PRELIMINARY INVESTIGATION OF THE UTILITY OF THE N400 FOR PASSIVELY ESTIMATING SPEECH HEARING ABILITY IN NOISE

by

Caroline Jamison

Submitted in partial fulfillment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
March 2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Recommended procedures for hearing assessment include the evaluation of speech hearing in background noise. Speech in noise assessments rely heavily on behaviour and attention, making them unsuitable for many patient populations. Electrophysiological measures are increasingly being used by audiologists to circumvent the limits of behavioural testing. This project looked at the possibility of using a well-studied event-related potential, the N400, as an objective measure of speech comprehension in noise. The N400 is associated with the processing of meaningful stimuli, and has been of particular interest in the study of written and spoken language comprehension. N400 amplitude varies with several factors including the degree to which a word is expected based on the surrounding context.

The effect of varying levels of speech-frequency background noise on the N400 as elicited by semantically anomalous spoken sentences in the absence of attention was investigated in eleven adults with normal hearing. It was hypothesized that the magnitude of the difference in N400 amplitude between congruent and incongruent trials (known as the N400 effect) would vary systematically with intelligibility, decreasing in more adverse listening conditions. Five signal to noise ratios relative to the behavioural threshold (-2 dB, threshold, +1 dB, +2 dB, and +4 dB) were tested as well as a quiet condition. The amplitude of the N400 effect did vary with intelligibility however this effect was nonlinear, with the +1 dB and -2 dB conditions having significantly smaller N400 effect amplitudes than the other conditions as a group, as determined through a partial least squares (PLS) analysis. This effect appeared to be driven by changes in the responses to incongruent, rather than congruent, stimuli. The results are discussed in the context of related literature on the N400 in adverse listening conditions. Because of the complex nature of the effect of noise on the N400 to sentence-level expectancy violations, this paradigm does not appear to be immediately useful for application in clinical speech in noise audiometry.

Keywords: event-related potentials, N400, speech in noise audiometry, objective audiometry

LIST OF ABBREVIATIONS AND SYMBOLS USED

ABR	Auditory brainstem response
APD	Auditory processing disorder
ASHA	American Speech-Language and Hearing Association
CAEP	Cortical auditory evoked potential
CI	Confidence interval
CP	Cloze probability
CST	Connected Speech Test
dB	Decibel
dBA	Decibel, A-weighted
EEG	Electroencephalography
ELAN	Early left anterior negativity
ERP	Event-related potential
fMRI	Functional magnetic resonance imaging
HI	Hearing impairment
HINT	Hearing in Noise Test
Hz	Hertz
LAN	Left anterior negativity
LiSN-S	Listening in Spatialized Noise – Sentences test
LV	Latent variable
MMN	Mismatch negativity
ms	Milliseconds
N400	A negative ERP that occurs approximately 400 ms after a meaningful stimulus
P1-N1-P2	The first positive, first negative, and second positive ERPs
PET	Positron emission tomography
PLS	Partial least squares
PSP	Postsynaptic potential
RTS	Receptive threshold for sentences
s	Seconds
SIN	Speech in noise
SNR	Signal to noise ratio
SNR-50	Signal to noise ratio at which 50% of test materials can be repeated correctly
SPIN	Speech Perception in Noise test
SRT	Speech reception threshold
QuickSIN	Quick Speech in Noise test
WIN	Words in Noise test
WIPI	Word Identification by Picture Identification test
µV	microvolts

ACKNOWLEDGEMENTS

Firstly I would like to thank my supervisor, Dr. Lauren Petley, for her guidance and assistance in every step of this project. Thank you Lauren for taking the time to explain things to me in detail, for compelling me to produce better work, for always being available and open to discussion, and for tolerating my at times slothful work habits. I have learned a tremendous deal from you and strive to be as knowledgeable and productive.

Next I would like to thank my thesis committee. Dr. Steve Aiken, thank you for your substantial work on this project, for the use of your laboratory, and for spending hours with Lauren programming my experiment in LabVIEW. Without this help my graduation would certainly have been delayed by another several years. Dr. Michael Kiefte, thank you for your input on the project. Your ideas have been extremely valuable. Thank you as well for the use of your recording equipment and your work on my stimuli. Thank you Dr. Aaron Newman for agreeing to be my external examiner and offering your insight on this project.

I would also like to thank fellow student Nathan Crawford for his help with EEG data collection, as well as my friends and family who supported me throughout the years in various ways including but not limited to feeding, sheltering, encouraging, nagging, commiserating, and distracting.

CHAPTER I: INTRODUCTION

Hearing assessment

Traditionally, hearing assessment aims to evaluate the peripheral auditory system including the external, middle, and inner ear as well as the auditory nerve. Due to the many structures and processes involved in hearing, hearing assessment includes many procedures. According to the guidelines of the American Speech-Language and Hearing Association, a basic hearing assessment should include a case history, external ear examination, otoscopic examination, acoustic immittance procedures, measurement of air-conduction and bone-conduction pure-tone threshold measures with masking as required, measurement of speech reception thresholds or speech detection/awareness thresholds with and without masking, word recognition measures, and speech-language screening (ASHA, 2006). Some of these procedures, such as otoscopic examination and acoustic immittance measurements, are objective and easy to obtain regardless of the patient's level of attention or arousal. Others require the examinee to be alert and attentive to stimuli. Pure tone audiometry and speech audiometry, with or without masking, fit into the latter category. The current project focuses specifically on speech in noise audiometry as the topic of interest.

Speech audiometry with or without background masking noise has been a recommended component of audiologic assessment since the 1940's (Carhart, 1946). It is generally accepted that speech audiometry gives additional information on hearing performance that cannot be obtained using pure tone audiometry alone. Speech audiometry in quiet involves detecting or repeating spondees (bisyllabic words with equal stress on both syllables) at the lowest level possible to determine the speech awareness or speech reception threshold (SRT); the level at which 50% of the spondees can be detected or repeated back correctly. It may also involve suprathreshold testing using monosyllabic words to obtain word recognition scores.

Speech in noise (SIN) measures use words or sentences presented with background noise. The listener must listen attentively and attempt to repeat the speech back to the examiner. The aim of SIN tests is typically to determine the signal-to-noise ratio (SNR) at which the listener can repeat back the target speech correctly 50% of the time. This is the SNR-50, and represents the SNR at the receptive threshold for sentences (RTS). Alternatively, a percent correct score can be obtained for several SNRs. Using sentence SIN tests in the clinic might give the audiologist the best picture of the patient's real-world performance, since these tests attempt to create more naturalistic listening conditions (Carhart & Tillman, 1970).

Why is speech in noise audiometry important?

Compared to average listeners, people with hearing impairments experience disproportionately more difficulty when listening to speech in a noisy environment. Evaluating speech hearing in noise is therefore a logical component of audiologic assessment. Indeed, recommended protocols for hearing aid fitting include SIN testing in both unaided and aided conditions (Mueller, 2003). However, there is some debate as to the usefulness of these tests. According to *The Hearing Journal*'s 2003 survey of hearing aid dispensers, only 19% of American dispensers routinely used hearing in noise tests as a pre-fitting assessment tool, and 30% routinely took aided speech measures. In 2010, 33% of the dispensers said they routinely used the QuickSIN test, which was by far the most popular SIN test (Mueller, 2010). Taylor (2007) wrote a systematic review on the use of speech audiometry pre- and post- fitting to determine if there were any "real-world" outcome measures to support its continued use. He found that aided and unaided scores on sentence-level SIN tests were only weakly correlated with hearing aid benefit as measured on self-report scales, and that word-level tests were worse indicators of real-world subjective benefit than sentence-level tests. However, subjective ratings of satisfaction with hearing aids likely rely on a host of factors including the quality of the counseling a hearing aid user received, their habitual listening environments, and their degree and type of hearing loss. Although SIN tests have limited value for predicting subjective outcome measures, there are many good reasons to conduct SIN audiometry pre- and post- fitting. Mueller (2001, 2003) describes several of these reasons, summarized below.

1. To help establish candidacy for borderline patients. People who present with only a

mild hearing loss in quiet may have much more difficulty with speech in noise. Therefore this is a crucial area of assessment for these patients, who might not be provided with amplification unless their difficulty with listening in noise is identified.

2. To aid in selection of appropriate amplification technology, features, and settings. This is an often-cited reason to conduct SIN testing. A baseline SIN assessment can help determine if directional microphones or other noise-reducing technology would be beneficial, and how much benefit someone is likely to receive from hearing aids.
3. To help detect an auditory processing disorder (APD). The presence of APD would influence clinical decision making, and some SIN tests can be used to help diagnose it.
4. To measure the benefit the hearing aid user is receiving and test special features of the hearing aid. Aided measures assist with judging appropriateness of current technology, outcome measures, and counseling. It may be important to demonstrate that the aids are actually working.

Assessment tools for speech in noise audiometry

There are several word- and sentence-level SIN tests available for adult and pediatric populations. In general, these tests involve listening to speech stimuli presented with background noise and repeating the speech segment back to the examiner. Sentence-level SIN tests for use with adults include the Speech Perception in Noise test (SPIN; Kalikow, Stevens, & Elliott, 1977), the Connected Speech Test (CST; Cox, Alexander, & Gilmore, 1987), the Hearing in Noise Test (HINT; Nilsson, Soli, & Sullivan, 1994), and the Quick Speech in Noise test (QuickSIN; Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004).

The SPIN and CST are fixed signal-to-noise ratio (SNR) tests. Both tests use multi-talker babble as background noise. The SPIN uses sentences 5-8 words long, but the last word of each sentence is the only word scored. Half of the sentences have high predictability and the other half low predictability. The test is scored as a percent correct with separate scores for the high and low predictability words. The CST uses segments of speech 9 to 10 sentences long. The score is based on the percent correct of 25 key words in each segment.

Adaptive SNR tests vary the SNR. These may be more useful than fixed SNR tests because they largely eliminate floor and ceiling effects (Mueller, 2003). They include the QuickSIN and the HINT. The QuickSIN uses sentences presented with 4-talker babble. The level of the sentences is fixed while the noise level varies. Five key words are scored in each sentence. The HINT consists of sentences presented in groups of 10 with noise that has been synthesized to match the long-term average spectrum of speech. The noise can be presented at 0°, 90°, or 270° azimuth, while the speech is presented at 0° azimuth. The noise is fixed at 65 dB SPL, while the level of the sentences varies. All the key words of the sentence must be repeated for a response to be considered correct.

The Words in Noise test (Wilson, 2003) is a word-level adaptive SNR test that is commonly used for evaluating adults as well as children as young as 6 years old. It uses monosyllabic words presented with mutitalker babble at seven different SNRs. The noise level is held constant at 70dB SPL while the signal level varies.

SIN tests have also been developed specifically for the pediatric population. These are similar to adult tests in terms of design and task demands; however, the speech stimuli have been selected for use with children and the tests have pediatric norms. The newest and probably most effective pediatric SIN test currently available is the Listening in Spatialized Noise – Sentences test (LiSN-S; Cameron & Dillon 2007; Michel Comeau, personal communication, July 13, 2014).

The LiSN-S is an adaptive SNR test administered through headphones, with different SIN conditions created using head-related transfer functions, so the listening environment is perceived as three dimensional. The competing speech consists of looped children's stories and its level is held constant at 55 dB SPL while the level of the target sentences is adjusted by the tester. There are four listening conditions tested, with the vocal identity and the perceived spatial location of the competing speech being either the same or different from the target speech. The target speech is always presented at 0° azimuth. The LiSN-S was originally designed for APD testing, and can be used with children as young as 6 years (Phonak, 2009).

Other examples of pediatric speech in noise tests are the HINT-C (Nilsson, Soli, & Gelnett, 1996), the Bamford-Kowal-Bench speech-in-noise (BKB-SIN) test with 4-talker babble (Ng, Meston, Scollie, & Seewald, 2011); the Word Identification by Picture

Identification test (WIPI); and the Pediatric Speech Intelligibility test (PSI; Jerger & Jerger, 1982). Many of these use single word stimuli instead of sentences, and require the child to point to a picture instead of giving a verbal response. The PSI test can be carried out with children as young as 3 years developmental age.

Shortcomings of available speech in noise tests

All of the above SIN tests, whether designed for use with adults or children, rely heavily on behaviour and attention. The listener must be attentive and attempt to decode speech presented in a challenging listening situation. Due to the high task demands, there is a large patient population that cannot undergo SIN testing. This population includes children younger than 3 years developmental age (although the most effective SIN tests can only be carried out from 6 years of age), and people with attentional or behavioural difficulties due to developmental or acquired conditions. The pediatric population is of particular concern given that for optimal speech and language development in hearing impaired children, appropriate amplification should be provided as soon as possible (Downs & Yoshinaga-Itano, 1999; Sininger et al., 2009). SIN testing is also highly relevant in the pediatric population since children are often in extremely noisy environments. Various studies have found that typical classrooms have 48-69 dBA of background noise (Crandell & Smaldino, 2000).

Another problem with sentence-level speech in noise testing is the facilitating effect of semantic context. In some cases, the redundancy in the signal makes it easy to guess content that may not have been heard. If SIN tests do not take semantic cues into account, they may over-estimate performance (Kalikow et al., 1977). For example, the HINT has been criticized because it does not control for the effects of semantic cues. To address this concern, some speech in noise tests, such as the QuickSIN and SPIN, have been designed to take semantic cuing effects into account.

Perhaps the largest drawback of behavioural SIN testing is its inability to disentangle cognition and hearing. As will be discussed below, comprehension of speech in noise is a complex process that relies heavily on cognitive factors. In daily life, hearing and cognitive processes are always linked. However, the goal of hearing assessment is usually to assess the auditory system only, without reference to cognitive factors. It would be useful to disentangle the effects of hearing sensitivity from cognition and behaviour for both clinical and research purposes. For example, obtaining a differential diagnosis of APD in children is difficult because often the assessments (which include SIN tests) cannot adequately distinguish APD from other cognitive/behavioural disorders (Jerger & Musiek, 2000). To get around this issue, the auditory brainstem response (ABR), an objective measure of sound encoding in the brainstem, has been suggested as a tool to diagnose APD (Jerger & Musiek, 2000; Anderson & Kraus, 2010). To better understand the relationship between cognition and hearing, we can look more closely at SIN processing and the factors that affect it.

Effects of noise on speech processing

People with normal hearing are able to understand speech even in the presence of competing speech or background noise that is louder than the target speech. This is known as the "cocktail party phenomenon" (Cherry, 1953). In these situations, auditory scene analysis and stream segregation are necessary. Auditory scene analysis and stream segregation involve organizing the sounds in the environment and assigning source information to them (Bregman & McAdams, 1994).

Signal and noise properties, binaural effects, and cognitive factors all contribute to the complex process of auditory scene analysis (Bronkhorst, 2000). There is a high degree of signal redundancy or "extra information" in natural speech. This is evidenced by the fact that in quiet listening conditions, listeners are still able to accurately perceive speech that is severely degraded, e.g. by bandpass filtering through a narrow "spectral slit" (Warren, Riener, Bashford, & Brubaker, 1995). This redundancy likely helps speech perception in noise by making the signal more robust to degradation. Normally hearing people use the speaker's vocal characteristics such as fundamental frequency and timbre, and timing cues such as onsets and offsets, to identify a particular speaker and focus attention on them (Anderson & Kraus, 2010). Not surprisingly, therefore, SIN tasks are more difficult when the signal and competing sounds are more similar to each other (Sarampalis, Kalluri, Edwards, & Hafter, 2009).

There are two main types of masking: informational and energetic. Informational masking is caused by a competing speech stream, whereas energetic masking is caused by sounds that overlap in time and frequency with the target speech rendering portions of the target inaudible (Brungart, Simpson, Ericson, & Scott, 2001). Informational masking generally presents a more difficult listening situation than energetic masking (Sperry, Wiley, & Chial, 1997). Speech babble and noise that has been filtered to match the long-term spectrum of speech, as used in many SIN tests, are examples of energetic maskers. A high degree of energetic masking causes signal degradation at the auditory nerve and brainstem, forcing the listener to rely on higher-level processes to decode the degraded signal (Delgutte, 1980; Anderson, Skoe, Chandrasekaran, & Kraus, 2010). This higher level auditory processing is not well understood and several processes are likely to be involved. For example Ding and Simon (2013) suggest that slow temporal modulations help us to understand speech in noise because these lower-frequency neural responses to speech remain stable at noise levels when higher-frequency neural responses are degraded.

Auditory scene analysis allows the listener to focus attention on the signals of interest and pick them out of the acoustic background. The incoming signals must then be associated with meanings stored in the mental lexicon. Speech perception is complex, and involves many cognitive processes. These include perceptual grouping, lexical segmentation, categorical perception, and perceptual learning (reviewed in Davis & Johnsrude, 2007). Perceptual grouping is the process whereby distinct sounds such as frication noises and vowels are perceived as a single stream, while lexical segmentation involves separating this stream into individual meaningful units. These processes depend on both bottom-up and top-down processing. For example, "migration" is a phenomenon in which, if two different syllables are presented to the two ears, the two syllables are combined into a single new unit, which contains elements of both of the "true" syllables that were presented (Cutting, 1975). Migration is affected by bottom-up properties because it is more likely to happen with a more similar voice speaking both syllables, and with sources that are closer together in space. It is also affected by top-down processing because migrations that create words are more common than those that create pseudowords (Kolinsky & Morais, 1996). Thus, language-specific vocabulary knowledge influences the way the auditory system groups incoming signals. It has been suggested that in highly noisy conditions in which lexical information is unreliable, the system relies more on bottom-up processing (Davis, Marslen-Wilson, & Gaskell, 2002).

Categorical perception in the context of speech processing refers to the perception of many acoustic variants as the same underlying unit. For example, syllables synthesized with an ambiguous consonant (e.g. somewhere between /ba/ and /pa/) are perceived as one syllable or the other, with the boundary between categories sensitive to top-down influences. Norris, McQueen, and Cutler (2003) presented words that ended with an ambiguous fricative between /s/ and /f/. Listeners who heard word stems biased in favor of the /f/ interpretation were subsequently more likely to judge isolated presentations of the ambiguous sound as an /f/ compared to listeners who heard it in /s/-biased contexts, or in unbiased contexts. The flexibility in phonetic boundaries as illustrated by this study provides some of the most compelling evidence for the top-down influence of lexicality on categorical perception (Davis & Johnsrude, 2007).

