nelson et al. 1995; Collins et al. 1997; Zeng et al. 2002). There is also typically a narrow dynamic range between the minimum current required for threshold detection and the maximum current of comfortable listening. As such, CI users are typically restricted to a much narrower dynamic range of loudness, including the number of discernible loudness steps (Nelson et al. 1996; Zeng et al. 1998). Despite such differences between CI and NH listening, CI users may achieve a good level of hearing performance from their device, with some reaching near-normal levels of speech recognition in quiet (Wilson & Dorman 2007). Whilst it is possible for CIs to provide enough information for some users to achieve such impressive outcomes (Wilson et al. 2011), it should be noted that many users struggle with speech understanding in everyday noisy environments (Fu et al. 1998), music appreciation (McDermott 2004), understanding prosody (Meister et al. 2009), and speaker characteristics (such as gender; Fu et al. 2004), as well as from issues arising from increased listening effort and fatigue (with some variability between unilateral and bilateral implantation; Hughes & Galvin 2013). As such, there is great research interest in improving many aspects of CI performance.

The exact output of the electrode array in response to an incoming sound is determined by the parameters of the speech processor, which extracts specific acoustic features of the sound. To reflect tonotopy in the cochlear, implants implement a bank of filters to divide sound signals into different frequency bands, with low to high frequency bands allocated to apical to basal electrodes. However, the processing strategies to extract, encode, and deliver stimulation vary between devices and settings. Whilst early processing strategies aimed to extract specific features from sounds (such as speech formants), modern strategies attempt to represent the envelopes of incoming sounds without cue-specific weighting (see Zeng 2004; McDermott 2004; Wilson & Dorman 2008 for general reviews). Continuous-Interleaved Sampling (CIS; Wilson et al. 1991) was designed to extract and deliver temporal envelope cues. Here, the envelope of each frequency band of the signal is extracted and compressed (in order to fit the large dynamic range of the acoustic world into the small dynamic range of a CI). The variations in the envelopes of each frequency band are then represented in corresponding electrodes using modulated trains of biphasic electrical pulses (Wilson & Dorman 2008). To avoid electrode interactions from stimulating one or more electrodes at the same time, pulse trains on different electrodes are interleaved, so only a single electrode is active at any given time. HiResolution (HiRes; Koch et al. 2004) is similar to CIS but with higher stimulation rates and higher cut-off frequencies used for envelope measurement, and a form of this (HiRes120) is the default for Advanced Bionic implants. MED-EL implants use a form of Fine Structure Processing (FSP; Hochmair et al. 2006) where groups of pulses (or can be single) are presented at the zero crossings, as compared to CIS where single pulses are presented at peaks of the envelope (Wilson et al. 1991). In other strategies, such as n-of-m (Wilson et al. 1988), Spectral Peak (SPEAK; Skinner et al. 1994), and Advanced Combination Encoder (ACE; Kiefer et al. 2001; ACE is the default in Cochlear CIs), only those filter bands with the highest energy (n) of all the bands (m) are selected to stimulate their corresponding electrodes. These strategies are similar to CIS apart from the channel selection. The idea is that by selecting only the channels with the highest energy, low amplitude channels are not stimulated, which reduces masking and interference across electrodes. In n-of-m and ACE, n is fixed; whereas in SPEAK n varies depending on the input. Furthermore, SPEAK also has a lower stimulation rate than n-of-m and ACE. The default programming strategy for each CI manufacturer is nearly always used for fittings. However, other strategies are available in each CI and can be used for an individual if necessary at the discretion of the audiologist (Wilson & Dorman 2008).

In addition to the method of sound processing, the manner in which charged-balanced pulses are delivered can be varied; either monopolar (in which the return electrode is typically located outside the cochlear) or bipolar (the return electrode is an adjacent intra-cochlear electrode). With monopolar stimulation the current has to travel to the return electrode outside the cochlear, and this distance causes a large spread of excitation, leading to poor spectral resolution. Bipolar stimulation attempts to reduce the distance the current needs to travel by having the return electrode situated inside the cochlear, adjacent to the stimulating electrode. However, all current CIs default to monopolar stimulation, as in practice it produces performance that is just as good as bipolar, but requires less current, and so battery power. To further improve performance, tripolar stimulation (only in the research domain) has been investigated. Tripolar stimulation is a variation of the bipolar mode in which two intra-cochlear return electrodes are used to limit spread on either side of the stimulating electrode. However, tripolar stimulation requires a lot of current and it cannot always support the whole dynamic range or full loudness growth of the CI (Wu & Luo 2013). In order to improve this, research is ongoing into partial tripolar stimulation where only part of the current is returned to the adjacent electrodes and the rest is returned to a ground electrode on the outside of the cochlear (Wu & Luo 2013).

The speech processing strategy and the parameters of this strategy are fitted to each individual user by clinicians. The clinician may also identify non-functioning or poorly functioning electrodes (i.e. those positioned outside of the cochlear) for deactivation. Whilst research studies can provide insight into theoretically optimum stimulation strategies (e.g. Bierer & Middlebrooks 2004) and electrode deactivation criteria (e.g. Bierer 2010), in reality the fitting parameters that achieve the best speech performance and listening comfort for an individual are hard to predict. As such, and given the limited clinical resources for individual fitting, there is considerable interest in understanding and improving the manner in which CIs are fitted to an individual.

Individual CI users have a wide range of speech recognition outcomes, with some achieving word recognition scores of over 95% after a year of being fitted, whereas others struggle to improve beyond 20% (Helms et al. 1997; Wilson & Dorman 2007). The reason for these performance differences between seemingly similar patients is not fully understood. A CI user’s auditory pathway may be degraded if they had a long period of auditory deprivation prior to implantation (Shepherd & Hardie 2001; Shepherd et al. 2006), or a late age of implantation (Buckley & Tobey 2011; Giraud & Lee 2007; Lee et al. 2001; Sharma et al. 2002). Auditory deprivation may result in cross-modal plasticity in the cortex. That is, processing in the auditory cortex may become reassigned to other modalities, such as vision (e.g. Lee et al. 2001; Kral & Eggermont 2007). Such factors mean that some patients have greater access to speech processes in the brain than others (Wilson et al. 2011).

Pre-lingually deafened CI users suffer from the effects of auditory deprivation, however, unlike post-lingually deafened CI users, they also lack neural representations of speech from before they lost their hearing. If infants do not receive language input in the first year of life their language development (e.g. babbling) begins to delay, and the longer they do not receive language exposure, the more delayed they become (Purves et al. 2001). By the time they hit puberty this delay may be irreversible (i.e. case studies of feral children; e.g. Curtiss et al. 1974). From an auditory perspective, this sensitive period in early life stems from plasticity in the auditory cortex and changes in this plasticity as the cortex develops (Kral 2013). Therefore, even if a pre-lingually deafened adult is able to receive auditory stimulation via a CI, they may struggle to use this information for communication as they do not have the necessary experience in early life, when the auditory cortex and linguistic systems were developing. Thus, pre-lingually deafened adults generally do not perform as well as post-lingually deafened adults when they receive CIs. Although post-lingually deafened CI users may still have access to neural representations of speech from before they lost their hearing, duration of deafness is still a strong predictor of implant success (van Dijk et al. 1999; Kelly et al. 2005; Sandmann et al. 2015). However, patients with identical durations of deafness can still display great variability in performance (Wilson et al. 2011).

3.1.1 Auditory Evoked Potentials in Cochlear Implant Users

This variability of CI outcomes is a current on-going topic of research. In this field, auditory evoked potentials (AEPs) are useful for those in the CI population who are unable to perform behavioural methods of auditory performance as are traditionally used in audiology, such as speech intelligibility tests (i.e. children and non-verbal adults; Scheperle & Abbas 2015b). AEPs have been successfully measured in both paediatric and adult CI populations (e.g. Hoppe et al. 2010; Alvarenga et al. 2012; He et al. 2014; Abbas & Brown 2015). Currently, the electronically evoked compound action potential (ECAP) is used to measure whether or not the stimulation from the electrodes elicits a response in the auditory nerve

(Brown et al. 2000; Hughes et al. 2000a, 2000b). Similarly, the electrically-evoked ABR (a subcortical measure) can be used to broadly estimate behavioural thresholds for signal detection (e.g. Brown et al. 2000). Such information is important for assessing the functioning of the electrode array, as well as the auditory nerve and subcortical structures. However, these measures are not informative of limitations at higher levels of processing, and so can only explain certain peripheral factors that influence speech perceptual performance. For successful speech intelligibility, users need to accurately encode acoustic details, such as spectral information and the temporal patterns conveyed by the amplitude fluctuations in speech. However, currently there are no electrophysiological measures that assess the perception of these characteristics of speech that are used routinely in clinical practice (Scheperle & Abbas 2015b). Furthermore, whilst it is important that users receive adequate acoustic detail from the periphery, it has been proposed that individual differences along the entire auditory pathway mean that there could be benefit to a more top down, cognitive approach to CI programming and fitting, rather than the current bottom up technique based primarily upon signal detection (Wilson et al. 2011). To this end, the current study investigated electrophysiological measures of both isolated spectral changes in speech and entrainment of cortical neural oscillations to the amplitude envelope of continuous speech, to explore their use in predicting speech perception performance in CI users.

Objective measures that can predict variability in CI outcomes is currently a research topic of interest. Due to the implanted device, there is limited scope for MEG and MRI imaging techniques, and so EEG has become the most common measure for studying brain activity evoked from CI stimulation. Certain CAEPs have been obtained successfully in CI users, and there is some evidence of cortical measures of discrimination being predictive of behavioural performance (e.g. Hoppe et al. 2010; He et al. 2014; Abbas & Brown 2015). Furthermore, CAEPs have been successfully elicited with phonemic contrasts in speech-like stimuli (e.g. Friesen & Tremblay 2006; Martin 2007). The mismatch negativity (MMN) has been used in CI studies with tonal deviations (Kileny et al. 1997; Kelly et al. 2005; Zhang et al. 2011; Obuchi et al. 2012), musical feature deviations (Timm et al. 2014) and speech stimuli deviations (Kraus et al. 1993; Groenen et al. 1996a; Kileny et al. 1997; Salo et al. 2002). In these cases, a larger mismatch negativity (MMN) response has been consistently found to be related to better CI performance (Groenen et al. 1996a; Roman et al. 2004; Zhang et al. 2011; Turgeon et al. 2014; Timm et al. 2014). For example, Roman et al. (2004) measured the MMN to frequency-deviant acoustically presented tones in CI users, and compared the results to performance on a word discrimination task. The discrimination task involved a pair of monosyllabic French words bain (/be/), and pain (/pe/). Subjects had to identify the word and an adaptive forced-choice procedure was used to measure a threshold (the sound pressure level of the word was the tracking variable). The authors reported correlations between the MMN latency and pure-tone detection thresholds, and performance on the behavioural discrimination task. This shows that neural activity in the auditory cortex can be indicative of individual differences in CI performance. As such, the MMN has potential to be used to measure speech recognition proficiency in CI populations that cannot perform behavioural tasks of speech recognition (e.g. non-verbal populations). However, other CAEPs such as the P1-N1-P2 onset complex or the ACC have advantages over the MMN, such as greater time-efficiency (Iverson et al. 2016), larger magnitudes, and greater test-retest reliability in individuals (Martin & Boothroyd 1999; Tremblay et al. 2003), at least in NH populations.

The P1-N1-P2 onset complex has been successfully recorded in CI adults (although often the P1 is not measured in these studies; Makhdoum et al. 1998; Firszt et al. 2002; Maurer et al. 2002; Roman et al. 2004; Kim et al. 2009; Timm et al. 2012; Abbas & Brown 2015; Brown et al. 2015). Furthermore, some characteristics of the onset complex have been found to relate to speech processing ability (Firszt et al. 2002; Maurer et al. 2002; Roman et al. 2004; Kelly et al. 2005). For example, Kelly et al. (2005) found that the CI users who performed worse on the HINT speech test (Hearing in Noise Test: Nilsson et al. 1994) also showed longer P2 latencies to acoustically presented pure tones. Groenen et al. (1996b) split seven CI users into two groups, 'good' and 'moderate' performers, and found that the amplitudes and latencies of the electrically evoked N1 and P2 to tone bursts were comparable to normal in the good group, whereas they had lower amplitudes and longer latencies in the moderate group. The authors suggested that this could indicate disrupted tonotopic organisation of the auditory cortex in the moderate performers (Groenen et al. 1996b). Maurer et al. (2002) measured evoked responses to an electrode pulse train presented through a research interface. Participants were split into 'good' and 'poor' performers based on a speech recognition task. Good performers exhibited shorter P2 latencies than poor performers, and other amplitude and latency measures were not significant. They concluded that some properties of the onset complex differed between participant groups, but that differences were more apparent in brainstem measures. Alvarenga et al. 2012 measured acoustically-evoked CAEPs in children using a /ba/ speech stimulus. The study aimed to characterise the P1 component only, as this component often dominates the overall early cortical response in children. The latency of the P1 component correlated with duration of sensorineural hearing deprivation, but not with duration of CI use. When participants were split into ‘good’ and ‘poor’ performer groups, good performers displayed significantly shorter P1 latencies. Such P1-N1-P2 results give promise to the ACC also being able to predict CI performance. However, as noted previously, the ACC and P1-N1-P2 complex do not always show the same patterns (e.g. Chapter 2; Jones et al. 1998).

The ACC predominately encodes the basic stimulus properties such as frequency, intensity, and phase (Jerger & Jerger 1970; Näätänen & Picton 1987; Martin et al. 2007; Ross et al. 2007; Dimitrijevic et al. 2008), as well as more naturalistic changes, such as spectral and intensity changes in speech or speech like sounds (Ostroff et al. 1998; Martin & Boothroyd 1999, 2000). Furthermore, some aspects of the ACC can be modulated by top down cognitive processes, such as selective attention (Hillyard et al. 1973; Näätänen & Picton 1987; Woldorf & Hillyard 1991), and classification or categorisation of auditory stimuli (Garcia-Larrea et al. 1992; Novak et al. 1992). Therefore, the ACC is a measure of sensitivity to acoustic changes in speech, and cognitive speech perception processes, beyond simple signal detection. Furthermore, aspects of the ACC can potentially be used to predict SIN ability in NH listeners (Chapter 2). Therefore, this measure could conceivably be useful also for the prediction of CI performance.

Recently, there have been some successful studies of the ACC recorded in CI users using direct stimulation. Direct stimulation refers to the use of a specifically made research interface that bypasses the speech proccesor and directly controls the output of the electrode array. This means that a carefully controlled pulse train can be sent to the listener. As the exact pattern of electrode output from a speech processor is not known to the reseacher, direct stimulation is needed when presenting pulses to specific electrodes at specific amplitudes, durations, rates, etc.

The ACC elicited from direct stimulation can be reliably recorded for changes in level (Kim et al. 2009), changes in stimulating electrode with the further apart the electrodes are the larger the ACC (Brown et al. 2008; Hoppe et al. 2010; He et al. 2014), and changes of frequency location of spectral peaks within a ripple-noise stimulus (Scheperle & Abbas 2015a). Due to the spread of electrical current assoicated with electrial stimulation, the neural populations stimulated by separate electrodes often overlap (He et al. 2014). Any overlapping stimulation will reduce the discriminability between stimulation from different electrodes. As such, several studies have found that psychophysical measurement of electrode discrimination is correlated with speech perception performance (e.g. Throckmorton & Collins 1999; Henry et al. 2000), although some studies have not (e.g. Hughes & Abbas 2006; Anderson et al. 2011). In a related paradigm, the ACC can be recorded to changes in the stimulating electrode (i.e. where larger responses are indicative of greater channnel independence). Such indexes of electrode discrimination measured with the ACC have been found to be consistent with behavioural thresholds (in children with ANSD; He et al. 2014). Furthermore, the magnitude of the ACC to changes in stimulating electrode has been found to predict speech perception performance in both children and adults (He et al. 2014; Scheperle & Abbas 2015b). Scheperle & Abbas (2015b) assessed speech performance and also recorded eCAPs and ACC responses to changes in both spatial (electrode pairs) and spectral (rippled noise) stimulus changes. Subjects were tested with three distinct MAPs, each designed to manipulate the likelihood of electrode interactions. Using a mixed-model analysis which factored multiple measures for each particpant, the authors observed that electrophysiological measures were significantly related to each other and to the speech scores. Whilst the eCAP measures proved to be the best predictor of speech scores in this model, the authors proposed that a combination of peripheral (eCAP) and central (ACC) measures of performance may combine to offer a better predictor of speech performance.

The ACC recorded by direct stimulation shows potential for predicting speech performance in CI users, and this method of stimulus presentation allows for the precise control of the stimuli being sent. However, whilst performance differences in CI users do arise from peripheral factors (e.g. electrode discriminability), the way in which their individual implant is fitted and set is also an important factor in determining outcomes. Although CI research using free field stimulus presentation results in less control over what each individual receives, it has the important benefit of measuring performance in a real-world listening situation. From acoustic presentation (via loud speaker), the ACC has been recorded in adult CI users with speech stimuli. Frieson & Tremblay (2006) measured the ACC in eight adult CI users at the onset of the vowel in /si/ and /ʃi/ tokens. The latency of the ACC was increased compared to a NH response. They found that the response could distinguish between the two tokens, with earlier latencies for /ʃi/, and that individuals had a good test-retest reliability (r = 0.63 to 0.88). Martin (2007) recorded the ACC to tokens of /ui/ with nine different second formant (F2) changes in a case study. She found agreement between the ACC and behavioural detection thresholds; and that the ACC response was larger when the participant attended to the stimuli than when they ignored it.

Only one study has explored the relationship between the ACC to acoustically presented stimuli and individual speech perception performance (Brown et al. 2015). These authors measured the ACC to changes in synthesized vowels (/ui/ and /iu/) in 10 Nucleus hybrid users, and tested their speech perception abilities using an adaptive speech in noise task (the SNR of 12 spondees in noise was adapted to reach 50% correct identification), and a closed set consonant recognition task. They did not find a significant relationship between the ACC and the speech perception scores. Hybrid devices are designed to preserve acoustic stimulation at low frequencies (i.e. apical cochlear region) whilst delivering electric stimulation to cochlear regions corresponding to higher frequncies (i.e. basal region). Such devices are favoured for those with residual low-frequency hearing. Brown et al. (2015) compared three different programming methods of the Nucleus hybrid device. Participants used each programming strategy for four weeks prior to testing to acclimatise to the new stimulation parameters. Their behavioural speech recognition performance was measured using a closed set consonant recognition task, and an adaptive test that measured the SNR that resulted in 50% correct identification of a set of 12 spondees presented in background noise. The participants listened to the vowel changes /ui/ and /iu/ and both the P1-N1-P2 and ACC were measured. The ACC was recorded successfully in all users and showed similar characteristics to those recorded from NH listeners. However, the response did not predict which programming strategy gave the best behavioural performance. Furthermore, compared to using electrical stimulation alone, combined acoustic and electrical stimulation improved consonant recognition performance for all subjects, however, this again was not predicted by the ACC (ACC responses were larger with the acoustic and electrical stimulation, but not significantly so). The authors concluded that the differences in performance between the programming strategies were not large enough for the ACC to predict (i.e. the ACC is not sensitive enough to indicate small improvements in performance). In Chapter 2, using multiple speech stimuli, some aspects of the ACC were able to predict relatively small differences in SIN performance in NH listeners. This may have been due to the rapid collection paradigm allowing for a large number of trials to be presented, and/or due to the presentation of non-monotous stimuli changes reducing adapation effects. As such, one aim of this study was to investigate whether the ACC to a broader range of speech stimuli than has been used in the past can be used to predict CI performance.

In this study, thirteen CI users with various processors and implants from various manufacturers listened to concatenated chains of vowels and fricatives (identical to the stimuli used in Chapter 2, Experiment 2, but presented in quiet only), and sections of naturally spoken stories via a loudspeaker. EEG was used to measure the ACC to the vowels and fricatives (Experiment 4: section 3.3), and the participants' phase coherence to the stories (Experiment 5: section 3.4). The participants also completed behavioural tasks of sentence recognition, and vowel and consonant identification (Experiment 3: section 3.2). The aims of this study were to compare the ACC recorded in the CI users to the ACC recorded in the NH listeners (Chapter 2), and to assess whether the ACC could predict CI users' behavioural speech perception performance. Furthermore, the study aimed to investigate whether phase coherence could be measured in CI users, and to see whether this could predict the CI users' behavioural speech perception performance (see section 3.4).

3.2 Experiment 3: Behavioural Measures of Speech Perception

A sentence recognition task, a consonant identification (ID) task, and a vowel ID task were used to measure the participants' speech perception abilities. The results from these tasks are later used alongside the ACC and phase coherence experiments to assess the ability of the objective measures to predict speech perception performance. Three behavioural tasks were employed to provide a broad measure of individual speech perception abilities.

3.2.1 Participants

Thirteen post lingually deafened, right handed CI users, with one or two CIs contributed to the experiment (age range: 23-72, mean age = 51.8). Participants were recruited through the Home Counties Cochlear Implant Users' Group and from advertising on Facebook CI user groups. The participants used CIs from a variety of manufactures: eight used Cochlear devices (Nucleus Freedom, 5, or 6 processors), three used Advanced Bionic, and two used MED-EL. The duration of implant use spanned from 1 to 23 years (mean implant duration = 11.2 years). All participants completed a questionnaire about their hearing loss and CIs, details of which are provided in Table 3.1. The experiment and recruitment for this study was approved by the University College London ethics committee. All subjects provided their informed consent before beginning the experiment and were paid an honorarium for their time.

Of the 13 participants, eight were unilaterally implanted and five were bilaterally implanted. The procedure for unilateral and bilateral users was the same, as participants used their CI(s) with their everyday settings during the experiment. This meant that those with two implants used both during the experiment, and this ensured that the results reflected each individual's performance as close to their everyday experience. Recruitment and time constraints meant that it was useful to include both unilateral and bilateral users to reach a reasonably sized group. Furthermore, the aim of the study was to assess the ACC and phase coherence measures in a broad group of CI users, and how the measures related to overall speech perception abilities. Therefore, including bilateral users and allowing them to keep both CIs on allowed us to examine the characteristics of the ACC and phase coherence in a wide variety of users.

As the participants were recruited outside of the clinic, their clinical information (such as programming strategies, active electrodes, etc.) was not available. The CI group that was tested was fairly small, from a mixture of manufactures, a spread of ages and CI experience, some were bilateral and some were unilateral, and each implant would likely have had different settings. Taking this all into account, it would be unrealistic to expect to observe any meaningful information/correlations between clinical settings and performance on the tasks. As such, this was not considered in this study.

Participant details

CI user	Gender	Age (yrs)	Side of implant	Duration of deafness	Duration of implant use	Cause of hearing loss	Device
1	Female	37	Right	31	15	Mumps	Cochlear Nucleus 5
2	Female	61	Left	19	3.5	Unknown	Medel
3	Male	58	Both	24	23	Meningitis	Cochlear Nucleus 5
4	Female	72	Both	55	5 Left 3 Right	Ushers syndrome type III	Advanced Bionic Harmony
5	Female	67	Right	9	6	Rupture of malformed round window	Advanced Bionic
6	Female	23	Right	21.5	20	Meningitis	Cochlear Nucleus 22
7	Female	70	Both	18	10 Left 8 Right	Viral infections and Meniere's disease	Cochlear Nucleus 6
8	Male	36	Both	32	2 Left 19 Right	Genetic	Cochlear Nucleus 6 Left, Freedom Right
9	Female	36	Left	36	1	Branchiootorenal syndrome	Cochlear Nucleus 6
10	Female	34	Right	25	7	Genetic	Cochlear Nucleus 6
11	Female	67	Left	45	5 Left 3 Right	Unknown	Medel
12	Male	40	Right	40	14	Premature birth	Advanced Bionic
13	Female	72	Both	72	2	Genetic	Cochlear Nucleus 6

Table 3.1 Details of participants' hearing loss and CIs as provided by the participants themselves. Those who have reported a duration of deafness the same as (or close to) their age also reported enough acoustic hearing as an infant to develop language.

3.2.2 Stimuli, procedure, and apparatus

For all tasks, stimuli were presented in quiet via a Rogers LS3/5A monitor loudspeaker situated 1.73 m in front of the listener at 0º azimuth, at a level of 65 dB SPL at the listener's ear (verified with a Brüel & Kjaek 2231 sound level meter). All participants wore their CI processors on their usual everyday comfortable setting.

The combined three behavioural tests took no longer than 30 minutes in total. All recordings were made in a sound-proof booth with a 44.1 kHz sampling rate at 16 bits per sample.

For the sentence recognition task, each participant listened to five randomly selected IEEE (Rothauser et al. 1969) sentence lists (50 sentences per list and so a total of 250 sentences). The sentences were presented in a random order. The sentences were recorded from one male southern British English Speaker. During the experiment, the participant and the experimenter were seated in a sound-proof booth. The participants were required to repeat what they heard, and the experimenter recorded the number of key words in the sentence they correctly identified (each IEEE sentence contains five key words; e.g. The BIRCH CANOE SLID on the SMOOTH PLANKS). The task was created and presented using Praat (Boersma & Weenik 2013).

For the consonant ID task, each participant listened to 96 vowel-consonant-vowel (VCV) tokens, 48 of which were spoken by a female southern British English speaker and 48 by a male. The consonants were one of /b/, /d/, /f/, /g/, /k/, /l/, /m/, /n/, /p/, /s/, /ʃ/, /t/, /v/, /w/, /y/, and /z/. The start and end vowels were always matched, and were one of /ɒ/, /i/, and /u/. Each speaker recorded two of each VCV combination, and one token of each combination from each speaker was randomly selected. The combined 96 tokens (48 from each speaker) were then presented in a randomised order. Participants looked at a computer screen that displayed all of the 16 possible consonants. Once they heard a VCV token, they were asked to click on the consonant that they thought they heard. The task was created and presented using MATLAB (MathWorks 2013a).

For the vowel ID task, participants listened to 102 CVC words, 51 of which were recorded by a female southern British English speaker and 51 by a male. The words were “bad”, “bard”, “bared”, “bawed”, “bayed”, “bead”, “beard”, “bed”, “bid”, “bide”, “bird”, “bod”, “bode”, “booed”, “bowed”, “boyd” and “bud”. Each speaker recorded six of each of these words, and three tokens of each word from each speaker were randomly selected. The list of 102 words was then presented in a randomised order. Participants looked at a computer screen that displayed all of the 17 possible words. Once they had heard a word, they were asked to click on the word that they thought they had heard. The task was created and presented using MATLAB (MathWorks 2013a).

3.2.3 Results

Results from all three tasks are shown in Figure 3.1. For all three tasks, there was a wide range of performance, particularly for the IEEE task where the lowest performer scored 15.6% and the highest scored 96.4%. Each participant only completed each task once and so it is possible that the large range could partially be due to the test-retest reliability of the tasks. A task with low test-retest reliability could produce variable results. To my knowledge there is little or no data available exploring the test-retest reliability of the exact tests used in this study with CI users. However, the IEEE sentence lists have a high test-re-test reliability when used in the Speech in Noise (SIN) test (Etymotic Research 1993) in both normal-hearing listeners and hearing-impaired listeners (Bentler 2000; Killion et al. 2004). Furthermore, Başkent and Shannon (2004) tested CI users using similar Vowel and consonant tasks to the current experiment (ten monophthongs and two diphthongs presented in a /h/-vowel-/d/ context, and 20 consonants presented in an /a/-consonant/a/ context) and found the tasks had a high test-retest reliability. Therefore, although we do not have actual data for the test-retest reliability of the tasks used in this study, the materials used were standard speech perception materials that are generally considered to produce reliable results.

All three tasks were highly positively correlated (IEEE vs. consonant ID: $r = 0.95$, $t = 9.29$, $p < .001$; IEEE vs. vowel ID: $r = 0.93$, $t = 7.41$, $p < .001$; consonant vs. vowel ID: $r = 0.92$, $t = 6.82$, $p < .001$). The IEEE task showed the greatest range of performance whilst remaining highly correlated with the other two measures. Because of this, this measure was selected as the index of speech recognition in the ACC (section 3.3) and phase coherence (section 3.4) experiments.

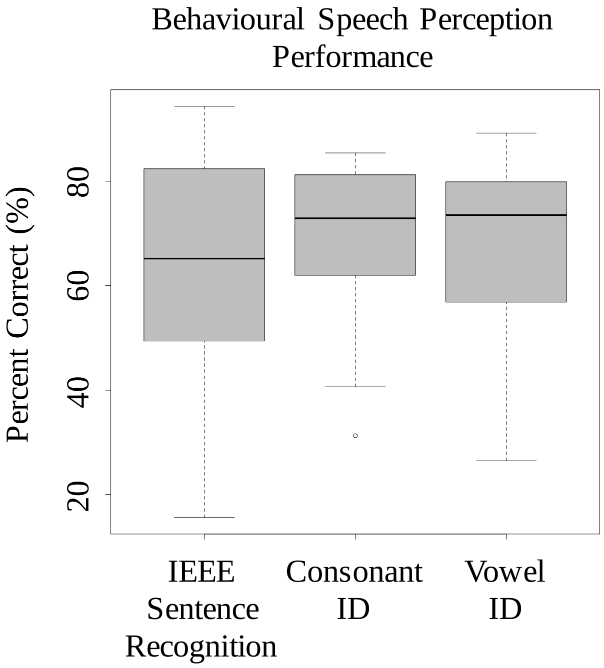

Figure 3.1 Boxplots showing the median (65.2, 72.92, 73.53), range (78.80, 44.79, 62.75) and interquartile ranges (33.00, 19.27, 23.04) for the IEEE sentence recognition, consonant identification (ID) and vowel identification (ID) tasks respectively.

3.3 Experiment 4: The ACC in Cochlear Implant Users

3.3.1 Participants

The same participants that were tested in the behavioural experiments (section 3.2) were also tested in the ACC experiment.

3.3.2 Stimuli

The stimuli were the same as in Chapter 2 (section 2.3 Experiment 2). The stimuli comprised of four vowels (/ɑ/, /i/, /ɔ/, /u/), as heard in the words *Bart*, *beat*, *bought*, and *boot*, and four fricatives (/ʃ/, /s/, /v/, /z/) as heard at the beginning of the words *short*, *sing*, *vine*, and *zip*, and silence. Sustained vowels and fricatives were recorded from a female Southern British English speaker. The recordings were made in a sound-proof booth with a 44.1 kHz sampling rate at 16 bits per sample. Speech stimuli manipulations were then conducted in Praat (Boersma & Weenink 2013). For each recorded sound, a one second segment that was deemed to be reasonably static (minimal spectral change) was extracted from the recording. The amplitude envelope of each segment was then flattened by calculating the envelope of the original recording (full rectification and a 50-Hz low-pass filter, and then normalized). The original recording was then divided by this envelope, and the overall amplitudes of all eight phonemes were rescaled to be equal. For the vowels and voiced fricatives, the pitch was flattened to the mean pitch of the recordings (200 Hz) using an overlap-add method (Boersma & Weenink 2013).

