May 2020

Cortical auditory processing of informational masking effects by target-masker similarity and stimulus uncertainty

Christopher Edward Niemczak
Syracuse University

Abstract

Purpose: Understanding speech in a background of other people talking is one of the most difficult listening challenges for hearing-impaired individuals, and even for those with normal hearing. Speech-on-speech masking, is known to contribute to increased perceptual difficulty over non-speech background noise because of *informational masking* provided over and above the energetic masking effect. While informational masking research has identified factors of similarity and uncertainty between target and masker that contribute to reduced behavioral performance in speech background noise, critical gaps in knowledge including the underlying neural-perceptual processes remain. By systematically manipulating aspects of similarity and uncertainty in the same auditory paradigm, the current study proposed to examine the time course and objectively quantify these informational masking effects at both early and late stages of auditory processing using auditory evoked potentials (AEPs) in a two-factor repeated measures paradigm.

Method: Thirty participants were included in this cross sectional repeated measures design. Target-masker similarity between target and masker were manipulated by varying the linguistic/phonetic similarity (i.e. language) of the talkers in the noise maskers. Specifically, four levels representing hypothesized increasing levels of informational masking were implemented: (1) No masker (quiet), (2) Mandarin (linguistically and phonetically dissimilar), (3) Dutch (linguistically dissimilar, but phonetically similar), and (4) English (linguistically and phonetically similar). Stimulus uncertainty was manipulated by task complexity, specifically target-to-target interval (TTI) of an auditory paradigm. Participants had to discriminate between English word stimuli (/bæt/ and /pæt/) presented in an oddball paradigm in each masker

condition at +3 dB SNR by pressing buttons to either the target or standard stimulus (pseudo-randomized between /bæt/ and /pæt/ for all participants). Responses were recorded simultaneously for P1-N1-P2 (standard waveform) and P3 (target waveform). This design allowed for simultaneous recording of multiple AEP peaks, including analysis of amplitude, area, and latency characteristics, as well as accuracy, reaction time, and d' behavioral discrimination to button press responses. Finally, AEP measurers were compared to performance on a behavioral word recognition task (NU-6 25-word lists) in the proposed language maskers and at multiple signal-to-noise ratios (SNRs) to further explore if AEP components of amplitude/area and latency are correlated to behavioral outcomes across proposed maskers.

Results: Several trends in AEP and behavioral outcomes were consistent with the hypothesized hierarchy of increasing linguistic/phonetic similarity from Mandarin to Dutch to English, but not all differences were significant. The most supported findings for this factor were that all babble maskers significantly affected outcomes compared to quiet, and that the native language English masker had the largest effect on outcomes in the AEP paradigm, including N1 amplitude, P3 amplitude and area, as well as decreased reaction time, accuracy, and d' behavioral discrimination to target word responses. AEP outcomes for the Mandarin and Dutch maskers, however, were not significantly different across all measured components. Outcomes for AEP latencies for both N1 and P3 also supported an effect of stimulus uncertainty, consistent with a hypothesized increase in processing time related to increased task complexity when target stimulus timing was randomized. In addition, this effect was stronger, as evidenced by larger effect sizes, at the P3 level of auditory processing compared to the N1. An unanticipated result was the absence of the expected additive effect between linguistic/phonetic similarity and stimulus uncertainty. Finally, trends in behavioral word recognition performance were generally

consistent with those observed for AEP component measures such that no differences between Dutch and Mandarin maskers were found, but the English masker yielded the lowest percent correct scores. Furthermore, correlations between behavioral word recognition and AEP component measures yielded some moderate correlations, but no common AEP components accounted for a majority of variance for behavioral word recognition.

Conclusions: The results of this study add to our understanding of auditory perception in informational masking in four ways. First, observable effects of both similarity and uncertainty were evidenced at both early and late levels of auditory cortical processing. This supports the use of AEPs to better understand the informational masking deficit by providing a window into the auditory pathway. Second, stronger effects were found for P3 response, an active, top-down level of auditory processing providing some suggestion that while informational masking degradation happens at lower levels, higher level active auditory processing is more sensitive to informational masking deficits. Third, the lack of interaction of main effects leads us to a linear interpretation of the interaction of similarity and uncertainty with an equal effect across listening conditions. Fourth, even though there were few and only moderate correlations to behavioral word recognition, AEP and behavioral performance data followed the same trends as AEP measures across similarity. Through both auditory neural and behavioral testing, language maskers degraded AEPs and reduced word recognition, but particularly using a native-language masker. The behavioral and objective results from this study provide a foundation for further investigation of how the linguistic content of target and masker and task difficulty contribute to difficulty understanding speech in noise.