Perceptual learning is especially relevant to SIN processing as it involves quickly learning to decode speech in unfamiliar (e.g. degraded or strongly accented) conditions. People with normal hearing can quickly learn to decode severely degraded speech through top-down influences. For example, if they are told that the signal they are hearing contains speech, listeners are more able to decode sine-wave speech, which lacks the amplitude comodulation and harmonic structure of natural speech (Remez, Rubin, Pisoni, & Carrell, 1981). Subjective intelligibility of speech in noisy conditions also increases with stimulus repetition, as does performance when identifying noise-vocoded speech (Jacoby, Allan, Collins, & Larwill, 1988; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005). However, in Davis et al. (2005), no improvement was seen for real words when subjects were trained using nonword stimuli. Therefore, the authors concluded that lexical feedback is required for perceptual learning. Davis and Johnsrude (2007) suggest that in addition to perceptual learning supported by lexical knowledge, pragmatic knowledge as well as the use of visual cues contribute to successful SIN decoding.

Several lines of research have also implicated more general effects of attention and cognition in SIN performance. For example, musicians have increased SIN performance as measured by the HINT and QuickSIN compared to non-musicians due partially to their increased auditory memory ability (Billings, McMillan, Penman, & Gille, 2013; Parbery-Clark, Skoe, Lam, & Kraus, 2009). Several studies have demonstrated that cognition is involved in SIN performance. For example, Sarampalis et al. (2009) found that in normally hearing listeners, increasing levels of background babble reduced preformance on the SPIN, and also reduced performance on simultaneous cognitive tasks (either a word memory task or a visual reaction time task), suggesting that increased cognitive effort was needed for the SPIN as demands increased, at the cost of the simultaneous tasks. When a noise-reducing algorithm was applied, performance on the SPIN did not improve but performance on the concurrent tasks did. The authors interpreted this as support for Hafter and Schlauch's (1992) theory that the algorithm performs functions akin to the system's natural process, which results in no intelligibility benefit but frees cognitive resources for other tasks. This may explain why, despite these algorithms' ineffectiveness at actually increasing speech intelligibility, hearing aid users often express greater perceived benefit and ease of listening when using them (Keidser, 1996).

Research on ageing is also informative on the association between cognition and SIN performance. Humes et al. (2012) reviewed the literature on age-related changes in auditory processing and cognition. In studies judged to be without a confound of presbycusis, significant decreases in SIN performance were commonly found in older age groups compared with younger ones. This implies that the decreased SIN performance was due to the decreases in cognitive ability that are typically seen in normal ageing. Indeed, the vast majority of studies investigating cognitive ability (as measured in myriad ways such as the Mini Mental State Examination, IQ tests, Wechsler Memory Scale, etc.) found that decreased cognitive ability was negatively correlated with SIN performance. It has also been suggested that the age-related changes in speech perception are due in part to reduced neural synchrony (Tremblay, Piskosz, & Souza, 2003).

In summary, successful comprehension of speech in noise depends on signal and noise properties, on specific auditory processing abilities (e.g., slow temporal modulations, use of binaural cues, phase locking), on specific cognitive processes (e.g. perceptual learning), and more general cognitive factors such as attention and memory.

As with most cognitive processes, there is a large amount of variability in speech in noise performance in the general population (Billings et al., 2013).

People with hearing impairments (HI) experience increased difficulty in situations where there is a lot of background noise. Hearing impairment poses a challenge for listening in background noise (1) because of reduced audibility and (2) because of increased difficulty with some aspect(s) of auditory scene analysis (Bronkhorst, 2000). Reduced audibility refers to the loss of hearing sensitivity in a frequency range important for hearing certain speech sounds. However, even when sounds are amplified so that audibility is “normal” in quiet conditions, background noise poses a challenge for people using hearing aids or cochlear implants. This may be attributable to signal distortion, loss of binaural cues, and auditory processing deficits related to the hearing loss. People with HI therefore rely on cognitive factors to compensate, as evidenced by their decreased performance on simultaneous cognitive tasks compared with individuals with normal hearing, in situations when both groups score the same for SIN performance (Rabbitt, 1991). This highlights how behavioural SIN testing is necessarily linked to cognitive function, which becomes even more important for SIN processing in people with hearing impairments. It could be argued that behavioural SIN tests are therefore certainly testing cognitive functions as well as the functioning of the auditory system.

In other areas of audiometry, such as infant hearing screening and pure tone testing, electrophysiological measures such as the auditory brainstem response are used to circumvent the inherent limits of behavioural testing. A SIN test using electrophysiological measures and which did not involve any behavioural task would serve to greatly reduce the contribution of top-down, cognitive influences on the test results. The purpose of the present work is to help determine whether a test of speech comprehension in noise could be developed using electroencephalography (EEG) rather than behavioural measures.

Electroencephalography

The human neocortex contains approximately 20 billion neurons. Neurons are constantly exchanging ions with the extracellular matrix, for example to propagate action potentials or maintain their resting potentials. This movement of ions creates minute electromagnetic fields. When a group of neurons fires in synchrony, these fields become larger. Due to the repulsion of like charges, the electricity spreads through the cerebral and extracerebral tissue, and can be measured at the scalp using electroencephalography (EEG). Developed in the 1920's-1930's, EEG detects electric field changes that are associated with synchronous postsynaptic potentials (PSPs) in groups of neurons, typically in the cerebral cortex. PSPs are changes in the membrane potentials of neurons when they receive synaptic signals. The voltage changes detected using EEG mostly originate from the extracellular current flow resulting from PSPs. The sum of the PSPs of neurons with similar orientations is commonly modeled as an equivalent current dipole (Luck, 2005).

The chief advantage of EEG over other functional neuroimaging methods (e.g. fMRI, PET) is that it measures neural activity with very high temporal resolution, on the order of milliseconds (ms). EEG is also much less expensive than other neuroimaging methods, making it more appropriate for widespread clinical use. However, there are some important drawbacks to this method. Any muscle or ocular movement creates artifacts in the signal that must be removed. Artifacts commonly arise from jaw movement, eye blinks, eye movements, and the heart. Because electrical resistance varies for the different tissues that the signal must travel though to reach the scalp, there is a large amount of distortion by the time the voltage changes are measured. Additionally, the distribution of currents from different neural sources can overlap, making it difficult to separate different simultaneous generator processes. These factors make it difficult to determine precisely where the activity of interest originates in the brain, an issue known as the "inverse problem". A variety of source modeling techniques exist to infer which cerebral sources are responsible for the EEG signal observed at the scalp, each of which has their own pros and cons with relation to accuracy and susceptibility to user bias.

Evoked responses

Starting in the 1930's, it was observed that reliable patterns of voltage deflection occurred in the EEG in response to specific stimulus types. These responses are called evoked responses, evoked potentials, or event-related potentials (ERPs). Evoked responses are typically very small compared to background brain activity, and are therefore obtained by averaging the responses to many stimuli together, effectively "averaging out" background brain activity. Evoked responses can be used to track sound processing in the ascending auditory pathway from the auditory vestibular nerve to higher-level cortical structures.

ERPs can be used for a variety of research and clinical purposes. In audiology, the use of the auditory brainstem response (ABR) is now standard practice for newborn hearing assessment, among other applications. Cortical auditory evoked potentials (CAEPs) are cortical ERPs that are reliably measured in response to specific auditory stimuli. They can be used to gain information about auditory processing at the cortical level, which the ABR does not provide. Some CAEPs that can be used in clinical audiology are the P1, N1, P2, MMN and P3b. These first three are sometimes collectively called the P1-N1-P2 complex.

The P1-N1-P2 complex reflects sound processing at the level of primary and secondary auditory cortex. It can be used to measure hearing sensitivity physiologically, estimating the behavioural threshold to within 5-10 dB (Hyde, 1997). In the 1960's it was commonly used for this purpose, before the use of the ABR largely replaced this method (Martin, Tremblay, & Korczak, 2008). The P1 is a small amplitude positive peak occurring about 50 ms after stimulus onset. It is a result of activity in the primary auditory cortex, hippocampus, planum temporale, and other lateral temporal regions (Geisler, Frishkopf, & Rosenblith, 1958; Wood & Wolpaw, 1982; Martin et al., 2008). The P2 occurs at 180 ms poststimulus and is also generated in multiple cortical auditory areas (Steinschneider & Dunn, 2002). It is less commonly studied than the P1 and N1 and less well understood (Martin et al., 2008).

Most research on the P1-N1-P2 complex focuses on the N1. The N1 is a large negative potential that peaks 80-120ms after stimulus onset (Davis, 1939). The N1 can be elicited using any auditory stimulus, but for audiology research clicks or synthesized speech syllables e.g. /ba/ are commonly used. The N1 is made up of three sub-components with neural generators in different parts of Heschl's gyrus and the planum temporale (Näätänen & Picton, 1987; Godey, Schwartz, De Graaf, Chauvel, & Liégeois-Chauvel, 2001). It is sensitive to a change in the acoustic properties of a stimulus, meaning the amplitude of the N1 decreases with successive identical stimuli, and increases when a novel stimulus is presented. N1 amplitude is also increased when attention is directed to the sounds, compared with passive conditions (Woldroff & Hillyard, 1991). Although the amplitude of the N1 does change with a change in stimulus properties, it is considered an "obligatory" response to stimulus and does not reflect discrimination of different sounds (Näätänen & Picton 1987).

Conversely, the mismatch negativity (MMN) is an ERP that does reflect discrimination of sounds. It is a negative wave that is typically observed shortly after the N1 in response to a stimulus that breaks an expected pattern, e.g. a 4000 Hz tone in a string of 500 Hz tones. It is not dependent on attention and is thought to reflect the mismatch between the novel (deviant) stimulus and the "sensory memory trace" of the previous stimuli (Näätänen, Jacobsen, & Winkler, 2005). Clinical studies have looked at the MMN in diverse patient populations. In audiology, the MMN can be used clinically for assessment of auditory neuropathy (Gabr, 2011).

The P3b is also seen in response to a deviant stimulus but it is dependent on attention and is best recorded when deviant stimuli are designated as actively-detected target stimuli. It is a large positive component that peaks around 300 ms post-onset of a deviant stimulus. It is recorded maximally from centroparietal electrode sites (Sutton, Braren, Zubin, & John, 1965; Picton, 1992; Martin et al., 2008). P3b amplitude and latency vary with the difficulty of the task, the improbability of the deviant stimulus, and attention (Picton, 1992). Due to the need for the listener to actively respond to stimuli, the P3b is a less useful component than the MMN for use in clinical audiology (Martin et al., 2008).

The P1, N1, P2, MMN and P3b can be used to look at sound detection, pure tone threshold estimation, and discrimination of sounds, but these components do not provide information on speech or language processing beyond the phoneme level. However, there are some ERPs that have been used to look at responses to language including the P600, early left anterior negativity (ELAN), left anterior negativity (LAN) and N400. These ERPs are not specific to the auditory modality and can be elicited using written or spoken language, as well as other language-related stimuli like pictures.

The ELAN is a negative deflection seen 125-250 ms after specific kinds of word order violations, recorded at anterior left electrode sites (Friederici, Pfeifer, & Hahne, 1993). Steinhauer and Drury (2012) reviewed the literature on the ELAN and concluded that the ELANs obtained in many auditory studies do not necessarily involve the same underlying cognitive activity and in many cases are likely to be artifacts, calling into question the validity of this ERP for studying auditory language processing.

The P600 and LAN are deflections seen in response to many kinds of syntactical errors in sentences. These include violations of verb tense or number, word order, and in the case of the P600, complex or ambiguous sentences that require re-evaluation (Gunter, Stowe, & Mulder, 1997; Osterhout & Holcomb, 1992; Gouvea, Phillips, Kazanina, & Poeppel, 2010). The LAN is a negative deflection seen 300-400 ms after the syntactic error, whereas the P600 is a positive deflection that peaks about 600 ms postistimulus (Gunter et al., 1997). There is uncertainly over whether the P600 is a language-specific response or a more general response to an unexpected stimulus similar to the P3b. P600-like late positivities have also been observed in nonlinguistic conditions, for example in response to violations of harmonic structure in music (Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Gouvea et al., 2010).

The most widely researched language-related component is the N400 (Luck, 2005), which is sensitive to semantic, rather than syntactic, violations. The N400 was first described by Kutas and Hillyard (1980) in a study where people read sentences that had unexpected or nonsensical final words. For example, if the sentence "I planted string beans in my sky" were presented, an enhanced N400 would occur in response to the word "sky" compared to if the expected final word "garden" had been presented. The N400 is a broad negative wave, peaking about 400 ms poststimulus, which is largest at central and parietal sites (Kutas & Hillyard, 1980). It is measured as slightly larger from right hemisphere recording sites in the visual modality, but in the auditory modality is more symmetrical, more frontal, and longer lasting (Kutas & Hillyard, 1982; Van Petten &

Luka, 2006; Kutas & Federmeier, 2011). It appears to be generated in left temporal regions (Kutas, Hillyard, & Gazzaniga, 1988; Luck, 2005; Lau, Phillips & Poeppel, 2008).

The N400 can be elicited by single words, phonologically legal nonwords (pseudowords), and nonlinguistic stimuli (Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008); however, it is most commonly associated with semantic tasks. Studies seeking to elicit N400s typically use either a semantic anomaly paradigm or a semantic priming paradigm (Lau et al., 2008). Kutas and Hillyard's 1980 study is an example of the semantic anomaly paradigm, in which sentences are presented with a semantically congruent or incongruent final word, and increased N400 amplitude is observed for the incongruent stimuli. A semantic priming paradigm presents pairs of single words, and an enhanced N400 is observed when the second word is semantically unrelated to the first. The N400s obtained from both these paradigms have similar latencies and scalp distributions and are therefore accepted to represent the same underlying activity, although the semantic anomaly paradigm generally results in larger amplitude N400s than the semantic priming paradigm (Kutas, 1993). Regardless of the experimental paradigm used to evoke the N400, the term "N400 effect" refers to the difference in N400 amplitude between congruent and incongruent nouns, or related and unrelated ones. Typically the averaged response to the congruent or related words is subtracted from the averaged response to the incongruent or unrelated words. This yields a difference wave that can be used to determine the magnitude of the N400 effect and compare its properties across experimental conditions.

In a semantic anomaly paradigm, N400 amplitude is influenced by the extent to which the final word violates expectation (Kutas & Hillyard, 1984). The extent to which a final word is expected is generally expressed by the Cloze probability (CP) of a sentence (Taylor, 1953). Cloze probabilities are established by having a large number of people complete a sentence with a word that will sensibly end the sentence. For example, in Bloom and Fishler's (1980) seminal study, the sentence "Captain Sheir wanted to stay with the sinking ______" was completed by 97% of participants with "ship" and by 2% with "raft." This is an example of a high (97%) CP sentence. Block and Baldwin (2010) defined high CP as a sentence having $\geq 67\%$ chance of being completed by a specific single word. Larger amplitude N400s are seen when higher Cloze probabilities are violated, even if the final word is not semantically incongruent (Kutas & Hillyard, 1984). For example, using the sentence above, an enhanced N400 would occur if “frigate” were presented compared with “ship,” but an even larger N400 would be expected to occur if “walnut” were presented.

The precise functional significance of the N400 remains unknown. Traditionally, there has been debate as to whether the N400 represents prelexical or postlexical processing. The postlexical view is that the N400 reflects the integration of the target word with the surrounding context, after word recognition has occurred. According to this theory, increased effort is needed to integrate words that do not fit well with the semantic context, leading to N400 amplitude increases for less expected words. This theory can account for the top-down effects on the N400 and its multimodal nature. In contrast, the lexical view states that the N400 reflects processing prior to the activation of a meaning in the semantic memory, or the facilitated activation of a word meaning from memory. This theory can account for the bottom-up influences on the N400 such as lexical frequency and neighbourhood size, as well as the fact that the N400 is observed even in response to meaningless stimuli such as pseudowords and nonwords (Lau et al., 2008; Kutas & Federmeier, 2011). It has also been suggested that the N400 reflects processing at the exact time of word recognition (Kutas & Federmeier, 2000). None of these theories can account for all of the top-down and bottom-up effects on the N400 that have been obtained empirically. For example, no theory can fully explain why responses to concrete and abstract words differ in incongruent conditions, but not in congruent ones Holcomb, Kounios, Anderson, & West, 1999). Similarly the substantial and immediate top-down effect of discourse-level context on the N400 is not adequately accounted for by these theories, which focus on a more narrow view of context and operate on the assumption that words must be recognized before they are integrated into context (Berkum, Hagoort, & Brown, 1999; Kutas & Federmeier, 2011). The above theories also cannot satisfactorily explain the results of the bodies of research looking at N400s obtained during sleep or the attentional blink, as well as N400s obtained using nonlinguistic stimuli (Ibáñez, Martin, Hurtado, & López, 2009; Kutas & Federmeier, 2011). Despite the uncertainty of its precise function, the N400 has been an extremely useful ERP for the study of language comprehension and its interaction with attention, perception, and memory (Kutas & Federmeier, 2011).

Effects of noise on CAEPs and the N400

Billings and colleagues (2009, 2013) looked at the effects of signal level and SNR on several CAEPs in the absence of attention by presenting tones and syllables at different levels and with differing levels of speech spectrum continuous noise. They found that signal level does not have a significant effect on CAEP amplitude or latency when signals are presented with background noise. N1 and P2 amplitude decreased significantly and latency increased significantly with decreasing SNR in the absence of attention. Billings et al. (2013) also determined participants' SRT behaviourally and found that they could predict this SRT to within 1 dB using the N1 amplitude and latency when the signal was presented at 70 dB and with an SNR of +5. This is promising for the development of SIN tests at the phoneme level based on electrophysiological measures rather than behavioural ones.

Relatively few studies have investigated the effect of noise and SNR on the N400 response and they vary widely in methodology and results. A handful of studies have looked at the N400 effect in response to sentences with various levels of acoustic distortion and found that the amplitude of the N400 effect decreases with increasing distortion (Strauß, Kotz & Obleser, 2013; Obleser & Kotz, 2011; Aydelott, Dick, & Mills, 2006). Connolly, Phillips, Stewart, and Brake (1992) used sentences from the SPIN and multi-talker babble to examine the effect on the N400. They found that masking babble significantly increased the latency of the N400 response for both high constraint and low constraint sentences, but had no significant effect on N400 amplitude. The authors did not state the SNR at which the stimuli were presented. Conversely, Daltrozzo, Wioland, and Kotchoubey (2005) found that N400 amplitude did vary with SNR for sentences presented with a pink noise masker: a high level of masking obliterated the N400, a moderate level of masking attenuated it, and a low level of masking slightly enhanced the N400 amplitude.

Jerger, Greenwald, Wambacq, Seipel, and Moncrieff (2000) attempted to create a speech in noise test using the N400 but were unsuccessful in eliciting the N400 response. Their procedure was quasi-dichotic: participants listened to a narrative broadcast from two speakers, but the second speaker lagged about 20s behind the first. Both channels contained periodic semantic and syntactic violations to elicit the N400, but participants were asked to keep track of how many they detected from only one side. The researchers observed a slight negativity at 400 ms and a large positivity at 900 ms to attended incongruencies but did not detect a discernable N400. Their difficulties in evoking an N400 may have been due to the choice of stimuli. The incongruent stimuli were a mix of syntactic and semantic violations, which are known to affect evoked responses in different ways. Also, there was no measurement or attempt to control the Cloze probability of their sentences.