The four vowels, four fricatives, and portions of silence were then concatenated into a random order without replacement (the same stimulus could not appear twice in a row) to create two-minute sequences that included all possible pair transitions (i.e. 72 possible transitions including each sound to silence and vice versa). Phonemes were concatenated by splicing each transition with 50 ms raised-cosine ramps at each overlapping transition (i.e. the phoneme was ramped down at the same time as the new phoneme was ramped up). The duration of each stimulus was randomly jittered from 350-450 ms, with the segment randomly selected from the longer, one-second recording. This minimised the possibility of any spectral discontinuity within the recorded stimuli from being consistently time-aligned to the spectral changes. After the phonemes were concatenated into a two-minute sequence, the entire amplitude envelope was flattened to correct for amplitude fluctuations during the brief overlapping transitions. Each two-minute sequence was only used once.

3.3.3 Apparatus

All stimuli were presented binaurally via Praat (Boersma & Weenink 2013). Stimuli were presented using a Fireface UC external audio interface (44.1 kHz sampling rate, 16 bits per sample) connected to a loud speaker situated 1.73 m in front of the listener at 0º azimuth. Stimuli were presented at 65 dB SPL at the listener's ear (verified with a Brüel & Kjaer 2231 sound level meter). Each participant was asked to adjust their CI processors to their usual everyday comfortable setting.

An additional audio channel was used to provide stimuli timing information to the EEG system. A custom converter was used to convert the audio signal into TTL pulses, which was recorded as a time-aligned trigger by the EEG system. Testing occurred in an electromagnetically shielded sound-proof booth that was kept at 19^{o}C to minimize sweat artefacts. Within the booth, mains electricity was turned off during testing to reduce electrical artefacts as far as possible.

Evoked responses were recorded with a BioSemi ActiveTwo EEG system. Sixty-four surface electrodes were placed in accordance with the international 10-20 system; however, the electrodes directly over the CI were not used. This was typically two or three electrodes per implant (i.e. 4-6 in the case of the bilateral users). Seven additional electrodes were placed on the left and right mastoid, left and right canthus, above and below the left eye, and on the tip of the nose. Responses were recorded at a sampling rate of 2048 Hz.

3.3.4 Procedure

The participants listened to 30 minutes of the stimuli (i.e. 15 two-minute sequences). This meant that in total, each participant heard approximately 63 repetitions of each possible pair (approximately 4356 trials in total). They listened to the ACC audio files and stimuli for the phase coherence experiment (see section 3.4) in two-minute blocks in a random order. EEG testing lasted one hour and they were given the opportunity for a break every 10 minutes. During testing, participants watched an animated film without audio.

3.3.5 Analysis

For each participant, data were processed offline in MATLAB (MathWorks 2013a), using the Fieldtrip (Oostenveld et al. 2011), EEGlab (Delorme & Makeig 2004), ERPlab (Lopez-Calderon & Luck 2014), and NoiseTools (de Cheveigné & Simon 2007, 2008a, 2008b; de Cheveigné 2010, 2012, 2016; de Cheveigné & Parra 2014; de Cheveigné & Arzounian 2015) toolboxes. For each electrode channel, data were first high pass filtered at 0.1 Hz (Butterworth filter, as implemented by the ERPlab plugins within EEGlab), and then referenced to the nose electrode. Using Fieldtrip, the recording was segmented into epochs spanning from 100 ms before to 350 ms after each stimulus change onset. Epochs were baseline corrected to a 100 ms window before each stimulus change onset. Data were low pass filtered at 30 Hz (Butterworth filter, as implemented by Fieldtrip), and downsampled to a 512 Hz sampling rate. The unused channels positioned over the implant(s) were dropped, and then DSS was run using NoiseTools. Here, the first 50 ms after the stimulus onset was used to identify DSS components in which the apparent CI artefact was more dominant than the response. For each individual, the DSS components were inspected, and the first one or two components that were identified as being primarily artefactual (e.g. Figure 3.2A) were removed from the full length epoched data. This step was then repeated once and the first one or two components were again removed if artefactual components remained. We then ran DSS again on the full length epoched data in which these artefactual components were removed (so that the response was dominant in the DSS components), and projected the first one, two or three components back into the data (components were inspected and selected for each individual, see Figure 3.2B). Epochs in which the signal amplitude exceeded

150 µV were rejected due to their high noise level, and the remaining epochs were averaged.

Statistical analysis was run using R (R Core Team 2013). For linear mixed-model analyses the R 'lme4' package (Bates et al. 2015) was used, and type II analysis-of-variance tables were calculated using the 'CAR' package (Fox &Weisberg 2011).

3.3.6 Dropped participants

Whilst evoked potentials have been successfully elicited CI listeners, it is often difficult to separate the implant artefacts from the response (e.g. Martin 2007). Implant artefacts arise primarily from the fact that the electrical pulses delivered by the array are detected by EEG, and are much larger in amplitude than the neural signal of interest. Additionally, electrical artefacts pose particular challenges when measuring AEPs that show stimulus-following characteristics, such as the ASSR (e.g. Hofmann & Wouters 2010). Whilst CAEPs do not mirror the temporal characteristics of the stimuli in this manner, CIs nevertheless produce an electrical stimulus-related artefact that occurs for the entire duration of the sound (Shallop, 1993), and a large artefact overlaps the signal of interest and so the response is hard to interpret. In this study, DSS was used to identify the artefact and then remove it from the rest of the recording (See section 3.3.5).

For most participants, isolating and removing the artefact using DSS was seemingly successful. However, for two participants this was not the case. Figure 3.2 shows topographic maps and traces of A) a typical artefactual component from the first stage of DSS, and B) a typical resulting ACC component after the artefact removal. Figure 3.2 C) shows an attempt to isolate the artefact in participant number 4. When this artefact was removed and the first components were projected back into the data, the overall response was flat. This is possibly because the artefact for this participant had a long duration (about 200 ms) and so it overlapped with the ACC response. Figure 3.2 D) shows the topography of the artefact component for participant 12. This shows that the energy of the artefact is in the fronto-central region, which is removed from the location of the implant and is instead where the ACC response is expected to be observed. Again, when this artefact was removed and the first components were projected back into the data, the overall response was flat. Therefore, we were unable to isolate the artefact or the response for both of these participants. As such, these two participants were removed from further analysis.

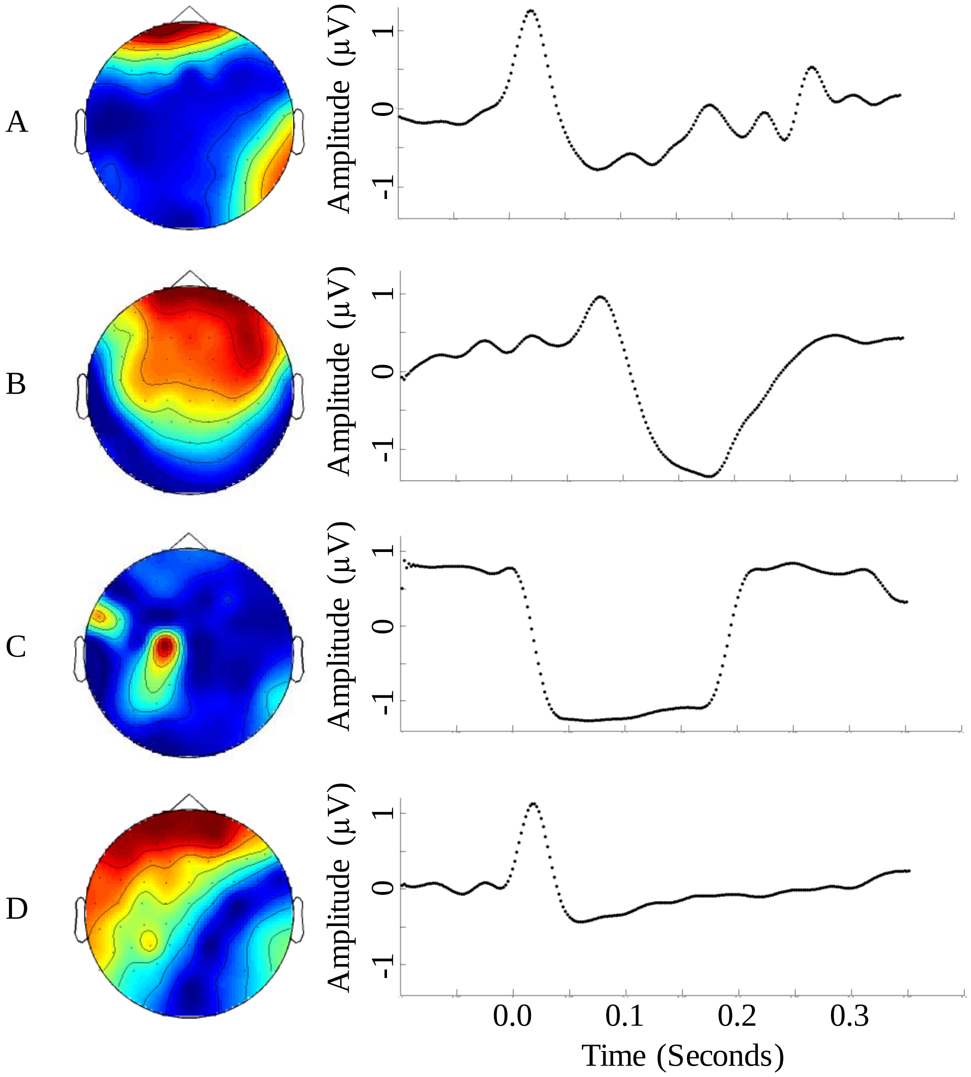

Figure 3.2 A) The topographic map and trace of a typical artefactual component from the first step of DSS. B) A typical ACC component from the second stage of DSS. C) The first component in the first stage of DSS for participant 4. D) The artefact component in the first stage of DSS for participant 12.

3.3.7 Results

Figure 3.3 shows the grand mean ACC response from all good CI participants (i.e. with participants 4 and 12 removed) and the normal-hearing (NH) response averaged over the quiet trials in the ACC in noise experiment (Chapter 2, Experiment

2). Figure 3.3 shows that the NH participants have much clearer P1 and P2 peaks and their response is generally more positive, whereas because the CI participants' response is dominated by the N1, their response is overall more negative. It also appears as if the CI N1 peak is slightly delayed and their P2 peak is further delayed. The mean data from the CI group partially conceals important individual variability. In Figure 3.4, individual ACC responses are plotted, and from this graph it is notable that whilst all subjects displayed an apparent N1 peak, both P1 and P2 prominence were more variable amongst individuals.

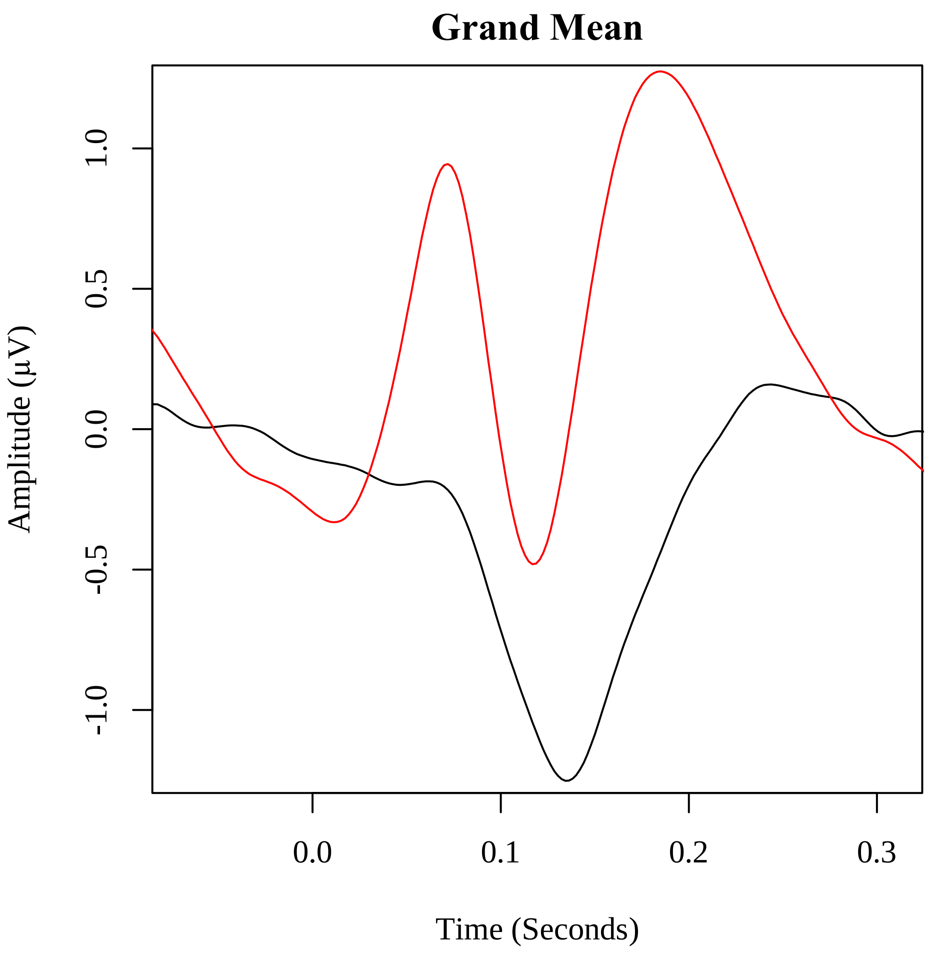

Figure 3.3 The grand-averaged ACC response across all good participants and all trials at FCz for the CI participants in black and the NH participants in red. The stimulus onset is at 0 s.

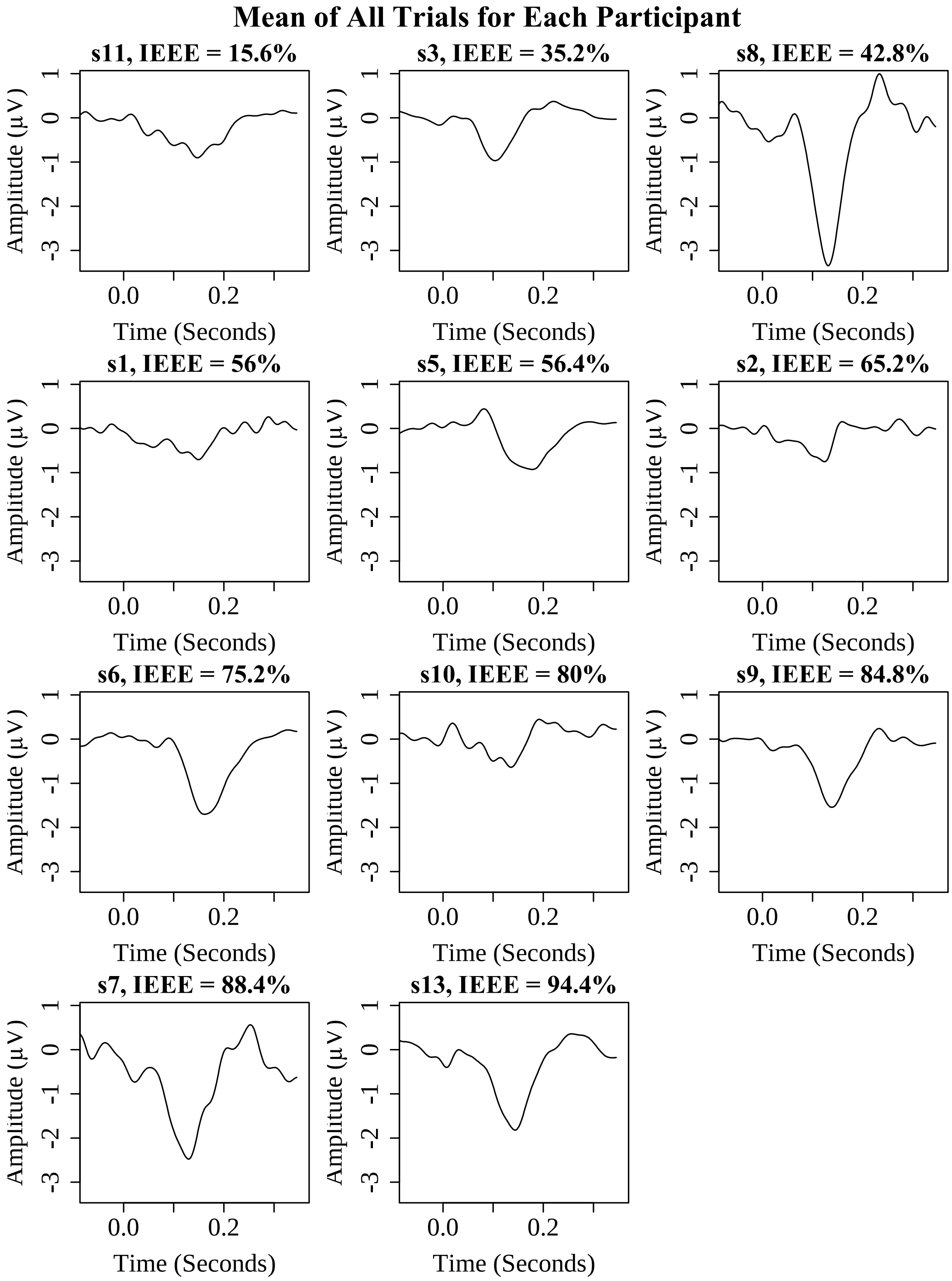

Figure 3.4 The response averaged across all trials at FCz for each participant ordered by IEEE percent correct performance. The stimulus onset is at 0 s.

The peak amplitudes for each participant were calculated by finding the maximum (P1 and P2) or minimum (N1) point within windows taken from the relative peaks in the grand mean response (Figure 3.3), then taking an average of the waveform 7 ms around that point (i.e. 3.5 ms either side of the peak). The individual responses for the CI participants (Figure 3.4) shows that there was variability in the baseline, meaning that at onset (0 ms) some participants' responses were already relatively negative (e.g. s7, Figure 3.4). In order to ensure the peak values were representative of the response and not of the baseline, the average of a 7 ms window around the minimum point 50 ms before the P1 peak latency was subtracted from each peak amplitude. The peak latencies were estimated by taking the time from the stimulus change onset to the maximum/minimum points used to calculate the peak amplitudes.

Descriptive statistics of the peak amplitudes and latencies for both groups are shown in Figure 3.5 and Table 3.2. A linear mixed models analysis was run, with by-subject and by-pair as random intercepts. The peak amplitude was the dependent measure (with P1, N1, and P2 peaks analysed in three separate models). The group (CI or NH) was a fixed factor. Results of the mixed model analysis and the descriptive statistics showed that the P1 and P2 peaks were significantly larger in the NH group (P1: $\chi^2(1) = 12.15$, $p < .001$; P2: $\chi^2(1) = 16.38$, $p < .001$), whereas there was no significant difference between the N1 amplitude in the two groups ($\chi^2(1) = 1.48$, $p = .22$). Latency analysis was conducted with the same model design (with peak latency as the dependent measure). Here, the N1 and P2 peaks were significantly delayed in the CI group (N1: $\chi^2(1) = 4.66$, $p = .031$; P2: $\chi^2(1) = 49.43$, $p < .001$), but the P1 peak latency was significantly shorter in the CI group ($\chi^2(1) = 4.46$, $p = .035$). However, this result may be partially attributable to the P1 latency being difficult to estimate accurately in the CI group, as the response was very small. Peak latencies are difficult to look at confidently because they are measured by using just one point within a window where the response is at maximum or minimum. The amplitude measurements, on the other hand, are more reliable, as an average over a 7 ms window is taken, meaning that the value represents not only the point at maximum or minimum, but also how broad the peak is.

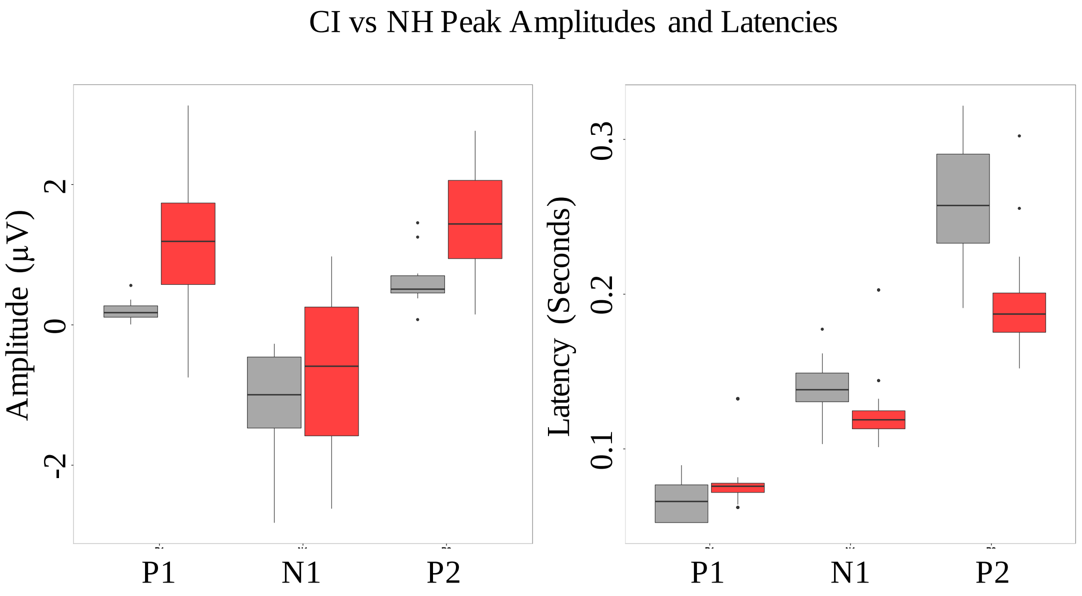

Figure 3.5 Boxplots showing the P1, N1 and P2 peak amplitudes on the left and latencies on the right at FCz for the CI participants (grey) and the NH participants (red).

Descriptive Statistics of the Peak Amplitudes and Latencies

		Amplitudes			Latencies		
		Mean (μV)	sd	N	Mean (μV)	sd	N
P1	CI	0.20	0.16	11	0.07	0.01	11
	NH	1.12	0.99	29	0.08	0.03	29
N1	CI	-1.12	0.76	11	0.14	0.02	11
	NH	-0.69	1.02	29	0.12	0.02	29
P2	CI	0.65	0.39	11	0.26	0.04	11
	NH	1.48	0.72	29	0.19	0.03	29

Table 3.2 Means and standard deviations of the P1, N1 and P2 peak amplitudes and latencies for the CI and NH participants.

To explore how the CI participants' ACC responses related to their IEEE sentence recognition scores from section 3.2, Figure 3.4 shows the individual responses of the CI users ordered by their IEEE task performance (percentage of keywords correct). For most individuals, a clear N1 was present, however the P2 was less defined than the N1, and the P1 was the hardest to identify of all the three peaks. Because the peak latencies are less reliable to measure than the peak amplitudes, latencies are not considered in the analysis of individual differences. Figure 3.6 and Table 3.3 show the scatter plots and correlation statistics for the IEEE sentence scores versus the P1, N1 and P2 peak amplitudes. The sample size was relatively low for a correlation analysis, and so bootstrapping using 2000 samples was used to calculate 95% confidence intervals for all correlations (Table 3.3). There were no strong correlations between the IEEE scores and the peak amplitudes, and all the bootstrapped 95% confidence intervals crossed zero (Table 3.3), which means whilst there were some apparent trends in the correlation, they were not significantly different from a correlation of 0 (IEEE vs. P1: r = 0.046, *2000* *bootstrapped 95% CI* = -0.80 to 0.63; IEEE vs. N1: *r* = -0.15, *2000 bootstrapped 95% CI* = -0.71 to 0.62; IEEE vs. P2: *r* = 0.12, *2000 bootstrapped 95% CI* = -0.61 to 0.61).

From Figure 3.6, which shows the IEEE score vs. the N1 and P2 peak amplitudes, it appears that one outlier may be unduly influencing the statistics, particularly in the case of the N1 analysis. If we remove this outlier (participant 8) the correlations become much stronger (IEEE vs. N1: *r* = -0.69, *p* = 0.029; IEEE vs. P2: *r* = 0.55, *p* = 0.096). Whilst this result is interesting, there is no priori reason to remove this outlier from the data.

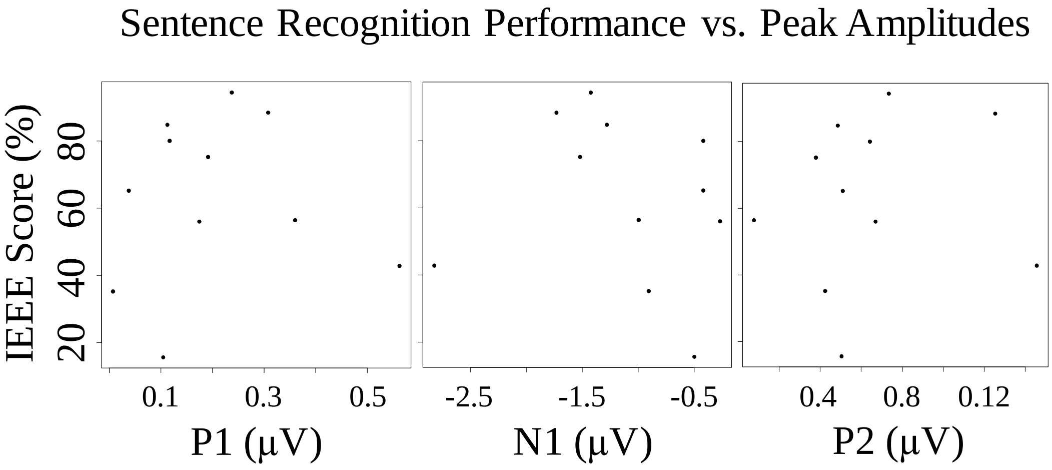

Figure 3.6 Scatter plots showing IEEE sentence score vs. P1, N1, and P2 peak amplitudes. Refer to Table 3.3 for the corresponding statistics.

Correlation Statistics of IEEE Score vs. Peak Amplitudes

			r	p	confidence intervals
Amplitudes	IEEE score vs.	P1	0.046	.98	-0.8049, 0.6299
		N1	-0.15	.60	-0.7147, 0.6220
		P2	0.12	.65	-0.6092, 0.6146

Table 3.3 Correlation coefficient, p value, and 2000 bootstrapped 95 % confidence intervals for the IEEE sentence scores vs. the P1, N1, and P2 peak amplitudes.

In summary, the results indicate that P1 and P2 peaks are smaller in CI participants compared to NH participants. CI users showed an N1 of a similar magnitude to NH listeners, suggesting that the CI users were successfully receiving the acoustic information of the spectral changes (Sharma & Dorman 1999; Martin & Boothroyd 2000; Sharma et al. 2000; Hoonhorst et al. 2009). However, their response was overall more negative than the NH response, possibly indicating that attention may have influenced the peak magnitudes (Hansen & Hillyard 1980; Näätänen & Picton 1987). For the CI users, the N1 and P2 peaks were delayed in comparison to NH responses. However, it should be noted that P1 and P2 peak latencies were difficult to estimate in CI users. Correlation analysis shows that there was no relationship between the individuals' peak amplitudes and speech perceptual performance. These observations are considered further in the discussion provided in section 3.5.

3.4 Experiment 5: Phase Coherence in Cochlear Implant Users

The entrainment of low frequency neural oscillations to the amplitude envelope of speech can be measured in the auditory cortex (e.g. Ahissar et al. 2001; Luo & Poeppel 2007). This entrainment or phase coherence to speech has been reliably shown to be affected by selective attention, indicating that it is not necessarily purely a response to acoustic information alone (Kerlin et al. 2010; Gomez-Ramirez et al. 2011; Ding & Simon 2012; Horton et al. 2014). Ding and Simon (2012) used MEG to record listeners' neural responses when participants listened to spoken narratives by two competing talkers whilst being instructed to pay attention to just one of them. Using the neural responses, they reconstructed the temporal envelope and compared this to the envelopes of the two talkers and also the envelope of the two talkers combined. The reconstructed envelope correlated more strongly with the attended speech envelope than the background speech envelope. Furthermore, they varied the intensity of the two talkers separately over an 8 dB range and found that the neural representation of the attended speech adapted only to the intensity of that speaker but not to the intensity of the background speaker (Ding & Simon 2012). They conclude that auditory objects that overlap in the spectral and temporal domains are individually encoded in the auditory cortex. Furthermore, their results provide support for the neural adaptation being driven by the acoustic properties of speech as well as being modulated by selective attention.

There have been a few studies that have investigated phase-locked responses to speech that varied in intelligibility, using MEG, EEG and ECoG. Results have varied, with some studies finding that as the intelligibility of the speech decreased so did the phase coherence (Peelle et al. 2013), and others finding that there was no difference between the phase coherence to intelligible versus unintelligible speech (Howard & Poeppel 2010; Millman et al 2015). Peelle et al. (2013) tested 16 adults using MEG. Participants listened to sentences that were noise vocoded to vary intelligibility by limiting the spectral information but preserving the amplitude envelope. They found that phase locking of the neural activity in the theta range (4-7 Hz) occurred to all speech, but was enhanced when the speech was intelligible. The authors suggest that coherence was enhanced when the speech was intelligible because the listeners were able to extract higher linguistic information in the speech signal, and that higher-level processes aided the low-level tracking of the amplitude envelope of the speech. Howard & Poeppel (2010) used MEG to measure phase coherence between neural oscillations and the amplitude envelope of intelligible sentences and their time-inverted counterparts. They found no difference between the observed phase coherence with the intelligible versus the unintelligible sentences, and so concluded that the discrimination of speech stimuli based on phase coherence depends on acoustics but not comprehension (Howard & Poeppel 2010). Peelle & Davis (2012) suggest that this result occurred because the speech intelligibility manipulation that Howard and Poeppel performed did not preserve the amplitude envelope of the stimuli, and so the relationship between the acoustic characteristics of the speech, the phase coherence, and the intelligibility of the stimuli is uncertain. On the other hand, Millman et al. (2015) did preserve the temporal envelope of their sentences that were tone-vocoded to make them intelligible or unintelligible, and they also found no difference in theta-band phase coherence between the intelligible and unintelligible speech. Rimmele et al. (2015) combined attention and intelligibility in their study where they played pairs of competing talkers of natural or noise vocoded speech to listeners who performed a detection task whilst ignoring the other speaker. They recorded the participants' MEG and found that phase coherence to the natural, but not the vocoded, speech was enhanced by attention, whereas for the unattended speech there was no difference between natural and vocoded. The authors reported that their findings suggest that when speech is attended, more precise phase locking is related to the spectral detail of the speech, whereas when it is unattended, the spectral detail is not processed to the same level and so the phase locking to this speech reflects tracking of the temporal envelope alone (Rimmele et al. 2015).

In summary, it seems that the majority of neural entrainment to the amplitude envelope of speech is driven by acoustic factors, however in a competing talker situation it is also modulated by attention. Furthermore, there is a possibility that intelligibility of the speech may influence phase coherence in the theta-band, however this relationship is not yet fully understood. It is not clear whether the larger phase coherence to intelligible speech compared to unintelligible speech that was found in some studies was purely driven by acoustics (more spectral detail), or whether it is driven at least in part by higher-level linguistic processes related to intelligibility.