Cortical auditory processing of informational masking effects by target-masker similarity and stimulus uncertainty

By

Christopher Edward Niemczak, Au.D.
B.S. University of Vermont, 2013
Au.D. Syracuse University, 2018

Dissertation

Submitted in partial fulfillment of the requirement
for the degree of Doctor of Philosophy in Audiology

Syracuse University
May 2020

Acknowledgements

There are so many people in my life who have made this dissertation possible and I want to say thank you to everyone who believed in me and pushed me to reach my potential.

First and foremost, thank you Dr. Kathy Vander Werff for your expertise, encouragement, and mentorship throughout my entire education at Syracuse University. That mentorship has given me the foundation and guided me along a path that I am proud to walk every single day. I also owe you a special thanks for your help on this dissertation, as well as helping me realize that science is a balance, which may or may not include Spartan races. Thank you, Kathy.

I would like to also thank my family and friends for giving me the utmost support, especially my wife Caitlin, who deserves an award for the sheer number of electrodes placed on her head. You are truly my number one test subject in every experiment and I thank you for that. Most importantly, I want to thank you for your support at home, never wavering in the belief that I would earn this degree; giving me support throughout my entire education from the day we met to this moment and into the future. I love you.

Many thanks to the members of my dissertation committee: Dr. Beth Prieve, Dr. Karen Doherty, Dr. Jonathan Preston, and Dr. Curtis Billings, who graciously gave their time, advice, and notably their critiques of this dissertation. This document, and more crucially, the way I view auditory science has been enriched because of you.

Thank you to Dr. Lauren Calandruccio, Dr. Susanne Brouwer, and Dr. DeLiang Wang for your assistance in creating language maskers, normalizing those maskers, and controlling for energetic masking, respectively. I would also like to thank Dr. Naomi Bramhall, Dr. Melissa Papesh, and Dr. Erick Gallun for helping me think about my dissertation ideas when they were still in their infancy. You have shown me that collaboration in this field of auditory science is unmatched. I look forward to collaborating with you in the future.

I would like also acknowledge all relationships between family, friends, professors, and students I have made along this journey. That includes my immediate and extended family, all of my classmates at Syracuse, colleagues at the NCRAR, students in CSD212, and so many more. Even though most of them are still confused about what I actually do, they support me nonetheless.

This research was supported in part by the Gerber Funds for Auditory Research at Syracuse University and the Council for Academic Programs in Communication Sciences and Disorders.

On the other hand, if this dissertation is not a success, I would like to dedicate it to Sigma Plot and Covid-19, which inexplicably added to the immense amount of stress graphing and writing the results and discussion section of this dissertation.

Finally, every person mentioned above, I would also like to blame all of you. Writing this dissertation has been an exercise in sustained suffering. The casual reader may excuse themselves from excessive guilt, but for those of you who have played the larger role in prolonging my agonies with your encouragement and support, well…you know who you are…Thanks.

Table of Contents

List of Figures

List of Tables

1.0 Introduction

1.1 Overview

In everyday listening situations, understanding speech in a noisy environment is a challenge, even for individuals with normal hearing. These background noises can be composed of any sound that covers or masks a target signal, including variable noises such as other people talking. This complex interaction of signal and noise creates difficulty comprehending speech, and is also one of the most common complaints of those with hearing impairment (Abutan et al. 1993; Oticon 2016) and even those with normal hearing. In addition, increased speech in noise difficulties have been documented for individuals with central auditory processing disorder (CAPD), head injury, and suspected cochlear synaptopathy, all of whom may have hearing within normal limits on the audiogram (Fausti et al. 2009; Festen et al. 1990; Kujawa et al. 2015; Putter-Katz et al. 2008). While extensive research has documented the prevalence of difficulty with this critical perceptual task, understanding the nature of the deficits across these populations remains incomplete. In addition, the specific acoustic factors of noise that degrade speech recognition and how cognitive factors, such as attention, modulate this process are not fully understood.

Background noise comprised of competing speech has generally been shown to be the most detrimental to speech understanding compared to continuous noises such as white noise or speech-shaped noise (continuous noise spectrally shaped to speech), even at similarly challenging signal-to-noise ratios (SNRs) (Brungart et al. 2001; Helfer et al. 2008). It is also in the presence of competing speech that those with auditory deficits have the most difficulty understanding a target speaker (Killion et al. 2004). Speech and non-speech competing background noise have been generally categorized as providing two kinds of masking to the target speech – energetic and informational. Energetic masking, the type provided by continuous white noise and speech-shaped noise, is caused by physical interactions between the target signal and masker at the auditory periphery. Informational masking, which is more detrimental to speech perception, not only physically interferes at the periphery, but also *perceptuall*y interferes with speech understanding at higher top-down levels (Durlach, Mason, Shinn-Cunningham, et al. 2003). That is, informational masking includes energetic components, but increases perceptual difficulty due to the addition of relevant auditory content. When background noise is other people talking, it can be difficult for the listener to attend to and separate out the target speaker they are trying to listen to from the competing speech signals. Energetic and informational masking, therefore, appear to operate at different levels of processing within the auditory pathway. Deficits in speech recognition in speech background noise are part of a complex interaction of target speech, masker speech, and listener related factors involved in this combination of energetic and informational masking.