In another study, Romei, Wambacq, Besing, Koehnke, and Jerger (2011) studied the N400 in multi-talker babble using a priming paradigm. They presented a series of 3 words in 20-talker babble at a SNR of +9 dB. Participants were asked to judge if the third word was semantically related to either of the first two. Analyses were done on the evoked responses to the first and second words, which were related to each other 50% of the time. They found that in babble the N400 amplitude was increased at the right central region, and that babble had slightly differing effects on the responses to related vs. unrelated words.

In summary, relatively few studies have examined the effects of noise and SNR on the N400. In general, these studies have found that the N400 is preserved in noise, but only Daltrozzo and colleagues (2005) systematically varied listening conditions to observe effects on the N400 in noise. According to their results, the N400 decreased in amplitude as intelligibility decreased. However, no study has explored the two crucial features for the design of an N400-based SIN test to replace behavioural measurements: 1) the relationship between the amplitude of the N400 and the behaviourally-measured RTS, and 2) whether these effects are observable under conditions of passive stimulation.

The current study

The aim of this project was to determine the effect of background noise on the N400 response to determine whether the N400 could be used as a physiological measure of speech understanding in noise. This kind of test would be useful for the pediatric population and other populations in which behavioural SIN testing is not possible. Using the N400 as an objective approach to testing hearing in noise in children would be appropriate because the N400 is seen in school-aged children (Holcomb, Coffey, & Neville, 1992). N400-like responses have also been observed in 36 and 48 month olds (Silva-Pereyra, Rivera-Gaxiola, & Kuhl, 2005), and even in 19-24 month olds (Friedrich & Friederici, 2005). Attention is not necessary for N400 elicitation (Relander, Rämä, & Kujala, 2009), and even in 1-2 year old children, N400-like responses can be evoked without taking measures to draw attention to the evoking speech (Friedrich & Friederici, 2005). Another advantage of using the N400 is that the method would depend on incongruent sentences rather than congruent ones, eliminating the possible effect of semantic cueing.

The current study used sentences with high Cloze probability, and presented them simultaneously with speech frequency noise, similarly to the HINT. SNR was varied across experimental conditions. This allowed observation of changes in the N400 across intelligibility conditions. Importantly, experimental procedures were used to direct attention away from the evoking speech. This design was intended to help determine whether a passive speech in noise test can be designed using N400. The results of this test were compared to the RTS as determined using a HINT-like procedure to determine the correspondence between physiological speech measurement and conventional speech reception thresholds.

Hypotheses:

1. The N400 will be elicited by sentences presented with speech-frequency noise.
2. N400 amplitude will vary with intelligibility such that decreases in amplitude will occur with decreased intelligibility.

3. The N400 effect will vary systematically with respect to signal to noise ratio relative to the behavioural threshold, with obliteration of the N400 effect at or near the behavioural threshold, and increasing N400 effect amplitude with increased intelligibility.

CHAPTER II: PILOT STUDY

Purpose

To evoke the N400, lists of sentences with high Cloze probability were created from Block and Baldwin's (2010) list of high CP sentences. One randomly selected sentence list was used to measure SNR-50, while the remaining lists were used for evoking the N400. The purpose of the pilot study was to determine whether these sentence lists were equivalent for intelligibility in noise, to ensure that the behaviourally measured SNR-50 would be valid for the sentence lists that were used to obtain the N400 in the electrophysiological portion of the experiment.

Methods

Participants

Six female adults (age 25-29; mean = 26.5, SD = 1.4) participated in the pilot study. All participants had no reported history of hearing loss or neurological problems, reported their first language to be English. Normal hearing status was verified via a pure tone hearing screen in soundfield at 20 dB HL at octave frequencies from 250-8000 Hz. Participants provided informed consent in accordance with the Research Ethics Board at Dalhousie University and were compensated ten dollars.

Stimuli

One hundred and fifty sentences with a minimum Cloze probability of 0.89 were selected from Block and Baldwin (2010). To compose stimuli for the incongruent condition, the terminal words of these 150 sentences were rearranged such that the congruent ending for one sentence became the incongruent ending for another sentence, in which it was contextually unpredictable. This served to control for lexical frequency. In all cases, incongruent words were matched for word class (i.e., nouns were replaced only with nouns), and plurality (i.e., plural words were replaced with plurals), and the initial phonemes of the congruent and incongruent words were different. This resulted in a total of 300 sentences (150 congruent, 150 incongruent), which were randomly divided into 15 lists of 20 such that no sentence was presented in both the congruent and incongruent form within the same list. Each list consisted of 10 congruent and 10 incongruent sentences, and a one-way ANOVA verified that Cloze probability was not significantly different across lists [$F(14) = 0.89$, $p = 0.57$].

All sentences were recorded by a young female speaker, and were spoken at a natural rate. Voice recording was performed using a Marantz PMD671 in a sound-shielded audiometric booth. Incongruent and congruent versions of each sentence were recorded separately. Full sentences, including the terminal words, were equalized for loudness using RMS normalization and speech-frequency noise was generated in MATLAB (The Mathworks, Natick, MA) using the spectrum of the recorded speech. For the pilot, only the sentence stems were used (i.e., the congruent or incongruent terminal words were omitted). This was done to reduce the well-known effect of sentence context, in which high contextual probability increases the intelligibility of speech in noise (Kalikow et al., 1977).

Procedure

Participants were seated in a sound-shielded audiometric booth with two speakers positioned at 45° and 315° azimuth. Stimuli were presented using a PC via a GSI 61 audiometer (Grason-Stadler, Eden Prairie, MN). A procedure similar to that of the HINT, but using stimuli developed for this study, was used to determine each participant's SNR-50 for each of the 15 lists. The standard HINT procedure is as follows: First, the starting speech level of the test must be determined. The first sentence of a randomly selected list is presented at 60 dB HL, while the noise is presented at 65 dB HL. If it is not correctly repeated, the sentence level is increased in 4 dB steps until it is correctly repeated. This is the starting SNR for the test. The next sentence is then presented at the starting SNR. If it is correctly repeated, the level of the next sentence is decreased by 4 dB. If it is incorrectly repeated, the level of the next sentence is increased by 4 dB. This procedure continues until the fifth sentence has been presented. Starting after the fifth sentence, 2 dB steps rather than 4 dB steps are used. This procedure continues until all the sentences have been presented. For 20 sentence presentation, 2 HINT lists of 10 sentences are used. For the slightly modified procedure used in this study, the level of the sentences was held at 70 dB HL and the noise level began at 75 dB HL. The level of the noise was then varied in 4 and 2 dB steps, in the same manner as the speech level is varied in the HINT. Participants were instructed to listen to the sentences stems in noise and repeat them back to the examiner as well as possible. Importantly, participants were warned that the sentences did not contain a final word. As in the HINT, for a response to be considered correct, all of the words in the sentence had to be correct except for the/a substitutions. List order was randomized across participants.

Data analysis

A one-way repeated measures ANOVA was performed to compare the mean SNR-50 (SNR at which a sentence is repeated correctly 50% of the time) across lists.

Results

The effect of word list on mean SNR-50 approached, but did not reach statistical significance [$F(14,70)=1.8$, $p = 0.06$]. Mean SNR-50 ranged from -8.3 (list 6) to -6.1 (list 7). The mean SNR-50 across all sentence lists is illustrated in Figure 1.

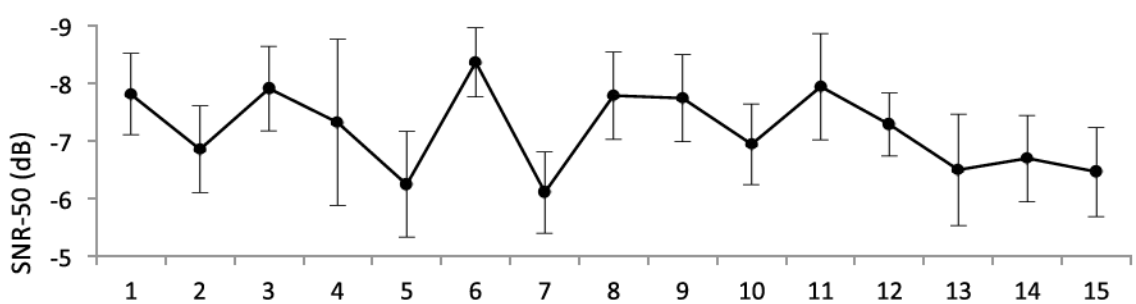

Figure 1. Mean SNR-50 for each list. Error bars represent standard error. A more negative SNR-50 indicates the list was more intelligible in noise.

CHAPTER III: ELECTROPHYSIOLOGICAL STUDY

Methods

Participants

Fourteen adults participated in this study. Data from three participants were not analyzed due to low data quality, e.g. due to too much movement artifact or poor electrode connections. Data from 11 participants (4 female; ages 21 – 35, mean = 27, SD = 3.9) were retained for analysis. All participants were right-handed, as confirmed by a modified version of the Edinburgh handedness questionnaire (Cohen, 2008), had normal or corrected to normal vision, normal hearing, had no known neurological problems, and their first language was English. Normal hearing status was verified via a pure tone hearing screen in soundfield at 20 dB HL at octave frequencies from 250-8000 Hz. All participants were paid for their participation. Participants provided informed consent in accordance with the Research Ethics Board at Dalhousie University.

Stimuli

The stimuli were the 15 lists of 20 sentences described in the pilot section. Each list contained 10 sentences with congruent terminal words and 10 sentences with incongruent terminal words. The sentences had Cloze probabilities of at least 0.9 and were presented in their entirety, including the terminal words. The close probabilities of the lists were not significantly different, nor were their intelligibilities, as determined in the pilot study.

Procedure

Behavioural threshold measurement

Participants were seated in a sound-shielded audiometric booth with two speakers positioned at 45° and 315° azimuth. Stimuli were presented using a PC via a GSI 61 audiometer (Grason-Stadler, Eden Prairie, MN). At the beginning of the experimental session, following pure tone screening, each participant's SNR-50 was determined using the standard HINT (20 sentence presentation), following the HINT procedure as described in the pilot methods. In brief, 4 dB and 2 dB steps in sentence level were used to determine the SNR-50 while noise level was held constant at 65 dB HL. Following SNR-50 measurement using the HINT, the participant's SNR-50 for the experimental stimuli was determined using one of the 15 recorded lists, according to the procedure described in the pilot study. A random list was selected for each participant, and this list was excluded from the rest of the experimental conditions. As in the pilot study, terminal words were not included in the behavioural test to reduce the effect of context on intelligibility. The results of this test determined the participant's threshold, which was used to determine the appropriate signal to noise ratios for all experimental conditions.

Experimental conditions

In individuals with normal hearing, a 1 dB increase in signal to noise ratio (SNR) using short sentences presented with speech-frequency background noise results in an 11-19% improvement in speech intelligibility performance, with the greatest performance increase per dB SNR increase close to the behavioural threshold (Plomp & Mimpen, 1979; Nilsson et al., 1994). Based on this relationship, five intelligibility conditions were established: SNR was lowered 2 dB below the SNR-50 for the very low intelligibility condition, kept at SNR-50 for the threshold condition, raised by 1 dB for the moderate intelligibility condition, raised by 2 dB for the moderately high intelligibility condition, and raised by 4 dB for the high intelligibility condition. Following presentation of each of the intelligibility conditions, all sentences were presented again in quiet. The purpose of this was twofold: first, to examine the N400 in quiet; and second, to create quiet analogs of the intelligibility conditions. The quiet analogs ("pseudoconditions") were examined to ensure that N400 differences between conditions were not due to differences in the efficacy of the sentences used to evoke the N400. Five blocks of stimuli, each consisting of one list of sentences, were presented for each experimental condition. These were selected randomly with replacement from the 14 remaining sentence lists with the constraint that no list was used more than twice. The order of presentation of conditions / blocks was randomized across participants.

EEG data recording and storage

The EEG was continuously recorded using a BioSemi Active-Two biopotential system (BioSemi Instrumentation, 2006). Participants wore an elastic cap with 128 active Ag/AgCl electrodes and ten additional electrodes adhered to the mastoids and face. Additional sites included the mastoids, in front of the tragus, the cheekbones, lateral to the outer canthi, and under the center of the lower eyelids. During EEG acquisition, participants watched a silent movie without subtitles and were instructed to ignore all auditory stimuli. Congruent and incongruent sentences in noise were presented under all five intelligibility conditions, and following that, the quiet condition. Sentences were presented at 70 dB for all experimental conditions, while noise level varied to achieve the required SNR. The noise, rather than the speech signal, was varied in loudness to ensure that any modulation of the ERP response could be attributable only to SNR, and not to the loudness of the speech stimuli. Stimuli were presented using a custom virtual instrument designed in LabVIEW (National Instruments, Austin TX), and played by a National Instruments PXI 4461 Dynamic Signal Acquisition Card (National Instruments, Austin, TX) routed through a GSI 61 audiometer (Grason-Stadler, Eden Prairie, MN). A random pause was delivered between the offset and onset of each sentence with an average of 2 s. This resulted in a total time of approximately 1.5 hours for EEG acquisition. After every 4-5 blocks of stimuli, participants were offered a short break. Data were recorded at a sampling rate of 2048 Hz and stored using Actiview 7.0 Software (Biosemi, Amsterdam, Netherlands).

ERP derivation and measurement

Data were processed offline to produce evoked responses in BrainVision Analyzer 2.0 (Brain Products GmbH, 2014). The EEG was first downsampled to 1024 Hz, then band-pass filtered in the 0.1 - 20 Hz range and re-referenced to an average of the two mastoids. Ocular artifacts were identified using a slope algorithm for blink detection and corrected using independent component analysis (fast ICA). The ICA components were visually inspected to determine which components corresponded to ocular artifact. The continuous EEG was then segmented into discrete epochs time-locked to the onset of the terminal word of the sentence. Epoch duration was 1100 ms including a 100 ms pre-stimulus baseline. Epochs were baseline corrected using the entire 100 ms prestimulus baseline and epochs that contained artifacts (e.g., from movement) were then automatically discarded using a threshold of ±75 μV for maximum deflection size. Epochs were then averaged separately for each recording site, experimental condition, and stimulus type (congruent and incongruent) for each subject. Table 1 shows the mean and standard deviation of the number of trials that were averaged for each stimulus type and intelligibility condition. Difference waves were calculated for each subject by subtracting the average of the congruent trials from the average of the incongruent trials for each intelligibility condition at each electrode site. The peak of the N400 was automatically selected at the most negative point in the 250 – 650 ms poststimulus period at Cz, separately for congruent, incongruent, and difference waveforms. N400 amplitude was then calculated as the average voltage in a ± 5 sample window surrounding the peak (11 ms). Quiet analogues (pseudocondtions) of each condition were created by taking the average of the responses in quiet to the same sentence lists that were used for any given intelligibility condition for each subject. For example, if lists 2, 5, 8, 10, and 14 were used for the +4 condition for subject 8, the average of those lists in quiet would be used to construct the +4 pseudocondition data for subject 8.

Table 1. Mean and standard deviation of number of trials contributing to the average for each stimulus type and experimental condition

	Congruent		Incongruent	
	Mean	SD	Mean	SD
Quiet	136	4.6	136	4.2
+4 dB	48	2.7	48	2.3
+2 dB	48	3.1	49	2.4
+1 dB	49	2.1	49	2.4
Threshold	49	2.4	49	1.7
-2 dB	49	0.7	49	1.6

Statistical analyses

Analyses were carried out separately for difference waves, congruent and incongruent trials. This was done because the N400 effect (i.e., difference wave) was the measure of interest for the study outcome, while the congruent and incongruent data were statistically analyzed to confirm that the experimental effect was attributable to changes in the incongruent, and not the congruent, waveforms. Repeated measures ANOVAs, least-squares comparisons, and a partial least squares analyses were used to determine the effects of intelligibility condition on N400 amplitude and latency. For the purpose of the ANOVAs and waveform figures, data from the 128 scalp electrode locations were pooled into 9 sites: frontal (F), frontal right (FR), frontal left (FL), central (C), central right (CR), central left (CL), parietal (P), parietal right (PR) and parietal left (PL). Two-way repeated measures ANOVAs with the factors intelligibility condition (-2, threshold, +1, +2, +4, quiet) and site (F, FR, FL, C, CR, CL, P, PR, PL) were performed separately for N400 amplitude as measured in the congruent, incongruent, and difference waveforms, and the pseudocondition difference waves. To compare N400 latency across conditions, a one-way repeated measures ANOVA with the factor intelligibility condition was performed on the latency of the N400 peak at the central site. Least-squares comparisons restricted to the central site were used post-hoc to compare specific conditions that appeared to warrant further analysis. A paired t-test was also used to compare the behavioural threshold measurements obtained with the standard HINT and with the stems-only procedure.

A secondary, data-driven analysis of the effect of intelligibility on the N400 was also carried out using partial least squares (PLS; McIntosh, Bookstein, Haxby, & Grady, 1996). Task PLS analyzes the covariance between brain activity and experimental design. It uses singular value decomposition (SVD) of the covariance matrix to identify latent variables (LVs) that represent the differences between experimental conditions (McIntosh et al., 1996). PLS gives electrode saliences, design scores, and scalp scores that are used to describe the contrasts expressed by an LV. Electrode saliences are the numerical weights at each time point and electrode site. The saliences show which time points are most related to differences expressed by the LV. Design scores (DS) indicate the degree to which an experimental condition contributes to the effect described by the LV. Scalp scores indicate how strongly individual subjects express the effect of the LV. Plotting scalp scores by design scores shows which conditions are maximally distinguished (Lobaugh, West, & McIntosh, 2001).

The significance of each LV is determined using a permutation test in which the association of data with corresponding experimental conditions is broken, and the data are reanalyzed. This process is repeated for a large number of permutations, and the statistics obtained from the permuted data are compared to the original results. A probability value is then assigned to the LV based on the frequency that a statistic from the permuted data is larger than that of the original data (McIntosh et al., 1996). This is a non-parametric method for significance testing, and provides a test against randomness. A second test is needed to establish the stability of the results. This is done using bootstrapping. The data are resampled with replacement, and the standard error of a large number of bootstrap samples allows estimation of the standard error of the saliences (Lobaugh et al., 2001; Krishnan, Williams, McIntosh, & Abdi, 2011).