It has been suggested that phase coherence in the theta range (4-7 Hz) is not thought to predict individual speech recognition scores, whereas phase coherence in the delta range (1-4 Hz) is (Ding & Simon 2014); however this has not been studied extensively. Ding et al. (2014) tested 12 normal hearing adults using MEG. The participants listened to a story (Alice in Wonderland) either in quiet or in spectrally matched stationary noise at +3 dB SNR. The speech was either unprocessed or noise vocoded through a 4- or 8-channel vocoder. Participants were asked to rate the intelligibility of the speech and to answer some comprehension questions every 50 s (ratings and comprehension scores were highly correlated). Noise affected coherence more for the vocoded speech than for the unprocessed speech. In quiet, as the spectral resolution of the stimuli was reduced using vocoding, theta-band coherence decreased. They did not find a relationship between the participants' intelligibility rating scores and their theta-band phase coherence measure. However, they did find a positive correlation between the participants' intelligibility rating scores and the phase coherence in the delta range for the 4-band vocoded speech in quiet and the 8-band vocoded speech in quiet and in noise (Ding et al. 2014). The authors suggest that when using the same stimulus, theta-band entrainment can sometimes correlate with speech intelligibility, whereas delta-band entrainment does correlate with speech intelligibility, when measured in individual listeners (Ding & Simon 2014). The relationship between speech intelligibility and delta- and theta-band entrainment is unclear, and so in this study we considered neural entrainment in a range of 2-8 Hz, which encompasses both the delta- and theta-band ranges.

As far as I am aware there have been no studies conducted with CI users using phase coherence measures. Therefore, the aim of this part of the study was to see if the phase coherence to the amplitude envelope of continuous speech could be measured in CI users and to see if it could be used to predict speech perception performance. So far there is limited data on the relationship between behavioural measures of a listener's speech processing abilities and the phase coherence of their neural oscillations to speech. However, it is possible that phase coherence is reduced when the speech is less intelligible, and so conceivably poorer CI performers may show reduced cortical entrainment.

3.4.1 Participants

The same participants that were tested in the behavioural speech perception tasks and the ACC experiment were tested in this phase coherence experiment. The same participants were dropped from analysis as in the ACC experiment (participants 4 and 12), as we were unable to isolate and remove their CI artefacts (see section 3.3.6).

3.4.2 Stimuli

Portions of the stories ‘The Secret Garden’ (Frances Hodgson Burnett) and ‘Lazy Jack’ (Joseph Jacobs) were recorded from one female southern British English speaker. The recordings were made in a sound-proof booth with a 44.1 kHz sampling rate at 16 bits per sample. The stories were then split into two-minute segments that started at the beginning of a phrase or sentence and finished at the end of a phrase or sentence.

3.4.3 Apparatus

The set-up was the same as was used in the ACC experiment (section 3.3).

3.4.4 Procedure

The participants listened to 15 of the two-minute story segments, presented in a random order with the two-minute ACC sequences from section 3.3. In total, the EEG recording lasted one hour and they were given the opportunity for a break every 10 minutes. During testing, participants watched an animated film without audio.

3.4.5 Analysis

Data were processed offline in MATLAB (MathWorks 2013a), using the Fieldtrip (Oostenveld et al. 2011), EEGlab (Delorme & Makeig 2004), ERPlab (Lopez-Calderon & Luck 2014), and NoiseTools (de Cheveigné & Simon 2007, 2008a, 2008b; de Cheveigné 2010, 2012, 2016; de Cheveigné & Parra 2014; de Cheveigné & Arzounian 2015) toolboxes. Data were high pass filtered at 0.1 Hz and low pass filtered at 40 Hz (Butterworth filters, as implemented by the ERPlab plugins within EEGlab), and then referenced to the nose electrode. The channels directly over the implants were dropped and then using NoiseTools, denoising source separation (DSS) was run. In order to isolate the artefact, the data from the ACC experiment (section 3.3) was used. To begin, the first 50 ms after the stimulus onset (in the ACC epochs) was used to create DSS components of the CI artefact by selecting linear combinations of electrodes that maximised the repeatability of the artefact in individual trials. For each individual, the DSS components were inspected and the first one or two components that were artefactual were removed from the full length story data. This step was then repeated once and the first one or two components were removed if the artefact was still present. We then ran DSS again on the full length story data (which had had the artefacts removed) to extract the components where the response was phase-locked to the corresponding speech signal for each subject, and the response was taken as the projection of the first three components back into the sensor space.

The amplitude envelopes of the stories were obtained by full-wave rectifying the signals and filtering them with low pass (0.1 Hz) and high pass (40 Hz) Butterworth filters (i.e. the same filters that were used for the EEG data). They were then downsampled to 2048 Hz (to match the EEG data). The read stories contained segments of silence (i.e. pauses between sentences and phrases, etc.), so the EEG data and the amplitude envelope of the audio signal were segmented into two-second epochs with any periods of silence in the stories removed from both signals. A Fast Fourier transform (FFT) was used to transform the epoched EEG and the amplitude envelope of the speech signal into the frequency domain. To calculate the coherence between the two signals (i.e. EEG and amplitude envelope of the stories), the cross-spectrum of the two transformed signals (FFT data) was calculated and then normalised by the power spectra of each signal. The coherence values were averaged across frontocentral electrodes (F, FC, and C electrodes) and used for further analysis. Statistical analysis was run using R (R Core Team 2013).

The main aim of this portion of the study was purely to investigate whether phase coherence could be measured in CI users, and so to be comparable to previous studies, a broad-spectrum envelope rather than a cochlear-model was used (e.g. Peelle et al. 2013). Furthermore, most of the energy in the speech signal comes from voicing, and so most of the modulations in the signal are coming from the same source (i.e. voicing). Since the modulations are mainly coming from the same source, they will be coherent with each other across the frequency bands. Therefore, whether the envelope is broad-spectrum or passed through a cochlear model, the coherence across frequency bands are unlikely to be greatly affected.

3.4.6 Results

The phase coherence between the EEG response and the stories is plotted in Figure 3.7. Both the coherence before and after the CI artefacts were removed using

DSS are shown, to display that the coherence measured is not merely a representation of the implant's entrainment to the audio signal, but rather the listener's neural oscillations. Before the artefacts were removed (black line in Figure 3.7) there was an unusual peak in the low frequencies (< 1 Hz). This peak was removed after DSS was performed to remove the CI artefacts, leaving a much more typical coherence plot (red line in Figure 3.7). There is a peak in the theta range, which is the correct region for phase coherence (between 4 and 8 Hz; Chi et al. 1999; Chandrasekaran et al. 2009; Elliott & Theunissen 2009; Peelle & Davis 2012).

For each participant, a Monte Carlo analysis was run, where the coherence between the neural data was run with 1000 randomised acoustic envelopes. The coherence of the randomised trials was then compared to the coherence of the neural data with the true acoustic envelope. The results showed that all of the participants showed a significant coherence peak in the theta range ($p < .05$). Although the mean coherence plot looks as expected and the CI artefact removal using DSS seems to be largely successful, the results should still be viewed with caution as any residual artefact could be boosting the coherence values for individual participants. However, the results are promising for the measurement of phase coherence in CI users.

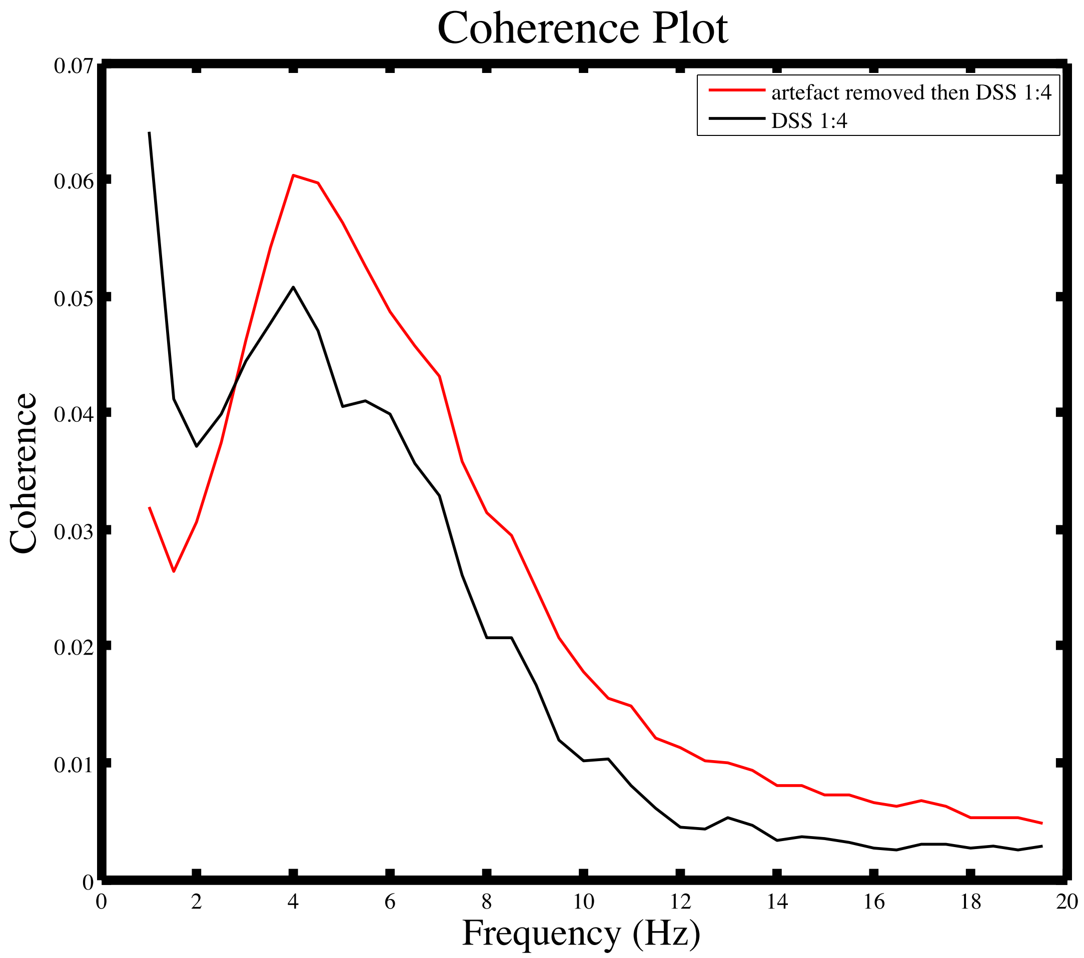

Figure 3.7 Phase coherence between the EEG response and the stories in the frequency range of 1.5 to 20 Hz. The black response is the coherence before removing the CI artefacts using DSS and the red line is the coherence after they have been removed.

In order to look at individuals' phase coherence and how this related to their speech perception abilities and their ACC responses, a phase coherence value was calculated for each person by taking an average of the coherence between 2 and 8 Hz. Figure 3.8 shows the scatter plots of the phase coherence values for individuals against their ACC P1, N1 and P2 amplitudes, and their IEEE sentence score (sections 3.3 and 3.2). Bootstrapping using 2000 samples was used to calculate 95% confidence intervals for all correlations (Table 3.4). There were no strong correlations between the phase coherence values and the peak amplitudes or the IEEE sentence scores. Furthermore, all the bootstrapped 95% confidence intervals crossed zero (Table 3.4), which means that they were not significantly different from a correlation of 0 (coherence vs. P1: *r* = 0.38, *2000 bootstrapped 95% CI* = -0.73 to 0.82; coherence vs. N1: *r* = -0.13, *2000 bootstrapped 95% CI* = -0.74 to 0.35; coherence vs. P2: *r* = -0.35, *2000 bootstrapped 95% CI* = -0.91 to 0.76; coherence vs. IEEE: *r* = -0.29, *2000 bootstrapped 95% CI* = -0.69 to 0.37). There was one participant with a particularly high coherence value (participant 5). This could be normal variability, however it is worth viewing this participant's result with caution, as, although the CI artefact seemed to be removed using DSS, it is possible that it still had an influence on the data, resulting in a high coherence value.

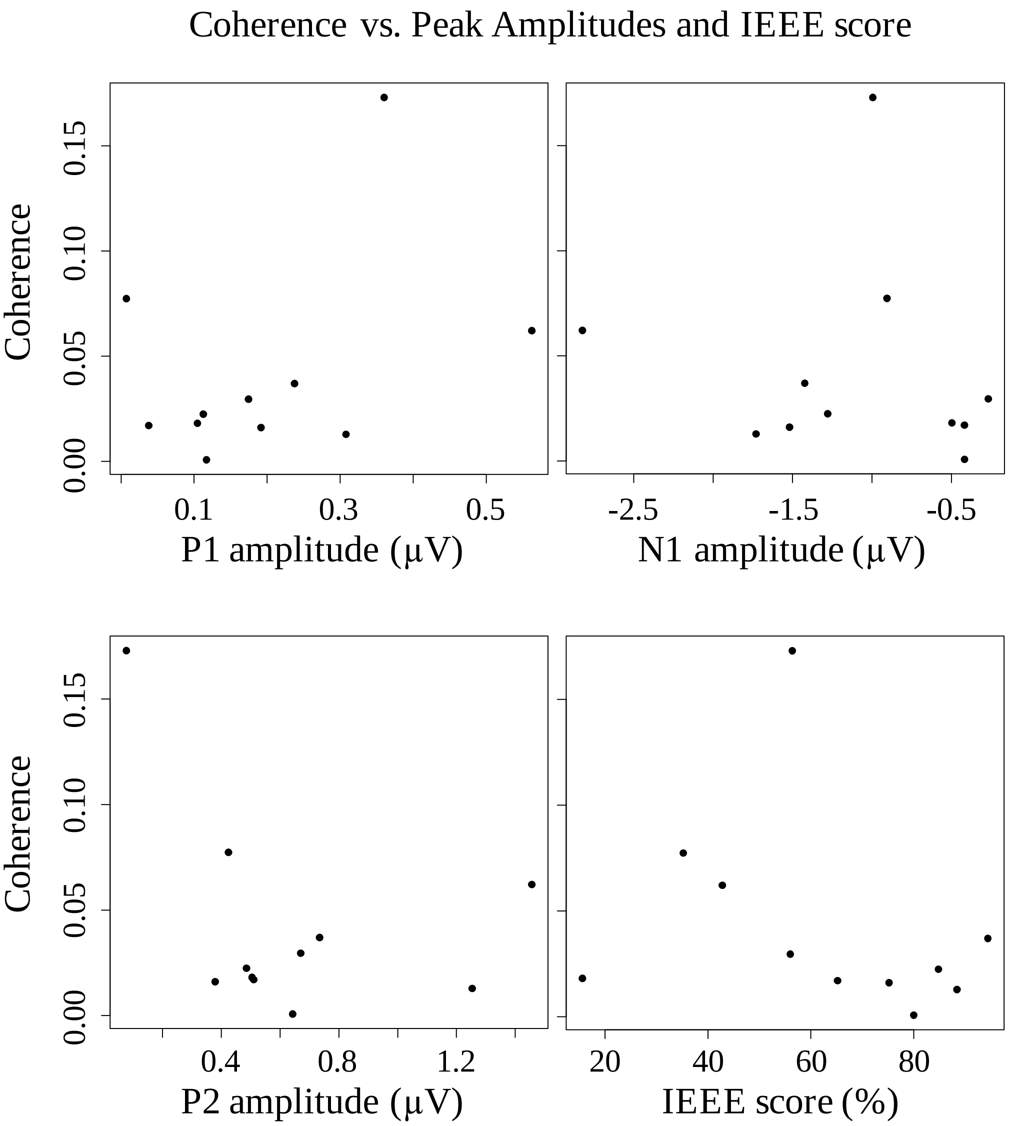

Figure 3.8 Scatter plots showing phase coherence vs. P1, N1 and P2 amplitudes, and IEEE sentence score. The peak values are from the ACC response recorded in section 3.3 and the IEEE score is from section 3.2. Refer to Table 3.4 for the corresponding statistics.

Correlation Statistics of Coherence value vs. Peak Amplitudes and IEEE Score

		r	p	confidence intervals
coherence value vs.	P1	0.38	.24	-0.7299, 0.8174
	N1	-0.13	.69	-0.7353, 0.3505
	P2	-0.35	.29	-0.9070, 0.7581
	IEEE	-0.29	.39	-0.6938, 0.3708

Table 3.4 Correlation coefficient, p value, and 2000 bootstrapped 95 % confidence intervals for the coherence values vs. the P1, N1 and P2 peak amplitudes and the IEEE percent correct score.

3.5 Discussion

The results of this study found that a speech-evoked ACC response was observed in 11 out of 13 CI users (two participants were excluded because of apparent artefactual issues, as discussed further later). For the ACC response, the P1 and P2 peaks were reduced in amplitude, and the N1 and P2 peaks were delayed in comparison to NH listeners. We investigated whether the properties of this response may predict individual behavioural speech perception performance (IEEE sentence recognition, vowel ID, consonant ID). However, there was no significant relationship between CI users' speech perception ability and the amplitudes of ACC response components to speech. Furthermore, phase coherence was successfully measured in CI users, but this did not appear to be related to their ACC morphology or speech perception performance. These findings will be considered in turn throughout this discussion section.

Artefact Reduction Technique

This study demonstrated that the substantial electrical artefact associated with CI stimulation via a speech processor can be successfully removed from the majority of CI users, with DSS; for both ACC and phase coherence measurements. By analysing a short time window immediately after the onset or change of a stimulus, DSS components relating to the CI artefact can be identified and then removed from the rest of the EEG data. In this time window, components related to early cortical responses should be absent. Two participants were excluded from analysis because we were unable to remove the artefact from their data satisfactorily. Interestingly, both of these participants used Advanced Bionic devices; however, three Advanced Bionic users were tested in total, and so artefactual effects do not appear to be fully consistent across devices (although it should be noted that the one participant with a large coherence value was the third Advanced Bionic user). Whilst electrical artefacts continue to pose problems for EEG measures of CI listening, DSS appears to be a useful tool for minimizing such issues. Although the current method of artefact removal was largely successful, it is worth noting that one cannot be absolutely certain that the entire artefact was removed, and that a different method, or a combination of methods, may work better for different processors, MAPs, and individuals. To this end, further research is still required to further develop techniques of artefact removal.

Comparisons of the ACC between Normal Hearing and Cochlear Implant Participants

No previous studies have directly compared ACC responses between NH and CI participants. Nevertheless, certain relevant literature has compared both the MMN and the onset P1-N1-P2 complex between CI and NH participants. Results have shown that the MMN is delayed in CI users (Obuchi et al. 2012; Turgeon et al. 2014), and the onset N1 and P2 have a lower amplitude in CI users (Kelly et al. 2005; Timm et al. 2012). Additionally, the whole P1-N1-P2 onset complex is delayed in children with CIs in comparison to NH (Munivrana & Mildner 2013). In summary, there is good evidence that CAEPs are often delayed, and/or of a lower amplitude in CI users in comparison to NH listeners.

The results of the current study are broadly consistent with these observations. When the ACC observed from CI participants is compared to that from NH participants, both the P1 and P2 peaks had lower amplitudes, and the N1 and P2 peaks were delayed. The amplitude of the N1 component did not differ between groups. Concerning the differences in peak amplitudes between the CI and NH groups, the N1 is thought to reflect the encoding of time-varying aspects of a sound, such as amplitude, spectral, and voice onset time changes (Sharma & Dorman 1999; Martin & Boothroyd 2000; Sharma et al. 2000; Hoonhorst et al. 2009). Therefore, one simple interpretation of the finding that CI and NH participants displayed similar N1 amplitudes is that both groups received and encoded salient acoustic time-varying changes in a comparable manner. If so, this would suggest that the observed differences in behavioural speech perception ability reflected differences in more central processes. This interpretation would seem unlikely however, as whilst the N1 amplitudes were not significantly different between the participant groups, the morphology of the global response was clearly different (i.e. latencies and P1 and P2 amplitudes). This suggests that although whilst central aspects of speech processing may well contribute to differences in speech perception ability, it remains likely that early cortical encoding is impaired in CI users, which may influence how peripheral spectral change information is encoded in the auditory cortex.

The amplitude of the N1 is known to increase with selective attention (Hillyard et al. 1973; Näätänen & Picton 1987; Woldorf & Hillyard 1991). Due to the reduced spectral resolution in the periphery, it has been proposed that CI users require greater listening effort than NH listeners in order to understand speech (Hughes & Galvin 2013; Pals et al. 2013). Therefore, it is plausible that comparable N1 amplitudes between groups may reflect a trade-off between reduced encoding and greater listening effort in the CI group. Concerning the global response in the CI group, reduced P1 and P2 amplitudes (but a normal sized N1 amplitude) could reflect a combined effect of a broad negativity that occurs with attention (Hansen & Hillyard 1980; Näätänen & Picton 1987), making the P1, N1, and P2 all appear more negative (Näätänen & Michie 1979; Näätänen & Picton 1987; Michie et al. 1990, 1993).

The CI group had a range of ages from 23 to 72 years (mean = 51.8 years). It is possible that for some of the older listeners there were ageing effects taking place that could affect the morphology of the ACC. For example, there has been a fairly consistent finding in the literature that the amplitude of the N1 is larger in elderly populations (Tomé et al. 2015; although this is not entirely consistent, see Čeponienė et al. 2008 for a review). It has been suggested that this is owing to the elderly having inhibitory deficits (e.g. Anderer et al. 1996; Amendo & Díaz 1999) or a reduced ability to inhibit the processing of information irrelevant to the task. This could also lead to a higher level of attention during tasks and so increased N1 amplitude (Čeponienė et al. 2008; Tomé et al. 2015). In the current study, the N1 was the most defined of the three peaks in the CI group, and was the only peak that wasn't significantly smaller than the NH group. It is possible that ageing effects, such as inhibitory deficits and attentional differences, could have affected the morphology of the ACC in the CI group.

Concerning latency differences, the N1 and P2 peaks were found to be delayed in CI users compared to NH participants. Research on auditory development in implanted children has suggested latencies and amplitudes of the N1-P2 complex may be associated with auditory system maturation and deprivation. Here, studies demonstrate significant changes in response latency and amplitude well into adolescence (Courchesne 1978, 1990; Ponton et al. 1996a, 1996b; Ponton et al. 2000; Sharma et al. 1997). Whilst this may suggest that the current CI group displayed such characteristics of auditory deprivation, it is important to note that the current participants were all of adult age, and with different histories of deafness and implant use. The increased latencies currently observed more likely reflect the decreased stimuli discrimination ability in the CI group. For example, Oviatt & Kileny (1991) measured P3 responses to deviant tonal stimuli in NH and CI groups. Response latencies were significantly longer in the CI group for 500-1000 and 500-2000 Hz changes, but not for 500-3000 Hz changes. The authors proposed that latency may be a reliable indicator of signal detection and discrimination ability in CI users (see also Okusa et al. 1999). In child CI listeners, Kileny et al. (1997) presented tonal or speech stimuli in a conventional P1-N1-P2 (and oddball) paradigm. The authors observed that the type of stimulus appeared to have a consistent (but non-significant) effect on peak latencies, and responses to the speech stimuli exhibited the most delayed latencies. They speculated that this trend might have reflected the greater complexity involved in processing the speech stimulus contrast compared with the simpler loudness or frequency contrasts. In the current study, complex speech stimuli were used, for which the salience of spectral changes would likely be reduced for the CI listeners (owing to reduced spectral resolution in the periphery). As such, it is perhaps unsurprising that the CI participants displayed greater N1-P2 latencies than the NH participants.

Whilst a global comparison of averaged responses is informative of how ACC responses broadly differ between NH and CI groups, it is also important to consider how individual responses differ, and whether such variations are predictive of speech perception performance. To this end, the next section of this discussion considers the relationship between the ACC and speech perception performance in CI users.

The Relationship between the ACC and Speech Perception Performance

There have been several previous reports that certain CAEPs may be predictive of CI users' speech perception scores, but generally these studies have used acoustically simpler stimuli than in the current study (Groenen et al. 1996b; Maurer et al. 2002; Roman et al. 2004; Alvarenga et al. 2012). Whilst there have been some measurements of the ACC in CI participants, much work with this paradigm has been used to measure peripheral frequency selectivity. Here, using direct stimulation via a research platform, pulse trains are presented on two sequentially stimulated electrodes (Brown et al. 2008; Hoppe et al. 2010; He et al. 2014). The ACC observed from the change in presenting electrode has been considered an objective measure of electrode discrimination, as the size of the ACC is partially dependent upon the degree with which the two electrodes stimulate nonoverlapping neural populations. Consistent with this notion, ACC amplitude generally increases as the electrode pairs are increasingly separated (Brown et al. 2008; He et al. 2014). Significant correlations have been observed between such ACC responses and behavioural measures of electrode discrimination (Hoppe et al. 2010; He et al. 2014), and speech perception in children identified with auditory neuropathy spectrum disorder (He et al. 2014). Furthermore, Scheperle & Abbas (2015b) found that both the ACC to changes in the stimulating electrode and the ACC to spectral changes using rippled noise were predictive of the speech perception performance of their 11 adult CI users.

Concerning the current results, it is important to note that stimuli were presented acoustically, and participants listened through their speech processor on its clinically fitted, every-day setting. As such, variability in the ACC response between subjects likely stems from multiple causes; from differences in the extent of surviving auditory processes after deafness, from differences in the peripheral effectiveness of the implanted array (i.e. insertion depth, electrode positioning, etc.), and differences in the optimisation of the participants' clinical fittings (i.e. number of channels, stimulation rate, pulse duration, etc.). As such, the current measure likely reflects individual differences across these related areas, and so provides a global assessment of performance in an everyday listening context.

Brown et al. (2015) measured the ACC to changes in synthesized vowels (/ui/ and /iu/) presented via a loudspeaker in 10 Nucleus hybrid users, but did not find a significant relationship with speech perception scores. In both Brown et al's (2015) study and the current study, correlations between CAEP measures and behavioural performance were not significant. Note however that Brown et al. (2015) tested participants with experimental MAPs, as opposed to those used in everyday listening by the subject (with a four week acclimatisation period prior to testing). Brown et al. (2015) tested 10 hybrid users, and the current study had 11 CI users (after removing two subjects). In both cases, unless a strong relationship is present, it is difficult to observe significant correlations with such sample sizes. Nonetheless, whilst the correlations observed in the current study were not significant, there do appear to be certain trends within the data set. There is a general pattern for which the amplitudes of the N1 and P2 increased with improved speech perception performance. This would be consistent with the previously observed CAEP results described above, and with the notion that worse performers received/encoded less spectral-change information than their better performing peers. ACC response morphologies have been shown previously to change when the spectral-change information received differs, as certain aspects of all three of the peaks relate to the acoustic characteristics of the stimuli (Chapter 2; Shtyrov et al. 1998; Sharma & Dorman 1999; Martin & Boothroyd 2000; Sharma et al. 2000; Ceponiene et al. 2005; Hoonhorst et al. 2009; Anderson 2010). As such, a reduction in ACC amplitude for poorer CI performers may represent poorer detection/encoding of the spectral changes in the stimuli.

Differences in the relationship between the ACC and Speech Perception Performance for NH and CI Groups

Chapter 2 detailed an experiment in which NH listeners were presented with the same vowel fricative sequences as was used in the current study, in either quiet or in the presence of speech-shaped noise. The participants who performed worse at a behavioural SIN task showed larger P1 responses in quiet. Furthermore, for all participants, as the noise increased, the amplitude of the ACC decreased as the spectral information was masked. The larger sample size in Chapter 2 meant that we could more confidently investigate individual differences in the ACC, including the morphology of the P1. On the other hand, in the current study, the P1 in CI users was less reliably observed, and the global morphology of the ACC response was different to that of NH listeners.

Looking at the results of Chapter 2 and this CI study together, it is plausible that the differences in response morphologies between CI and NH listeners may reflect the combination of different aspects of auditory cortical processing. In Chapter 2, the tested NH participants were screened for normal-level audiometric thresholds. As such, the group should have had reasonably similar peripheral processing (barring any undetected differences, such as hidden hearing loss). The NH ACC results suggested that those with poorer SIN ability may apply greater pre-conscious attention and cognitive effort to the speech stimuli, and potentially have reduced neural efficiency, proposing a relationship between higher level processing and obligatory cortical activity. To clarify, because the participants had normal hearing, differences in peripheral hearing, whilst potentially present, should be relatively small, and therefore the individual differences in the morphology of the ACC are likely representative of higher level processing in these listeners (such as cognitive effort and neural efficiency). In contrast, in the current CI study, although there are likely to be individual differences in central speech processing, the individual differences in their peripheral auditory pathway (i.e. neural survival, electrode positioning, etc.) are likely to be considerable. As such, whilst the current ACC measure in CI users provides a meaningful index of performance in an everyday listening context, the effect on the ACC morphology of an interaction between peripheral and central factors is likely to be more heavily weighted towards peripheral factors in CI users than in NH participants.

Phase Coherence

The final part of the study saw that phase coherence could successfully be measured in CI users. There were no correlations seen between phase coherence and the behavioural measures of speech perception or with the ACC response, although again the sample size is quite small for a correlation analysis. However, even the poor performers (IEEE % correct scores as low as 15.6%) showed entrainment to the amplitude envelope of the speech. One might expect that the very poor performers would have lower phase coherence as it is likely that the majority of the stories would have been unintelligible, and phase coherence can sometimes be reduced with unintelligible stimuli (Peelle et al. 2013).

The participants were not asked to attend to the stories and they continued watching the animated film without audio as the stories were played. Due to reduced spectral resolution, in order for CI users to understand speech it is possible that they require more listening effort than NH listeners (Hughes & Galvin 2013; Pals et al. 2013). Without actively listening to the stories, it could be that for the majority of the participants the stimuli would have been unintelligible. Neural oscillations do entrain to the amplitude envelope of unintelligible speech; however entrainment can be greater for intelligible speech (e.g. Peelle et al. 2013). If the stories were unintelligible to all the participants in this study, it is unlikely that the phase coherence would not have predicted speech perception performance even if it potentially could. Furthermore, telling participants not to concentrate on the stories and to watch an animated film does not necessarily mean that they will have all successfully ignored the stories. Phase coherence to speech is effected by selective attention (Kerlin et al. 2010; Gomez-Ramirez et al. 2011; Ding & Simon 2012;

Horton et al. 2014), and so conceivably, some of the variation in the group could have been caused by attentional differences between individuals.

A study similar to this one but where participants had to attend to the speech and answer comprehension questions would mean that the intelligibility of the stories could be measured and attention could be controlled. This would allow for intelligibility to be used as a factor in the analysis, and effects of selective attention could be minimised. Furthermore, a design that controls intelligibility could be useful, by ensuring that the stories were unintelligible to all (e.g. by using tone vocoding as in Millman et al. 2015) or intelligible to all (e.g. by using a sample of CI users with fairly high performance, but still with a range). Although there are still many questions to be answered, this experiment shows that phase coherence can be measured in the majority of CI users and so research in this area is promising.

3.5.1 Conclusion

This study was designed to investigate CI users' performance as close to as when in the real world as possible. Participants were accepted with a variety of processors and fittings, and the study used a broad range of speech stimuli presented through a loud speaker. Participants were also asked to keep their processors on their usual everyday comfortable setting. As such, performance variability in the behavioural, ACC, and phase coherence measures likely stems from three causes; differences in auditory processing associated with the effects of deafness, differences in the peripheral effectiveness of the implanted array (i.e. electrode positioning), and differences in the optimisation of the participant's clinical fitting (i.e. number of channels, stimulation rate, pulse duration, etc.). These three factors are undoubtedly closely linked; for example, a well-fitted implant for any individual will need to factor in the peripheral functioning of the array (e.g. extra-cochlear electrodes), as well as individual preferences for certain MAP parameters (e.g. processing strategies), which while not fully understood, may be linked to both peripheral and central processing abilities. Whilst any CAEP measure alone will be unable to disentangle such variables, if the aim of the measure is to assess optimum fitting parameters for everyday use, then a useful next step would be to investigate whether CAEPs are affected by MAP parameters, and if so, whether changes in CAEPs are indicative of listening preference and speech perception performance.