For the current study, mean-centered task PLS was performed separately for difference waves, congruent, and incongruent data using plsgui in MATLAB (Baycrest, 2011). The analysis used data from 132 electrode sites and the 6 intelligibility conditions. The permutation test used 500 permutations, and the bootstrapping test used 500 bootstrapping samples with a 99% confidence interval. Resulting salience maps, visualized on the scalp, were plotted in VisionAnalyzer2.0. These salience maps were thresholded by bootstrap ratio such that only saliences that met the 99% CI threshold for bootstrap ratio were visualized.

Results

HINT and stems-only SNR-50 tests

SNR-50 on the stems-only test was significantly lower than the HINT [t(13) = 5.5, $p < 0.001$]. This is likely partly due to practice effects, since all subjects were tested with the HINT first. It may also be partially due to the increased sentence length in the stems-only test, increasing the listener's ability to use context to predict or interpret misheard content, despite the lack of a terminal word.

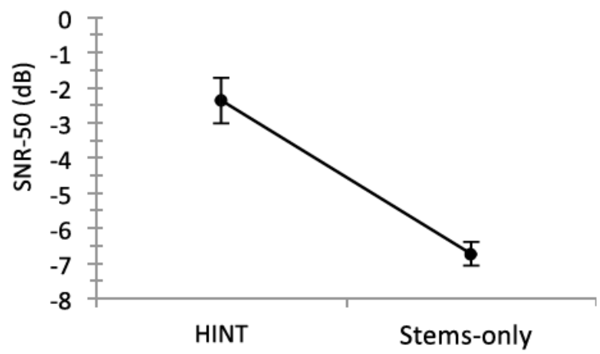

Figure 2. Mean SNR-50 for the HINT and stems-only test. Error bars represent standard error.

N400 effects

N400 difference waves are displayed on the scalp in Figure 3, as the average in 100 ms intervals from stimulus onset to 1000 ms post-stimulus. In quiet, the N400 effect peaked in the 500-600 ms period and was largest at central and parietal sites. Compared to the quiet condition, the N400 effect was reduced in the +4 dB condition, and continued to decrease as SNR decreased, reaching a minimum in the +1 dB condition. At threshold, the N400 effect appeared to increase again in amplitude, but peaked later in the epoch (around 700-800 ms), with a scalp distribution that appeared more frontal and right lateralized. In the -2 dB condition, the N400 effect was again attenuated. This trend can be seen in the grand average difference wave scalp maps (Figure 3) and waveform figures (Figure 6). The trend was similar, but less apparent, for incongruent stimuli as shown in

Figures 5 and 8. The trend was not observed for congruent waves (Figure 4), which showed a reduced N400 compared to incongruent waveforms, and an effect of intelligibility in which a late positivity appeared to emerge at parietal regions, particularly for the +4 dB condition.

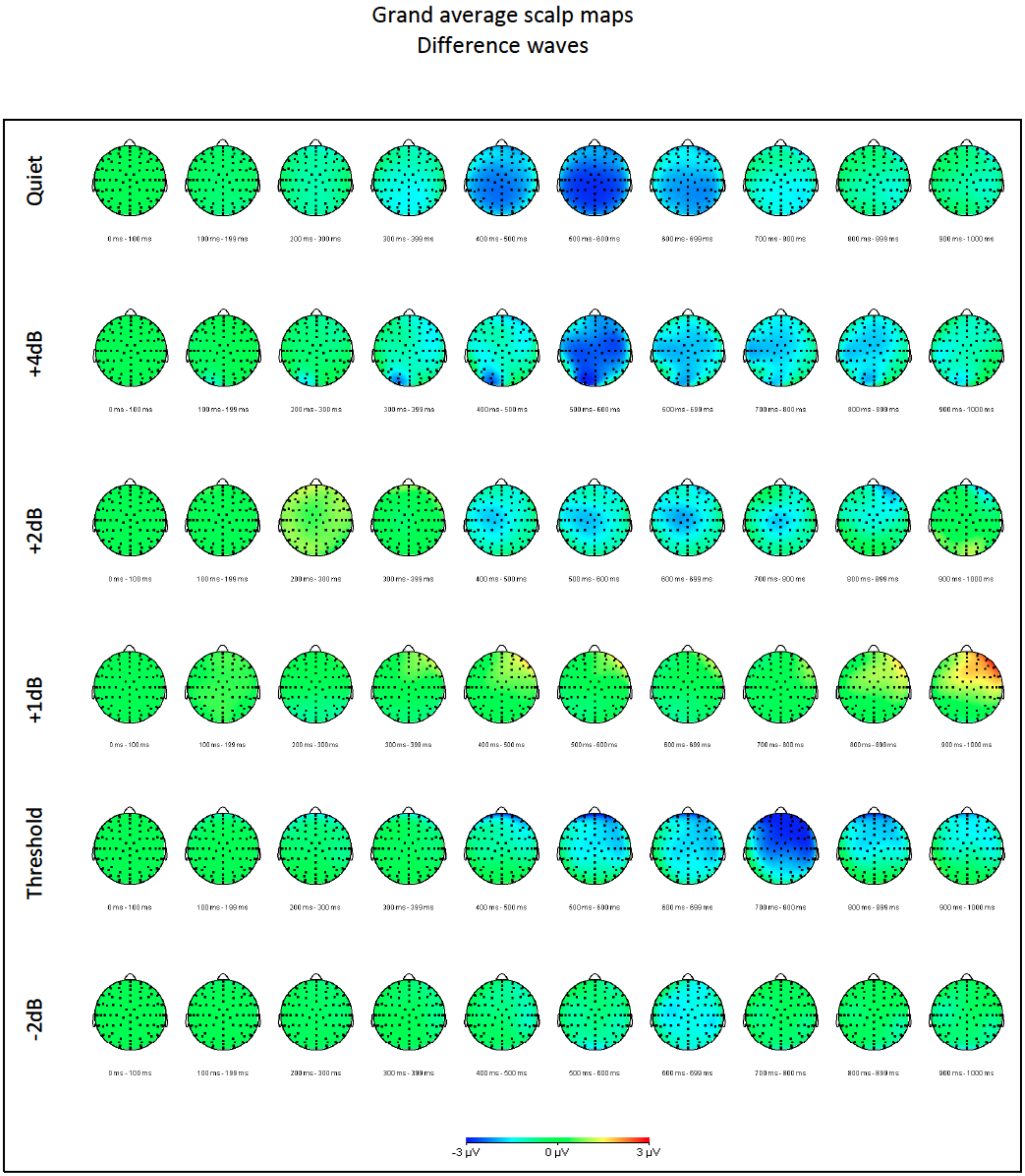

Figure 3. Scalp maps showing the grand average of the difference wave averaged over 100 ms periods for all intelligibility conditions.

Grand average scalp maps
Congruent

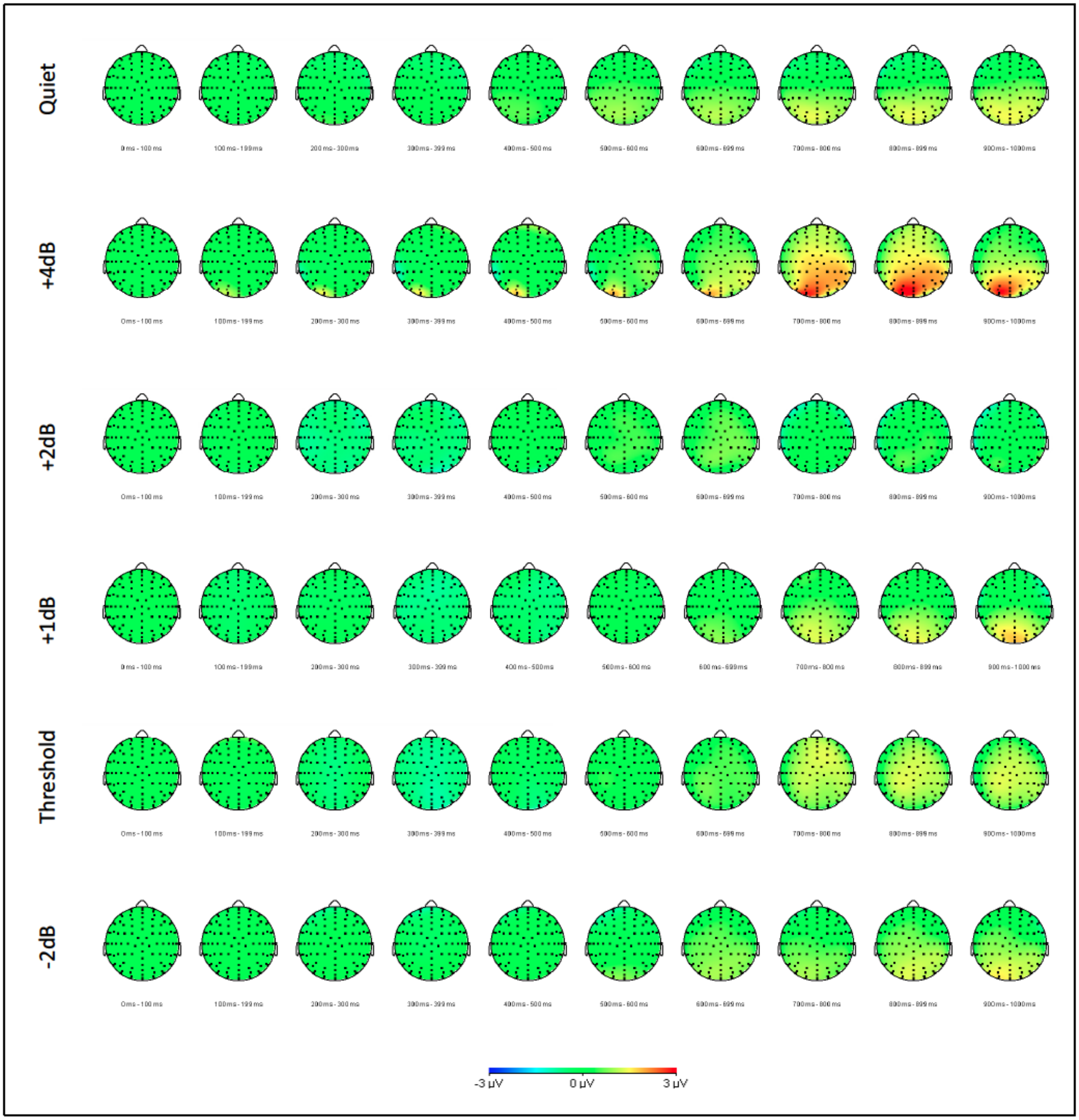

Figure 4. Scalp maps showing the grand average of the congruent trials averaged over 100 ms periods for all intelligibility conditions.

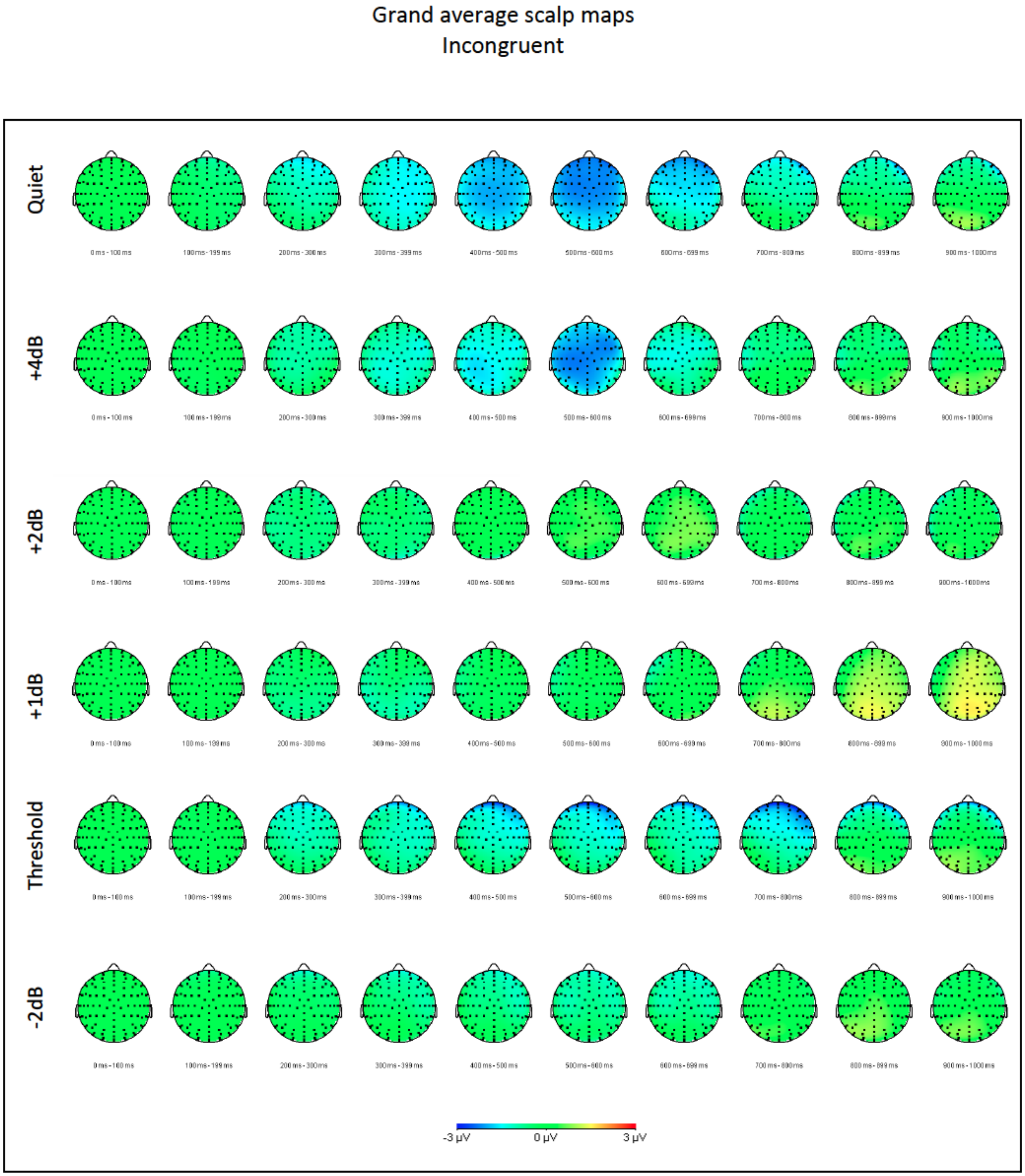

Figure 5. Scalp maps showing the grand average of the incongruent trials averaged over 100 ms periods for all intelligibility conditions.

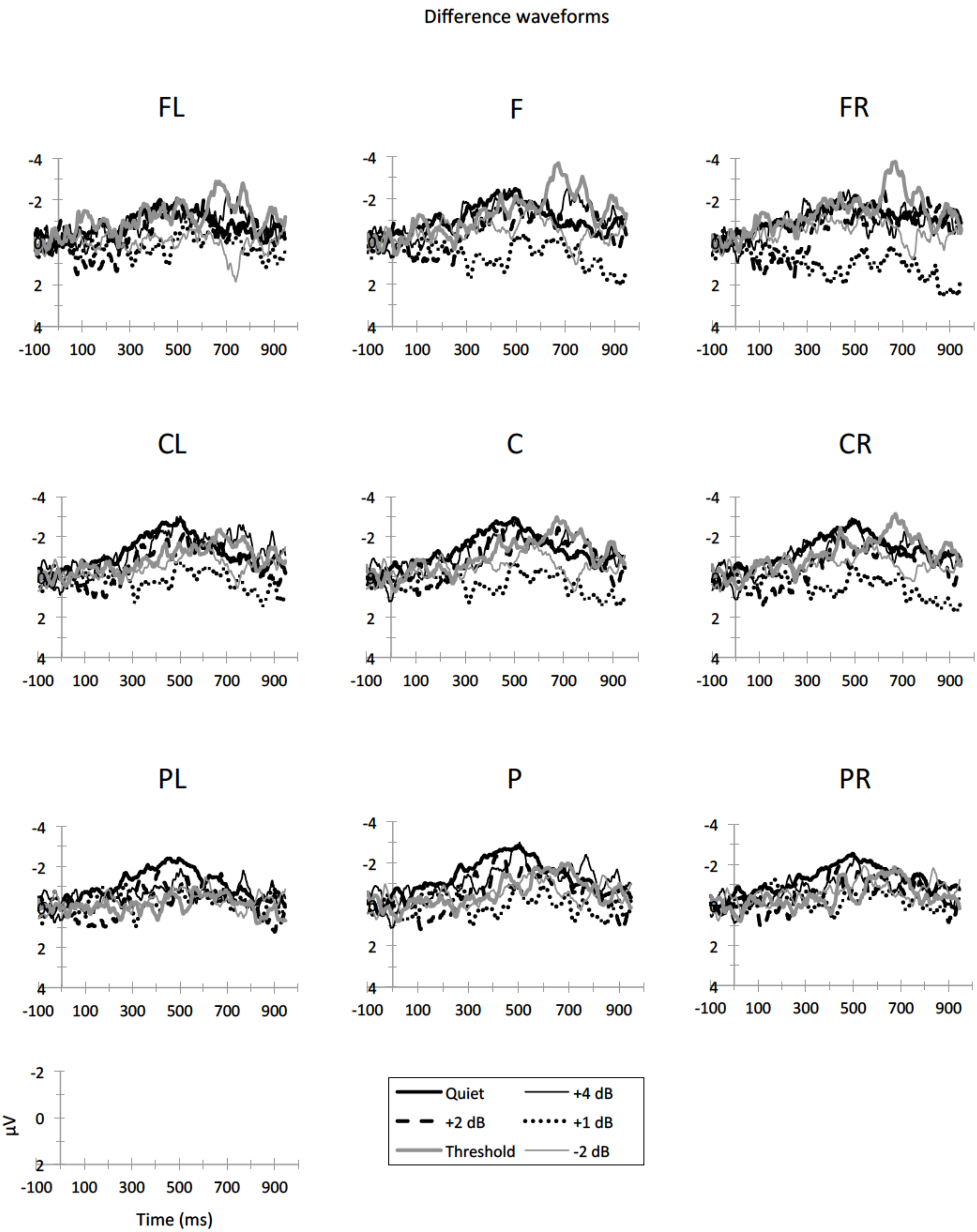

Figure 6. Grand average waveforms for the incongruent minus congruent difference waves at 9 pooled electrode sites for all intelligibility conditions.

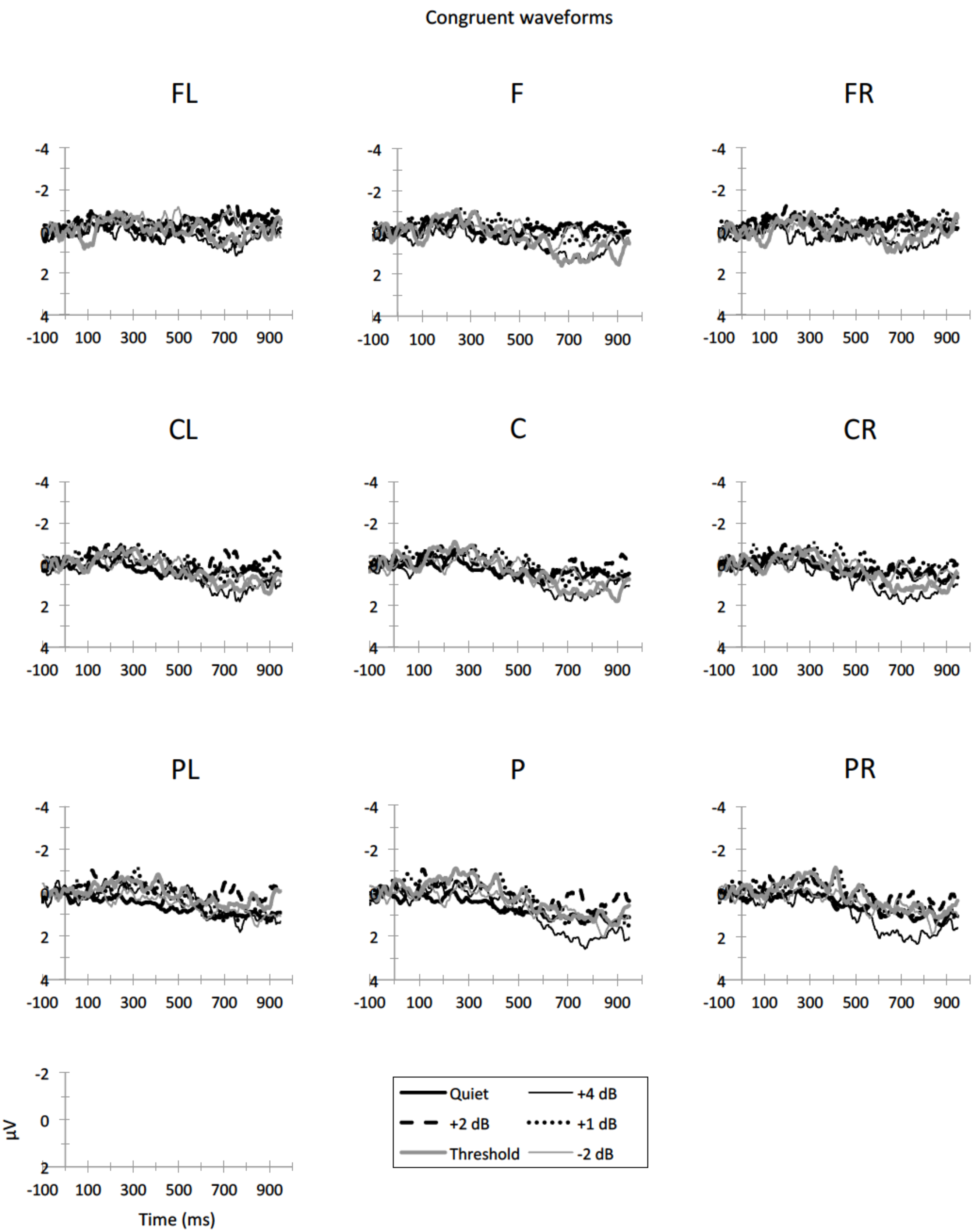

Figure 7. Grand average waveforms for congruent stimuli at 9 pooled electrode sites for all intelligibility conditions.

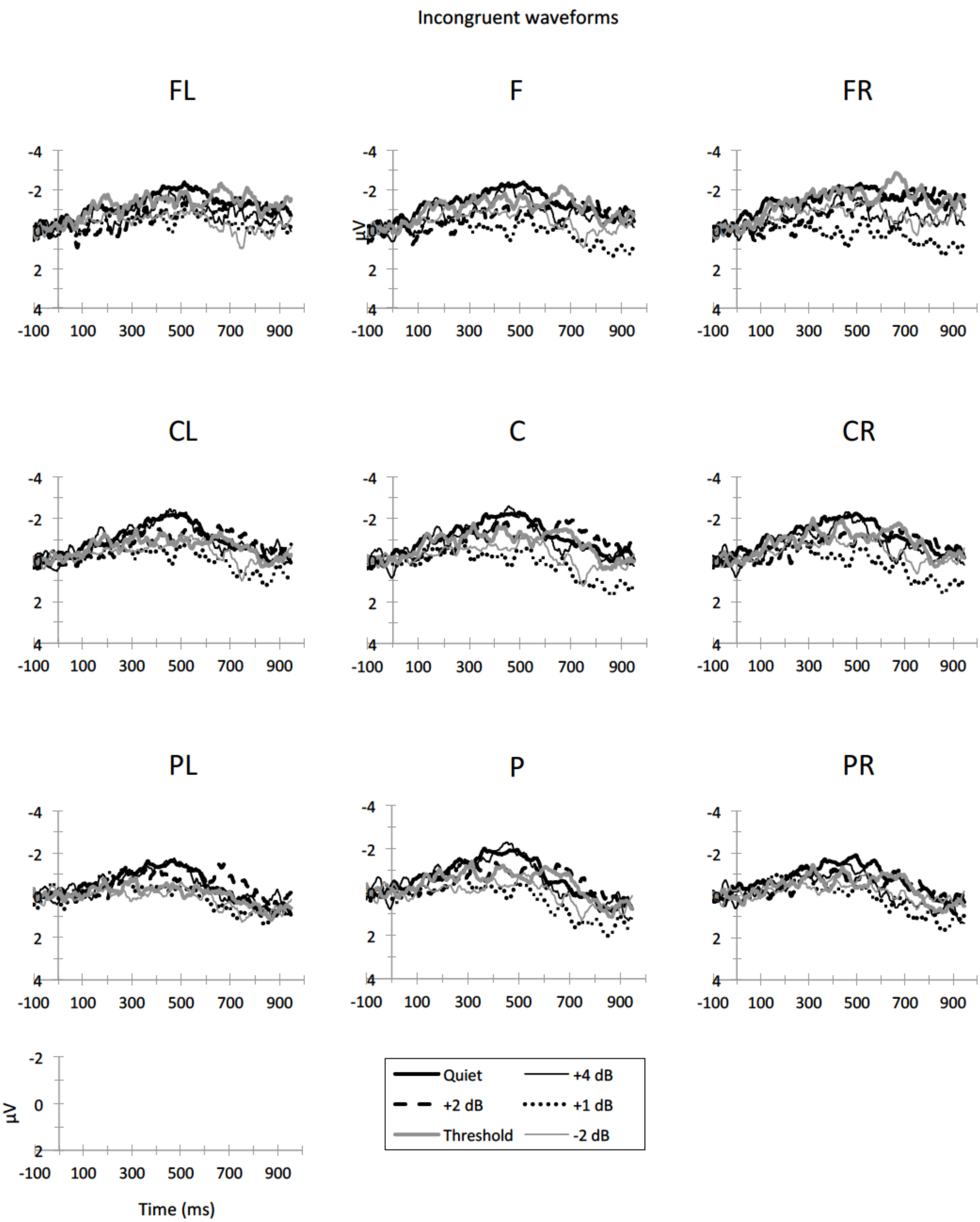

Figure 8. Grand average waveforms for incongruent stimuli at 9 pooled electrode sites for all intelligibility conditions.

Statistical results: ANOVAs and post-hoc comparisons

The two-way ANOVAs for intelligibility and site demonstrated no significant main effect of intelligiblity condition, and no significant interaction of intelligibility and site for difference waves ($p > 0.1$), congruent data ($p > 0.1$), incongruent data ($p > 0.1$), or pseudocondition difference waves ($p > 0.1$). In all cases there was a significant main effect of site (difference waves: [$F(8,80) = 3.6$, $p = 0.001$]; congruent: [$F(8,80) = 6.4$, $p < 0.001$]; incongruent: [$F(8,80) = 5.4$, $p < 0.001$]). Tukey post-hoc tests demonstrated that, for the difference waves, the N400 effect measured at the central site was significantly larger than at the frontal right and parietal left and right sites. For congruent data, the N400 measured at parietal left and right sites was significantly smaller than at frontal, frontal left, central, and central left sites. Analysis for incongruent data showed that the N400 was significantly smaller at the parietal right site than at the frontal, central, and central left sites; and significantly smaller at the parietal left site than at the frontal, frontal right, central, central right, and central left sites.

Despite the non-significance of the omnibus ANOVAs, the clear effects of intelligibility condition observed in the scalp maps and ERP waveforms suggested that post-hoc statistical testing was warranted. These effects are illustrated in Figure 9, which shows modulation of mean N400 amplitude by intelligibility condition for each pooled site, and each wave type (congruent, incongruent, difference wave). Clear effects of intelligibility, in which N400 amplitude is reduced in the +1 dB and -2 dB conditions, can be observed for the difference and incongruent waveforms, while for congruent waveforms there appears to be a slight trend of increasing N400 amplitude between quiet and +2 dB. Therefore, least-squares comparisons were performed using data from the central site only. For the N400 effect measured in the incongruent minus congruent difference wave, a least-squares comparison between the +4 dB condition and the +1 dB condition yielded a significant difference ($p < 0.05$). For the congruent waveforms, a least-squares comparison showed a significant difference between the quiet condition and all other conditions, except +4 dB ($p < 0.05$).

ANOVAs looking at latency of the peak N400 amplitude at the central site did not detect significant differences for any of the conditions. However, the clear effect of intelligibility condition on the latency of the N400 in the scalp maps suggested that post-hoc testing was again warranted. A least-squares comparison showed that the latency of the difference wave in quiet was significantly shorter than for the other conditions ($p < 0.05$). Mean peak latency was 514 ms in quiet, and ranged from 597 ms – 644 ms for the other conditions.

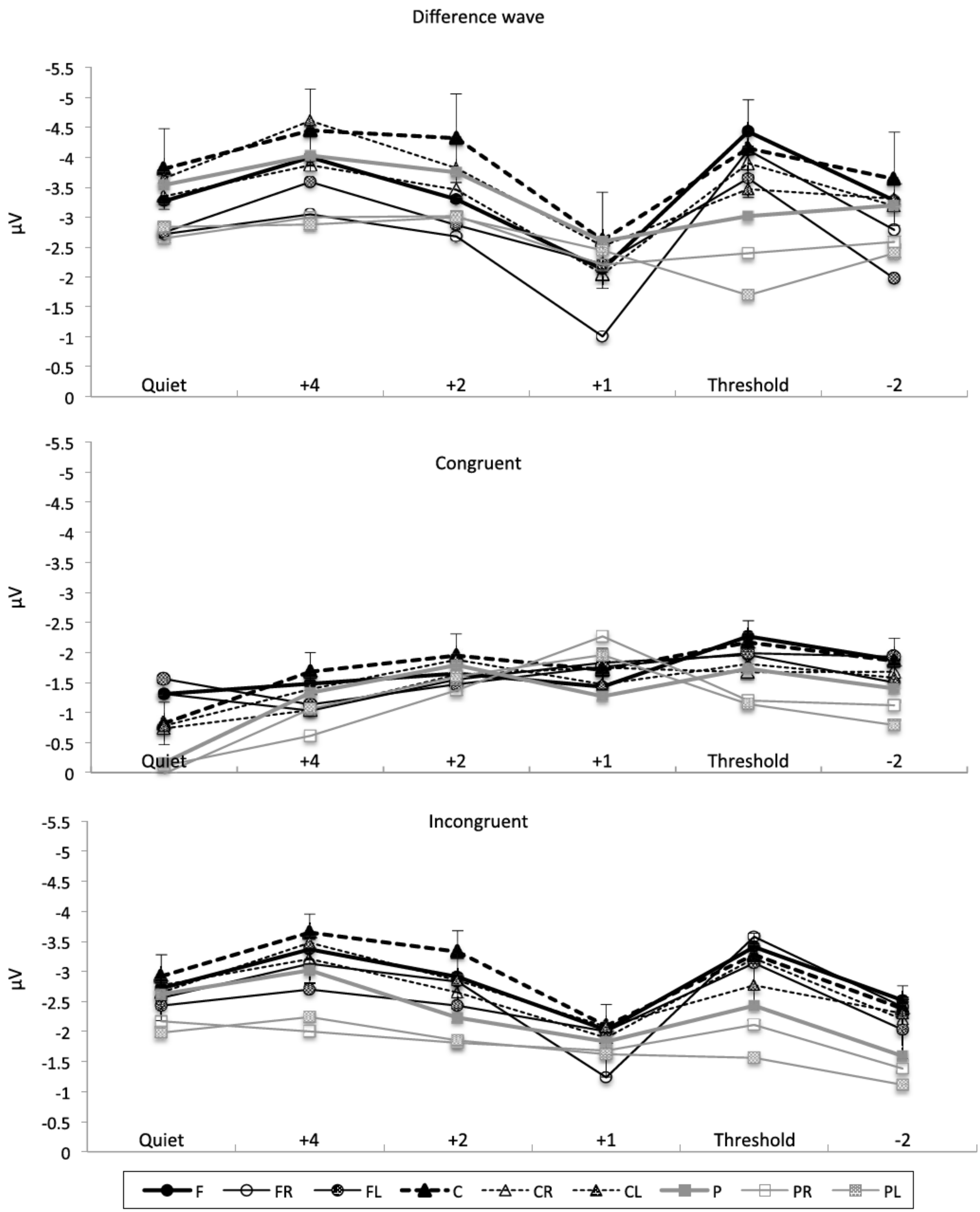

Figure 9. Mean N400 amplitude by SNR and pooled electrode site. Error bars represent standard error at the central pooled site.

Statistical results: Data-driven PLS

For the PLS analysis of the difference waves, one significant latent variable (LV) was identified ($p < 0.01$). Figure 11 shows the design scores, scalp scores, and saliences for this LV. The design scores demonstrate that the difference waveforms for the quiet, +4 dB, +2 dB, and threshold conditions were significantly different as a group from the +1 and -2 dB conditions. However, the relatively small design score for -2 dB versus the +1 dB condition indicates that the difference between these groups of conditions was primarily driven by the +1 dB condition. Part C of Figure 11 shows brain saliences plotted across the scalp, thresholded by a bootstrap ratio of 3.09 (roughly equivalent to a 99% CI threshold). The salience plots indicate that the quiet, +4 dB, +2 dB, and threshold conditions tended to show more negative voltages than the +1 and -2 dB conditions, particularly in frontal right regions in the 475-550 ms latency range, which is consistent with the observed latency of the N400 effect.

For the PLS analysis of the congruent waves, there was one significant LV ($p <$ 0.005. Design scores, scalp scores, and saliences are shown in Figure 12. The design scores indicate that the +2 dB and threshold conditions, as a group, are significantly different from the quiet, +4, +1, and -2 dB conditions. The relative design scores for these conditions indicate that these differences were primarily driven by the +4 dB condition, while the -2 dB condition contributed little to the LV. The salience plot indicates a stable negative difference in voltage at posterior sites in the 750-850 ms range, and in the 950 – 1000 ms range. This late negative difference could indicate that the N400 in the congruent waveform was longer-lasting at posterior sites for the threshold and +2 dB conditions.

Like the PLS analyses for difference and congruent waveforms, the PLS analysis for incongruent waveforms yielded one significant LV ($p < 0.01$). Figure 13 shows the design scores, brain scores, and saliences for the significant LV. The design scores indicate that the quiet, +2 dB and threshold conditions were significantly different as a group from the -2, +1, and +4 dB conditions. However, the very small design score for the +4 dB condition indicates that it contributed very little to this effect. The salience maps in Part C of Figure 13 demonstrate that the most robust effect is a more negative voltage for the quiet, +2 dB, and threshold conditions in frontal right regions between 500 and 550 ms, consistent with a brief portion of the latency range for the N400.

Responses at the single subject level

To examine the stability of the effect of intelligibility condition on the amplitude of the N400 effect for potential clinical applications, Figure 10 shows N400 effect amplitude at the central site, by intelligibility condition, individually for each subject. Values along the Y-axis are incremented for each subject to separate the curves vertically. Overall, the tendency for the N400 effect to reach a minimum at the +1 dB (and sometimes additionally the -2 dB) condition was observable in six subjects. Another group of three subjects displayed a slightly different pattern of responses, in which the minimum N400 effect amplitude was observed at +2 dB instead of +1 dB.

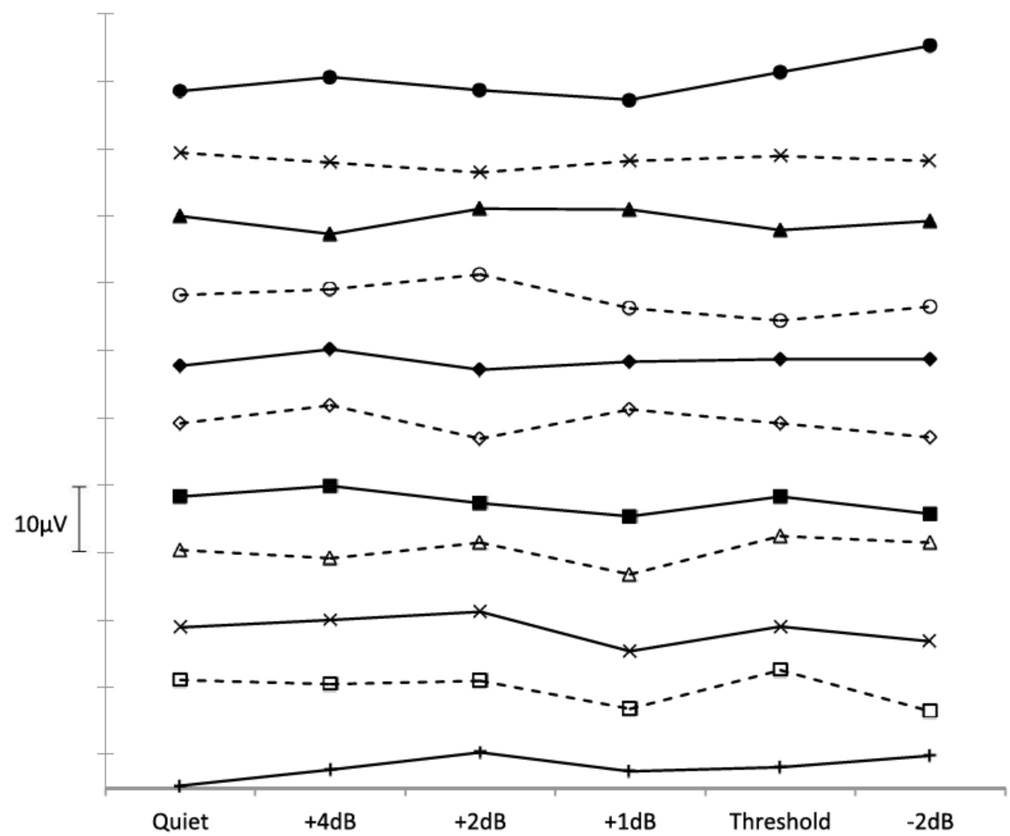

Figure 10. Individual subjects' peak N400 effect (difference wave) amplitude at the central pooled electrode site for all intelligibility conditions. Values along the Y-axis are incremented for each subject to separate the curves vertically.