The results of this study, add to the growing body of research measuring cortical neural activity in the CI population. Results suggest that there are differences between the way the spectral change information is encoded in the auditory cortex between NH and CI listeners, with the ACC in CI listeners likely being more strongly influenced by individual variations in the spectral information encoded in the periphery. However, there may also be some aspect of attention modulating the CI ACC response. Entrainment of low frequency neural oscillations to speech in CI users was successfully measured, giving potential to further research using this technique in this group. Notably, the CI artefact was able to be removed using DSS in all but two participants from all of the EEG data, giving promise to the broader use of electrophysiology in CI research.

4 Native Language Experience and the ACC

4.1 Introduction

So far, this thesis has explored the influence of peripheral auditory factors on the ACC. The current chapter controls the peripheral auditory factors (i.e. normal-hearing, speech in quiet) and addresses a more central factor of speech perception, by measuring responses from listeners that differ in their native language experience. During infancy, listeners start to become attuned to their native language (L1) speech sounds, so that by the time they reach the age of six months, individuals have started to become specialised to detect differences between L1 phonemes (Kuhl et al. 1992; Kuhl 1992, 1998; Werker 1994). It has been suggested that this specialisation can interfere with L2 speech perception later in life, resulting in speech perception difficulties with certain non-native speech sound contrasts (e.g. Ringbom 1992; Best 1995; Flege 1995; Kuhl & Iverson 1995; Hattori & Iverson 2009; Iverson et al. 2012). Exactly where in the speech processing pathway this L1-L2 interference occurs is unknown, and often debated.

Traditionally, theories of L2 speech processing have focussed on the effect of L1-L2 interference at the level of phonological categorisation. For example, Best's (1995) Perceptual Assimilation Model suggests that listeners' perception of L2 speech sounds is driven by how they assimilate these L2 sounds into their L1 phonological categories. Whereas, Flege's (1995) Speech Learning Model suggests that a listener has separate subsystems of phonetic processing for L1 and L2 categories that exist in a common phonological space, and difficulty occurs when an L2 category that the listener is trying to learn is too close to an existing L1 category. However, recent L2 speech processing research has suggested that native language categories can interfere at a lower, more peripheral, level of speech processing than was previously thought. For example, cross language differences have been found as early as the brainstem. When listeners are presented with Mandarin tones, native Chinese listeners show a more robust FFR than English listeners, whereas when they are presented with language neutral variations in pitch, no cross language differences occur (Krishnan et al. 2005, 2009a, 2009b). Encoding at the level of the brainstem is primarily thought to preserve or enhance the spectral and temporal signal transmitted by the auditory nerve, with little high-level processing occurring at this stage in the pathway. Therefore, a cross language effect at this level suggests that differences in speech perception between listeners with different language backgrounds may stem from a more peripheral auditory level than is often thought.

Evidence from neurophysiological studies using the MMN has shown that cross language differences can also be seen in the auditory cortex (Dehaene-Lambertz 1997; Näätänen et al. 1997; Winkler et al. 1999). For example, Näätänen et al. (1997), found a larger MMN response when Finnish participants were presented with the Finnish (and Estonian) vowel /e/ as a standard and the Finnish (and Estonian) vowel /ö/ as a deviant than when the deviant was the Estonian-only vowel /õ/. On the other hand, the MMN elicited for the Estonian participants was about equal in both conditions. Furthermore, Winkler et al. (1999) found that Hungarian native listeners without any experience of Finnish could not behaviourally discriminate between the Finnish vowels /e/ and /ä/, and no MMN response was observed to these stimuli in these subjects, whereas the opposite was found for

Hungarian natives who had learnt Finnish. Dehaene-Lambertz (1997) demonstrated that the MMN is larger for L1 stimuli than for foreign-language phonemes. She found that an MMN response was elicited when French native listeners were exposed to French phoneme contrasts, yet no MMN was found when using a contrast of Hindi phonemes at the same amplitude (Dehaene-Lambertz 1997). Overall, these results support the notion of a language-specific element to the MMN response.

Earlier cortical measures, however, including the onset P1-N1-P2 complex and the ACC are currently not considered to be affected by language experience (Sharma & Dorman 2000; Elangovan et al. 2011; An et al. 2013; Wagner et al. 2013). An et al. (2013) used an /iri/-/ili/ continuum with changing third formant frequency for an identification task and for ACC recording at the change from the initial vowel to the consonant (about 300 ms into the token). They tested English (n = 5), Korean (n = 5), and Japanese (n = 5) participants and didn't find any difference in the ACC in central locations between groups. Wagner et al. (2013) tested the onset P1-N1-P2 complex to word onsets that were legal in the Polish language and illegal in English, in English (n = 12) and Polish (n = 12) listeners. Again they found no difference between the two language groups. They did, however, find a later onset negativity (88-280 ms after onset) that was larger in the Polish group compared to the English group, which they attributed to an attentional effect, such as the broad processing negativity that occurs with selective attention (Hansen & Hillyard 1980; Näätänen & Picton 1987; Näätänen 1990; Alho et al. 1994).

Research so far has shown that cross language differences can be found at the level of the brainstem (FFR; Krishnan et al. 2005, 2009a, 2009b) and at the level of the auditory cortex (MMN: Dehaene-Lambertz 1997; Näätänen et al. 1997; Winkler et al. 1999). Furthermore, there may be cross-language attentional effects that can be measured in the early auditory cortex (Wagner et al. 2013). These results suggest that L1-L2 interference may span the length of the speech perception pathway, beginning with differences with neural encoding of the acoustic characteristics of speech, and ending with differences in perceptual categorisation of speech sounds. However, results of previous studies have shown that earlier cortical measures, such as the ACC are not affected by language experience (Sharma & Dorman 2000; Elangovan et al. 2011; An et al. 2013; Wagner et al. 2013).

Chapters 2 and 3 suggested that when the effect of peripheral differences in detecting spectral changes were controlled (i.e. normal-hearing listeners), the morphology of the ACC was sensitive to higher level processing, such as attention, cognitive effort and neural efficiency. Studies that have investigated the effect of language experience on the ACC have used a narrower range of speech sounds than in this thesis so far. Iverson et al. (2016) used a multi-pair design (as in Chapters 2 and 3), that allowed for observation of acute sensitivities of the ACC to phonological importance of stimuli and modulation of attention, that would go unmeasured with designs that involve just one or two contrasts. They were able to view their ACC results on a two dimensional Euclidean space using multidimensional scaling (MDS). Furthermore, they were able to measure asymmetries in their ACC responses, where the response was different depending on which order the pair of stimuli were presented. The current study explored the scope of this design for revealing cross language differences in the early auditory cortex, using both MDS and order asymmetries.

Iverson et al. (2016) used two stimulus sets, the first contained four voiced and four voiceless fricatives, and the second contained eight vowels, to measure the ACC for all possible pairs within each stimulus set. They then successfully applied multidimensional scaling to the results to produce two dimensional perceptual spaces that related to voicing and place of articulation for fricatives, and height and front-back articulation for vowels. Multidimensional scaling (MDS) is a tool used to view the relative similarity of independent items. With Iverson et al's (2016) solutions, the closer together two fricatives (or two vowels) were on the MDS space, the smaller the ACC response was to that pair and vice versa. They found that the P2 peak amplitude minus the P1 peak amplitude provided the best fit into a two-dimensional Euclidean space for both the vowels and the fricatives. They compared the P2-P1 MDS solutions to solutions based on the cochlear-scaled spectral difference between their stimululus pairs. They found that the MDS solutions were fairly predictable, in that they produced phonetically realistic perceptual spaces (e.g. the P2-P1 MDS vowel spaces looked much like a traditional vowel quadrilateral). However, the P2-P1 MDS spaces had greater representation of place of articulation (for fricatives) and front-back articulation (for vowels) on the horizontal plane, than the MDS solutions based on the spectral difference between the stimuli. The authors suggest that the P2-P1 perceptual spaces were scaled more by the phonological importance of these dimensions than by the pure acoustic difference (Iverson et al. 2016).

Iverson et al. (2016) also found that the ACC response recorded to a pair of phonemes presented in one order was met with an asymmetry when the pair was presented in the opposite order. They describe it as a broad negativity overlapping N1, similar to the processing negativity found in selective attention experiments (Hansen & Hillyard 1980; Näätänen & Picton 1987; Näätänen 1990; Alho et al. 1994; Iverson et al. 2016). They suggest that this asymmetry effect was driven by the low frequency peaks in the stimuli, in that a greater asymmetry was seen when the contrast consisted of one stimulus with a relatively flat spectrum and one with stronger low frequency peaks (i.e. for fricatives this is largely accounted for by voicing as voiced fricatives have greater low frequency peaks than voiceless fricatives). They suggest that these strong low frequency peaks implicitly modulate attention giving rise to a greater negativity when these stimuli occur. Order asymmetries have also been found in MMN experiments where with a pair of vowels, the amplitude of the MMN is larger or smaller depending on which vowel is the deviant and which vowel is the standard (Eulitz & Lahiri 2004; Ylinen et al. 2010; Scharinger et al. 2012). Furthermore, there are consonant (e.g. Kuhl et al. 2006) and vowel asymmetry differences often seen in infant and adult behavioural discrimination data (e.g. Kuhl et al. 1992; Polka & Bohn 2003, 2011), and in word learning studies with consonants (e.g. Altvater-Mackensen & Fikkert 2010; Altvater-Mackensen et al. 2014). Polka & Bohn (e.g. 2011) have found consistent asymmetries where discrimination is easier for vowel pairs going in the direction from the centre to the periphery of the F1-F2 vowel space than the other way around. They suggest that this is related to these peripheral vowels being perceptually salient or stable. Furthermore, they suggest that these asymmetries only occur in adults for non-native contrasts, whereas for contrasts where native language experience requires the listener to perceive the vowels phonemically, the asymmetry effect is small or absent.

The aim of this study was to investigate the effect of language experience on the ACC, using the P1, N1, and P2 peak amplitudes, and the asymmetry magnitude as seen in Iverson et al. (2016). For this thesis, language experience was chosen, as peripheral hearing can be controlled (i.e. normal hearing listeners), whilst speech perception is manipulated without using peripheral factors (i.e. speech in quiet), in order to compliment the experiments in Chapters 2 and 3 (i.e. speech in noise and CI users). In this study, eight voiceless fricatives, ranging from bilabial to velar, were concatenated into sequences that included every possible pair, and were played to English (n = 15), Polish (n = 15), and Finnish (n = 15) native listeners. The ACC response was recorded for each stimulus change. The ACC magnitude was used to create similarity matrices that were analysed by non-metric MDS, and an acoustic analysis of the fricative stimuli was performed to enable the spectra of the stimuli to be compared to the ACC. P1, N1, P2, and asymmetry measures, and their relationship to the spectra of the stimuli were used to investigate cross language differences between the groups.

The three groups of listeners were selected for their varying native language experience of fricatives. Finnish speakers have two voiceless and no voiced fricative sounds in their native language (voiceless: /f/, /s/, and /h/), English speakers have five voiceless and four voiced fricative sounds in their native language (voiceless: /f/, /θ/, /s/, /ʃ/, and /h/; voiced: /v/, /ð/, /z/, and /ʒ/), and Polish have five voiceless and five voiced fricative sounds (voiceless: /f/, /s/, /ʂ/, /ɕ/, /x/; voiced: /v/, /z/, /ʐ/, /ʑ/, and /ɣ/; Campbell & King 2011). Polish has the largest fricative inventory of the three languages, and Finnish has the smallest. Cross language differences in CAEPs are usually realised as a larger response for native stimuli compared to non-native stimuli (e.g. using the MMN; Dehaene-Lambertz 1997; Näätänen et al. 1997; Winkler et al. 1999), however the onset P1-N1-P2 complex and the ACC are currently not considered to be affected by language experience (Sharma & Dorman 2000; Elangovan et al. 2011; An et al. 2013; Wagner et al. 2013). Therefore, we expect that there will be no ACC magnitude differences between the three language groups. On the other hand, Iverson et al. (2016) found an ACC asymmetry effect using this design, and according to Polka & Bohn's (2011) theory, listeners show larger asymmetries to contrasts that are not in their native language. Therefore, although we may not see ACC magnitude differences between the three language groups, it is possible that the separate language groups will show larger asymmetries to contrasts that do not involve sounds that are in their native language.

4.2 Experiment 6: The ACC to Fricatives in English, Finnish, and Polish Listeners

4.2.1 Participants

Fifteen English, 15 Finnish, and 15 Polish speaking, normally hearing adults contributed to the experiment (age range: 19-39). Native speakers were defined as those whose first language was English, Finnish, or Polish, with no other language being spoken in their household before they started school. All participants passed a hearing screening test by demonstrating thresholds of ≤ 20 dB HL for pure-tones presented in quiet at the octave frequencies between 250 and 8000 Hz.

The experiment and recruitment for this study was approved by the University College London ethics committee. All subjects provided their informed consent before beginning the experiment and were paid an honorarium for their time.

4.2.2 Stimuli

Sustained fricatives were recorded from one female trained phonetician who is a native Polish speaker with a high fluency in English. The recordings were made in a sound-proof booth with a 44.1 kHz sampling rate at 16 bits per sample. Speech stimuli manipulations were then conducted in Praat (Boersma & Weenink 2013). For each recorded sound, a 750 ms segment that was deemed to be fairly static (minimal spectral change) was extracted from the recording. The amplitude envelope of each segment was then flattened by calculating the envelope of the original recording (full rectification and a 50-Hz low-pass filter, scaled to a maximum value of 1), dividing the original recording by this envelope, and rescaling the amplitude to be equal across phonemes.

The eight fricatives (Table 4.1) were then concatenated into a random order without replacement (the same stimulus could not appear twice in a row) to create five-minute sequences that included all possible pair transitions (i.e. 56 possible transitions). The concatenation was done by splicing stimuli using 50 ms raised-cosine overlapping transitions. The duration of each stimulus was randomly jittered from 450-500 ms, with the segment randomly selected from the longer 750 ms recordings. This minimised the possibility of any spectral discontinuity within the recorded stimuli from being consistently time-aligned to the spectral changes. After the phonemes were concatenated into a five-minute sequence, the entire amplitude envelope was flattened to correct for amplitude fluctuations during the brief overlapping transitions. Each five-minute sequence was used only once.

The acoustic characteristics of fricatives are sometimes described in terms of spectral moments (mean, variance, skewness, and kurtosis; Jongman et al. 2000). A spectral analysis of the stimuli was run using Praat (Boersma & Weenink 2013) to show the cochlear-scaled spectra and the four spectral moments for each of the fricatives (Figure 4.1 and Table 4.2). Figure 4.1 shows that the stimuli appear to fit into two groups, those with a flatter response who have higher low frequency energy (solid lines) and those with a peakier response who have lower low frequency energy with high frequency peaks (dotted lines).

Fricative Stimuli

	International Phonetic Alphabet	Place of Articulation	Present in Native English?	Present Native Polish?	Present in Native Finnish?
b	ɸ	Bilabial	No	No	No
f	f	Labiodental	Yes	Yes	South Western dialects only
th	θ	Dental non-sibilant	Yes	No	No
s	s	Alveolar sibilant	Yes	Yes	May be retracted (s̠)
ap	ɕ	Alveo-palatal	No	Yes	No
sh	ʃ	Palatal-alveolar sibilant	Yes	No	No
pal	ç	Palatal	No	No	No
x	x	Velar	No	Yes	No

Table 4.1 Details of the eight voiceless fricative stimuli.

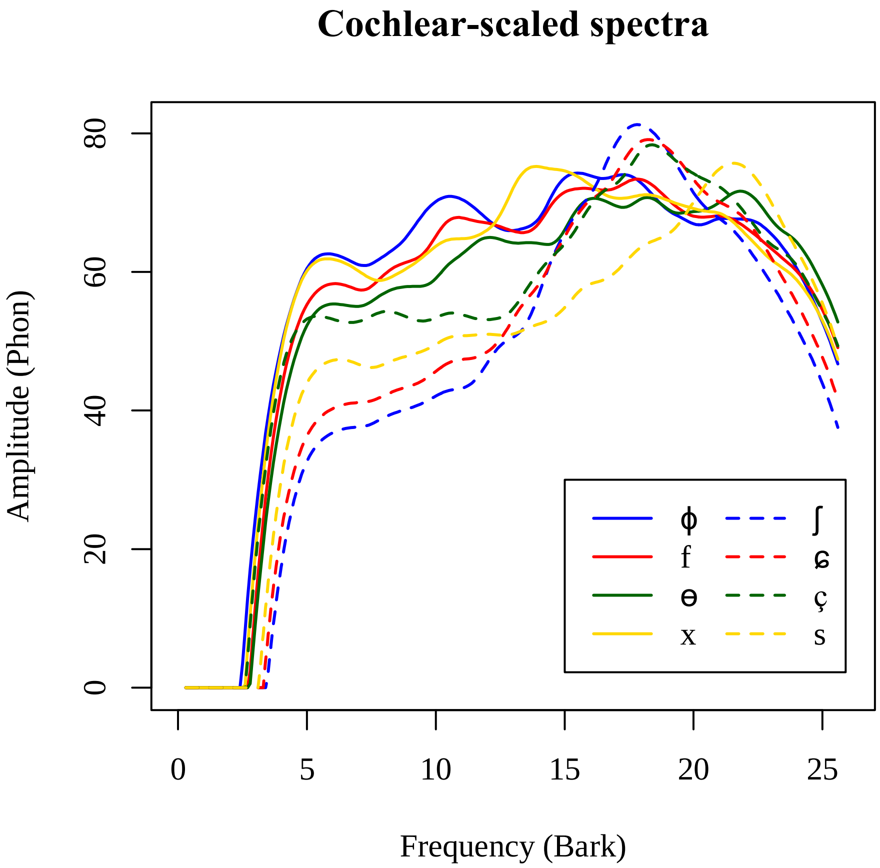

Figure 4.1 Graph showing the cochlear-scaled spectra of each of the eight stimuli. The lines are split into two groups, those with a flatter spectral response (solid lines) and those with a peakier spectral response (dotted lines).

	International Phonetic Alphabet	Centre of Gravity (mean; Hz)	Standard deviation (variance; Hz)	Skewness	Kurtosis
b	ɸ	3907.82	3048.44	1.11	0.34
f	f	4771.23	3431.77	0.95	0.02
th	θ	7354.42	3538.75	-0.01	-0.43
s	s	7755.23	1723.68	0.65	4.56
ap	ɕ	4967.38	1669.15	1.70	4.01
sh	ʃ	4317.18	1142.37	2.96	12.55
pal	ç	5864.44	2602.09	0.99	0.77
x	x	3644.58	2754.76	1.63	1.99

Table 4.2 The four spectral moments of the eight fricative stimuli.

4.2.3 Apparatus

All stimuli were presented binaurally via Praat (Boersma & Weenink 2013). A Fireface UC external audio interface (44.1 kHz sampling rate, 16 bits per sample) connected to Etymotic Research ER-1 insert earphones was used for stimulus presentation via a custom built headphone amplifier. Stimuli were presented at 65 dB SPL, and the sound level was verified with a Brüel & Kjaer 4157 artificial ear for insert earphones.

An additional audio channel was used to provide stimuli timing information to the EEG system. A custom converter was used to convert the audio signal into TTL pulses, which was recorded as a time-aligned trigger by the EEG system. Testing occurred in an electromagnetically shielded sound-proof booth that was kept at 19^{o}C to minimize sweat artefacts. Within the booth, mains electricity was turned off during testing to reduce electrical artefacts as far as possible.

Evoked responses were recorded with a BioSemi ActiveTwo EEG system. Sixty-four active surface electrodes were placed in accordance with the international 10-20 system. Seven additional electrodes were placed on the left and right mastoid, left and right canthus, above and below the left eye, and on the tip of the nose. Responses were recorded at a sampling rate of 2048 Hz.

4.2.4 Procedure

One block consisted of one five-minute sound file that was prepared before the session. This was repeated six times (a new sound file each time) giving a total of approximately 240 repetitions for each possible pair (120 in each direction). During testing, participants watched a silent animated film, and were given the opportunity for a short break every 10 minutes.

4.2.5 Analysis

Data were processed offline in MATLAB (MathWorks 2013a), using the Fieldtrip (Oostenveld et al. 2011), EEGlab (Delorme & Makeig 2004), ERPlab (Lopez-Calderon & Luck 2014), and NoiseTools (de Cheveigné & Simon 2007, 2008a, 2008b; de Cheveigné 2010, 2012, 2016; de Cheveigné & Parra 2014; de Cheveigné & Arzounian 2015) toolboxes. For each electrode channel, data were first high pass filtered at 0.1 Hz (Butterworth filter, as implemented by the ERPlab plugins within EEGlab), and then referenced to the mastoid electrodes. Using

Fieldtrip, the recording was then segmented into epochs spanning from 100 ms before to 350 ms after each stimulus onset. Epochs were then baseline corrected to a 100 ms window before the stimulus change. Data were then low pass filtered at 30 Hz (Butterworth filter, as implemented by Fieldtrip) and downsampled to a 512 Hz sampling rate. Using NoiseTools, DSS was used to extract the ACC components for each subject, using a method that increased the signal-to-noise ratio of the neural signals, by selecting linear combinations of electrodes that maximised the repeatability of the ACC activity. After visual inspection of the DSS components, the first three were selected and the response was taken as the projection of these three components back into sensor space at electrode FCz. Epochs in which the signal amplitude exceeded 150 µV were rejected due to their high noise level, and the remaining epochs were averaged for each pair transition.

Statistical analysis was run using R (R Core Team 2013). For linear mixed-model analyses the lmer function in the R package lme4 (Bates et al. 2015) was used, with type II analysis-of-variance tables calculated using the package CAR (Fox &Weisberg 2011). MDS solutions were calculated using the classical-MDS procedure cmdscale within R (R Core Team 2013), using two dimensions and including a constant. The plotted solutions were translated, rotated, reflected, and scaled, to produce plots that were visually comparable; as MDS solutions only display the relative similarity of items (i.e. the scaling is arbitrary).

4.3 Results

The scalp distributions in Figure 4.2 show that the average response across all pairs for each group is prominent in fronto-central locations, with the largest response being found at FCz. When looking at the average response at FCz for the English, Finnish and Polish groups (Figure 4.2), the English and Finnish responses look to be very similar, whereas the Polish response appears to be larger.

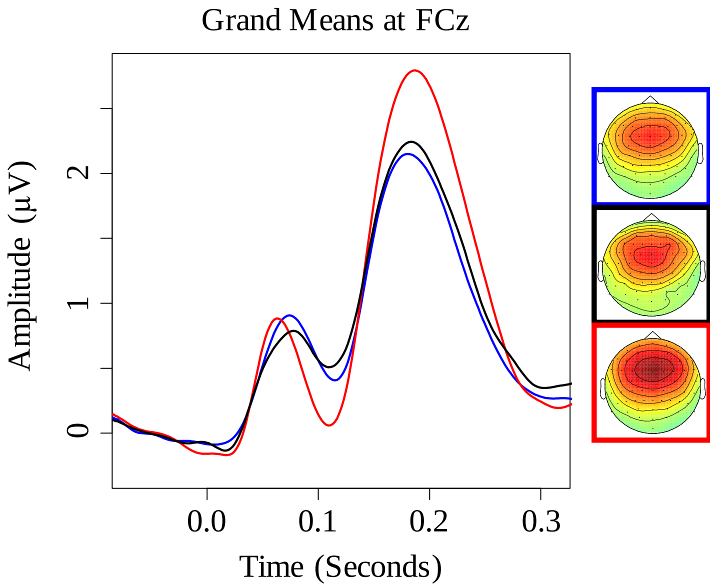

Figure 4.2 Grand mean responses at FCz and scalp distributions for the English (blue), Finnish (black) and Polish (red) groups. The stimulus onset is at 0 s. Scalp distributions are calculated for P2 (150-250 ms), and the colour ranges from -1.5 to +1.5 µV.

Figures 4.3, 4.4 and 4.5 show the ACC responses for each language group for each possible pair of stimuli. It is clear that for all the groups some of the pairs elicit a greater response (e.g. ɸ-s, ɸ-ʃ, f-ɕ and s-x) than others (e.g. ɸ-f, f-θ, f-x and ʃ-ɕ). The pairs that are eliciting a greater response appear to be those that are more spectrally different (Figure 4.1; e.g. /ɸ-s/). Furthermore, there seems to be an effect of presentation of order (as seen in Iverson et al. 2016) between the response of the pair presented in one order compared to in the opposite order, and this asymmetry effect appears to be greater for a pair of fricatives that are more spectrally different (i.e. more negative when going between fricatives with a flat spectrum to ones with high frequency peaks; Figure 4.1).

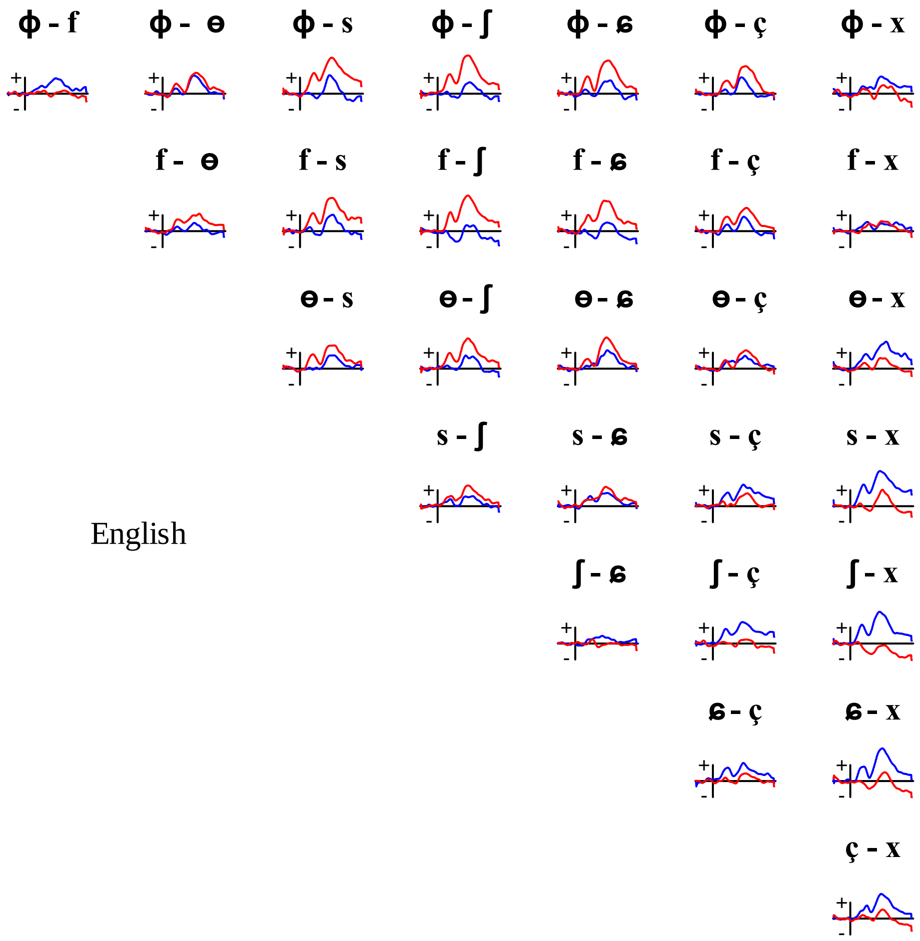

Figure 4.3 English matrix plot for both pair directions (blue = marked order, red = reverse order) at FCz. The y-axis shows +/- 2 µV, the x-axis shows time.

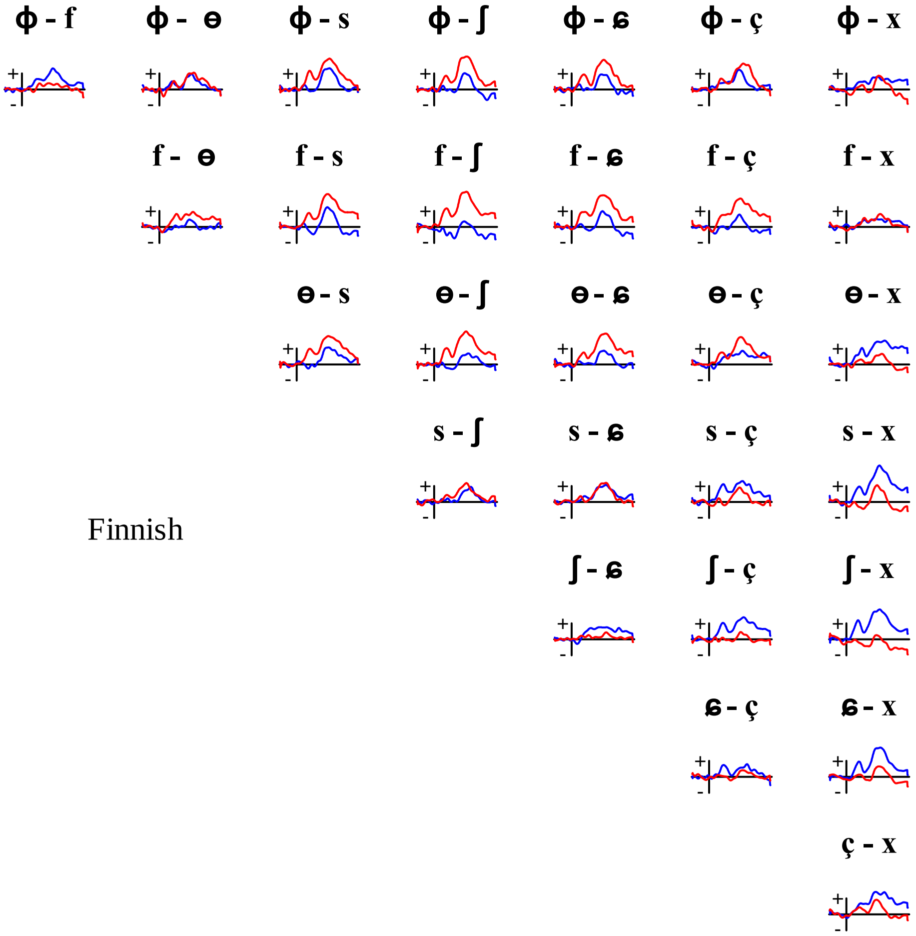

Figure 4.4 Finnish matrix plot for both pair directions (blue = marked order, red = reverse order) at FCz. The y-axis shows +/- 2 µV, the x-axis shows time.