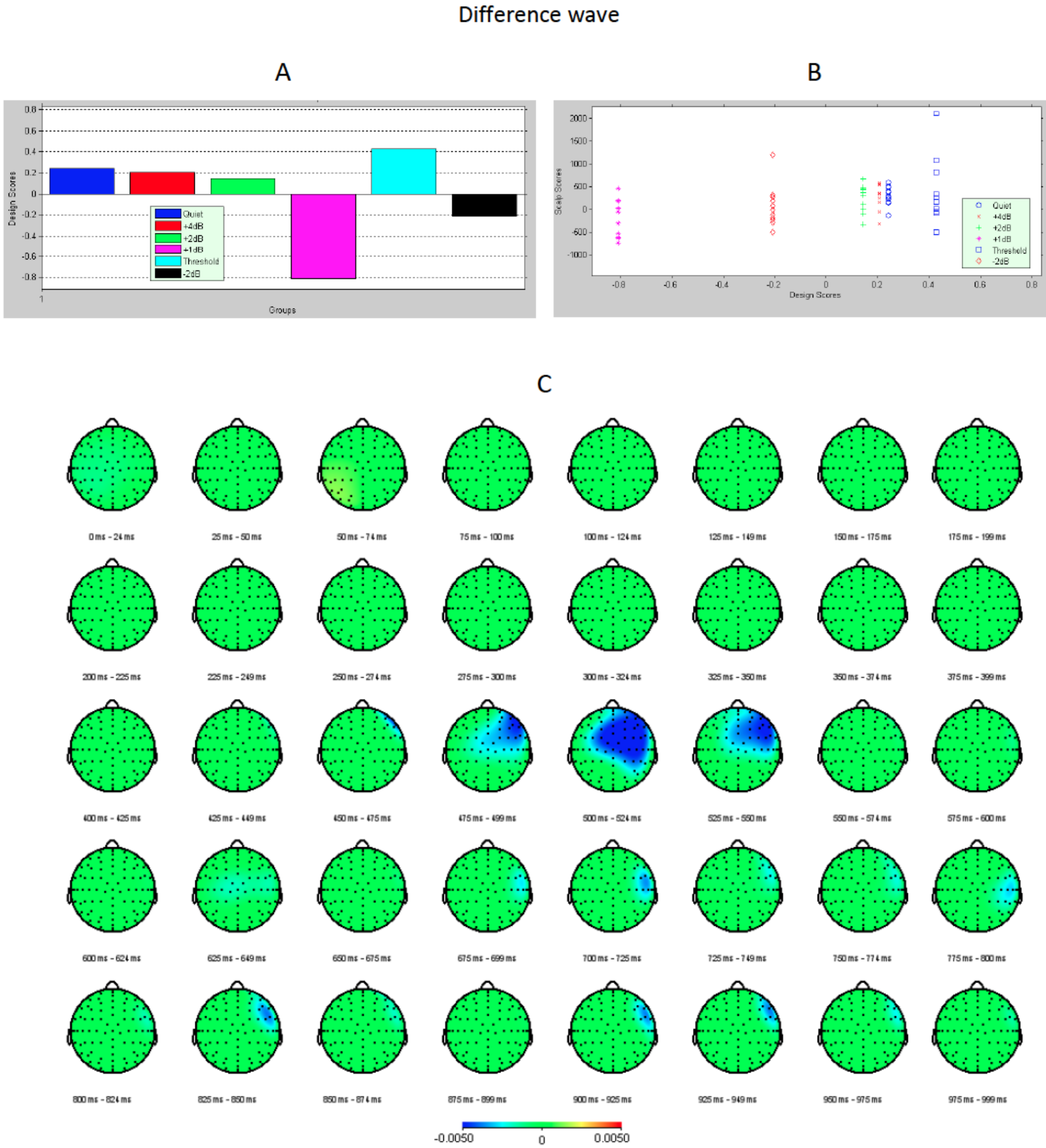

Figure 11. PLS results for difference waves. One significant LV, $p < 0.01$. Part A shows design scores by experimental condition. Part B shows scalp scores by design scores. Part C shows electrode saliences averaged over 25 ms periods, thresholded by a bootstrap ratio of 3.09.

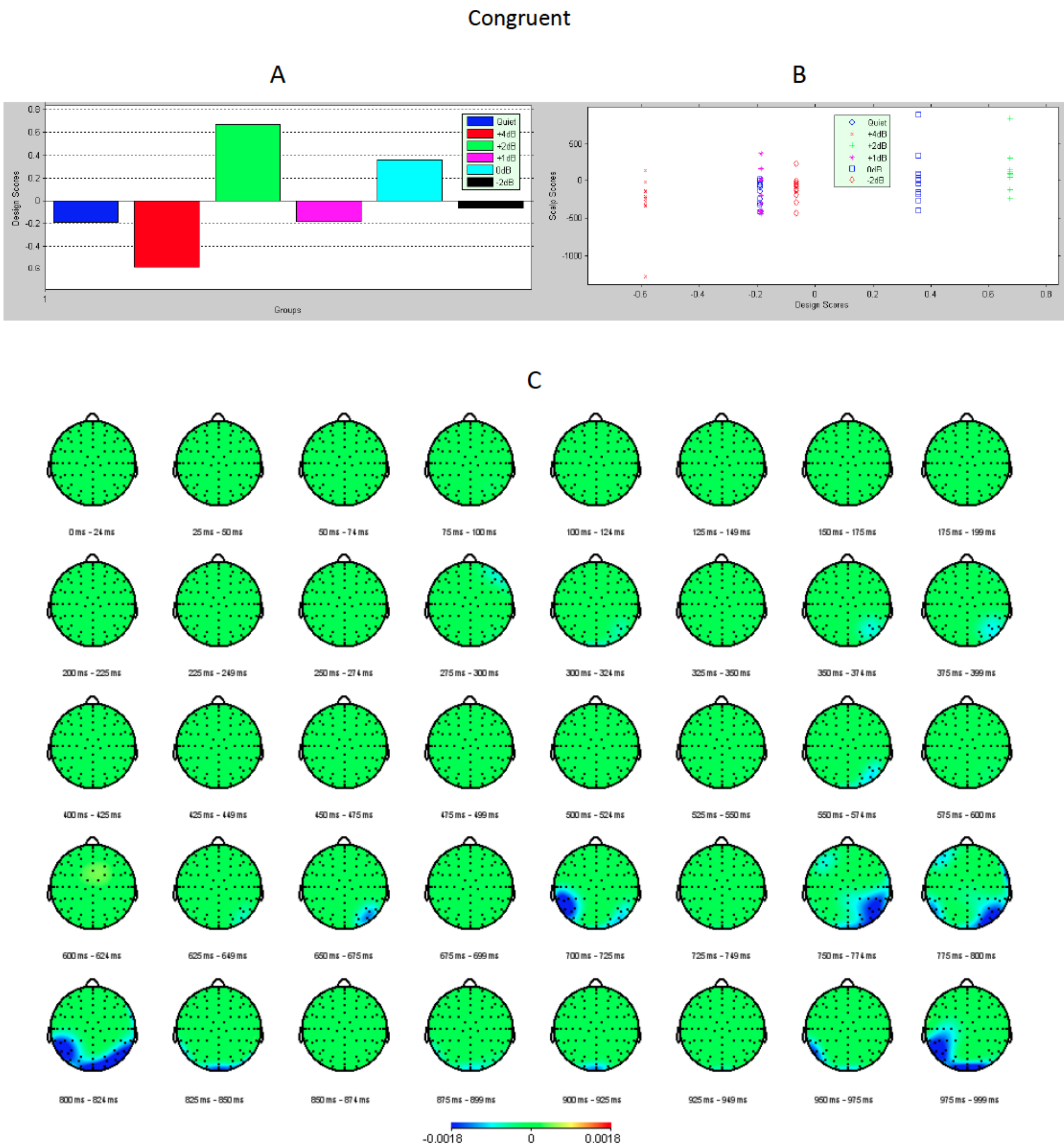

Figure 12. PLS results for congruent stimlui. One significant LV, $p < 0.005$. Part A shows design scores by experimental condition. Part B shows scalp scores by design scores. Part C shows electrode saliences averaged over 25 ms periods, thresholded by a bootstrap ratio of 3.09.

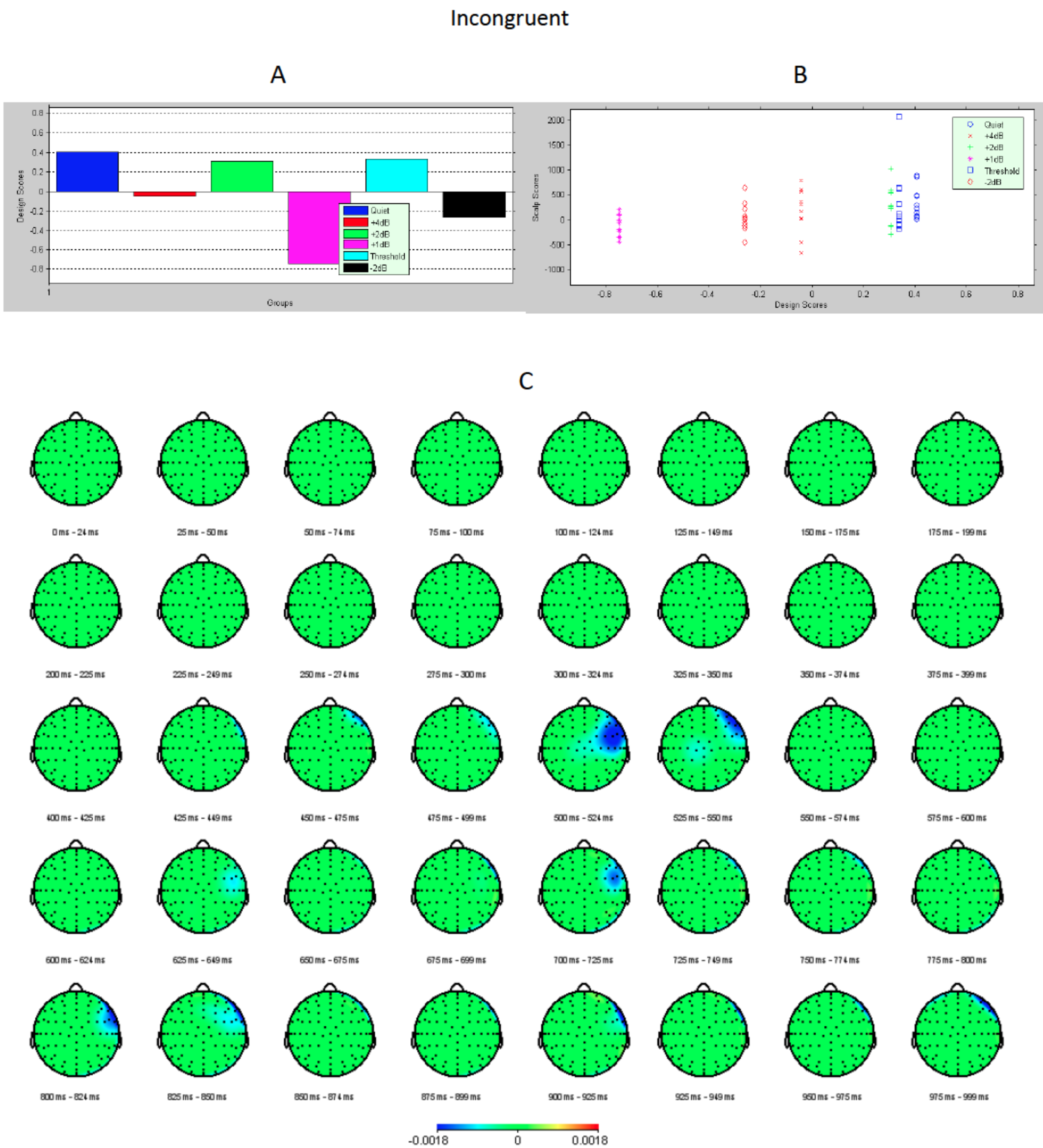

Figure 13. PLS results for incongruent stimlui. One significant LV, $p < 0.01$. Part A shows design scores by experimental condition. Part B shows scalp scores by design scores. Part C shows electrode saliences averaged over 25 ms periods, thresholded by a bootstrap ratio of 3.09.

CHAPTER IV: DISCUSSION

Pilot study

The purpose of the pilot study was to determine whether the lists of sentences used for N400 elicitation were equivalent for intelligibility. The results showed that there was variability in the mean intelligibility of the lists that approached statistical significance. In particular, lists 6, 3, and 11 yielded lower SNR-50s than other lists, meaning they were more intelligible. Using a larger sample size may have yielded a significant difference in intelligibility between lists.

When determining the SNR-50 with the HINT or stems-only test, 2 dB steps were used, resulting in a minimum error of ±1dB on the behavioural measurement. The magnitude of any real difference in intelligibility between lists is likely to be about 1-2 dB, based on the region of non-overlap of the error bars on Figure 1. In this study the standard deviation of SNR-50 for most lists was between 1 and 2 dB. In comparison, the standard deviations of the mean intelligibility of the HINT sentence lists are approximately 1 dB (Nilsson et al., 1994).

In this study, it is likely that the error introduced by individual factors and the inherent error of threshold measurement made it difficult to statistically detect differences in intelligibility between lists. However, it is always advantageous to limit extraneous sources of error, especially if the experiment depends on fine contrasts such as that between threshold and a +1 dB SNR. For future research using these lists, the pilot could be repeated with a larger sample size to identify problem lists, and the sentences could be redistributed or replaced and re-tested as necessary to attempt to equalize the lists for intelligibility.

Electrophysiological study

The aim of the project was to determine whether a passive speech in noise test could be designed using the N400. The hypotheses were that the N400 would be elicited by sentences presented with speech-frequency noise, that its amplitude would vary directly with intelligibility, and that the N400 effect (difference wave) would vary systematically with respect to signal to noise ratio relative to the behavioural threshold.

This study found that a broad, negative ERP consistent with the N400 was elicited in the presence of varying levels of background speech-frequency noise, and its mean amplitude was variable. However, the effect of intelligibility condition on N400 effect amplitude was not linear. The N400 effect appeared to decrease as noise level increased (as seen on the grand average scalp maps), but this difference was not statistically significant until SNR reached +1 dB relative to threshold. Unexpectedly, the amplitude of the N400 effect then increased again at threshold, such that its amplitude at threshold was not significantly different from the +2 dB, +4 dB and quiet conditions. Finally as SNR reached -2 dB relative to threshold, the N400 effect amplitude decreased again to a level not statistically different from the +1 dB condition. PLS yielded a significant difference between the +1 dB and -2 dB conditions and the other conditions in the 475 – 550 ms latency range at frontal right scalp sites. Additionally, a least-squares comparison confirmed that peak latency at the central site was significantly increased in conditions with noise compared with the quiet condition.

This trend appeared to be driven primarily by the change in the incongruent responses, rather than the congruent ones, since the incongruent results followed a similar pattern. The PLS analysis demonstrated that the -2 dB, +1 dB, and +4 dB conditions were, as a group, significantly different from the quiet, +2 dB, and threshold conditions. Given that the +4 dB condition contributed little to this latent variable, and that the contrast was most robust in frontal right regions in the 500-550 ms latency range, the incongruent results seem to roughly correspond to the results for the N400 effect in the difference wave. The congruent results, in contrast, did not mirror the N400 effect results because although the threshold and +2 dB conditions had significantly more negative response than the other conditions, this occurred in a later latency range (after 750 ms) and at posterior sites.

Examining the congruent results gives information on how noise affected normal sentence processing in this study, albeit using highly constraining sentences with predictable terminal words. The least-squares comparison results support the possibility that the noise did affect the congruent N400 since mean N400 amplitude for all congruent conditions with noise except +4, the least noisy intelligibility condition, were significantly different from the quiet condition at the central electrode site. This finding was not supported by the PLS results, however, which instead showed that the threshold and +2 dB conditions demonstrated a longer-lasting negativity at posterior sites compared to all other conditions, which may have reflected a prolonged N400.

The relationship between intelligibility condition and N400 amplitude and latency is apparently non-linear and not straightforward, with differing effects on congruent and incongruent terminal words. One difficulty in interpreting the results arises from the small number of studies looking at the effect of noise or other forms of signal degradation on the N400. These studies have used a variety of methods in terms of type of signal degradation or masking, experimental design, and task. Not surprisingly, results have also varied widely, probably in large part due to these methodological differences.

Some studies have found that as the signal was degraded, the amplitude of the N400 or N400 effect decreased. This was the case in Aydelott et al. (2006), Strauß et al. (2013), and Obleser and Kotz (2011). These studies did not use noise maskers; rather, they used acoustic distortion of the sentential context, target words, or both. This was accomplished by low-pass filtering the recorded speech as in Aydelott et al., or by using a vocoding algorithm to reduce the spectral information in the speech, as in the other studies. In Aydelott et al., the reduction of the N400 effect was due mostly to a decrease in N400 amplitude for incongruent words, as opposed to any change in the congruent N400. They attributed this amplitude decrease to the decreased availability of semantic information due to acoustic degradation. This is consistent with the current results, with the effect on the N400 effect driven primarily by the responses to the incongruent stimuli. They also found that distortion increased the peak latency for both congruent and incongruent trials, which is somewhat consistent with the results reported here, since latency was significantly longer in noise for the difference wave but not for congruent or incongruent responses separately. However, this study used degraded speech for only the context, not the target word (which was in medial sentence position), which likely caused overlap with the P3 component, and calls into question how comparable their results are to the current study, as the target words were all fully intelligible.

Obleser and Kotz (2011) and Strauß and colleagues (2013) presented sentence stimuli that were degraded using a vocoding algorithm. In Obleser and Kotz, simple sentences with more expected or less expected (yet congruent) terminal words were presented with 16 (fully intelligible), 4 (intermediate intelligibility) or 1 spectral band (essentially no spectral detail). In the 1-band condition, there was no significant N400 effect. In the 4-band condition the effect was reduced and peak latency was prolonged compared with the 16-band condition. They attributed this reduction in N400 effect to increased effort needed to integrate the congruent words. In Strauß et al., sentences were manipulated for Cloze probability as well as "context-based typicality" of terminal words, a measure of how common a verb-object pair is in a language. They played these sentences in unaltered form, in an 8-band condition, and in a 4-band condition. There was no robust N400 effect for 4-band speech regardless of constraint or typicality. For 8-band speech, latency was delayed compared to natural speech, and the effect of typicality was only significant when the signal was degraded. For natural speech, highly constrained low-typicality terminal words resulted in low amplitude N400s while in 8-band speech the same stimulus type resulted in N400s indistinguishable from the responses to low constraint, low typicality words. This suggests that in a degraded listening condition, factors that are less important in quiet (such as typicality or frequency of a word in the language) become more important. Both studies employed the same behavioural task: participants rated the intelligibility of each sentence on a four-point scale. The increase in latency reported in both studies is consistent with the current results. The decrease in amplitude may be consistent with that observed at +1 dB and -2 dB in the current results; however, there is no evidence that mirrors the present result in which amplitude increases at threshold. Interestingly, Obleser and Kotz and Strauß et al. used very similar methods but one study found a significant N400 effect in 4-band speech and the other did not. One explanation for this is the differing sentence stimuli used, since both studies had approximately the same number of trials per condition and used the same behavioural task. This highlights the variability of the N400 and its sensitivity to the factors that mediate its amplitude.

Amplitude and latency effects in studies using maskers rather than distortion had differing results. Daltrozzo et al. (2005) was the only study to use noise as a masker, while Romei et al. (2011) used 20-talker babble and Connolly et al. (1992) used 12-talker babble. Daltrozzo et al. found that a “high” level of masking with pink noise obliterated the N400 effect (as measured using sentences with congruent or incongruent terminal words). A “medium” level of masking using pink noise low pass filtered at 4000 Hz attenuated the N400 effect and resulted in a more frontocentral distribution. Finally, a “low” level of masking with pink noise low pass filtered at 2000 Hz slightly enhanced the N400 amplitude and resulted in a more parietocentral distribution. Much information regarding the methodology for this study has not been disseminated, including the exact SNR used, if and how sentences were controlled for Cloze probability, and whether any of the contrasts in amplitude between conditions were statistically significant.

Romei et al. (2011) used a priming paradigm with babble presented at an SNR of +9 dB, which corresponded to a mean word recognition score of 95% for their participants. N400 amplitude to both related and unrelated words was increased in babble, a finding the authors attributed to increased use of working memory and attention. For related words, this was significant at anterior, right and left central sites. For unrelated words, it was a significant increase at anterior and right central sites only. No analysis of the N400 effect was done in this study; however, these results imply small or no changes in the N400 effect at anterior and right central sites, with a possible decrease in the N400 effect at central left sites. Visual inspection of the ERP grand averaged waveforms more or less confirms this; the N400 effect appears slightly reduced at central sites and unchanged at anterior sites in babble.

Connolly et al. (1992) was the only study to include experimental conditions with no behavioural task (though subjects were instructed to pay attention to the speech stimuli). High and low constraint sentences (all congruent) from the SPIN were presented with babble at an SNR that resulted in an 11-19% error rate on a semantic judgment task (exact SNR is unknown). N400 amplitude was larger for low constraint sentences than high constraint sentences at all electrode locations except left temporal. They found no significant effect of babble on N400 amplitude, although visual inspection of the data shows a trend of smaller amplitudes for the masked conditions relative to the unmasked conditions and for the conditions without an active task compared to conditions with a behavioural task. Latency was significantly delayed in babble compared with quiet for both the semantic task condition and the no task condition. This increase in latency was attributed to the effect of the babble masker on attention and memory.

In summary, research on the effect of signal degradation and masking on the N400 has yielded varying results. The most consistent finding from these studies is an increase in peak latency of the N400 effect in the degraded or masked conditions, which is also reported here. There is also a commonality of differing effects of noise or distortion on congruent and incongruent words, as in Aydelott et al. and Romei et al., which was also the case in this study. Decreases in N400 effect amplitude with noise or distortion are also commonly reported, especially with high levels of distortion/masking, as in Daltrozzo et al., Obleser and Kotz, and Strauß et al. In the present study, reductions in N400 effect amplitude were seen in the +1 dB and -2 dB conditions. Because no study has compared the effects of differing maskers (e.g. pink noise vs. speech babble) on the N400, it is unknown how noise characteristics affect the relationship between intelligibility and N400 amplitude. Daltrozzo et al. and Romei et al. reported increases in N400 amplitude with low levels of babble or noise, when speech was nearly completely intelligible. These amplitude increases may have resulted from increased attention and effort on the part of subjects to complete the behavioural tasks in these studies. In fact, the purposes of the Romei et al. and the Connolly et al. studies were to mask speech at the cognitive level using babble, thereby forcing subjects to engage attention and memory, even when there was no associated behavioural task. This is in direct contrast to the purpose of the current study, in which the goal was to introduce perceptual masking during passive stimulation and not to engage additional cognitive resources.

Attention to visual or auditory stimuli is known to increase the amplitude of the N400 (Kutas & Federmeier, 2011). Relander et al. (2009) illustrated this by using a priming paradigm and comparing N400 amplitudes and latencies when subjects did a semantic task, a phonological task, or watched a movie with subtitles and ignored auditory stimuli. They found that in the passive condition the N400 was present but its amplitude was reduced compared to the conditions with active responding, which did not differ from each other. In the current study, increased covert attention to terminal words presented with noise could theoretically result in an N400 amplitude increase in both congruent and incongruent conditions. It is unlikely that the N400 amplitude changes in the incongruent conditions were due to differences in attention, because it is unclear why subjects would have increased attention in the quiet, +2 dB, and threshold conditions compared to the -2 dB, +1 dB, and +4 dB conditions. To isolate the effect of attention from the effect of word predictability on N400 amplitude, the congruent waveforms were analyzed alone. A planned comparison at the central site found increased N400 amplitude for all intelligibility conditions (excluding +4 dB) compared to quiet. This suggests a possible effect of attention; however, the PLS analysis did not support this finding. The participants were instructed to ignore auditory stimuli, and reported they had no trouble doing so, but it is still possible that attention affected the results. In future research, the use of an active distraction task could be used to compare performance metrics between intelligibility conditions and verify whether participants were dedicating more attention towards the speech stimuli under any conditions.

Perhaps the most difficult finding to explain in this experiment was the increased amplitude of the N400 effect in the threshold condition. The difference in SNR among these conditions was very small. Indeed, the 1 dB difference in noise level between the +1 dB and threshold conditions is the smallest difference that could theoretically be perceived; yet there was a significant difference in N400 effect amplitude between these conditions. No other study had experimental conditions that were so close together in intelligibility, and it is also likely that no other studies have examined the N400 close to the behavioural hearing threshold. Therefore there is little in the literature that can shed light on the obtained result.

Given the differences in scalp distribution and latency between the N400 effect observed at quiet and in the +4 dB condition, versus the threshold condition, it may be reasonable to assume that the processes underlying generation of the the N400 effect were different under these two conditions. One possible alternative mechanism by which the N400 effect could have been generated at threshold is a word-pseudoword paradigm. It is possible that at threshold the incongruent terminal words could not be reliably identified, and were instead processed more similarly to pseudowords (phonologically legal nonwords). In the context of lexical decision tasks, studies have shown that there is an enhanced N400 to single spoken pseudowords compared to single spoken words (e.g. Supp et al., 2004). It has been proposed that this difference arises from the additional processing that pseudowords undergo in an attempt to identify a match for them in the mental lexicon (Friedrich, Eulitz, & Lahiri, 2006; O'Rourke & Holcomb, 2002). In Friedrich et al. (2006) the difference in N400 amplitude between words and pseudowords was greatest at frontal electrode sites in the 500-1000 ms latency range. This finding corresponds roughly to the location and timeframe identified by the PLS analysis in this study. In an experiment without a behavioural task, subjects in O'Rourke and Holcomb (2002) listened to single words and pseudowords while paying close attention to the stimuli. N400 latencies for words and pseudowords were compared, but amplitudes were not, although visual inspection of the ERP waveforms suggests an enhanced N400 for pseudowords compared to words. The pseudoword N400s were longer lasting at all electrode sites and peak latency was delayed for the pseudowords. It is difficult to directly compare these results to the current study because of the additional influence of noise on the N400 latency.

While it is known that pseudowords generate a larger N400 than words, it is unknown how this effect is modulated by presentation in the context of a sentence. To date, no studies have been conducted using speech stimuli in which pseudowords were presented in a sentence anomaly paradigm. Only one study, in the visual modality, could be found using words and pseudowords in a sentence context. In Laszlo and Federmeier's (2009) study, subjects read high Cloze probability sentences that could end in an expected word, an unexpected word that was an orthographic neighbour of the expected word, a pseudoword orthographic neighbour, a pseudoword non-neighbour, or a nonword orthographic neighbour or non-neighbour. They found that the mean N400 amplitude was significantly larger for unexpected words and pseudowords compared with expected words, but that the responses to unexpected words and pseudowords did not differ from each other. Given that no significant difference was observed between the N400 effect observed to unexpected words and pseudowords, these findings may not support the theory that incongruent words in the present paradigm were processed similarly to pseudowords at threshold. Additional research in the auditory modality and in the absence of attention are necessary before any conclusion can be reached.

Despite the nonlinear nature of the relationship between SNR and N400 effect amplitude, the consistent occurrence of a minimum amplitude near-threshold could yield a clinically useful test. To examine the consistency of the effect across subjects, individual N400 effect amplitudes were plotted as a function of intelligibility condition in Figure 10. The effect was not consistent across subjects: for some individuals, the N400 effect reached a minimum at +1 dB, while in several others, it tended to reach a minimum at +2 dB. One possible source of this variation is inter-individual differences in the error of the measured SNR-50. Variability in the measured SNR-50 as a consequence of the sentence list that was used for the behavioural test was examined in the study pilot. Theoretically, if a more intelligible list were used for threshold determination, the threshold would be overshot and the intelligibility conditions would be set too difficult. This would result in a shift whereby the level set for the +2 dB condition might actually represent a +1 dB difference from the "true" threshold, the +1 dB level would actually represent the threshold, etc. By examining Figures 1 and 10 and the study records indicating which individual was tested with which list, such a shift could be identified. For example, it appears this could have happened with the 6th subject from the top on Figure 10, represented by a dashed line and open diamonds, because the inflection point in peak N400 effect amplitude was at +2 dB rather than +1 dB. For this individual, list 4 was used for the behavioural test. List 4 was average in terms of intelligibility, as shown in Figure 1, so the difference in the location of the inflection point for this subject is not likely attributable to SNR-50 measurement error. All subjects' records were examined for any potential effect of list intelligibility, but in no case was there reason to believe that this had an effect on the results.

Another consideration is the question of whether N400 amplitude changes were due to differing occurrence of the N400 or due to varying amplitude of individual N400s. The N400 is always observed by averaging many trials. In this study, approximately 50 trials were used for each intelligibility condition and stimulus type. Therefore reductions in N400 amplitude in the grand averages could have been due to either reduced N400 amplitude in all 50 trials or complete absence of the N400 in some trials with a "normal amplitude" N400 in others. Similarly, an N400 enhancement in the grand average could be due to actual amplitude increases in individual trials, or due N400s occurring in an increased number of trials. When the average of the trials was taken, both of these scenarios could result in the same average N400 amplitude. So long as the measurement of the N400 effect is performed in averaged waveforms, the underlying single-trial morphology of the response remains a theoretical consideration.

CHAPTER V: CONCLUSION

It is clear from this study that varying levels of speech-frequency noise affect the amplitude of the N400 effect as measured using a sentence anomaly paradigm, and that this effect is nonlinear. Comparison of these results with other studies investigating the effect of signal degradation or masking on the N400 yielded some commonalities between studies. Firstly, changes in the N400 effect were due to differences in the incongruent, rather than congruent waveforms, as found in Aydelott et al. (2006). The increase in N400 latency in masked conditions found in this study was also present in Obleser and Kotz (2011), Strauß et al. (2013), and Connolly et al. (1992). Finally, decreased N400 effect amplitude was observed in this study in certain intelligibility conditions, which was another finding commonly reported in the literature, e.g. by Daltrozzo et al. (2005) and Obleser and Kotz (2011).

However, there were also important contrasts between related studies and the current work. They often conflicted in terms of methods and purpose, e.g., to encourage rather than discourage increased use of attention, which is another factor that affects the N400. Most other studies also used active tasks whereas this study was passive. Unfortunately, no other research has explicitly investigated N400 morphology close to the behavioural hearing threshold, making interpretation of the relative amplitude increase at threshold found in this study more difficult. One possible reason for this increase could be a shift in processing strategy once intelligibility is close to the behavioural threshold (SNR-50). Suggested here is that this processing shift may result in incongruent words being processed more similarly to pseudowords; however, this is a speculative hypothesis that would require further research to investigate.

Before the N400 can be said to be useful for application in a physiological speech-in-noise test, further research into the effects of noise on the N400 effect must be done, including attempting to replicate the current results. Repeating a similar procedure with a more difficult subthreshold condition may also be useful, to examine at what point the N400 effect would be obliterated. A priming paradigm could also be used to investigate the effect of noise on the N400. The results of such a study would provide a comparison for the results presented here, and would be informative on the feasibility of an electrophysiology-based words-in-noise test. Another problem is that hearing tests are used on individuals, whereas ERP research is usually done at the group level. In this study patterns of N400 amplitude change with SNR did vary across individuals with normal hearing, so further research into individual responses and variability would need to be done before clinical application could be considered. Of course, research would also have to be done across age groups and hearing abilities.

The N400 component is complex and likely involves several sub-components that are distinct in terms of timing, distribution, and cognitive function (Nobre & McCarthy 1994). It appears that the combined effects of listening condition and congruency affected these underlying processes in a way that resulted in an unexpected pattern. This project showed that the relationship between listening condition and the N400 is complex and varies between individuals, and that the effect appears to be driven primarily by changes in the responses to the incongruent terminal words. With the current findings, the N400 does not appear to be immediately useful as an objective measure of speech comprehension in noise. However, the results of this study do not rule out this possibility if the relationship were better described through further research.

REFERENCES

American Speech-Language and Hearing Assocition. (2006). *Preferred Practice Patterns for the Profession of Audiology*. Retreived from http://www.asha.org/policy/PP2006-00274/#sec1.4.5

Anderson, S., & Kraus, N. (2010). Sensory-Cognitive Interaction in the Neural Encoding of Speech in Noise: A Review. *Journal of the American Academy of Audiology, 21*(9), 575-585.

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience,30*(14), 4922-4926.

Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, *43*, 454–464.

Berkum, J. V., Hagoort, P., & Brown, C. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657-671.

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., 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*(1-2), 15-24.

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods, 42*(3), 665-670.

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory and Cognition, 8,* 631-642.

Bregman, A. S., & McAdams, S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound. *The Journal of the Acoustical Society of America*, *95*, 1177-1178.

Bronkhorst, A. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica*, *86*, 117-128.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 110*(5), 2527-2538.

Cameron, S. & Dillon, H. (2007). Development of the Listening in Spatialized Noise – Sentences Test (LISN-S). *Ear and Hearing, 28(2),* 196–211.

Carhart, R. (1946). Selection of hearing aids. *Archives of Otolaryngology. 44*, 1-18.

Carhart, R. & Tillman, T. (1970). Interaction of competing speech signals with hearing loss. *Archives of Otolaryngology, 91*(3), 273-279.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975-979.

Cohen, M.S. (2008). *Handedness Questionnaire*. Retrieved from http://www.brainmapping.org/shared/Edinburgh.php#

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and language, 43,* 1-18.

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). Development of the connected speech test (CST). *Ear and hearing, 8,* 119-126.

Crandell, C. C., & Smaldino, J. J. (2000). Classroom Acoustics for Children with Normal Hearing and with Hearing Impairment. *Language, Speech, and Hearing Services in Schools, 31*(4), 362-370.

Cutting, J. E. (1975). Aspects of phonological fusion. *Journal of Experimental Psychology: Human Perception and Performance*, *104*(2), 105-120.

Daltrozzo J., Wioland, N., & Kotchoubey, B. (2005). *Explicit and implicit mechanisms of semantic priming: ERP effects to masked sentences.* [PDF poster]. Retreived from http://neurochem.u-strasbg.fr/posters/

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1), 132-147.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222-241.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 218-244.

Davis, P. A. (1939). Effects of acoustic stimuli on the waking human brain. *Journal of Neurophysiology*, *2*(6), 494-499.

Deacon, D., & Shelley-Tremblay, J. (2000). How automatically is meaning accessed: a review of the effects of attention on semantic processing. *Frontiers in Bioscience, 5*, 82-94.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *The Journal of the Acoustical Society of America, 68*(3), 843-857.

Ding, N., & Simon, J. Z. (2013). Acaptive temporal encoding leads to a background-insensitive cortical representation of speech. *The Journal of Neuroscience, 33*(13), 5728-5735.

Downs, M. P., & Yoshinaga-Itano, C. (1999). The efficacy of early identification and intervention for children with hearing impairment. *Pediatric Clinics of North America, 46,* 79-87.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183-192.

Friedrich, C. K., Eulitz, C., & Lahiri, A. (2006). Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions, 2*, 36-36. doi:10.1186/1744-9081-2-36

Friedrich, M., & Friederici, A. D. (2005). Semantic sentence processing reflected in the event-related potentials of one- and two-year-old children. *Neuroreport, 16,* 1801-1804.

Gabr, T. A. (2011). Mismatch negativity in auditory neuropathy/auditory dys-synchrony. *Audiological Medicine*, *9*(3), 91-97.

Geisler, C. D., Frishkopf, L. S., & Rosenblith, W. A. (1958). Extracranial responses to acoustic clicks in man. *Science, 128*(3333), 1210-1211.

Godey, B., Schwartz, D., De Graaf, J. B., Chauvel, P., & Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology, 112*(10), 1850-1859.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*(2), 149-188.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*(6), 660-676.

Hafter, E. R., & Schlauch, R. S. (1992). Cognitive factors and selection of auditory listening bands. In A. Dancer, D. Henderson, R. J. Salvi, & R. P. Hammernik (Eds.), *Noise-induced hearing loss* (303–310). Philadelphia: B.C. Decker.

Holcomb, P. J., Coffey, S. A., & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology, 8,* 203-241.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: an electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721-742.

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., ... & Wingfield, A. (2012). Central presbycusis: a review and evaluation of the evidence. *Journal of the American Academy of Audiology*, *23*(8), 635-666.

Hyde, M. (1997). The N1 response and its applications. *Audiology and Neuro-Otology, 2*(5), 281-307.

Ibáñez, A. M., Martin, R. S., Hurtado, E., & López, V. (2009). ERPs studies of cognitive processing during sleep. *International Journal of Psychology*, *44*(4), 290-304.

Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(2), 240-247.