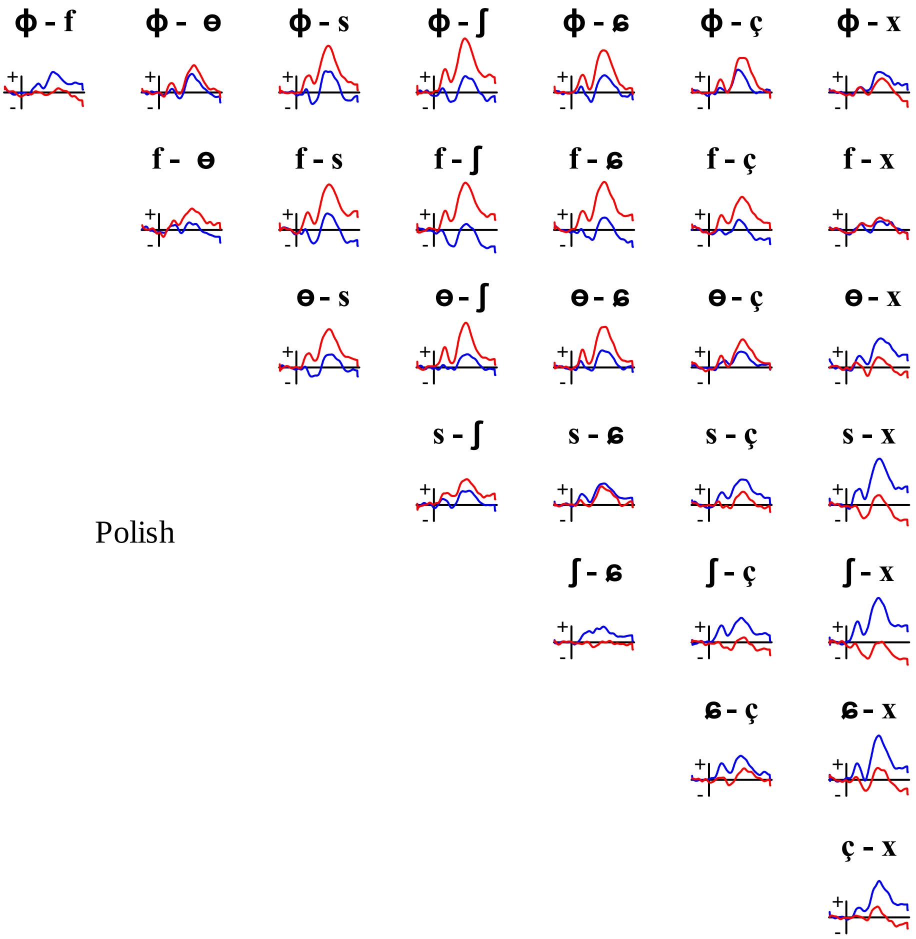

Figure 4.5 Polish matrix plot for both pair directions (blue = marked order, red = reverse order) at FCz. The y-axis shows +/- 2 μV, the x-axis shows time.

Perceptual maps generated by multidimensional scaling

To calculate the peak amplitudes for each participant the maximum (P1 and P2) and minimum (N1) points within a window taken from the relative peaks in the grand mean response (Figure 4.2) were identified, and the peak amplitude was estimated as the average amplitude of the waveform from a 7 ms window around that point (i.e. 3.5 ms either side of the peak).

Classical MDS (Torgerson 1958; Gower 1966) was used to map the perceptual spaces underlying the fricative stimuli. The further away two fricatives are on the space, the greater the magnitude of the ACC response was when measured at the change between these two fricatives. The P2 magnitude provided the best fit into a two-dimensional Euclidean space for all three groups (English: $R^2 = 0.81$; Finnish: $R^2 = 0.67$; Polish: $R^2 = 0.82$). P1 magnitudes provided slightly worse fits (English: $R^2 = 0.81$; Finnish: $R^2 = 0.62$; Polish: $R^2 = 0.80$), and N1 magnitude provided the worst fits of the three peaks (English: $R^2 = 0.62$; Finnish: $R^2 = 0.62$; Polish: $R^2 = 0.36$). Figure 4.6 displays the MDS solutions for all three language groups based on P2 magnitude, and for comparison, the MDS solution based on cochlear-scaled spectral differences between the fricative pairs ($R^2 = 0.99$). The cochlear-scaled spectral difference between the fricatives was calculated by taking the RMS difference between the cochlear-scaled spectra (i.e. Figure 4.1) of each pair of fricatives.

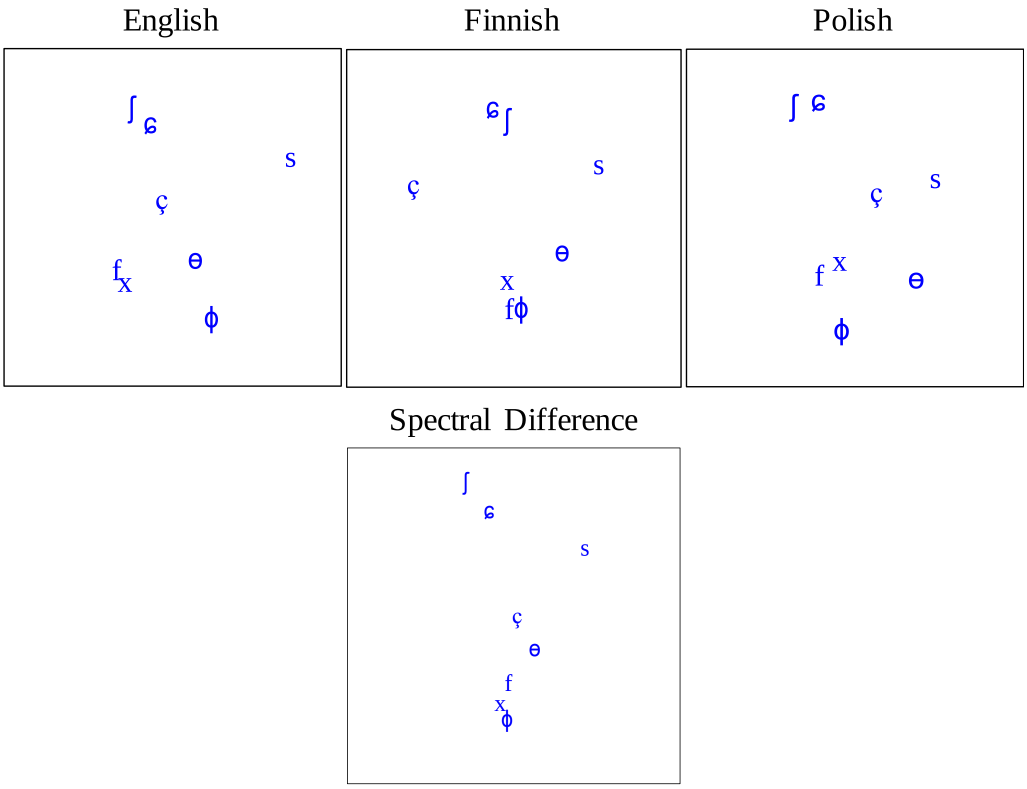

Figure 4.6 Top panel: MDS solutions based on P2 peak values (mean of wave ±7 ms around the peak) at FCz (English: $R^2 = 0.81$, Finnish: $R^2 = 0.67$, Polish: $R^2 = 0.82$). The solutions are based on the P2 peak difference in that the further away two fricatives are on the space, the greater the P2 value was when they were presented as a pair. Bottom panel: MDS solution based on the cochlear-scaled spectral difference between stimuli ($R^2 = 0.99$). The greater the spectral difference between two fricatives the greater the distance between them on the 2D space. There are no axes displayed on the MDS solutions as they are meaningless, and the orientation of the plots are arbitrary; it is the distances between the fricatives that is important and used as a visual representation of the data.

The MDS solutions in Figure 4.6 should be viewed using the distance between items (in this case, fricatives) on the 2D space. For the top panel in Figure 4.6, the further apart two fricatives are, the larger their P2 magnitude was when that pair of fricatives were presented in the ACC stimulus chains. For the bottom panel in Figure 4.6, the further apart two fricatives are, the larger the cochlear-scaled spectral difference is between those two fricatives. For example, in the Finnish MDS solution in the top panel of Figure 4.6, /ç/, /ɕ/, /s/, and /x/ form an approximate square on the space, meaning that they are all an equal distance from each other, and so when each pair of those four fricatives was played in the stimulus chain the resulting P2 magnitude was approximately equal. Whereas, in the same solution, /x/, /ʃ/, and /ɕ/ form a sharp triangle on the space, showing that /ʃ/, and /ɕ/ are approximately the same distance from /x/ (and so had the same size P2 magnitude), but are much closer to each other, showing that the resulting P2 magnitude when they were played in the stimulus chains was much smaller.

The P2 MDS solutions (Figure 4.6) show that for all groups the middle sibilant fricatives (/ʃ/ and /ɕ/) with peakier spectra (Figure 4.1) are towards the top of the space, /s/ and /ç/ are towards the centre of the space in the vertical plane, and /s/ is pulled out to the right in the horizontal plane. For all three groups the frontal /ɸ/, /f/, /θ/, which have flatter spectra (Figure 4.1) are all located towards the bottom of the space. These solutions support the earlier observation that as the difference between the spectra of a pair increases, the ACC magnitude increases. It appears that the general pattern formed by the position of the fricatives is similar for all three languages; however, the Finnish group shows tighter clustering with the frontal fricatives.

The MDS solution based on the cochlear-scaled spectral difference between the fricatives shows a similar differentiation on the vertical plane with the frontal fricatives at the bottom of the space and the sibilant fricatives towards the top. There is a greater differentiation along the vertical axis than the horizontal axis for the spectral difference solutions suggesting that, as for the P2 solutions, the vertical axis represents the spectra of the stimuli. However, there is more differentiation on the horizontal plane for the P2 solutions than the spectral difference solutions. It is not clear what the horizontal plane represents in the P2 MDS solutions. Iverson et al. (2016) found that the horizontal plane of their P2-P1 MDS solutions for fricatives represented place of articulation, however in the current study this does not appear to be the case. For example, /f/ has a labiodental articulation and /x/ has a velar articulation, and both these fricatives are situated close together on the horizontal plane. Alongside their vowel MDS solutions, Iverson et al. (2016) suggest that their solutions are scaled more by phonological importance rather than pure acoustic difference. Although it is not clear what horizontal plane relates to in the current study, there is more differentiation on the horizontal plane for the P2 solutions than the spectral difference solutions, which suggests that the P2 solutions are not purely a representation of the acoustic difference between the fricatives.

Relationship of P1 and P2 and spectral difference differs between language groups

The multiple-pair design of this experiment was used to further explore the relationships of the P1, N1, and P2 amplitudes with the spectral differences between the fricatives, between the three language groups. A mixed-effects model analysis was run using the lmer function in the R package lme4 (Bates et al. 2015). Models were created with a by-subject random intercept, with the peak amplitude averaged across presentation order as the dependent measure (P1, N1, and P2 in separate models), and with language group, stimulus pair, and their interaction as fixed factors. None of the models had a main effect of language (P1: $\chi^2(2) = 0.60$, $p = .74$; N1: $\chi^2(2) = 0.79$, $p = .67$; P2: $\chi^2(2) = 2.82$, $p = .24$), and they all had a main effect of pair (P1: $\chi^2(27) = 161.75$, $p < .001$; N1: $\chi^2(27) = 87.21$, $p < .001$; P2: $\chi^2(27) = 1050.35$, $p < .001$). Interestingly, there was a significant interaction between language and pair for the P2 amplitude ($\chi^2(54) = 84.22$, $p < .01$), suggesting that the P2 peak amplitude was not the same for all three languages for every pair.

To look at the interaction between language and stimulus pair for P2, a one-way Anova was run on each individual pair between the three groups. The results showed that the strongest interaction occurred where the Finnish group had a smaller P2 for /ɸ-s/ ($F = 5.80$, $p = .01$), the English has a smaller P2 for /f-ç/ ($F = 3.51$, $p = .04$), and the Finnish had a smaller P2 and the Polish had a larger P2 for /f-ɕ/ ($F = 3.55$, $p = .04$). There is also a milder cross language difference (although not significant) where the Finnish had a smaller P2 and the Polish had a larger P2 for /ɸ-ʃ/ ($F = 2.67$, $p = .09$), and where the Polish had a larger P2 for /θ-s/ and /ʃ-x/ ($F = 3.08$, $p = .06$; $F = 2.74$, $p = .08$). When looking at the pairs where the interaction is occurring it is not clear what is causing this cross language difference. It does not appear to be related to the different fricative inventories of the three language groups. For example, the interaction is not occurring only with contrasts that are native for one or two of the language groups or non-native for one or two of the language groups. These results are somewhat clarified by relating them to the P2 MDS solutions. The Finnish group had a smaller response for pairs involving the frontal fricatives than the other two language groups, and this is seen by tighter clustering in their MDS solution for the frontal fricatives. The Polish group's MDS, on the other hand, shows a greater differentiation between the fricatives, particularly the frontal fricatives, and this is relayed in the P2 amplitude cross language differences, where for certain contrasts (usually involving pairs of frontal and sibilant fricatives), the Polish response is larger than the other two groups. Furthermore, in general, the interactions are only occurring for pairs with a large spectral difference (i.e. for contrasts between fricatives at the top and bottom of the spectral difference MDS solution, and not for contrasts where the fricatives are both at the top or both at the bottom, Figure 4.6).

As these cross language differences appear to occur with pairs where there was a greater spectral difference, an analysis using the cochlear-scaled spectral difference between the fricatives was considered. As for the MDS solution, the cochlear-scaled spectral difference between the fricatives was calculated by taking the RMS difference between the cochlear-scaled spectra (i.e. Figure 4.1) of each pair of fricatives. Models were created with by-subject and by-pair random intercepts, with the peak amplitude as the dependent measure (P1, N1, and P2 in separate models), and with language group, cochlear-scaled spectral difference, and their interaction as fixed factors. For P1, there was no main effect of language ($\chi^2(2) = 0.59$, $p = .74$), but there was a main effect of spectral difference ($\chi^2(1) = 13.10$, $p < .001$), and there was a significant interaction between language and spectral difference ($\chi^2(2) = 13.75$, $p < .01$). For N1, there was a main effect of spectral difference ($\chi^2(1) = 8.33$, $p < .01$), but there was no main effect of language ($\chi^2(2) = 0.79$, $p = .67$), nor a significant interaction between language and spectral difference ($\chi^2(2) = 2.80$, $p = .25$). For P2, there was no main effect of language ($\chi^2(2) = 2.81$, $p = .25$), but there was a main effect of spectral difference ($\chi^2(1) = 30.72$, $p < .001$), and a significant interaction between language and spectral difference ($\chi^2(2) = 30.54$, $p < .001$).

The main effects of spectral difference on all three peaks represents a positive relationship between the size of the response and the cochlear-scaled spectral difference between the pair of fricatives (P1: $r = 0.46$; N1: $r = -.033$; P2: $r = 0.66$). Furthermore, of the three peaks, P2 showed the strongest relationship between spectral difference and peak amplitude. This supports Iverson et al. (2016) who found that, of the three peaks, the P2 showed the most consistent relationship with spectral difference between pairs of vowels and fricatives.

The interaction between language and spectral difference for the P1 and P2 peaks is explored in Figure 4.7. The interaction plots show that for both the P1 and P2, the Polish group has a steeper slope of peak amplitude and spectral difference than the other two groups. For the P1, when the spectral difference between the fricatives in the pair was larger, the Finnish group had a smaller P1 amplitude than the other two groups, but when the spectral difference was smaller, the Polish had a smaller P1 than the other two groups. For the P2, when the spectral difference between the fricatives in the pair was small, all three groups had a similar sized P2, whereas when the spectral difference was large, the Polish group had a bigger P2 than the other two groups. To explore the interaction further, Pearson correlations were used to describe the relationships between the P1 and P2 peak amplitudes and the cochlear-scaled spectral difference between the fricatives, averaged across subjects for each of the three language groups. Significance was tested using linear mixed-models with by-subject and by-pair random intercepts, the peak as the dependent measure, and cochlear-scaled spectral difference as the independent measure. A different model was calculated for each language group. Both P1 and P2 were significantly correlated with spectral distances for the English (P1: $r = 0.47$, $\chi^2(1) = 39.37$, $p < .001$; P2: $r = 0.69$, $\chi^2(1) = 24.15$, $p < .001$), Finnish (P1: $r = 0.25$, $\chi^2(1) = 7.46$, $p < .01$; P2: $r = 0.65$, $\chi^2(1) = 19.74$, $p < .001$), and Polish groups (P1: $r = 0.70$, $\chi^2(1) = 26.62$, $p < .001$; P2: $r = 0.77$, $\chi^2(1) = 39.37$, $p < .001$). The Polish group's P1 and P2 responses had a stronger relationship to spectral difference than the other two groups' (P1: vs English $Z = -0.027$, $t = -2.62$, $p = .0090$, vs Finnish $Z = -0.037$, $t = -3.58$, $p < .005$; P2: vs English $Z = -0.052$, $t = -4.00$, $p < .001$, vs Finnish

$Z = -0.069$, $t = -5.30$, $p < .001$). Whereas, there was no significant difference when comparing the Finnish group's to the English group's P1 and P2 responses' relationship to spectral difference (P1: $Z = -0.010$, $t = -0.97$, $p = .33$; P2: $Z = -0.017$, $t = -1.31$, $p = .19$).

In the pair wise analysis, the Finnish group had smaller P2 responses for pairs involving the frontal fricatives and their MDS solution showed tighter clustering of the frontal fricatives than the other two language groups. However, this effect seems to be independent from the spectral difference between the pairs, as for the P2 peak, the English and the Finnish groups have similar slopes, whereas it is the Polish group that is creating the interaction (Figure 4.7). On the other hand, the pair wise result for the Polish group can be related to the interaction between language and spectral difference for the P2. The Polish group showed a greater differentiation between fricatives in their P2 MDS solutions (particularly the frontal fricatives), they showed a larger P2 amplitude than the other two groups for certain contrasts (mainly between frontal and sibilant fricatives where there is a greater spectral difference), and they had a larger P2 amplitude than the other two language groups when the spectral difference between the fricatives was large.

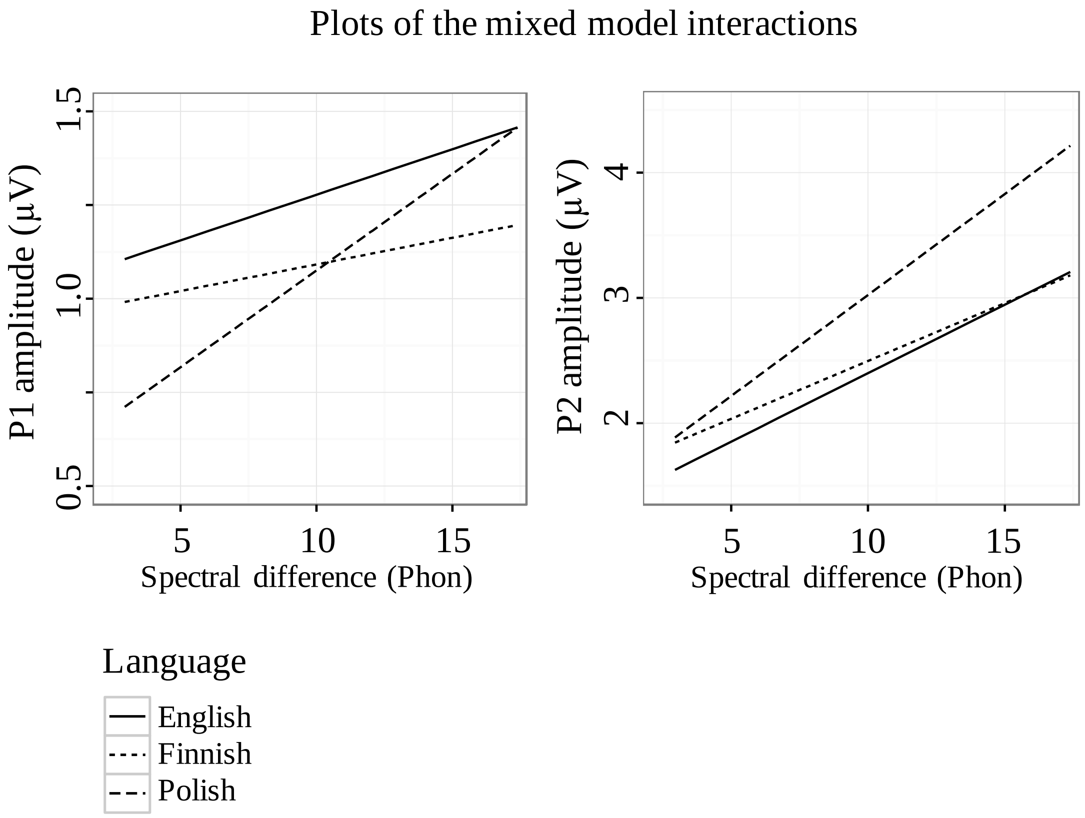

Figure 4.7 Plots showing the interactions between language group and spectral difference between the fricatives for the P1 (left), and P2 (right) amplitudes, by showing linear regression lines of spectral difference versus P1/P2 peak amplitude for each language group.

Relationship of asymmetry and spectral peakiness differs between language groups

To explore the effect of presentation order seen in figures 4.3, 4.4, and 4.5, the asymmetry was calculated by taking the mean difference of the response when the pair was presented in one order compared to the other. For this analysis, the cochlear-scaled spectral difference between the fricatives could not be used as a measure of the acoustic characteristics of the stimuli as it is a value without a direction (i.e. the mean loudness of the frequency components across the whole spectrum for a fricative on its own has little meaning, whereas it is the difference between the cochlear-scaled spectra of two fricatives that is interesting). The acoustic characteristics of fricatives are sometimes described in terms of spectral moments (mean, variance, skewness, and kurtosis; Jongman et al. 2000). The values of these moments alone have meaning. For example, a fricative with a low spectral standard deviation and a high kurtosis has a peakier spectrum than one with a high standard deviation and a low kurtosis. The values of the spectral moments for each of the fricatives are described in the stimulus section (Table 4.2). The spectral moments can be used in analysis with the asymmetry values, as the difference between the spectral moments between a pair of fricatives has direction. For example, a negative difference in standard deviation for a pair of fricatives would mean that the first fricative had a lower standard deviation than the second, and then when presented in the opposite order the difference in standard deviation would be positive.

Linear mixed models were created with the magnitude of the asymmetry as the dependent measure, each spectral measure as independent factors in separate analyses, and by-subject and by-pair random intercepts. There were significant effects for centre of gravity ($\chi^2(1) = 4.46$, $p = .035$), standard deviation ($\chi^2(1) = 85.93$, $p < .001$), skewness ($\chi^2(1) = 4.53$, $p = .033$), and kurtosis ($\chi^2(1) = 36.63$, $p < .001$). In all models, there was a greater negativity when the stimuli changed from a relatively flat spectrum (e.g. /f/ or /ɸ/) to one with a peakier spectrum (e.g. /ʃ/ or /s/). The peakiness of the fricatives can be described using the spectral standard deviation, or kurtosis, in that the smaller the standard deviation and the larger the kurtosis, the peakier the fricative. When these four models were compared, the best-fitting model had standard deviation as the independent measure ($\chi^2(0) = 63.34$, $p < .001$) and so this spectral moment was used for further analysis.

To explore any cross language differences in the asymmetries, a model was created with a by-subject random intercept, with the asymmetry magnitude as the dependent measure, and with language group, spectral standard deviation, and their interaction as fixed factors. There was no main effect of language ($\chi^2(2) = 1.39$, $p = .50$), but there was a significant main effect of spectral standard deviation ($\chi^2(1) = 1165.45$, $p < .001$), and a significant interaction between language and standard deviation ($\chi^2(2) = 38.40$, $p < .001$).

A Pearson correlation showed that the main effect of spectral standard deviation on the asymmetry magnitude represents a strong negative relationship ($r = -0.85$, $t = -14.51$, $p < .001$). This means that the more positive the asymmetry, the more negative the difference in spectral standard deviation between the stimulus pair is (i.e. the response is more negative when going from a fricative with high standard deviation to one with low standard deviation than the other way around). The interaction between language and standard deviation for the asymmetry magnitude is explored in Figure 4.8. The interaction plot shows that the English and Finnish groups have slopes that are very similar, whereas the Polish group's shows a different, steeper slope between asymmetry magnitude and standard deviation. To explore the interaction further, Pearson correlations were used to describe the relationship between the asymmetry magnitude and the standard deviation between each pair of stimuli across subjects for each of the three language groups. Significance was tested using linear mixed-models with by-subject and by-pair random intercepts, asymmetry as the dependent measure, and standard deviation as the independent measure. A different model was calculated for each language group. Asymmetry and standard deviation were significantly correlated for all three of the language groups, with the Polish having the strongest relationship between asymmetry and standard deviation of the three groups (English: *r* = -0.84, $\chi^2(1)$ = 66.57, *p* < .001; Finnish: *r* = -0.86, $\chi^2(1)$ 83.35, *p* < .001; Polish: *r* = -0.87, $\chi^2(1)$ = 89.53, *p* < .001).

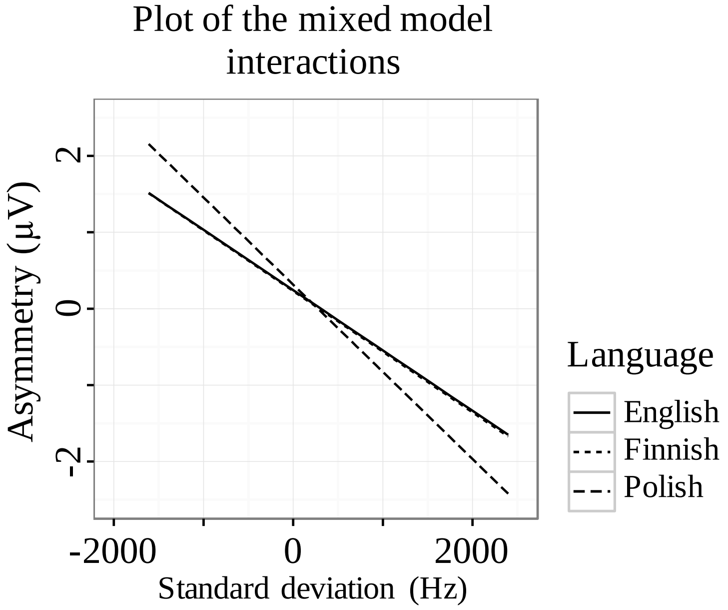

Figure 4.8 Plot showing the interactions between language group and the spectral standard deviation between the fricatives for the asymmetry magnitude, by showing linear regression lines of spectral standard deviation versus asymmetry magnitude for each language group.

4.4 Discussion

Results of this study showed that the amplitudes of the P1, N1, and P2 were all related to the cochlear-scaled spectral difference between fricatives, with the P2 having the strongest relationship. Previously, the ACC was not thought to be effected by language experience (Sharma & Dorman 2000; Elangovan et al. 2011; An et al. 2013; Wagner et al. 2013), however in the current study, the relationships of the P1 and P2 to the spectral difference between fricatives were affected by the language background of the listener. Furthermore, there was an order asymmetry effect, in that the ACC response from a fricative with a flatter spectrum (higher spectral standard deviation) followed by a fricative with a peakier spectrum (lower spectral standard deviation) was more negative than the response to the same fricatives presented in the opposite order. Furthermore, there was also a cross language difference for the asymmetry effect.

Fricative inventory size

In general, the Polish group were responsible for the cross language differences. Compared to the other two language groups, the Polish group showed a greater differentiation between fricatives (particularly the frontal fricatives) in their P2 MDS solutions, they showed a larger P2 than the other two groups for certain contrasts (mainly between frontal and sibilant fricatives where there is a greater spectral difference), they had a larger P2 than the other two language groups when the spectral difference between the fricatives was large, and they showed a stronger relationship between asymmetry and standard deviation than the other two groups. It is possible that this is because the Polish language has a larger fricative inventory than the other two language groups. The Polish language includes five voiceless fricatives ranging from labiodental (/f/) to velar (/x/) and a further five voiced fricatives ranging from front to back. Although the English language also has five voiceless fricatives (including /h/), these generally have a relatively frontal place of articulation, in that apart from /h/ they are all articulated in front of the hard palate. Furthermore, the English language only contains another four voiced fricatives, which again are articulated in front of the hard palate. The Finnish language has the smallest fricative inventory size of the three language groups, containing only three voiceless fricatives, including /h/, one of which is only used in some southern dialects or in loan words (/f/; Campbell & King, 2011).

If fricative inventory size is driving the cross language differences in this study, one would expect the Finnish group to have shown smaller responses where the Polish group showed larger ones. For some contrasts, the Finnish group did have a smaller P2 than the other two language groups (in particular, contrasts containing frontal fricatives), however this effect was not consistent. A possible reason for this would be that while Finnish listeners only have a small number of fricatives in their native language, in reality, most adult native Finnish speakers have vast English experience and a high level of proficiency. So although Finnish listeners have a small native fricative inventory, overall, they have experience nearing that of a native English speaker. This means that their behaviour may lean towards being more like a bilingual, rather than a monolingual with English as an L2 (Segalowitz & Hulstijn 2005).

Implications for the P1, N1, and P2

Previously, the P1 has not been thought to be sensitive to the magnitude of acoustic differences (Čeponienė et al. 2005) and it is often described as purely an obligatory or gating response to the onset of a sound (e.g. Pratt 2012). However, in this study, the amplitude of the P1 increased with increasing spectral difference between fricatives. Furthermore, there was an interaction between language group and spectral difference for the P1 amplitude. When the spectral difference between the fricatives in the pair was larger, the Finnish group had a smaller P1 amplitude than the other two groups, but when the spectral difference was smaller, the Polish group had a smaller P1 than the other two groups. It is possible that the P1 is not purely at a level of general auditory perception as is often thought (e.g. Picton & Hillyard 1974; Pratt 2012) and this cross language difference gives evidence of speech specific processes occurring in the early auditory cortex. Furthermore, results from Chapter 2 suggested that the P1 could be affected by cognitive effort and attention, suggesting that the P1 is susceptible to modulation by higher level cognitive processes. However, cross language differences have been seen as early as the brainstem using FFR (Krishnan et al. 2005, 2009a, 2009b), suggesting that differences in speech perception between listeners with different language backgrounds may stem from a more peripheral auditory level than is often thought. Therefore, this cross language difference with the P1 could also be a reflection of these peripheral auditory cross language differences. Although it is not clear exactly where the cross language differences with the P1 amplitude are stemming from, the results of this study show that with a broad set of speech stimuli, these early speech specific differences can also be measured in the early auditory cortex before linguistic processes such as phonetic categorisation have occurred or completed.

The N1 did increase in amplitude with increasing spectral difference, however, this relationship was the weakest of the three peaks, and the MDS solutions created using the N1 had the poorest fits. This is surprising, as it is generally accepted that the N1 is driven by the magnitude of the acoustic change (Näätänen & Picton 1987; Picton 1990; Steinschneider & Dunn 2002), so it would be expected that at least compared to the P1, it would show a stronger relationship to spectral difference. Furthermore, again since the P1 showed cross language differences, it would be expected that the N1 would also show these effects as it occurs later in the response (so further along the pathway), but this was not the case. The asymmetry effect seen in the current experiment resembled that seen by Iverson et al. (2016), and they suggested that the asymmetry effect interfered with the relationship between the N1 and the magnitude of spectral change. They suggest this, because the asymmetry effect resembled the processing negativity found in selective attention experiments (Näätänen 1990; Alho et al. 1994; Iverson et al. 2016), and since the amplitude of the N1 is known to be modulated with attentional focus and the predictability of the stimuli (Näätänen & Picton 1987; Hillyard et al., 1973; Woldorff & Hillyard 1991), conceivably the asymmetry effect could have interfered with the relationship between N1 and the magnitude of spectral change. In both Iverson et al. (2016) and the current study, the stimuli were presented in a random order and so it is unlikely that predictability of the stimuli would have suppressed the N1 response. However, if the asymmetry effect is driven to some extent by a modulation of attention, it could have interfered with the relationship between N1 and the magnitude of spectral change.

Iverson et al. (2016) suggest that the P2 can be used to generate phonetically realistic multidimensional perceptual maps. The results of this study support their finding using voiceless fricatives. The solutions largely represented the spectral difference between the stimuli. However there was more differentiation on the horizontal plane for the P2 solutions than the spectral difference MDS solutions, suggesting that the P2 solutions are not purely a representation of the acoustic difference between the fricatives. Previous research has suggested that the P2 may have a role in the classification or categorisation of auditory stimuli, as it can be modulated by attention in order to identify a stimulus as a target (Garcia-Larrea et al. 1992; Novak et al. 1992). The results of the current study support the notion of the P2 being a measure that is affected by linguistic aspects of the stimuli rather than purely auditory processing, as it had a strong language interaction, in that for pairs with a larger spectral difference, the Polish group had a larger P2. Furthermore, the results of this study show that the P2 was the most consistent of the three peaks with a stronger relationship to spectral difference (supporting Iverson et al. 2016) and also with stronger cross language differences than the P1 and N1.

Order asymmetries

There was an order related negativity (the asymmetry effect) whereby a response from a fricative with a flatter spectrum (i.e. high spectral standard deviation and low kurtosis) followed by a fricative with a peakier spectrum (i.e. low spectral standard deviation and high kurtosis) was more negative. Iverson et al. (2016) suggest that the asymmetry effect was acoustically driven, in that certain types of spectra may be inherently more likely to draw attention (in particular those with low frequency spectral peaks). However, in the current study, this order related negativity also showed a cross language effect, in that when the difference in spectral standard deviation between a pair of fricatives was large, the Polish group showed a larger asymmetry than the other two groups. This could suggest that the asymmetry effect is not purely caused by acoustic differences between the stimuli that modulate attention, but rather higher level linguistic processes. However, as previously mentioned, a cross language difference does not automatically rule out peripheral acoustic factors, as cross language differences can be seen as early as the brainstem (Krishnan et al. 2005, 2009a, 2009b). On the other hand, it is likely that the cross language difference seen here is not purely due to acoustical factors, but rather to do with the modulation of attention. If in this study attention is modulated by the perceptual salience of the fricatives, it more likely that the level of attention modulation is different between language groups, rather than the way in which the listeners are receiving the acoustic information of the spectral changes is different between language groups.

Polka & Bohn (e.g. 2011) have found asymmetries in both adults and infants where discrimination is easier for vowel pairs going in the direction from the centre to the periphery of the F1-F2 vowel space than the other way around. They suggest that this is related to the peripheral vowels being perceptually salient or stable. The asymmetries seen in the current study are driven by the spectral peakiness (standard deviation and kurtosis) of the fricatives, in that when a fricative with a peakier spectrum follows one with a flatter spectrum, the fricative with the peakier spectrum is more perceptually salient and so inherently draws more attention, causing a more negative ACC response. Therefore, it seems possible that the asymmetries seen with the ACC to speech are similar to the asymmetries seen with vowel discrimination, as they are both driven by the perceptual salience of the speech sounds. However, Polka & Bohn (2011) have also found that adults' asymmetries fade or disappear for native contrasts and are maintained or enhanced for non-native contrasts. Based on Polka & Bohn's (2011) theory, one would have expected that the separate language groups would have shown larger asymmetries to contrasts that do not involve sounds that are in their native language. However, this was not the pattern seen in the current study. The cross language difference did not only occur for contrasts that are native or non-native for one of the language groups. In fact, for the contrasts where an asymmetry effect occurred, this effect occurred for all three language groups; it was the size of the asymmetry between language groups that caused the cross language differences, not whether or not the asymmetry was present. Therefore, it seems that the vowel discrimination asymmetries and the fricative ACC asymmetries are not necessarily the same. However, aspects of Polka & Bohn's approach to investigating the level of speech processing that is causing their asymmetries could be applied to the ACC asymmetries.

Polka & Bohn (2011) suggest that the asymmetry effect in adults only occurs with tasks that tap into phonetic or phonemic processing of speech and not for tasks that facilitate general auditory processing of speech. For example, they conducted a series of experiments with English and German adults using the German vowel contrasts /u/–/y/ and /ʊ/–/ʏ/ (Polka et al. 2005; Polka & Bohn 2011). They used a go/no go task where the vowels were presented in /dVt/ syllables and the listeners had to respond at the end of each trial whether the vowel in a sequence of four had changed or not. They found that English adults showed asymmetries in their performance, whereas the Germans' performance was symmetrical. They then reduced the inter stimulus interval (ISI) from 2.0 s to 500 ms, as reducing the ISI changed the task from one that facilitated phonetic processing of speech to one that increased access to auditory memory and reduced demands on phonetic encoding (Werker & Logan 1985; Werker & Tees 1984; Polka & Bohn 2011). They then found that the asymmetries in the English group disappeared (i.e. neither group showed an asymmetry effect) and so they conclude that the asymmetry effect seen with vowel discrimination is not one of general auditory processing, but rather accesses phonetic processing of speech. In the current study, it is not clear where the asymmetry effect is stemming from, whether it is an acoustically driven effect where the cross language differences reflect acoustic cross language differences (as measured in the brainstem; Krishnan et al. 2005, 2009a, 2009b), or whether it is an effect of phonetic or phonemic processing of speech and the cross language differences reflect this. A study investigating the language background of listeners using the ACC with a multiple contrast design, where the stimuli varied in spectral peakiness (i.e. standard deviation and kurtosis), but could not be categorised as phonemes of speech, could be useful in exploring the origins of the asymmetry effect and the cross language differences seen in this chapter.

Behavioural perception and MDS solutions

MDS solutions, such as the ones in this study, can be created using behavioural similarity rating data. This is where the listener hears every possible pair of the stimuli, and rates (on a fixed scale) how similar they think the stimuli in the pair sound. An example of how behavioural similarity data can be used to create MDS solutions that represent listeners' perception of speech stimuli can be found in Iverson et al. (2003). These authors ran an experiment using the English /r/-/l/ contrast in Japanese speakers (who have well documented problems with this contrast), German speakers (who do not have any problem with it but do not have an English /r/ in their language), and American English speakers. Stimuli were created along a continuum with a changing F2 and F3, to create a two-dimensional grid of speech tokens with equal spacing (based on the Mel scale). Listeners identified each of the stimuli in terms of their own language phoneme categories and rated whether they were a good exemplar of that category from 1 (bad) to 7 (good). They were also played all possible pairs of the stimuli and asked to rate how similar they were from 1 (dissimilar) to 7 (similar). The results were then analysed using MDS. The MDS solutions showed that the American English two-dimensional space was pulled apart in the middle, at the category boundary, with tokens that were identified as /r/ grouped together and tokens identified as /l/ grouped together, despite the physical acoustic distances between the tokens being equal. On the other hand, the Japanese MDS solution showed that the listeners were identifying nearly all the tokens as the same phoneme (alveolar flap) and so the distances between the stimuli on the MDS space were all very similar, much more like the physical stimulus grid. It this case, the MDS solutions were used to demonstrate how the listeners' perception of the stimuli was warped based on their native language.

In the current study, in order to create behavioural MDS solutions, a task could be run where listeners would hear every possible pair of fricatives and would be required to rate how similar the two fricatives sound. The results could then be plotted on a two-dimensional space, where the more similar the fricatives sounded to the listeners, the closer they would sit on the space and vice versa. The behavioural MDS solutions could then be compared to the MDS solutions created using the P2 magnitude and the acoustic differences between the stimuli (Figure 4.6). If the listeners are using the overall cochlear-scaled spectral difference between the stimuli to judge how similar they sound then it would be expected that the behavioural MDS solutions would look similar to the solutions in Figure 4.6, with the fricatives with the peakier spectra grouped together, and the fricatives with the flatter spectra grouped together. On the other hand, it is possible that the listeners would judge the similarity of the fricatives using other cues, such as place of articulation, where, for example, frontal fricatives would be grouped together as would the palatal fricatives. A behavioural measure of similarity in this study would be useful to see if the ACC responses related to the listeners' behavioural perception of the stimuli, or if the cues that drive perception are different to those that drive the ACC response.

4.4.1 Conclusion

The broad stimulus set and multi-pair design of this study has revealed cross language differences in the ACC that have not been previously seen. In particular, the P2 showed a consistent relationship to the spectral difference between fricatives, and this relationship was modulated by native language experience. Furthermore, an asymmetry effect was seen that appears to be a result of attention modulation to the salient acoustical features of speech sounds, and this effect was also influenced by native language experience. The findings of this study indicate that there is potential for the ACC with a multiple contrast design to be used to explore the underlying neural processes behind cross language differences.

5 General Discussion

The aim of this thesis was to investigate how the speech-evoked ACC is affected by manipulations of listening condition, and to explore the potential benefit of the measure for wider use in research and in the clinic. The work presented in this thesis described three experiments that investigated the ACC in normal-hearing listeners in noise, cochlear implant users, and second language listeners.

Chapter 2 detailed a study using a stimulus combination of four vowels and four fricatives in quiet and in noise to investigate how the ACC and the P1-N1-P2 complex in normal-hearing listeners are affected by background noise, and how they relate to individual differences in SIN performance. The results of this study showed that the underlying mechanisms of the ACC provided a more detailed representation of SIN than the P1-N1-P2 onset response. Findings suggested that noise affects the various cortical peaks in a differential manner and that the general acceptance of the early cortical responses being purely obligatory may not be definitive. Furthermore, results indicated that larger P1 amplitudes may reflect aspects of greater pre-conscious attention and cognitive effort, and lower neural efficiency in poorer SIN performers. Whereas, larger N1 amplitudes may reflect a greater resilience to noise and a better cortical representation of the acoustic information in the speech signal by better SIN performers.

Chapter 3 described a study that presented cochlear implant (CI) users with the same fricative vowel stimuli as in Chapter 2, but presented in quiet only. Here, the ACC was measured in a group of CI users and was compared to their behavioural speech perception performance. The CI users' neural entrainment to continuous speech was also measured and compared to their speech perception performance and

ACC responses. The results of this study add to the growing body of research on the use of measuring cortical neural activity in the CI population. Results suggested that there are differences between the way the spectral change information is encoded in the auditory cortex between NH and CI listeners, with the ACC in CI listeners being dominated by the spectral information they are receiving. However, there may also be some aspect of attention modulating the CI ACC response. Entrainment of low frequency neural oscillations to speech in CI users was successfully measured, giving potential to research using this technique in CI users. Most notably, the CI artefact was able to be removed using DSS in all but two of the participants, from all of the EEG data, giving promise to the broader use of electrophysiology in CI research.

Chapter 4 detailed an ACC study using a stimulus combination of eight voiceless fricatives played to native English, Finnish and Polish speakers. The ACC magnitude was used to create similarity matrices that were analysed by non-metric MDS, and an acoustic analysis of the fricative stimuli was performed to enable the spectra of the stimuli to be compared to the ACC. P1, N1, P2, and asymmetry measures, and their relationship to the spectra of the stimuli were used to investigate cross language differences between the groups. The results revealed cross language differences in the ACC that have not been previously seen. In particular, the P2 showed a consistent relationship to the spectral difference between fricatives, and this relationship was modulated by native language experience. Furthermore, an asymmetry effect was seen that appeared to be a result of attention modulation to the salient acoustical features of the speech sounds, and this effect was also influenced by native language experience.

5.1 Implications for the P1, N1, and P2

Results of this thesis have reinforced the notion that the cortical peaks represent different aspects of auditory processing (Anderson et al. 2010; Zendel et al. 2015), with different patterns being seen between the peaks in all three studies.

P1

Typically, the P1 is small in adults and so is often left out of analyses (e.g. Kaplan-Neeman et al. 2006; Parbery-Clark et al. 2011; McCullagh et al. 2012). However, the design of the studies in this thesis meant that many trials could be recorded in a short space of time, meaning that the P1 could be reliably measured and included in analyses (the response was very small in CI users, but still fairly reliable). In this thesis, the P1 as part of the ACC has shown a graded response to the reduction in signal to noise ratio, and has shown graded sensitivity to spectral differences between fricatives, challenging views that the P1 is purely a gating response (e.g. Čeponienė et al. 2005; Pratt 2012). Furthermore, it has shown sensitivity to SIN performance and native language experience of the listeners. As the P1 is the smallest of the three peaks, it is possible that these sensitivities cannot be seen without a multiple contrast stimulus set and a large number of trials.

In Chapter 2, the P1 magnitude in quiet was *larger* for poor SIN performers than better SIN performers. Based on results from studies with musicians (Lee et al. 2009; Strait et al. 2009; Zendel et al. 2015) and listening effort (Rao et al. 2010), it is possible that this greater P1 is due to the poor SIN performers subconsciously paying more attention and using more cognitive effort to listen to the acoustic cues in speech compared to the higher SIN performers, in order to aid them in SIN perception. It could be that pre-conscious cognitive effort and attentional processes were effecting the top-down modulation of responses, even though listeners were not actively attending to the stimuli. Of the three peaks, the N1 is most often associated with an increase in amplitude with attention (Hillyard et al. 1973; Näätänen & Picton 1987; Woldorf & Hillyard 1991); however the results of Chapter 2 suggest that the P1 may also be modulated by certain aspects of subconscious attention or effort. It is possible that these effects are not observed in studies where attention is modulated explicitly because they are masked by the broad processing negativity that can occur with attention (Hansen & Hillyard 1980; Näätänen & Picton 1987; Näätänen 1990; Alho et al. 1994).

However, the effect seen on the P1 in Chapter 2 was somewhat contradicted in the following chapter. In Chapter 3, CI listeners showed a smaller P1 (and P2) than NH listeners, and it was suggested that it could have been an effect of an attentional broad negativity modulating the amplitude of P1. Due to reduced spectral resolution, in order for CI users to understand speech they require more listening effort than NH listeners (Hughes & Galvin 2013; Pals et al. 2013). A reduced P1 and P2 amplitude (but a normal sized N1 amplitude) could reflect a combined effect of a broad negativity that occurs with attention (Hansen & Hillyard 1980; Näätänen & Picton 1987), making the P1, N1 and P2 all appear more negative (Näätänen & Michie 1979; Näätänen & Picton 1987; Michie et al. 1990, 1993).

On the other hand, in Chapter 4 the P1 findings of Chapter 2 are supported. In Chapter 4, the amplitude of the P1 increased with increasing spectral difference between fricatives and this relationship was modulated by language experience. In general, the cross language differences seen in Chapter 4 were caused by the Polish having larger responses (P1, P2, and asymmetries) for fricative contrasts with larger spectral differences, which is possibly due to the Polish language having a larger fricative inventory than the English and Finnish languages. However, when the spectral difference between fricatives was small (i.e. contrasts between fricatives with a similar cochlear-scaled spectra) the Polish group had a smaller P1 than the other two groups. Whereas for the same contrasts (i.e. where the spectral difference between the fricatives is small) all three groups had the same sized P2. This P1 result could be explained in a similar way to in Chapter 2 where the P1 was smaller for *better* SIN performers. Perhaps when the spectral difference between the fricatives is small, the Polish listeners use less cognitive effort than the other two groups because their larger native fricative inventory means that they have a greater sensitivity to small spectral differences in fricatives. This reduction in cognitive effort compared to the other two language groups is then reflected in the P1. If this is the case, it indicates that when compared to NH listeners, the CI users in Chapter 3 did not have a smaller P1 because of greater cognitive effort or attention causing a broad negativity that increases the size of the N1 and reduced the sizes of the P1 and P2. The differences in ACC morphology between CI and NH listeners could instead be due to purely peripheral acoustic factors, due to the differences between receiving sound through a CI and through a NH ear. The P1 findings of this thesis as a whole suggest that it is possible that the P1 is not purely driven by general auditory perception as is often thought (e.g. Picton & Hillyard 1974; Pratt 2012), but rather higher level cognitive processes involved in cognitive effort, attention, and neural efficiency are able to modulate its magnitude.

N1

The N1 is thought to reflect the encoding of time-varying aspects of a sound, such as amplitude, spectral, and voice onset time changes (Sharma & Dorman 1999;

Martin & Boothroyd 2000; Sharma et al. 2000; Hoonhorst et al. 2009) and its amplitude varies with the magnitude of the acoustic change (Näätänen & Picton 1987; Picton 1990; Steinschneider & Dunn 2002). Accurate cortical neural representations of the acoustic information in a speech signal are important for speech perception in noise (Anderson et al 2010; Parbery-Clark et al. 2011), and in Chapter 2, the magnitude of the N1 was larger for better SIN performers. This may suggest better cortical representation of the acoustic information in the speech signal for better SIN performers (Parbery-Clark et al. 2011). Furthermore, the N1 was the same magnitude in CI users as in NH listeners, which may suggest that the users were successfully receiving a certain level of the acoustical information of the spectral changes required for successful speech perception.

The N1 was the most reliable peak to measure from CI users in Chapter 3, and it predicted some aspects of SIN ability in Chapter 2. However, in Chapter 4, its relationship with spectral change in fricatives was the weakest of the three peaks, despite there being consistent evidence that it is driven by the magnitude of the acoustic change (Näätänen & Picton 1987; Picton 1990; Steinschneider & Dunn 2002). Furthermore, even though both the P1 and P2 were affected by the language experience of the listener, the N1 was not. This result could be due to salient acoustical features of the stimuli modulating attention, which in turn overshadows complex acoustic and language N1 effects. With selective attention, a broad negativity can occur (Hansen & Hillyard 1980; Näätänen & Picton 1987). In Chapter 4, there was an order related negativity (the asymmetry effect) whereby a response from a fricative with a flatter spectrum (high spectral standard deviation and low kurtosis) followed by a fricative with a peakier spectrum (low spectral standard deviation and high kurtosis) was more negative than the response when they are presented in the opposite order. This asymmetry effect, in part, could be due to the salient acoustic features of the fricatives (i.e. peakiness) being inherently more likely to draw attention meaning that when changing to a peakier fricative, the ACC response is more negative (Iverson et al. 2016). Therefore, these negativity effects of attention in Chapter 4 could have overshadowed any complex acoustic and language N1 effects, meaning that of the three peaks it showed the weakest relationship to the magnitude of the spectral change between fricatives, and it was not modulated by the native language experience of the listener. As it seems possible that implicit attention modulation can occur from the acoustical characteristics of the stimuli, N1 measurement and effects in future research may need to be handled with extra caution.

P2

This thesis supports that the P2 magnitude is graded with acoustic difference. In Chapter 2, the P2 magnitude (when measured in the ACC, not as part of the P1-N1-P2 onset complex) showed a graded reduction in size as the SNR was decreased, and in Chapter 4, the P2 had a stronger relationship to the cochlear-scaled spectral difference between fricatives than the P1 and the N1 magnitudes. Furthermore, the P2 magnitude was used to create MDS solutions that largely represented the spectral difference between the stimuli, however when compared to MDS solutions based on the cochlear-scaled spectral difference between the fricatives, there was more differentiation on the horizontal plane for the P2 solutions. This suggests that the P2 solutions are not purely a representation of the acoustic difference between the fricatives, but are perhaps more phonologically relevant, as with the MDS solutions created by Iverson et al. (2016). Moreover, the P2 magnitude’s relationship to spectral difference was modulated by native language experience, suggesting that the P2 is a measure that is affected by linguistic aspects of the stimuli rather than pure auditory processing.

The results of this thesis have shown that the ACC is sensitive to language experience and SIN ability, and that it can be measured using a variety of speech stimuli in CI users. This indicates its potential as a tool for examining speech perception for a variety of different conditions. Furthermore, using multiple speech sounds and efficient stimulus presentation (i.e. sounds played in a random order changing every 350-400 ms), the ACC can be used to measure individual differences in normal-hearing listeners, and potentially in CI users.

5.2 Hearing Loss Beyond the Audiogram

Currently, there is a lot of interest in tools that can assess hearing performance beyond the audiogram and can assess people's performance in real world listening conditions. Even a mild hearing loss can make speech perception challenging, particularly in noise, a problem further increased with the extent of hearing loss. As hearing thresholds increase, a listener's ability to extract information from a speech signal decreases and so their speech recognition ability decreases (Boothroyd 1978; Walden 1984). Hearing loss can usually be identified through an audiogram, however often speech processing difficulties arise from problems not associated with a loss of hearing. For example, elderly listeners with normal audiometric thresholds and matched cognitive performance show reduced temporal encoding in comparison to younger listeners (Füllgrabe et al. 2015). Furthermore, certain younger listeners demonstrate complicated listening difficulties that are not predicted by audiometric performance. For example, children with auditory processing disorders (APD) that appear to be uncertain about following oral instructions, may have difficulties listening in background noise, and understanding fast or degraded speech, even though their peripheral hearing is normal (Jerger & Musiek 2000).

Another example of auditory problems that do not show on an audiogram is the recently talked about ‘hidden hearing loss’ (Schaette & McAlpine 2011). Recent research has identified a form of ‘hidden’ hearing loss, which is a line of evidence that suggests that there may not be a distinction between a hearing loss as measured using conventional audiometric thresholds, and a hearing loss that affects performance in complex listening tasks, such as SIN. In summary, animal work has identified that noise exposure that induces a temporary hearing threshold shift may also kill auditory nerve fibres with high firing thresholds and low spontaneous rates (Kujawa & Liberman 2009). Even after the temporary hearing threshold shift has returned to normal, the death of these cells has discreet effects on performance, such as reduced ABR Wave I amplitude at high levels. The loss of high-threshold fibres is currently thought to lead to a remapping of auditory encoding to preserve the dynamic range of normal hearing. This remapping may, however, harm temporal processing, and so also harm such processes as speech recognition in noise. Work on this topic remains on-going, but could potentially account for hearing difficulties experienced by subjects with normal audiometric thresholds (e.g. elderly and APD populations, amongst others). Whilst the existence and effects of hidden hearing loss in humans remains to be established, there is currently renewed interest in developing and understanding hearing measures that may account for the phenomenon (For a review see Plack et al. 2014).

In this thesis, the ACC has proved sensitive to SIN performance differences in normal-hearing listeners. Ageing, APD, and hidden hearing loss are all areas of research that could benefit from a robust cortical measure of speech perception that is sensitive to such differences. The ACC with its efficient measurement, its sensitivity to language background and SIN performance of the listener, and its test-retest reliability (Tremblay et al. 2003), alongside other speech measures, could prove useful in these areas of speech and hearing research that are not yet fully understood.

5.3 In the Clinic

For those with identified hearing loss, common treatments include hearing aids for those with a mild to severe loss, and cochlear implants for those with a profound loss. However, even in quiet, those listening through hearing devices suffer unique challenges. For example, loss of directional information and reduced dynamic range/increased loudness recruitment affects hearing aid users, and cochlear implants are limited by poor frequency selectivity, greatly reduced dynamic range, an inability to provide fine temporal structure, amongst other factors. Challenges to speech perception can come from a broad range of issues, and for those using hearing devices, understanding SIN is commonly reported as very challenging.

Currently in the clinic, speech perception post hearing aid fitting or cochlear implantation is measured using speech testing (e.g. Boothroyd 1968). This involves the patient wearing their device, listening to words, and repeating them to the audiologist. The audiologist can then compare their scores to pre-fitting. This is a useful way of measuring a patient's speech perception benefit from their device;

however, it involves patient cooperation. Some patients are unable to undertake this kind of task and so it remains unclear how much speech they are able to perceive with their device. Furthermore, it can take a significant amount of time before a patient is getting full benefit from their device. For example, it can take up to two years for a CI patient to reach their peak speech perception scores (Oh et al. 2003). Therefore, an objective task that is able to predict speech recognition abilities would be very useful in the clinic.

The ACC is objective, and shows some potential for measuring speech recognition abilities in CI users, however, larger subject numbers and further study is needed. For example, it is not yet known how the ACC changes from immediately post implantation to multiple years post implantation. Furthermore, there appears to be large individual variation, where by a large ACC response does not necessarily mean high speech perception performance. However, it is possible that the ACC could be used to track an individual's development, or used to ensure that the individual's fitting is providing enough acoustic detail to allow for speech perception improvement. A speech ACC response could be recorded immediately post implantation, and after each mapping. Using the individual as their own control in this way, the ACC could be a useful objective measure of CI users' development and performance.

5.4 Conclusion

This thesis has shown that the language experience of the listener and how a normal-hearing listener performs in a SIN task modulates certain aspects of the ACC. This supports recent literature suggesting that the ACC is not merely a measure of general auditory detection as it is often described, but rather is a measure at the border between auditory and linguistic processing in the auditory cortex. Furthermore, results have shown the potential of cortical measures of speech perception for CI users. Using multiple speech sounds and efficient stimulus presentation, the ACC can be used to measure individual differences in normal-hearing listeners, and potentially in CI users. The possible scope of the ACC beyond this is vast. Investigation into its behaviour with more extensive language groups, normal-hearing listeners with speech processing difficulties, and clinical populations could prove fruitful for many areas of speech and hearing research.

Bibliography

Abbas, P. J., & Brown, C. J. (2015). Assessment of Responses to Cochlear Implant Stimulation at Different Levels of the Auditory Pathway. *Hearing Research, 322,* 67-76.

Adler, G., & Adler, J. (1989). Influence of Stimulus Intensity on AEP Components in the 80- to 200-millisecond Latency Range. *Audiology, 28*(6), 316-324.

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mehncke, H., & Merzenich, M. M. (2001). Speech Comprehension is Correlated with Temporal Response Patterns Recorded From Auditory Cortex. *98*(23), 13367-13372.

Alain, C., Quan, J., McDonald, K., & Van Roon, P. (2009). Noise-Induced Increase in Human Auditory Evoked Neuromagnetic Fields. *European Journal of Neuroscience, 30,* 132-142.

Alho, K., Teder, W., Lavikainen, J., & Näätänen, R. (1994). Strongly Focused Attention and Auditory Event-related Potentials. *Biological Psychology, 38*(1), 73-90.

Altvater-Mackensen, N., & Fikkert, P. (2010). The Acquisition of the Stop-fricative Contrast in Perception and Production. *Lingua, 120,* 1898-1909.

Altvater-Mackensen, N., van der Feest, S. V., & Fikkert, P. (2014). Asymmetries in Early Word Recognition: The Case of Stops and Fricatives. *Language Learning and Development, 10*(2), 149-178.

Alvarenga, F. K., Amorim, R. B., Agostinho-Pesse, R. S., Costa, O. A., Nascimento, L. T., & Bevilacqua, M. C. (2012). Speech Perception and Cortical Auditory Evoked Potentials in Cochlear Implant Users with Auditory Neuropathy Spectrum Disorders. *International Journal of Pediatric Otorhinolaryngology, 76,* 1332-1338.

Amenedo, E., & Díaz, F. (1999). Ageing-related Changes in the Processing of Attended and Unattended Standard Stimuli. *NeuroReport, 10*(11), 2383-2388.

An, L. J., Martin, A. B., & Long, G. R. (2013). Effects of Phonetic Experience on Neural Processing of English /r/ and /l/ by Korean and Japanese Listeners.

Proceedings of Meetings on Acoustics. 19. Montreal: Acoustical Society of America.

Anderer, P., Semlitsch, H. V., & Saletu, B. (1996). Multichannel Auditory Event-related Brain Potentials: Effects of Normal Ageing on the Scalp Distribution of N1, P2, N2 and P300 Latencies and Amplitudes. *Electroencephalography and Clinical Neurophysiology, 99*(5), 458-472.

Anderson, E. S., Nelson, D. A., Krefy, H., Nelson, P. B., & Oxenham, A. J. (2011). Comparing Spatial Tuning Curves, Spectral Ripple Resolution, and Speech Perception in Cochlear Implant Users . *Journal of the Acoustical Society of America, 130*, 364-375.

Anderson, S., Chandrasekaran, B., Yi, H.-G., & Kraus, N. (2010). Cortical-Evoked Potentials Reflect Speech-in-Noise Perception in Children. *European Journal of Neuroscience, 32*, 1407-1413.

Appleby, S. V., McDermick, P., & Scott, J. W. (1963). The Sound Evoked Cerebral Response as a Test of Hearing. *Electroencephalography and Clinical Neurophysiology, 15*, 1050-1055.

Başkent, D., & Shannon, R. V. (2004). Frequency-place Compression and Expansion in Cochlear Implant Listeners. *Acoustical Society of America, 116*, 3130-3140.

Barnet, A. B. (1975). Auditory Evoked Potentials During Sleep in Normal Children From Ten Day to 3 Years of Age. *Electroencephalography and Clinical Neurophysiology, 39*(1), 29-41.

Barnet, A. B., & Lodge, A. (1966). Diagnosis of Deafness in Infants with the use of Computer-averaged Electroencephalographic Responses to Sound. *The Journal of Paediatrics, 69*, 753-758.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear Mixed-effects Models Using Eigen and S4. *R Package Version 1.1-9.*

Beagley, H. A., & Kellogg, S. E. (1969). A Comparison of Evoked Response and Subjective Auditory Thresholds. *International Journal of Audiology, 8*, 345-353.

Bentler, R. A. (2000). List Equivalency and Test-Retest Reliability of the Speech in Noise Test. *American Journal of Audiology, 9*, 84-100.

Best, C. T. (1995). A Direct Realist View of Cross-language Speech Perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 171-204). Baltimore: York Press.

Bierer, J. A. (2010). Probing the Electrode-neuron Interface with Focussed Cochlear Implant Stimulation. *Trends in Amplification, 14*(2), 84-95.

Bierer, J. A., & Middlebrooks, J. C. (2004). Cortical Responses to Cochlear Implant Stimulation: Channel Interactions. *Journal of the Association for Research in Otolaryngology, 5*(1), 32-48.

Billings, C. J., McMillan, G. P., Penman, T. M., & Gille, S. M. (2013). Predicting Perception in Noise Using Cortical Auditory Evoked Potentials. *Journal of the Association for Research in Otolaryngology, 14*, 891-903.

Billings, C. J., Papesh, M. A., Penman, T. M., Baltzell, L. S., & Gallun, F. J. (2012). Clinical Use of Aided Cortical Auditory Evoked Potentials as Measure of Physiological Detection or Physiological Discrimination. *International Journal of Otolaryngology, 2012*, 14 pages.

Billings, C. J., Tremblay, K. L., Stecker, G. C., & Tolin, W. M. (2009). Human Evoked Cortical Activity to Signal-to-Noise Ratio and Absolute Signal Level. *Hearing Research, 254*, 15-24.

BioSemi. (n.d.). *BioSemi Layout 64+2 Electrodes.* Retrieved August 30, 2012, from www.biosemi.com: http://www.biosemi.com/pics/cap_64_layout_medium.jpg

Boersma, P., & Weenink, D. (2013). *Praat: Doing Phonetics by Computer*. Retrieved 2013, from (Computer program): http://www.praat.org/

Boothroyd, A. (1968). Developments in Speech Audiometry. *British Journal of Audiology, 2*(1), 3-10.