Jerger, J., Greenwald, R., Wambacq, I., Seipel, A., & Moncrieff, D. (2000). Toward a more ecologically valid measure of speech understanding in background noise. *Journal of the American Academy of Audiology, 11,* 273-282.

Jerger, J., & Musiek, F. (2000). Report of the consensus conference on the diagnosis of auditory processing. *Journal of the American Academy of Audiology*, *11*(9), 467-474.

Jerger, S., & Jerger, J. (1982). Pediatric speech intelligibility test: Performance-intensity characteristics. *Ear and Hearing, 3,* 325-334.

Jerger, S., Lewis, S., Hawkins, J., & Jerger, J. (1980). Pediatric speech intelligibility test. I. generation of test materials. *International Journal of Pediatric Otorhinolaryngology, 2,* 217-230.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61,* 1337-1351.

Keidser, G. (1996). Selecting different amplification for different listening conditions. *Journal of the American Academy of Audiology*, *7*, 92–104.

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. *The Journal of the Acoustical Society of America, 116,* 2395-2405.

Kolinsky, R., & Morais, R. J. (1996). Migrations in speech recognition. *Language and cognitive processes*, *11*(6), 611-620.

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage*, *56*(2), 455-475.

Kutas, M. (1993). In the company of other words: electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes, 8*(4), 533-572.

Kutas, M., Hillyard, S. A., & Gazzaniga, M. S. (1988). Processing of semantic anomaly by right and left hemispheres of commissurotomy patients. evidence from event-related brain potentials. *Brain: A Journal of Neurology, 111*, 553-76.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences 4*(12), 463–470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621-47.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203-205.

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia, 20,* 579-590.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161-163.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language, 61*(3), 326-338.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920-933.

Lobaugh, N. J., West, R., & McIntosh, A. R. (2001). Spatiotemporal analysis of experimental differences in event‐related potential data with partial least squares. *Psychophysiology*, *38*(3), 517-530.

Luck, S. J. (2005). An introduction to the event-related potential technique. Cambridge: MIT press.

Martin, B. A., Tremblay, K. E., & Korczak, P. (2008). Speech evoked potentials: from the laboratory to the clinic. *Ear and Hearing, 29*(3), 285-313.

McArdle, R. A., Wilson, R. H., & Burks, C. A. (2005). Speech recognition in multitalker babble using digits, words, and sentences. *Journal of the American Academy of Audiology, 16,* 726-739.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage*, *3*(3), 143-157.

Mueller, H. G. (2001). Speech audiometry and hearing aid fittings: Going steady or casual acquaintances? *The Hearing Journal, 54*(10), 19-29.

Mueller, H. G. (2003). Fitting test protocols are "more honored in the breach than the observance". *The Hearing Journal, 56*(10), 19-26.

Mueller, H. G. (2010). Three pre-tests: What they do and why experts say you should use them more. *The Hearing Journal, 63*(4), 17-24.

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*(4), 375-425.

Ng, S. L., Meston, C. N., Scollie, S. D., & Seewald, R. C. (2011). Adaptation of the BKB-SIN test for use as a pediatric aided outcome measure. *Journal of the American Academy of Audiology, 22,* 375-386.

Nilsson, M. J., Soli, S. D., & Gelnett, D. (1996). Development of the hearing in noise test for children (HINT-C). *House Ear Institute*, 1–9.

Nilsson, M., Soli, 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. *The Journal of the Acoustical Society of America, 95,* 1085-1099.

Nobre, A. C., & McCarthy, G. (1994). Language-Related ERPs: Scalp Distributions and Modulation by Word Type and Semantic Priming. *Journal of Cognitive Neuroscience, 6*(3), 233-255.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, *47*(2), 204-238.

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, *55*(2), 713-723.

O'Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology, 60*(2), 121-150. doi:10.1016/S0301-0511(02)00045-5

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-804.

Parbery-Clark A, Skoe E, Lam C, Kraus N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing, 30*, 653–661.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733

Phonak. (2009). *Listening in Spatialized Noise – Sentences Test*. [PDF document] Retreived from https://www.phonakpro.com/content/dam/phonak/b2b/C_M_tools/Professional_Tools/Diagnostic/LiSN-S-User-Manual.pdf

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*(4), 456-479.

Plomp, R., & Mimpen, A. M. (1979). Improving the reliability of testing the speech reception threshold for sentences. *Audiology: official organ of the International Society of Audiology, 18,* 43-52.

Rabbitt, P. A. (1991). Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. *Acta Otolaryngologica Supplementum*, *476,* 167–176.

Relander, K., Rämä, P., & Kujala, T. (2009). Word Semantics Is Processed Even without Attentional Effort. *Journal of Cognitive Neuroscience*, *21*(8), 1511-1522.

Romei, L., Wambacq, I. J. A., Besing, J., Koehnke, J., & Jerger, J. (2011). Neural indices of spoken word processing in background multi-tasker babble. *International Journal of Audiology, 50,* 321-333.

Sarampalis, A., Kalluri, S., Edwards, B., & Hafter, E. (2009). Objective Measures of Listening Effort: Effects of Background Noise and Noise Reduction. Journal of Speech, Language, and Hearing Research, 52(5), 1230-1240.

Silva-Pereyra, J., Rivera-Gaxiola, M., & Kuhl, P. K. (2005). An event-related brain potential study of sentence comprehension in preschoolers: Semantic and morphosyntactic processing. *Cognitive brain research, 23,* 247-58.

Sininger, Y. S., Martinez, A., Eisenberg, L., Christensen, E., Grimes, A., & Hu, J. (2009). Newborn hearing screening speeds diagnosis and access to intervention by 20-25 months. *Journal of the American Academy of Audiology, 20,* 49-57.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037-2057.

Sperry, J. L., Wiley, T. L., & Chial, M. R. (1997). Word recognition performance in various background competitors. *Journal of the American Academy of Audiology, 8*(2), 71-80.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and language*, *120*(2), 135-162.

Steinschneider, M. & Dunn, M. (2002). Electrophysiology in developmental neuropsychology. In S. Segalowitz, & I. Rapin (Eds.), *Handbook of neuropsychology* (91–146). Amsterdam: Elsevier.

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed Expectancies under Degraded Speech: Revisiting the N400. *Journal Of Cognitive Neuroscience*, *25*(8), 1383-1395. doi:10.1162/jocn_a_00389

Supp, G., Schlögl, A., Gunter, T., Bernard, M., Pfurtscheller, G., & Petsche, H. (2004). Lexical memory search during N400: Cortical couplings in auditory comprehension. *Neuroreport, 15*(7), 1209-13.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science, 150*(3700), 1187-8.

Taylor, B. (2007). Predicting real-world hearing aid benefit with speech audiometry: an evidence-based review. *Audiology Online.* Retreived from http://www.audiologyonline.com/articles/predicting-real-world-hearing-aid-946

Taylor, W. L. (1953). 'Cloze procedure': A new tool for measuring readability. *Journalism Quarterly,* 415-433.

Tremblay, K., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology, 114*, 1332–1343.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279–293.

Warren, R. M., Riener, K. R., Bashford, J. A., & Brubaker, B. S. (1995). Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits. *Perception & Psychophysics*, *57*(2), 175-182.

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology, 14*(9), 453-70.

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.

Wood, C. C., & Wolpaw, W. J. (1982). Scalp distribution of human auditory evoked potentials. II. evidence for overlapping sources and involvement of auditory cortex. *Electroencephalography and Clinical Neurophysiology, 54*(1), 25-38.

APPENDIX: SENTENCE STIMULI

Sentence stem	Congruent ending	Incongruent ending	CP (Block & Baldwin, 2010)	List number (congruent)	List number (incongruent)
She could tell he was mad by the tone of his	Voice	Air	0.99	1	2
She went to the bakery for a loaf of	Bread	Art	0.98	11	5
Bob proposed and gave her a diamond	Ring	Bank	0.98	9	3
The dentist recommends brushing your teeth twice a	Day	Band	0.98	1	3
He loosened the tie around his	Neck	Bike	0.97	1	15
Dan was asked to be the new coach of the	Team	Bet	0.97	12	3
They paid for their meals but forgot to leave a	Tip	Broom	0.97	13	14
To pay for tuition she took out two student	Loans	Dogs	0.97	13	10
She didn't have her watch so she asked for the	Time	Book	0.97	4	14
Sherry had to read lips because she was	Deaf	Green	0.97	14	8
They sat together without speaking a single	Word	Bread	0.96	14	5
After hitting the iceberg the ship began to	Sink	Bloom	0.96	6	2
She went to the beauty parlor to perm her	Hair	Book	0.96	1	6
Without her sunglasses the sun hurt Erika's	Eyes	Cake	0.96	10	11
Joe did not like his outfit and decided to	Change	Swim	0.96	9	7
The limping horse was obviously in much	Pain	Car	0.96	9	6
To prevent football injury all players must wear shoulder	Pads	Change	0.96	11	13
To hang the picture Ted needed a hammer and	Nail	Cake	0.96	7	13
After the argument Ann went to her room and slammed the	Door	Class	0.96	5	8
Father carved the turkey with a	Knife	Cold	0.95	12	9
Water and sunshine help plants	Grow	Cook	0.95	13	5
The kids fed the ducks some stale	Bread	Luck	0.89	5	7
He mailed the letter without a	Stamp	Cry	0.95	4	11
She wore a colorful scarf around her	Neck	Door	0.95	14	4
The cheap pen ran quickly out of	Ink	Day	0.95	10	4
He had a long day and was in a bad	Mood	Heat	0.95	10	5
She graduated at the top of her	Class	Board	0.95	8	1
The athlete enjoyed lifting weights at the	Gym	Door	0.95	15	12

Sentence stem	Congruent ending	Incongruent ending	CP (Block & Baldwin, 2010)	List number (congruent)	List number (incongruent)
Expecting Jeff's call she waited for the phone to	Ring	Fit	0.95	11	7
When she got out of the car she closed the	Door	Gym	0.95	10	7
The package was sent through the	Mail	Check	0.95	4	14
In the shower he washed his skin with	Soap	Eggs	0.95	5	6
You would need a raincoat to avoid getting	Wet	Deaf	0.95	10	3
They turned in their project on the date it was	Due	Hot	0.95	1	8
For his date Tom bought a long stemmed	Rose	Farm	0.95	7	6
The genie promised the man he would grant one	Wish	Fault	0.95	2	6
One year after her death Bill visited his mother's	Grave	Fire	0.95	4	14
The children went outside to	Play	Grow	0.94	12	10
The teacher wrote the problem on the	Board	Fish	0.94	10	9
At night the old woman locked the	Door	Tour	0.94	7	9
Her job was easy most of the	Time	Gas	0.94	9	14
When you go to bed turn off the	Light	Fly	0.94	2	1
After dinner he washed his hands with	Soap	Fire	0.94	7	10
In the quiet movie theater, Kim's phone	Rang	Walked	0.94	7	2
Spring was Jo's favorite season of the	Year	Grave	0.94	6	2
It was dark in the room so she turned on the	Light	Door	0.94	3	4
The kitten played with the ball of	Yarn	Ink	0.94	14	1
The farmer spend the morning milking his	Cows	Hair	0.94	8	10
After every meal it's good to brush your	Teeth	Hand	0.94	8	10
He brought his bait to the lake to catch	Fish	Heat	0.94	2	8
It was cold in the room so they turned on the	Heat	Knife	0.94	6	12
The room was loud so I had to yell to be	Heard	Due	0.94	11	7
Bill went to the dentist to check all his	Teeth	Time	0.94	3	5
It was windy enough to fly a	Kite	Leash	0.94	9	12
Ellen enjoys poetry, painting, and other forms of	Art	Hair	0.94	6	10
He turned to channel 13 to watch the daily	News	Kite	0.94	11	12
Jenny lit the candles on the birthday	Cake	Stars	0.94	7	9
Although Keith bowled well he did not have the highest	Score	Heat	0.94	15	5
When the alarm rang the firefighter slid down the	Pole	Knife	0.94	4	8

Sentence stem	Congruent ending	Incongruent ending	CP (Block & Baldwin, 2010)	List number (congruent)	List number (incongruent)
The little girl left Santa a plate of cookies and	Milk	Eyes	0.94	8	15
In the night sky it is easier to see all the	Stars	Knees	0.94	5	8
The dirty dishes were piling up in the	Sink	Leaves	0.94	2	13
She could not buy the shirt because it did not	Fit	Light	0.94	4	11
He cashed his new paycheck at the	Bank	Light	0.94	15	13
The baby birds were ready to leave the	Nest	Lights	0.94	6	2
Cold weather outside meant it was time to turn on the	Heat	Loans	0.94	15	8
He was so sure the racehorse would win he made a	Bet	Tune	0.94	3	8
At first the woman refused, but she changed her	Mind	Lunch	0.93	9	11
She went to the salon to color her	Hair	Trees	0.93	6	13
Success is often just a matter of hard	Work	Milk	0.93	11	15
The gambler had a streak of bad	Luck	Brooms	0.89	3	11
Most cats see very well at	Night	Score	0.93	14	5
Since driving at night decreases visibility it is better to turn on the	Lights	Nail	0.93	11	10
A flat tire forced Katie to pull to the side of the	Road	Name	0.93	8	13
To learn about their ancestors they drew a family	Tree	Mood	0.93	14	13
The wonderful waitress received a generous	Tip	Neck	0.93	8	9
She loved playing the guitar so she joined the	Band	Neck	0.93	12	3
He wore a heavy jacket because it was	Cold	News	0.93	2	14
Because there was lightning she could not go to the pool to	Swim	Nest	0.93	8	15
On Valentine's Day the woman received a single red	Rose	Night	0.93	13	15
The princess would someday become a	Queen	Desk	0.93	12	11
The Parkers' baby could already say three	Words	Pads	0.93	10	13
The grandmother left everything to her son in her	Will	Pain	0.93	3	13
The fluffy white clouds are high up in the	Sky	Paint	0.93	9	10
To promote their album the band went on	Tour	Pan	0.93	13	9
Katie put the flowers in an expensive	Vase	Road	0.93	9	12
The maid dusted the books on the	Shelf	Name	0.93	15	15
The birthday card was funny and made me	Laugh	Play	0.93	12	1
Bill jumped in the lake and made a big	Splash	Pole	0.93	12	9
The boat passed easily under the	Bridge	Name	0.89	15	5

Sentence stem	Congruent ending	Incongruent ending	CP (Block & Baldwin, 2010)	List number (congruent)	List number (incongruent)
To keep the dogs out of the yard he put up a	Fence	Mind	0.89	3	15
Jane hung the colorful painting up on the	Wall	Rain	0.92	8	14
Sarah saw animals from around the world at the	Zoo	Seeds	0.92	13	14
For his birthday Jan baked a	Cake	Ring	0.92	7	1
When driving you should keep your eyes on the	Road	Ring	0.92	2	4
After inhaling smoke from the fire she needed fresh	Air	Pants	0.92	7	1
The college student went to the library to read a	Book	Road	0.92	5	6
Cid needed a belt to hold up his	Pants	Rose	0.92	12	10
She washed the dirty dishes in the	Sink	Rose	0.92	4	8
She married just for money and not for	Love	School	0.92	1	9
To grow a garden you must first plant	Seeds	Mail	0.92	15	7
After playing in the sun all day, his face was badly	Burnt	Heard	0.92	1	12
Walking through the dark room I accidentally stubbed my	Toe	Shark	0.92	3	5
When babies are hungry they may often	Cry	Work	0.92	11	9
Surfers are scared of getting bitten by a	Shark	Beard	0.92	4	6
At dinner he cut his food with a	Knife	Sign	0.92	2	12
The announcer on the radio told the breaking	News	Sink	0.92	5	10
Dan gathered more wood for the	Fire	Sink	0.92	6	3
After raking the yard Pat jumped into the pile of	Leaves	News	0.92	12	15
He eats out because he is a lousy	Cook	Size	0.92	5	11
Every autumn leaves fall from the	Trees	Splash	0.92	11	9
Wanting color in the room, he bought a can of	Paint	Sky	0.92	10	15
Jessie ran the race at a slower	Pace	Smell	0.92	13	15
John felt sorry, but it was not his	Fault	Soap	0.92	12	5
To pay for the car, Al simply wrote a	Check	Shelf	0.91	5	3
The lecture should last about one	Hour	Sky	0.91	3	11
They raised pigs on their	Farm	Stamp	0.91	3	12
The wealthy child attended a private	School	Sword	0.91	6	4
Her new shoes were the wrong	Size	Swim	0.91	2	4
Instead of receiving at the holidays it is better to	Give	Pace	0.91	13	3
Bradley prefers cats over	Dogs	Tea	0.91	11	7

Sentence stem	Congruent ending	Incongruent ending	CP (Block & Baldwin, 2010)	List number (congruent)	List number (incongruent)
My favorite part of spring is when the flowers	Bloom	Team	0.91	15	8
For breakfast Jim wanted bacon and	Eggs	Teeth	0.91	2	6
The baby bird was ready to learn to	Fly	Give	0.91	13	1
The knight readied for battle and drew his	Sword	Time	0.91	8	4
The piano sounded awful and was out of	Tune	Tip	0.91	7	1
The full moon lit up the night	Sky	Toe	0.91	14	7
She could not drink the coffee because it was too	Hot	Down	0.91	1	14
When Colin saw smoke he called 911 to report a	Fire	Work	0.91	14	6
I realized I had no umbrella as it began to	Rain	Laugh	0.91	14	4
The exit was marked by a large	Sign	Soap	0.91	15	1
He was afraid to work the night	Shift	Tree	0.9	14	6
George must keep his pet on a	Leash	Love	0.9	6	11
John swept the floor with a	Broom	Vase	0.9	8	1
The old house will be torn	Down	Voice	0.9	4	3
Benny tried to make and keep a new resolution each	Year	Gym	0.9	13	14
A good way to exercise is to ride a	Bike	Wall	0.9	6	13
She made herself a sandwich and chips for	Lunch	Wet	0.9	1	7
The movie was so sad it made the audience	Cry	Will	0.9	15	3
John took his dog out for a	Walk	Wish	0.9	9	7
She preheated the oven and greased the	Pan	Word	0.9	7	2
He turned the page of his favorite	Book	Hour	0.9	10	2
I would drive but my car is low on	Gas	Words	0.9	3	12
The accountant ironed a shirt before going to	Work	Melt	0.9	5	2
The garbage from last week had a foul	Smell	Yarn	0.9	5	4
Amber went to the dealership to purchase a new	Car	Year	0.9	9	2
She had the flu and needed to drink some hot	Tea	Year	0.9	10	11
The princess was only permitted to marry a	Prince	Zoo	0.9	4	2
While shipwrecked, the sailor grew a long	Beard	Shift	0.9	2	4
In order to study Karen sat down at her	Desk	Tip	0.9	1	12