Boothroyd, A. (1978). Speech Perception and Sensorineural Hearing Loss. In M. Ross, & T. G. Giolas (Eds.), *Auditory Management in Hearing-Impaired Children* (pp. 117-144). Baltimore: University Park Press.

Bronkhorst, A. W. (2000). The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-talker Conditions. *Acta Acustica united with Acustica, 86*(1), 117-128.

Brown, C. J., Etler, C., He, S., O'Brien, S., Erenberg, S., Kim, J.-R., . . . Abbas, P. J. (2008). The Electrically Evoked Auditory Change Complex: Preliminary

Results from Nucleus Cochlear Implant Users. *Ear & Hearing, 29*(5), 704-717.

Brown, C. J., Hughes, M. L., Luk, B., & et al. (2000). The Relationship Between EAP and EABR Thresholds and Levels used to Program the Nucleus 24 Speech Processor: Data from Adults. *Ear & Hearing, 21*, 151-163.

Brown, C., Jeon, E. K., Chiou, L.-K., Kirby, B., Karsten, S. A., Turner, C. W., & Abbas, P. J. (2015). Cortical Auditory Evoked Potentials Recorded From Nucleus Hybrid Cochlear Implant Users. *Ear & Hearing, 36*(6), 723-732.

Bruneau, N., Roux, S., Guerin, P., Barthelemy, C., & Lelord, G. (1997). Temporal Prominence of Auditory Evoked Potentials (N1 Wave) in 4-8-year-old Children. *Psychophysiology, 34*(1), 32-38.

Buchsbaum, M. (1976). Self-regulation of Stimulus Intensity: Augmenting/reducing and the Average Evoked Response. In G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-regulation* (pp. 101-135). New York: Plenum Publishing Corp.

Buckley, K. A., & Tobey, E. A. (2011). Cross-modal Plasticity and Speech Perception in Pre- and Postlingually Deaf Cochlear Implant Users. *Ear and Hearing, 32*(1), 2-15.

Calandruccio, L., & Smiljanic, R. (2012). New Sentence Recognition Materials Developed Using a Basic Non-native English Lexicon. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1342-1355.

Campbell, G. L., & King, G. (2011). *The Concise Compendium of the World's Languages* (2nd ed.). New York: Routledge.

Carcagno, S., & Plack, C. J. (2011). Subcortical Plasticity Following Perceptual Learning in a Pitch Discrimination Task. *Journal of the Association for Research in Otolaryngology, 12*, 89-100.

Čeponienė, R., Alku, P., Westerfield, M., Torki, M., & Townsend, J. (2005). ERPs Differentiate Syllable and Nonphonetic Sound Processing in Children and Adults. *Psychophysiology, 42*, 391-406.

Čeponienė, R., Rinne, T., & Näätänen, R. (2002). Maturation of Cortical Sound Processing as Indexed by Event-Related Potentials. *Clinical Neurophysiology, 113*, 870-882.

Čeponienė, R., Westerfield, M., Torki, M., & Townsend, J. (2008). Modality-specificity of Sensory Aging in Vision and Audition: Evidence from Event-related Potentials. *Brain Research, 1215*, 53-68.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The Natural Statistics of Audiovisual Speech. *PLoS Computational Biology, 5*(7), e1000436.

Cherry, C. (1953). Some Experiments on the Recognition of Speech, With One and With Two Ears. *Journal of the Acoustical Society of America, 25*(5), 975-979.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999). Spectro-Temporal Modulation Transfer Functions and Speech Intelligibility. *Journal of the Acoustical Society of America, 106*, 2719-2732.

Cody, D. T., & Bickford, R. G. (1965). Cortical Audiometry: An Objective Method of Evaluating Auditory Acuity in Man. *Mayo Clinic Proceedings, 40*, 273-287.

Collins, L. M., Zwolan, T. A., & Wakefield, G. H. (1997). Comparison of Electrode Discrimination, Pitch Ranking, and Pitch Scaling Data in Postlingually Deafened Adult Cochlear Implant Subjects. *Journal of the Acoustical Society of America, 101*(1), 440-455.

Courchesne, E. (1978). Neurophysiological Correlates of Cognitive Development: Changes in Long-latency Event-related Potentials from Childhood to Adulthood. *Enectroencephalography and Clinical Neurophysiology, 45*, 468-482.

Courchesne, E. (1990). Chronology of Postnatal Human Brain Development: Event-related Potentials, Positron Emission Tomography, Myelogenesis, and Synaptogenesis Studies. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-related Brain Potentials. Basic Issues and Apllications* (pp. 210-241). New York: Oxford University Press.

Crowley, K. E., & Colrain, I. M. (2004). A Review of the Evidence for P2 being an Independent Component Process: Age, Sleep and Modality. *Clinical Neurophysiology, 115*, 732-744.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic Responses to Speech in Noise in Children with Learning

Problems: Deficits and Strategies for Improvement. *Clinical Neurophysiology, 112,* 758-767.

Curtiss, S., Fromkin, V., Krashen, S., Rigler, D., & Rigler, M. (1974). The Linguistic Development of Genie. *Language, 50*(3), 528-554.

Dau, T. (2003). The Importance of Cochlear Processing for the Formation of Auditory Brainstem and Frequency Following Responses. *Journal of the Acoustical Society of America, 113,* 936-950.

Dau, T., Wegner, O., Mellert, V., & Kollmeier, B. (2000). Auditory Brainstem Responses with Optimized Chirp Signals Compensating Basilar-membrane Dispersion. *Journal of the Acoustical Societ of America, 107*(3), 1530-1540.

Davis, A. C. (1989). The Prevalence of Hearing Impairment and Reported Hearing Disablity Among Adults in Great Britain. *International Journal of Epidemiology, 18*(4), 911-917.

Davis, H. (1966). Validation of Evoked Response Audiometry (ERA) in Deaf Children. *International Journal of Audiology, 5,* 77-81.

Davis, H., Hirsch, S. K., Shelnott, J., & Bowers, C. (1967). Further Validation of Evoked Response Audiometry (ERA). *Journal of Speech, Language, and Hearing Research, 10,* 717-732.

Davis, P. A. (1939). Effects of Acoustic Stimuli on the Waking Human Brain. *Journal of Neurophysiology, 2,* 494-499.

de Cheveigné, A. (2010). Time-shift Denoising Source Separation. *Journal of Neuroscience Methods, 189,* 113-120.

de Cheveigné, A. (2012). Quadratic Component Analysis. *Neuroimage, 59,* 3838-3844.

de Cheveigné, A. (2016). Sparse Time Artefact Removal. *Journal of Neuroscience Methods, 262,* 14-20.

de Cheveigné, A., & Arzounian, D. (2015). Scanning for Oscillations. *Journal of Neural Engineering, 12*(6), 066020.

de Cheveigné, A., & Parra, L. (2014). Joint Decorrelation: A Flexible Tool for Multichannel Data Analysis. *Neuroimage, 98,* 487-505.

de Cheveigné, A., & Simon, J. Z. (2007). Denoising Based on Time-shift PCA. *Journal of Neuroscience Methods, 165,* 297-305.

de Cheveigné, A., & Simon, J. Z. (2008a). Denoising Based on Spatial Filtering. *Journal of Neuroscience Methods, 171*, 331-339.

de Cheveigné, A., & Simon, J. Z. (2008b). Sensor Noise Suppression. *Journal of Neuroscience Methods, 168,* 195-202.

Dehaene-Lambertz, G. (1997). Electrophysiological Correlates of Categorical Phoneme Perception in Adults. *NeuroReport, 8,* 919–924.

Delorme, A., & Makeig, S. (2004). EEGLAB: An Open Source Toolbox for Analysis of Single-trial EEG Dynamics. *Journal of Neuroscience Methods, 134*, 9-21.

Dempster, F. N. (1991). Inhibitory Processes: A Neglected Dimension of Intelligence. *Intelligence, 15*, 157-173.

Dempster, F. N. (1992). The Rise and Fall of the Inhibitory Mechanism: Toward a Unified Theory of Cognitive Development and Aging. *Developmental Review, 12,* 45-75.

Dimitrijevic, A., Michalewski, H. J., Zeng, F.-G., Pratt, H., & Starr, A. (2008). Frequency Changes in a Continuous Tone: Auditory Cortical Potentials. *Clinical Neurophysiology, 119*, 2111–2124.

Ding, N., & Simon, J. Z. (2012). Emergence of Neural Encoding of Auditory Objects While Listening to Competing Speakers. *PNAS, 109*(29), 11854-11859.

Ding, N., & Simon, J. Z. (2014). Cortical Entrainment to Continuous Speech: Functional Roles and Interpretations. *Frontiers in Human Neuroscience, 8*(311), 1-7.

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust Cortical Entrainment to the Speech Envelope Relies on the Spectro-temporal Fine Structure. *NeuroImage, 88*, 41-46.

Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of Age and Mild Hearing-loss on Speech Recognition in Noise. *Journal of the Acoustical Society of America, 76*(1), 87-96.

Elangovan, S., & Stuart, A. (2011). A Cross-linguistic Examination of Cortical Auditory Evoked Potentials for a Categorical Voicing Contrast. *Neuroscience Letters, 490*(2), 140-144.

Elliott, T. M., & Theunissen, F. E. (2009). The Modulation Transfer Function for Speech Intelligibility. *PLos Computational Biology, 5*(3), e1000302.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience, 10*, 590-604.

Etymotic Research. (n.d.). The SIN Test (Compact Disk). 61 Martin Lane, Elk Grove Village, IL, 60007.

Eulitz, C., & Lahiri, A. (2004). Neurobiological Evidence for Abstract Phonological Representations in the Mental Lexicon During Speech Recognition. *Journal of Cognitive Neuroscience, 16*(4), 577-583.

Füllgrabe, C., Moore, B. C., & Stone, M. A. (2015). Age-group Differences in Speech Identification Despite Matched Audiometrically Normal Hearing: Contributions from Auditory Temporal Processing and Cognition. *Frontiers in Aging Neuroscience, 6*(347).

Firszt, J. B., Chambers, R. D., & Kraus, N. (2002). Neurophysiology of Cochlear Implant Users II: Comparison Among Speech Perception, Dynamic Range, and Physiological Measures. *Ear & Hearing, 23*, 516-531.

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech Perception and Language Experience: Issues in Cross-language Research.* Baltimore, MD: York Press.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (2nd Edition ed.). Thousand Oaks, California: Sage.

Friesen, L. M., & Tremblay, K. L. (2006). Acoustic Change Complexes Recorded in Adult Cochlear Implant Listeners. *Ear & Hearing, 27*, 678-685.

Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech Recognition in Noise as a Function of the Number of Spectral Channels: Comparison of Acoustic Hearing and Cochlear Implants. *Journal of the Acoustical Society of America, 110*(2), 1150-1163.

Frisina, D. R., & Frisina, R. D. (1997). Speech Recognition in Noise and Presbyacusis: Relations to Possible Neural Mechanisms. *Hearing Research, 106*, 95-104.

Fruhstorfer, H. (1971). Habituation and Dishabituation of the Human Vertex Response. *Electroencephalography and Clinical Neurophysiology, 30*(4), 306-312.

Fu, Q. J., Chinchilla, S., & Galvin, J. J. (2004). The Role of Spectral and Temporal Cues in Voice Gender Discrimination by Normal-hearing Listeners and Cochlear Implant Users. *Journal of the Association for Research in Otolaryngology, 5*(3), 253-260.

Fu, Q. J., Shannon, R. V., & Wang, X. (1998). Effects of Noise and Spectral Resolution on Vowel and Consonant Recognition: Acoustic and Electric Hearing. *Journal of the Acoustical Society of America, 104*(6), 3586-3596.

Garcia-Larrea, L., Lukaszewicz, A. C., & Mauguiere, F. (1992). Revisiting the Oddball Paradigm. Non-target vs. Neutral Stimuli and the Evaluation of ERP Attentional Effects. *Neuropsychologia, 30*(8), 723-741.

Garner, W. R. (1974). *The Processing of Information and Structure.* Pontomac, MD: Erlbaum.

Geisler, C. D., Frishkopf, L. S., & Rosenblith, W. A. (1958). Extracranial Responses to Acoustic Clicks in Man. *Science, 128*, 1210-1211.

Giraud, A.-L., & Lee, H.-J. (2007). Predicting Cochlear Implant Outcome from Brain Organisation in the Deaf. *Restorative Neurology and Neuroscience, 25,* 381-390.

Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The Frequency Following Response (FFR) may Reflect Pitch-bearing Information but it is not a Direct Representation of Pitch. *Journal of the Association for Research in Otolaryngology, 12*, 767-782.

Gockel, H. E., Krugliak, A., Plack, C. J., & Carlyon, R. P. (2015). Specificity of the Human Frequency Following Response for Carrier and Modulation Frequency Assessed Using Adaptation. *Journal of the Association for Research in Otolaryngology, 16*(6), 747-762.

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., & Schwartz, T. H. (2011). Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electrocorticographic Investigation. *The Journal of Neuroscience, 31*(50), 18556-18567.

Gower, J. C. (1966). Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika, 53*(3/4), 325-328.

Greenberg, S., Marsh, J. T., Brown, W. S., & Smith, J. C. (1987). Neural Temporal Encoding of Low Pitch. I. Human Frequency-following Responses to Complex Tones. *Hearing Research, 25*(2), 91-114.

Groenen, P., Makhdoum, M., van den Brink, J. L., Stollman, M. H., Snik, A. F., & van den Broek, P. (1996b). The Relation Between Electric Auditory Brain Stem and Cognitive Responses and Speech Perception in Cochlear Implant Users. *Acta Oto-Laryngologica, 116*, 785-790.

Groenen, P., Snik, A., & van den Broek, P. (1996a). On the Clinical Relevance of Mismatch Negativity: Results from Subjects with Normal Hearing and Cochlear Implant Users. *Audiology & Neuro-Otology, 1*, 112-124.

Grunwald, T., Boutros, N. N., Pezer, N., von-Oertzen, J., Fernandez, G., Schaller, C., & Elger, C. E. (2003). Neuronal Substrates of Sensory Gating Within the Human Brain. *Biological Psychiatry, 53*, 511-519.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP Measure of Syntactic Processing. *Language and Cognitive Processes, 8*, 439-483.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous Brain Potentials Associated with Selective Auditory Attention. *Electroencephalography and Clinical Neurophysiology, 49*(3-4), 277-290.

Hari, R., Pelizzone, M., Makela, J. P., Hallstrom, J., Leinonen, L., & Lounasmaa, O. V. (1987). Neuromagnetic Responses of the Human Auditory Cortex to On- and Offsets of Noise Bursts. *Audiology, 26*(1), 31-43.

Hassaan, M. R. (2015). Auditory Evoked Cortical Potentials with Competing Noise in Children with Auditory Figure Deficit. *Hearing, Balance and Communication, 13*, 15-23.

Hattori, K., & Iverson, P. (2009). English /r/-/l/ Category Assimilation by Japanese Adults: Individual Differences and the Link to Identification Accuracy. *Journal of the Acoustical Society of America, 125*(1), 469-479.

He, S., Grose, J. H., Teagle, H. F., & Buchman, C. A. (2014). Objective Measures of Electrode Discrimination With Electrically Evoked Auditory Change Complex and Speech-Perception Abilities in Children With Auditory Neuropathy Spectrum Disorder . *Ear & Hearing, 35*, e63-e74.

Helms, J., Miller, J., Schön, F., Moser, L., Arnold, W., Janssen, T., . . . Hochmail-Desoyer, I. (1997). Evaluation of Performance with the COMBI 40 Cochlear Implant in Adults: A Multicentric Clinical Study. *ORL, 59*, 23-35.

Henry, B. A., McKay, C. M., McDermott, H. J., & Clark, G. M. (2000). The Relationship Between Speech Perception and Electrode Discrimination in Cochlear Implantees. *Journal of the Acoustical Society of America, 108*(3 Pt 1), 1269-1280.

Hillyard, S. A., & Picton, T. W. (1978). On and Off Components in the Auditory Evoked Potential. *Perception & Psychophysics, 24*(5), 391-398.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science, 182*(4108), 177-180.

Hiraumi, H., Nagamine, T., Morita, T., Naito, Y., Fukuyama, H., & Ito, J. (2008). Effect of Amplitude Modulation of Background Noise on Auditory-Evoked Magnetic Fields. *Brain Research, 1239*, 191-197.

Hochmair, I., Nopp, P., Jolly, C., Schmidt, M., Schösser, H., Garnham, C., & Anderson, I. (2006). MED-EL Cochlear Implants: State of the Art and a Glimpse into the Future. *Trends in Amplification, 10*(4), 201-219.

Hofmann, M., & Wouters, J. (2010). Electrically Evoked Auditory Steady State Responses in Cochlear Implant Users. *Journal of the Association for Research in Otolaryngology, 11*(2), 267-282.

Hoonhorst, I., Serniclaes, W., Collet, G., Colin, C., Markessis, E., Radeau, M., & Deltenre, P. (2009). N1b and Na Subcomponents of the N100 Long Latency Auditory Evoked-potential: Neurophysiological Correlates of Voicing in French-speaking Subjects. *Clinical Neurophysiology, 120*(5), 897-903.

Hoppe, U., Wohlberedt, T., & Danilkina, G. (2010). Acoustic Change Complex in Cochlear Implant Subjects in Comparison with Psychoacoustic Measures. *Cochlear Implants International, 11*(1), 426-430.

Horton, C., Svinivasan, R., & D'Zmura, M. (2014). Envelope Responses in Single-Trial EEG Indicate Attended Speaker in a 'Cocktail Party'. *Journal of Neural Engineering, 11*.

Howard, M. A., Volkov, I. O., Mirksy, R., Garell, P. C., Noh, M. D., Granner, M., . . . Brugge, J. F. (2000). Auditory Cortex on the Human Posterior Superior Temporal Gyrus. *Journal of Comparative Neurology, 416*, 79-92.

Howard, M. F., & Poeppel, D. (2010). Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension. *The Journal of Neurophysiology, 104*, 2500-2511.

Hughes, K. C., & Galvin, K. L. (2013). Measuring Listening Effort Expended by Adolescents and Young Adults with Unilateral or Bilateral Cochlear Implants or Normal Hearing. *Cochlear Implants International: An Interdisciplinary Journal, 14*(3), 121-129.

Hughes, M. L., & Abbas, P. J. (2006). The Relation Between Electrophysiologic Channel Interaction and Electrode Pitch Ranking in Cochlear Implant Recipients. *Journal of the Acoustical Society of America, 119*, 1527-4537.

Hughes, M. L., Abbas, P. J., Brown, C. J., & et al. (2000a). Using Electrically Evoked Compound Action Potential Thresholds to Facilitate Creating MAPs for Children with the Nucleaus CI24M. *Advances in Oto-Rhino-Laryngology, 57*, 260-265.

Hughes, M. L., Brown, C. J., Abbas, P. J., & et al. (2000b). Comparison of EAP Thresholds with MAP Levels in the Nucleus 24 Cochlear Implant: Data From Children. *Ear & Hearing, 21*, 164-174.

Huotilainen, M., Winkler, I., Alho, K., Escera, C., Virtanen, J., Ilmoniemi, R. J., . . . Näätänen, R. (1998). Combined Mapping of Human Auditory EEG and MEG Responses. *Electroencephalography and Clinical Neurophysiology, 108*, 370-379.

Hyde, M. L. (1997). The N1 Response and its Aplications. *Audiology and Neuro-Otology, 2*, 281-307.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A Perceptual Interference Account of Acquisition Difficulties for Non-native Phonemes. *Cognition, 87*, B47-B57.

Iverson, P., McCarthy, K., Brint, E., Tuomainen, J., & Wagner, A. (2016). Mapping the perception of vowel and fricative systems using auditory P1-N1-P2. *Manuscript submitted for publication*.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory Training for Experienced and Inexperienced Second-language Learners: Native French Speakers Learning English Vowels. *Applied Psycholinguistics, 33*(1), 145-160.

Jerger, J., & Jerger, S. (1970). Evoked Responses to Intensity and Frequency Change. *Archives of Otolaryngology, 91*, 433-436.

Jerger, J., & Musiek, F. (2000). Report of the Consensus Conference on the Diagnosis of Auditory Processing Disorders in School-Aged Children. *Journal of the American Academy of Audiology, 11*(9), 467-474.

Jones, S. J., Longe, O., & Vaz Pato, M. (1998). Auditory Evoked Potentials to Abrupt Pitch and Timbre Change of Complex Tones: Electrophysiological Evidence of 'Streaming'? *Electroencephalography and Clinical Neurophysiology, 108*, 131-142.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic Characteristics of English Fricatives. *Journal of the Acoustical Society of America, 108*(3 Pt 1), 1252-1263.

Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., & Muchnik, C. (2006). Identification of Syllables in Noise: Electrophysiological and Behavioural Correlates. *Acoustical Society of America, 120*(2), 926-933.

Kaukoranta, E., Hari, R., & Lounasmaa, O. V. (1987). Responses of the Human Auditory Cortex to Vowel Onset After Fricative Consonants. *Experimental Brain Research, 69*, 19-23.

Kelly, A. S., Purdy, S. C., & Thorne, P. R. (2005). Electrophysiological and Speech Perception Measures of Auditory Processing in Experienced Adult Cochlear Implant Users. *Clinical Neurophysiology, 116*, 1235-1246.

Kenemans, J. L., Verbaten, M. N., Roelofs, J. W., & Slangen, J. L. (1989). "Initial-" and "Change-orienting Reactions": An Analysis Based on Visual Single-trial Event-Related Potentials. *Biological Psychology, 28*(3), 199-226.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a "Cocktail Party". *The Journal of Neuroscience, 30*(2), 620-628.

Kiefer, J., Hohl, S., Stürzebecher, E., Pfennigdorff, T., & Gstöettner, W. (2001). Comparison of Speech Recognition with Different Speech Coding Strategies (SPEAK, CIS, and ACE) and their Relationship to Telemetric Measures of COmpound Action Potentials in the Nucleus CI 24M Cochlear Implant System. *Audiology, 40*(1), 32-42.

Kileny, P. R., Boerst, A., & Zwolan, T. (1997). Cognitive Evoked Potentials to Speech and Tonal Stimuli in Children with Implants. *Otolaryngology–Head and Neck Surgery, 117*, 161-169.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a Quick Speech-in-noise Test for Measuring Signal-to-noise Ratio Loss in Normal-hearing and Hearing-impaired Listeners. *Acoustical Society of America, 116*, 2395–2405.

Kim, J.-R., Ahn, S.-Y., Jeong, S.-W., Kim, L.-S., Park, J.-S., Chung, S.-H., & Oh, M.-K. (2012). Cortical Auditory Evoked Potential in Aging: Effects of Stimulus Intensity and Noise. *Otology & Neurotology, 33*, 1105-1112.

Kim, J.-R., Brown, C. J., Abbas, P. J., Etler, C. P., & O'Brien, S. (2009). The Effect of Changes in Stimulus Level on Electrically Evoked Cortical Auditory Potentials. *Ear & Hearing, 30,* 320-329.

Kisley, M. A., Polk, S. D., Ross, R. G., Levisohn, P. M., & Freedman, R. (2003). Early Postnatal Development of Sensory Gating. *Neuroreport, 14,* 693-697.

Knight, R. T., Hillyard, S. A., Woods, D. L., & Neville, H. J. (1980). The Effects of Frontal and Temporal-parietal Lesions on the Auditory Evoked Potential in Man. *Electroencephalography and Clinical Neurophysiology, 50*(1-2), 112-124.

Koch, D. B., Osberger, M. J., Segal, P., & Kessler, D. (2004). HiResolution and COnventional Sound Processing in the HiResolution Bionic Ear: Using Appropriate Outcome Measures to Assess Speech Recognition Ability. *Audiology and Neurotology, 9*(4), 214-223.

Kral, A. (2013). Auditory Critical Periods: A Review from System's Perspective. *Neuroscience, 247,* 117-133.

Kral, A., & Eggermont, J. J. (2007). What's to Lose and What's to Learn: Development Under Auditory Deprivation, Cochlear Implants and Limits of Cortical Plasticity. *Brain Research Reviews, 56*(1), 259-269.

Kraus, N., Micco, A. G., Koch, D. B., McGee, T., Carrell, T., Sharma, A., . . . Weingarten, C. Z. (1993). The Mismatch Negativity Cortical Evoked Potential Elicited by Speech in Cochlear-implant Users. *Hearing Research, 65*, 118-124.

Krishnan, A. (2008). Frequency-following Response. In R. F. Burkard, M. Don, & J. J. Eggermont, *Auditory Evoked Potentials: Basic Principles and Clinical Application* (pp. 313-333). Baltimore: Lippincott Williams & Wilkins.

Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009a). Experience Dependent Neural Representation of Dynamis Pitch in the Brainstem. *NeuroReport, 20*, 408-413.

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009b). Experience Dependent Enhancement of Linguistic Pitch Representation in the Brainstem is not Specific to a Speech Context. *Journal of Cognitive Neuroscience, 21*, 1092-1105.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of Pitch in the Human Brainstem is Sensitive to Language Experience. *Cognitive Brain Research, 25*, 161-168.

Kuhl, P. K. (1992). Psychoacoustics and Speech Perception: Internal Standards, Perceptual Anchors, and Prototypes. In L. A. Werner, & E. W. Rubel (Eds.), *Developmental Psychoacoustics* (pp. 293-332). Washington, DC: American Psychological Association.

Kuhl, P. K. (1998). The Development of Speech and Language. In T. C. Carew, R. Menzel, & C. J. Shatz (Eds.), *Mechanistic Relationships Between Development and Learning* (pp. 53-73). New York: Wiley.

Kuhl, P. K., & Iverson, P. (1995). Linguistic Experience and the 'Perceptual Magnet Effect'. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 121-154). Baltimore: York Press.

Kuhl, P. K., & Iverson, P. (1995). Linguistic Experience and the 'Perceptual Magnet Effect'. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 121-154). Baltimore: York Press.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants Show a Facilitation Effect for Native Language Phonetic Perception Between 6 and 12 Months. *Developmental Science, 9*(2), F13-F21.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science, 255*, 606-608.

Kujawa, S. G., & Liberman, M. C. (2009). Adding Insult to Injury: Cochlear Nerve Degeneration After "Tempory" Noise-induced Hearing Loss. *Journal of Neuroscience, 29*(45), 14077-14085.

Kuruvilla-Mathew, A., Purdy, S. C., & Welch, D. (2015). Cortical Encoding of Speech Acoustics: Effects of Noise and Amplification. *International Journal of Audiology, 54,* 852-864.

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., & Näätänen, R. (2002). Maturation of the Auditory Change Detection Response in Infants: a Longitudinal ERP Study. *NeuroReport, 13*(15), 1843-1848.

Kutas, M., & Hillyard, S. A. (1983). Event-related Brain Potentials to Grammatical Errors and Semantic Anomalies. *Memory & Cognition, 11*(5), 539-550.

Lee, D. S., Lee, J. S., Oh, S. H., Kim, S.-K., Kim, J.-W., Chung, J.-K., . . . Kim, C. S. (2001). Cross-modal Plasticity and Cochlear Implants. *Nature, 406*, 149-150.

Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective Subcortical Enhancement of Musical Intervals in Musicians. *The Journal of Neuroscience, 29*(18), 5832-5840.

Lenth, R. V. (2016). Least-squares Means: The R Package lsmeans. *Journal of Statistical Software, 69*(1), 1-33.

Liégeois-Chauvel, C., De Graaf, J. B., Laguitton, V., & Chauvel, P. (1999). Specialisation of Left Auditory Cortex for Speech Perception in Man Depends on Temporal Coding. *Cerebral Cortex, 9*, 484-496.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked Potentials Recorded from the Auditory Cortex in Man: Evaluation and Topography of the Middle Latency Components. *Electroencephalography and Clinical Neurophysiology, 92,* 204-214.

Lightfoot, G., & Kennedy, V. (2006). Cortical Electric Response Audiometry Hearing Threshold Estimation: Accuracy, Speed, and the Effects of Stimulus Presentation Features. *Ear & Hearing, 27*, 443-456.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An Open-source Toolbox for the Analysis of Event-related Potentials. *Frontiers in Human Neuroscience, 8*(213).

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron, 54*(6), 1001-1010.

Makhdoum, M. J., Groenen, P. A., Snik, A. F., & van den Broek, P. (1998). Intra- and Interindividual Correlations Between Auditory Evoked Potentials and Speech Perception in Cochlear Implant Users. *Scandinavian Audiology, 27*, 13-20.

Martin, B. A. (2007). Can the Acoustic Change Complex be Recorded in an Individual with a Cochlear Implant? Separating Neural Responses from Cochlear Implant Artifact. *Journal of the American Society of Audiology, 18*, 126-140.

Martin, B. A., & Boothroyd, A. (1999). Cortical, Auditory, Event-related Potentials in Response to Periodic and Aperiodic Stimuli with the Same Spectral Envelope. *Ear & Hearing, 20*, 33-44.

Martin, B. A., & Boothroyd, A. (2000). Cortical, Auditory, Evoked Potentials in Response to Changes of Spectrum and Amplitude. *Journal of the Acoustical Society of America, 107*(4), 2155-2161.

Martin, B. A., & Stapells, D. R. (2005). Effects of Low-Pass Noise Masking on Auditory Event-Related Potentials to Speech. *Ear & Hearing, 26*, 195-213.

Martin, B. A., Boothroyd, A., Ali, D., & Leach-Berth, T. (2010). Stimulus Presentation Strategies for Eliciting the Acoustic Change Complex: Increasing Efficiency. *Ear & Hearing, 31*(3), 356-366.

Martin, B. A., Kertzberg, D., & Stapells, D. R. (1999). The Effects of Decreased Audibility Produced by High-Pass Noise Masking on N1 and the Mismatch Negativity to Speech Sounds /ba/ and /da/. *Journal of Speech, Language, and Hearing Research, 42*, 271-286.

Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The Effects of Decreased Audibility Produced by High-Pass Noise Masking on Cortical Event-Related Potentials to Speech Sounds /ba/ and /da/. *Journal of the Acoustical Society of America, 101*(3), 1585-1599.

Martin, B. A., Tremblay, K. L., & Korczak, P. (2008). Speech Evoked Potentials: From the Laboratory to the Clinic. *Ear & Hearing, 29*, 285-313.

Martin, B. A., Tremblay, K. L., & Stapells, D. R. (2007). Principles and Applications of Cortical Auditory Evoked Potentials. In R. F. Burkard, M. Don, & J. J. Eggermont, *Auditory Evoked Potentials: Basic Principles and Clinical*

Application (pp. 482-507). Philadelphia, PA: Lippincott Williams and Wilkins.

MathWorks. (2013a). MATLAB.

Maurer, J., Collet, L., Pelster, H., Truy, E., & Gallégo, S. (2002). Auditory Late Cortical Response and Speech Recognition in Digisonic Cochlear Implant Users. *Laryngoscope, 112*, 2220-2224.

McCandless, G. A., & Best, L. (1964). Evoked Responses to Auditory Stimuli in Man Using a Summing Computer. *Journal of Speech and Hearing Research, 7*, 193-202.

McCullagh, J., Musiek, F. E., & Shinn, J. B. (2012). Auditory Cortical Processing in Noise in Normal-Hearing Young Adults. *Audiological Medicine, 10*(3), 114-121.

McDermott, H. J. (2004). Music Perception with Cochlear Implants: A Review. *Trends in Amplification, 8*(2), 49-82.

Megela, A. L., & Teyler, T. J. (1979). Habituation and the Human Evoked Potential. *Journal of Comparative and Physiological Psychology, 93*(6), 1154-1170.

Meister, H., Landwehr, M., Pyschny, V., Walger, M., & Wedel, H. V. (2009). The Perception of Prosody and Speaker Gender in Normal-hearing Listeners and Cochlear Implant Recipients. *International Journal of Audiology, 48*(1), 38-48.

Melara, R. D., Rao, A., & Tong, Y. (2002). The Duality of Selection: Excitatory and Inhibitory Processes in Auditory Selective Attention. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 279-306.

Mendel, M. I., Hosick, E. C., Windman, T. R., Davis, H., Hirsh, S. K., & Dinges, D. F. (1975). Audiometric Comparison of the Middle and Late Components of the Adult Auditory Evoked Potentials Awake and Asleep. *Electroencephalography and Clinical Neurophysiology, 38*(1), 27-33.

Michalewski, H. J., Starr, A., Zeng, F.-G., & Dimitrijevic, A. (2009). N100 Cortical Potentials Accompanying Disrupted Auditory Nerve Activity in Auditory Neuropathy (AN): Effects of Signal Intensity and Continuous Noise. *Clinical Neurophysiology, 120*, 1352-1363.

Michie, P. T., Bearpark, H. M., Crawford, J. M., & Glue, L. C. (1990). The Nature of Selective Attention Effects on Auditory Event-related Potentials. *Biological Psychology, 30*(3), 219-250.

Michie, P. T., Solowij, N., Crawford, J. M., & Glue, L. C. (1993). The Effects of Between-source Discriminability on Attended and Unattended Auditory ERPs. *Psychophysiology, 30*(2), 205-220.

Millman, R. E., Johnson, S. R., & Prendergast, G. (2015). The Role of Phase-locking to the Temporal Envelope of Speech in Auditory Perception and Speech Intelligibility. *Journal of Cognitive Neuroscience, 27*(3), 533-545.

Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973). Scalp-recorded Early Responses in Man to Frequencies in the Speech Range. *Electroencephalography and Clinical Neurophysiology, 35*(6), 665-667.

Muller-Gass, A., Marcoux, A., Logan, J., & Campbell, K. B. (2001). The Intensity of Masking Noise Affects the Mismatch Negativity to Speech Sounds in Human Subjects. *Neuroscience Letters, 299*(2001), 197-200.

Munivrana, B., & Mildner, V. (2013). Cortical Auditory Evoked Potentials in Unsuccessful Cochlear Implant Users. *Clinical Linguistics & Phonetics, 27*(6-7), 472-483.

Näätänen, R. (1990). The Role of Attention in Auditory Information Processing as Revealed by Event-related Potentials and Other Brain Measures of Cognitive Function. *Behavioral and Brain Sciences, 13*(2), 201-233.

Näätänen, R. (2001). The Perception of Speech Sounds by the Human Brain as Reflected by the Mismatch Negativity (MMN) and its Magnetic Equivalent (MMNm). *Psychophysiology, 38*, 1–21.

Näätänen, R., & Kreegipuu, K. (2012). The Mismatch Megativity (MMN). In S. J. Luck, & E. S. Kappenman, *The Oxford Handbook of Event-Related Potential Components* (pp. 143-157). New York: Oxford University Press.

Näätänen, R., & Michie, P. T. (1979). Early Selective-attention Effects on the Evoked Potential: A Critical Review and Reinterpretation. *Biological Psychology, 8*(2), 81-136.

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology, 24*, 375-425.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early Selective-attention Effect on Evoked Potential Reinterpreted. *Acta Psychologica, 42*(4), 313-329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Livonen, A., . . . Alho, K. (1997). Language-Specific Phoneme Representations Revealed by Electric and Magnetic Brain Responses. *Nature, 385*(6615), 432-434.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive Intelligence' in the Auditory Cortex. *TRENDS in Neurosciences, 24*(5), 283-288.

Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., & Javel, E. (1996). Intensity Discrimination as a Function of Stimulus Level with Electrical Stimulation. *Journal of the Acoustical Society of America, 100*(4), 2393-2414.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., & Levine, S. (1995). Electrode Ranking of "Place Pitch" and Speech Recognition in Electrical Hearing. *Journal of the Acoustical Society of America, 98*(4), 1987-1999.

Nilsson, M., Soil, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the Measurement of Speech Reception Thresholds in Quiet and in Noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099.

Novak, G., Ritter, W., & Vaughan, J. H. (1992). Mismatch Detection and the Latency of Temporal Judgements. *Psychophysiology, 29*(4), 398-411.

Oades, R. D., Zerbin, D., & Dittmann-Balcar, A. (1995). The Topography of Event-related Potentials in Passive and Active Conditions of a 3-tone Auditory Oddball Test. *International Journal of Neuroscience, 81*(3-4), 249-264.

Obuchi, C., Harashima, T., & Shiroma, M. (2012). Auditory Evoked Potentials Under Active and Passive Hearing Conditions in Adult Cochlear Implant Users. *Clinical and Experimental Otorhinolaryngology, 1*(1), S6-S9.

Oh, S.-h., Kim, C.-s., Kang, E. J., Lee, D. S., Lee, H. J., Chang, S. O., . . . Koo, J. W. (2003). Speech Perception After Cochlear Implantation Over a 4-Year Time Period. *Acta Oto-Laryngologica, 123*(2), 148-153.

Okusa, M., Shiraishi, T., Kubo, T., & Nageishi, Y. (1999). Effects of Discrimination Difficulty on Cognitive Event-related Brain Potentials in Patients with Cochlear Implants. *Otolaryngology–Head and Neck Surgery, 121*(5), 610-615.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience, 2011*.

Osterhout, L., & Holcomb, P. J. (1992). Event-related Potentials Elicited by Syntactic Anomaly. *Journal of Memory and Language, 31*, 785-806.

Ostroff, J. M., Martin, B. A., & Boothroyd, A. (1998). Cortical Evoked Response To Acoustic Change Within A Syllable. *Ear & Hearing, 19*(4), 290-297.

Oviatt, D. L., & Kileny, P. R. (1991). Auditory Event-Related Potentials Elicited from Cochlear Implant Recipients and Hearing Subjects. *American Journal of Audiology, 1*(1), 48-55.

Paetau, R., Ahonen, A., Salonen, O., & Sams, M. (1995). Auditory Evoked Magnetic Fields to Tones and Pseudowords in Healthy Children and Adults. *Journal of Clinical Neurophysiology, 12*(2), 177-185.

Paiva, T. O., Almeida, P. R., Ferreira-Santos, F., Vieira, J. B., Silveira, C., Chaves, P. L., . . . Marques-Teixeira, J. (2016). Similar Sound Intensity Dependence of the N1 and P2 Components of the Auditory ERP: Averaged and Single Trial Evidence. *Clinical Neurophysiology, 127*(1), 499-508.

Pals, C., Sarampalis, A., & Baskent, D. (2013). Listening Effort with Cochlear Implant Simulations. *Journal of Speech, Language and Hearing Research, 56*, 1075-1084.

Pang, E. W., & Taylor, M. J. (2000). Tracking the Development of the N1 from Age 3 to Adulthood: An Examination of Speech and Non-speech Stimuli. *Clinical Neurophysiology, 111*(3), 388-397.

Pantev, C., Eulitz, C., Hampson, S., Ross, B., & Roberts, L. E. (1996a). The Auditory Evoked 'Off' Response: Sources and Comparison with the 'On' and the 'Sustained' Responses. *Ear & Hearing, 17*(3), 255-265.

Pantev, C., Hoke, M., Lutkenhoner, B., & Lehnertz, K. (1991). Neuromagnetic Evidence of Functional Organization of the Auditory Cortex in Humans. *Acta Otolaryngology Supplement, 491*, 106-115.

Pantev, C., Roberts, L. E., Elbert, T., Ross, B., & Wienbruch, C. (1996b). Tonotopic Organization of the Sources of Human Auditory Steady-state Responses. *Hearing Research, 101*(1-2), 62-74.

Papesh, M. A., Billings, C. J., & Baltzell, L. S. (2015). Background Noise can Enhance Cortical Auditory Evoked Potentials Under Certain Conditions. *Clinical Neurophysiology, 126*, 1319-1330.

Parbery-Clark, A., Marmel, F., Julia, B., & Kraus, N. (2011). What Subcortical-Cortical Relationships tell us about Processing Speech in Noise. *European Journal of Neuroscience, 33*, 549-557.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex, 23*, 1378-1387.

Peelle, J., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm Through to Comprehension. *Frontiers in Psychology, 3*, Article 320.

Perl, E. R., Galambos, R., & Glorig, A. (1953). The Estimation of Hearing Threshold by Electroencephalography. *Electroencephalography and Clinical Neurophysiology, 5*(4), 501-512.

Picton, T. (1990). Auditory Evoked Potentials. In D. D. Daly, & T. A. Redley (Eds.), *Current Practice in Clinical Electroencephalography* (2nd Edition ed., pp. 625-678). New York: Raven Press Ltd.

Picton, T. W., & Hillyard, S. A. (1974). Human Auditory Evoked Potentials: II. Effects of Attention. *Electroencephalography and Clinical Neurophysiology, 36*, 191-199.

Pike, K. L. (1945). *The Intonation of American English.* Ann Arbor: University of Michigan Press.

Plack, C. J., Barker, D., & Prendergast, G. (2014). Perceptual Concequences of "Hidden" Hearing Loss. *Trends in Hearing, 18*.

Plomp, R. (1978). Auditory Handicap of Hearing Impairment and Limited Benefit of Hearing Aids. *Journal of the Acoustical Society of America, 63*(2), 533-549.

Polka, L., & Bohn, O.-S. (2003). Asymmetries in Vowel Perception. *Speech Communication, 41*, 221-231.

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) Framework: An Emerging View of Early Phonetic Development. *Journal of Phonetics, 39*(4), 467-478.

Polka, L., Bohn, O. -S., & Molnar, M. (2005). Natural Referent Vowels Guide the Development of Vowel Perception. *Journal of the Acoustical Society of America, 117*(4), 2398.

Ponton, C. W., Don, M., Eggermont, J. J., Waring, M. D., & Masuda, A. (1996b). Maturation of Human Cortical Auditory Function: Differences Between Normal-hearing Children and Children with Cochlear Implants. *Ear & Hearing, 17*(5), 430-437.

Ponton, C. W., Don, M., Eggermont, J. J., Waring, M. D., Kwong, B., & Masuda, A. (1996a). Auditory System Plasticity in Children After Long Periods of Complete Deafness. *Neuroreport, 8*, 61-65.

Ponton, C. W., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of Human Central Auditory System Activity: Seperating Auditory Evoked Potentials by Dipole Source Modeling. *Clinical Neurophysiology, 113*, 407-420.

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of Human Central Auditory System Activity: Evidence from Multi-channel Evoked Potentials. *Clinical Neurophysiology, 111*(2), 220-236.

Pool, K. D., Finitzo, T., Hong, C. T., Rogers, J., & Pickett, R. B. (1989). Infarction of the Superior Temporal Gyrus: A Description of Auditory Evoked Potential Latency and Amplitude Topology. *Ear & Hearing, 10*, 144-152.

Potts, G. F., Dien, J., Hartry-Speiser, A. L., McDougal, L. M., & Tucker, D. M. (1998). Dense Sensor Array Topography of the Event-related Potential to Task-relevant Auditory Stimuli. *Electroencephalography and Clinical Neurophysiology, 106*(5), 444-456.

Pratt, H. (2012). Sensory ERP Components. In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford Handbook of Event Related Potential Components* (pp. 89-114). New York: Oxford University Press.

Pratt, H., Aminoff, M., Nuwer, M. R., & Starr, A. (1999). Short-latency Auditory Evoked Potentials. In G. Deuschl, & A. Eisen (Eds.), *Recommendations for the Practice of Clinical Neurophysiology* (pp. 69-78). Amsterdam: Elsevier.

R Core Team. (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rao, A., Zhang, Y., & Miller, S. (2010). Selective Listening of Concurrent Auditory Stimuli: An Event-related Potential Study. *Hearing Research, 268*(1-2), 123-132.

Rapin, I., & Graziani, L. J. (1967). Auditoru-evoked Responses in Normal, Brain-damaged, and Deaf Infants. *Neurology, 17,* 881-894.

Reinke, K., He, Y., Wang, C., & Alain, C. (2003). Perceptual Learning Modulates Sensory Evoked Response During Vowel Segregation. *Cognitive Brain Research, 17,* 781-791.

Reite, M., Teale, P., Zimmerman, J., Davis, K., & Whalen, J. (1988). Source Localisation of a 50-msec-latency Auditory Evoked Field Component. *Electroencephalography and Clinical Neurophysiology, 70,* 490-498.

Rif, J., Hari, R., Hamalainen, M. S., & Sams, M. (1991). Auditory Attention Affects Two Different Areas in the Human Supratemporal Cortex. *Electroencephalography and Clinical Neurophysiology, 79*(6), 464-472.

Rimmele, J. M., Golumbic, E. Z., Schröger, E., & Poeppel, D. (2015). The Effects of Selective Attention and Speech Acoustics on Neural Speech-Tracking in a Multi-Talker Scene. *Cortex, 68,* 144-154.

Ringbom, H. (1992). On L1 Transfer in L2 Comprehension and L2 Production. *Language Learning, 42*(1), 85-112.

Ritter, W., Vaughan, H. G., & Costa, L. D. (1968). Orienting and Habituation of Auditory Stimuli: A Study of Short Term Changes in Averaged Evoked Responses. *Electroencephalography and Clinical Neurophysiology, 25,* 550-556.

Robaey, P., Cansino, S., Dugas, M., & Renault, B. (1995). A Comparative Study of ERP Correlates of Psychometric and Piagetian Intelligence Measures in Normal and Hyperactive Children. *Electroencephalography and Clinical Neurophysiology, 96,* 56-75.

Rogers, C. L., Lister, J. J., Febo, D. M., Besing, J. M., & Abrams, H. B. (2006). Effects of Bilingualism, Noise, and Reverberation on Speech Perception by Listeners with Normal Hearing. *Applied Psycholinguistics, 27,* 465-485.

Roman, S., Canévet, G., Lorenzi, C., Triglia, J.-M., & Liégeois-Chauvel. (2004). Voice Onset Time Encoding in Patients with Left and Right Cochlear Implants. *Neuroreport, 15*(4), 601-605.

Ross, B., & Tremblay, K. (2009). Stimulus Experience Modifies Auditory Neuromagnetic Responses in Young and Older Listeners. *Hearing Research, 248*(1-2), 48-59.

Ross, B., Tremblay, K. L., & Picton, T. W. (2007). Physiological Detection of Interaural Phase Differences. *Journal of the Acoustical Society of America, 121*, 1017-1027.

Rothauser, E. H., Chapman, W. D., Guttman, N., Hecker, M. H., Nordby, K. S., Silbiger, H. R., . . . Weinstock, M. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics, 17*, 225-246.

Russo, N., Zecker, S., Trommer, B., Chen, J., & Kraus, N. (2009). Effects of Background Noise on Cortical Encoding of Speech in Autism Spectral Disorders. *Journal of Autism and Developmental Disorders, 39*, 1185-1196.

Rust, J. (1977). Habituation and the Orienting Response in the Auditory Cortical Evoked Potential. *Psychophysiology, 14*(2), 123-126.

Salo, S., Peltola, M. S., Aaltonen, O., Johansson, R., Lang, A. H., & Laurikainen, E. (2002). Stability of Memory Traces for Speech Sounds in Cochlear Implant Patients. *Logopedics Phoniatrics Vocology, 27*, 132-138.

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory Frequency Discrimination and Event-related Potentials. *Electroencephalography and Clinical Neurophysiology, 62*(6), 437-448.

Sandmann, P., Plotz, K., Hauthal, N., de Vos, M., Schönfeld, R., & Debener, S. (2015). Rapid Bilateral Improvement in Auditory Cortex Activity in Postlingually Deafened Adults Following Cochlear Implantation. *Clinical Neurophysiology, 126*, 594-607.

Schaette, R., & McAlpine, D. (2011). Tinnitus With a Normal Audiogram: Physiological Evidence for Hidden Hearing Loss and Computational Model. *Journal of Neuroscience, 31*(38), 13452-13457.

Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2012). Asymmetries in the Processing of Vowel Height. *Journal of Speech Language and Hearing Research, 55*(3), 903-918.

Scheperle, R. A., & Abbas, P. J. (2015a). Peripheral and Central Contributions to Cortical Responses in Cochlear Implant Users. *Ear & Hearing, 36*, 430-440.

Scheperle, R. A., & Abbas, P. J. (2015b). Relationships Among Peripheral and Central Electrophysiological Measures of Spatial and Spectral Selectivity and Speech Perception in Cochlear Implant Users. *Ear & Hearing, 36*, 441-453.

Schoof, T. (2014). The Effects of Ageing on teh Perception of Speech in Noise. *A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*. UCL: Department of Speech, Hearing and Phonetic Sciences.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in Bilingualism and Second Language Learning. In J. Kroll, & A. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 371-388). Oxford: Oxford University Press.

Shallop, J. K. (1993). Objective Electrophysiological Measures from Cochlear Implant Patients. *Ear & Hearing, 14*(1), 58-63.

Sharma, A., & Dorman, M. F. (1999). Cortical Auditory Evoked Potential Correlates of Categorical Perception of Voice-Onset Time. *Journal of the Acoustical Society of America, 106*(2), 1078-1083.

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic Correlates of Cross-Language Phonetic Perception. *Journal of the Acoustical Society of America, 107*(5), 2697-2703.

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002). A Sensitive Period for the Development of the Central Auditory System in Children with Cochlear Implants: Implications for Age of Implantation. *Ear and Hearing, 23,* 532-539.

Sharma, A., Kraus, N., McGee, T. J., & Nicol, T. G. (1997). Developmental Changes in P1 and N1 Central Auditory Responses Elicited by Consonant-vowel Syllables. *Electroencephalography and Clinical Neurophysiology, 104*, 540-545.

Sharma, M., Purdy, S. C., Munro, K. J., Sawaya, K., & Peter, V. (2014). Effects of Broadband Noise on Cortical Evoked Auditory Responses at Different Loudness Levels in Young Adults. *NeuroReport, 25,* 312-319.

Shepherd, R. K., & Hardie, N. A. (2001). Deafness-Induced Changes in the Auditory Pathway: Implications for Cochlear Implants. *Audiology & Neuro-Otology, 6*, 305-318.

Shepherd, R. K., Meltzer, N. E., Fallon, J. B., & Ryugo, D. K. (2006). Consequences of Deafness and Electrical Stimulation on the Peripheral and Central Auditory System. In S. B. Waltzman, & J. T. Roland Jr (Eds.), *Cochlear Implants* (2nd Edition ed., pp. 25-39). New York: Thieme Medical Publishers.

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., & Näätänen, R. (1998). Background Acoustic Noise and the Hemispheric Lateralization of Speech Processing in the Human Brain: Magnetic Mismatch Negativity Study. *Neuroscience Letters, 251*, 141-144.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., . . . et al. (1994). Evaluation of a New Spectral Peak Coding Strategy for the Nucleus 22 Channel Cochlear Implant System. *American Journal of Otology, 15*(2), 15-27.

Skoe, E., & Kraus, N. (2010). Auditory Brainstem Response to Complex Sounds: A Tutorial. *Ear & Hearing, 31*(3), 302.

Stapells, D. R. (2002). Cortical Event-related Potentials to Auditory Stimuli. In J. Katz, R. F. Burkard, & L. Medwetsky (Eds.), *Handbook of Clinical Audiology* (5th Edition ed., pp. 378-406). Philadelphia: Lippincott, Williams, and Wilkins.

Stapells, D. R., Gravel, J. A., & Martin, B. A. (1995). Thresholds for Auditory Brainstem Responses to Tones in Notched Noise From Infants and Young Children with Normal Hearing or Sensorineural Hearing Loss. *Ear & Hearing, 16*, 361-371.

Stapells, D. R., Picton, T., & Durrieux-Smith, A. (1994). Electrophysiologic Measures of Frequency-specific Auditory Function. In J. T. Jacobson (Ed.), *Principles and Applications in Auditory Evoked Potentials* (pp. 251-284). Boston: Allyn and Bacon.

Steinschneider, M., & Dunn, M. (2002). Electrophysiology in Developmental Neuropsychology. In S. Segalowitz, & I. Rapin (Eds.), *Handbook of Neuropsychology* (2nd Edition ed., Vol. 8, pp. 91-146). Amsterdam: Elsevier.

Strait, D. L., Nina Kraus, Skoe, E., & Richard, A. (2009). Musical Experience and Neural Efficiency - Effects of Training on Subcortical Processing of Vocal Expressions of Emotion. *European Journal of Neuroscience, 29*, 661-668.

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The Maturation of Human Evoked Brain Potentials to Sounds Presented at Different Stimulus Rates. *Hearing Research, 236*(1-2), 61-79.

Suzuki, T., & Origuchi, K. (1969). Averaged Evoked Response Audiometry (ERA) in Young Children During Sleep. *Acta Otolaryngologica, 252*, 19-28.

Szymanski, M. D., Yund, E. W., & Woods, D. L. (1999). Phonemes, Intensity and Attention: Differential Effects on the Mismatch Negativity (MMN). *Journal of the Acoustical Society of America, 106*(6), 3492-3505.

Taguchi, K., Picton, T. W., Orpin, J. A., & Goodman, W. S. (1969). Evoked Response Audiometry in Newborn Infants. *Acta Otolaryngologica, 252*, 5-17.

Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, A. D. (2009). Top-Down Modulation of Auditory Processing: Effects of Sound Context, Musical Expertise and Attentional Focus. *European Journal of Neuroscience, 30*, 1636-1642.

The Development of Language: A Critical Period in Humans. (2001). In D. Purves, G. J. Augustine, D. Fitzpatrick, L. C. Katz, A.-S. LaMantia, O. J. McNamara, & S. M. Williams (Eds.), *Neuroscience* (2nd Edition ed.). Sunderland, MA: Sinauer Associates.

Throckmorton, C. S., & Collins, L. M. (1999). Investigation of the Effects of Temporal and Spatial Interactions on Speech-recognition skills in Cochlear-Implant Subjects. *Journal of the Acoustical Society of America, 105*(2 Pt 1), 861-873.

Timm, L., Agrawal, D., Viola, F. C., Sandmann, P., Debener, S., Büchner, A., . . . Wittfoth, M. (2012). Temporal Feature Perception in Cochlear Implant Users. *Plos One, 7*(9), e45375.

Timm, L., Vuust, P., Brattico, E., Agrawal, D., Debener, S., Büchner, A., . . . Wittfoth, M. (2014). Residual Neural Processing of Musical Sound Features in Adult Cochlear Implant Users. *Frontiers in Human Neuroscience, 8*, Article 181.

Tomé, D., Barbosa, F., Nowak, K., & Marques-Teixeira, J. (2015). The Development of the N1 and N2 Components in Auditory Odball Paradigms: A systematic Review with Narative Analysis and Suggested Normative Values. *Journal of Neural Transmission, 122*(3), 375-391.

Tonnquist-Uhlen, I., Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2003). Maturation of Human Central Auditory System Activity: The T-complex. *Clinical Neurophysiology, 114*(4), 685-701.

Torgerson, W. S. (1958). *Theory and Methods of Scaling.* New York: Wiley.

Trainor, L. J. (2008). Event-related Potential (ERP) Measures in Auditory Development Research. In L. A. Schmidt, & S. J. Segalowitz, *Developmental Psychophysiology: Theory, Systems and Methods* (pp. 69-102). New York: Cambridge University Press.

Traunmüller, H., & Eriksson, A. (1995). The Frequency Range of the Voice Fundamental in the Speech of Male and Female Adults. *Unpublished manuscript*.

Tremblay, K. L., Friesen, L., Martin, B. A., & Wright, R. (2003). Test Retest Reliability of Cortical Evoked Potentials Using Naturally Produced Speech Sounds. *Ear & Hearing, 24*(3), 225–232.

Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the Auditory Evoked P2 Response a Biomarker of Learning? *Frontiers in Systems Neuroscience 20, 8,* 28.

Tremblay, K., & Kraus, N. (2002). Auditory Training Induces Asymmetrical Changes in Cortical Neural Activity. *Journal of Speech, Language and Hearing Research, 45*, 564-572.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central Auditory Plasticity: Changes in the N1-P2 Complex After Speech-sound Training. *Ear & Hearing, 22*, 70-90.

Turgeon, C., Lazzouni, L., Lepore, F., & Ellemberg, D. (2014). An Objective Auditory Measure to Assess Speech Recognition in Adult Cochlear Implant Users. *Clinical Neurophysiology, 125*, 827-835.

Tyberghein, J., & Forrez, G. (1971). Objective (ERA) and Subjective (COR) Audiometry in the Infant. *Acta Oto-Laryngologica, 71*, 249-252.

van Dijk, J. E., van Olphen, A. F., Langereis, M. C., Mens, L. H., Brokx, J. P., & Smoorenburg, G. F. (1999). Predictors fo Cochlear Implant Performance. *Audiology, 38*, 109-116.

van Wijngaarden, S. J. (2001). Intelligibility of Native and Non-native Dutch Speech. *Speech Communication, 35*, 103-113.

van Zomeran, A. M., & Bouwer, W. H. (1994). *Clinical Neuropsychology of Aging.* New York: Oxford University Press.

Vaughan, H. G., & Ritter, W. (1970). The Sources of Auditory Evoked Responses Recorded from the Human Scalp. *Electroencephalography and Clinical Neurophysiology, 28,* 360-367.

Wagner, M., Shafer, V. L., Martin, B., & Steinschneider, M. (2013). The Effect of Native-Language Experience on the Sensory-Obligatory Components, the P1–N1–P2 and the T-complex. *Brain Research, 1522,* 31–37.

Walden, B. E. (1984). Speech Perception of the Hearing-impaired. In J. Jerger (Ed.), *Hearing Disorders in Adults* (pp. 263-309). San Diego: College-Hill Press.

Warrier, C. M., Johnson, K. L., Hayes, E. A., Nicol, T., & Kraus, N. (2004). Learning Impaired Children Exhibit Timing Deficits and Training-Related Improvements in Auditory Cortical Responses to Speech in Noise. *Experimental Brain Research*.

Werker, J. F. (1994). Cross-language Speech Perception: Development Change Does Not Involve Loss. In J. C. Goodman, & H. C. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 93-120). Cambridge, MA: MIT Press.

Werker, J. F., & Logan, J. S. (1985). Cross-language Evidence for Three Factors in Speech Perception. *Perception & Psychophysics, 37,* 35-44.

Werker, J. F., & Tees, R. C. (1984). Phonemic and Phonetic Factors in Adult Cross-language Speech Perception. *Journal of the Acoustical Society of America, 75,* 1866-1878.

Whiting, K. A., Martin, B. A., & Stapells, D. R. (1998). The Effects of Broadband Noise Masking on Cortical Event-Related Potentials to Speech Sounds /ba/ and /da/. *Ear & Hearing, 19*(3), 218-231.

Wible, B., Nicol, T., & Kraus, N. (2002). Abnormal Neural Encoding of Repeated Speech Stimuli in Noise in Children with Learning Problems. *Clinical Neurophysiology, 113,* 485-494.

Wilson, B. S., & Dorman, M. F. (2007). The Surprising Performance of Present-Day Cochlear Implants. *IEEE Transactions on Biomedical Engineering, 54*(6), 969-972.

Wilson, B. S., & Dorman, M. F. (2008). Cochlear Implants: Current Designs and Future Possibilities. *Journal of Rehabilitation Research & Developement, 45*(5), 695-730.

Wilson, B. S., Dorman, M. F., Woldorff, M. G., & Tucci, D. L. (2011). Cochlear Implants: Matching the Prothesis to the Brain and Facilitating Desired Plastic Changes in Brain Function. *Progress in Brain Research, 194*, 117-129.

Wilson, B. S., Finley, C. C., Farmer, J. C., Lawson, D. T., Weber, B. A., Wolford, R. D., . . . Schindler, R. A. (1988). Comparative Studies of Speech Processing Strategies for Cochlear Implants. *Laryngoscope, 98*(10), 1069-1077.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). Better Speech Recognition with Cochlear Implants. *Nature, 352*(6332), 236-238.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokosk, A., . . . Näätänen, R. (1999). Brain Responses Reveal the Learning of Foreign Language Phonemes. *Psychophysiology, 36*, 638–642.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of Early Auditory Processing During Selective Listening to Rapidly Presented Tones. *Electroencephalography and Clinical Neurophysiology, 79*(3), 170-191.

Wolpaw, J. R., & Penry, J. K. (1975). A Temporal Component of the Auditory Evoked Response. *Electroencephalography and Clinical Neurophysiology, 39*(6), 609-620.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical Experience Shapes Human Brainstem Encoding of Linguistic Pitch Patterns. *Nature Neuroscience, 10*(4), 420-422.

Wood, C. C., & Wolpaw, J. R. (1982). Scalp Distribution of Human Auditory Evoked Potentials: II. Evidence for Multiple Sources and Involvement of Auditory Cortex. *Electroencephalography and Clinical Neurophysiology, 54*, 25-38.

Woods, D. L., Knight, R. T., & Scabini, D. (1993). Anatomical Substrates of Auditory Selective Attention: Behavioral and Electrophysiological Effects of Posterior Association Cortex Lesions. *Brain Research. Cognitive Brain Research, 1*(4), 227-240.

Wu, C.-C., & Luo, X. (2013). Current Steering with Partial Tripolar Stimulation Mode in Cochlear Implants. *Journal of the Association for Research in Otolaryngology, 14,* 213-231.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., & Näätänen, R. (2010). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-language Users of English. *Journal of Cognitive Neuroscience, 22*(6), 1319-1332.

Zendel, B. R., Tremblay, C.-D., Belleville, S., & Peretz, I. (2015). The Impact of Musicianship on the Cortical Mechanisma Related to Separating Speech from Background Noise. *Journal of Cognitive Neuroscience, 27*(5), 1044-1059.

Zeng, F. G. (2004). Trends in Cochlear Implants. *Trends in Amplification, 8*(1), 1-34.

Zeng, F. G., Galvin III, J. J., & Zhang, C. (1998). Encoding Loudness by Electrical Stimulation of the Auditory Nerve. *NeuroReport, 9*(8), 1845-1848.

Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., & Segel, P. (2002). Speech Dynamic Range and its Effect on Cochlear Implant Performance. *Journal of the Acoustical Society of America, 111*(1), 377-386.

Zhang, F., Hammer, T., Banks, H.-L., Benson, C., Xiang, J., & Fu, Q.-J. (2011). Mismath Negativity and Adaptation Measures of the Late Auditory Evoked Potential in Cochlear Implant Users. *Hearing Research, 275,* 17-29.

Zhang, Q., Shi, J., Luo, Y., Zhao, D., & Yang, J. (2006). Intelligence and Information Processing During A Visual Search Task in Children: an Event-Related Potential Study. *Neuroreport, 17*(7), 747-752.

Zhao, F., & Stephens, D. (2007). A Critical Review of King-Kopetzky Syndrome: Hearing Difficulties, but Normal Hearing? *Audiological Medicine, 5*, 119-124.