Speech Recognition in Noise: Energetic Masking

Energetic masking can be defined to occur when target and masker energy are present at a similar time and frequency (Brungart et al. 2001), such that they directly compete at the auditory periphery. This phenomena has much in common with the neurophysiological concept of “line-busy” masking, in which the presence of an auditory signal does not increase the average response rate of the auditory nerve above the response rate elicited by the masker at a similar frequency (Culling 2013). Thus, it was hypothesized that speech can be masked by continuous noise that shares the same frequency spectrum at similar intensity levels. In the auditory system, a broadband steady-state noise masker is filtered into a series of narrowband signals, each with a bandwidth corresponding to the bandwidth of the auditory filter at that center frequency (Glasberg et al. 1990). Therefore, if the noise energy falls within the same bands as the target speech, masking will occur.

Across speech perception studies, is has been shown that the effect of purely energetic masking on speech recognition is determined not only by the level of spectral overlap, but also by the temporal fluctuations within that spectrum (Billings et al. 2011; Culling 2013; Hall et al. 2002; Stuart et al. 1995). If the temporal envelope of the noise is not completely steady, but has fluctuations, there will be periods of time where the target signal is not completely masked within each channel (Elhilali et al. 2003). The effects of spectro-temporally modulated maskers on speech recognition have generally shown a *release from masking* (i.e. improved speech recognition) with modulations of energetic masking (Brungart et al. 2001; Rosen et al. 2013; Stone et al. 2016). For example, Stone et al. (2012) manipulated the depth and bandwidth of the fluctuations of a white noise masker and found that listeners performance increased with the depth and duration of envelope fluctuation of the maskers. Cooke (2006) found that speech recognition scores in various temporally modulated masking conditions confirmed that sufficient information exists during “dips” in the noise modulation to support consonant identification and that the proportion of the dips in a noise source is a good predictor of intelligibility. The theoretical basis for listening during dips, termed *glimpsing*, is based on taking advantage of glimpses where the target signal is least affected by the background noise (Cooke 2006; Stone and Canavan 2016).

It might be expected, therefore, that background noise composed of speech, which contains temporal fluctuations, would also provide the listener *improved* recognition due to this advantage of listening during these glimpses. However, Brungart et al. (2001) found *poorer* sentence recognition in normal-hearing young adults on a speech recognition task of a target phrase in the presence of speech-on-speech masking (termed *babble* masking) compared to energetic speech-shaped noise with and without envelopes modulated to mimic the temporal variations in speech. Specifically, they found that in a binaural listening condition, 2- and 3-talker maskers resulted in ~30% worse speech understanding than a matched envelope modulated speech-shaped noise at the same SNR. That is, continuous noise with the same spectral shape and temporal fluctuations matched to mimic the fluctuation of babble maskers provided the listeners an advantage (masking release), but performance decreased in the babble conditions. The difference in performance between babble and modulated continuous noise was attributed to informational masking effects that were not present in the energetic masking of speech-shaped noise. Thus, if the spectral and temporal properties of the masker cannot explain the difficulty of speech in noise recognition when the background noise is also speech, what other factors make informational masking remarkably troublesome for listeners?

1.2 Speech Recognition in Noise: Informational Masking

Informational masking can be understood as masking effects that occur *in addition to and beyond* overlap on the auditory periphery, involving higher-order top-down auditory processing such as discrimination, attention, and memory. In the presence of informational masking noise, which is typically background speech, the listener must listen to and separate a particular person (the target speech) talking from a background of other people talking. The maskers therefore may be similar to the target speech in spectral and temporal features, may come from a similar location, may be more or less intense, and may also contain similar linguistic content – all of which may affect how much masking the listener experiences.

While Pollack (1975) has been credited with devising the term “informational masking”, the classic hypothesis of informational masking was defined by Durlach, Mason, Kidd, et al. (2003) stating that *similarity* and *uncertainty* of the target speech to the babble masker result in difficulty separating and, thus accurately perceiving the target speech. Although this definitional structure of informational masking remains broad, researchers have studied many aspects of similarity and uncertainty and how these phenomena affect speech recognition in noise (Brungart et al. 2001; Durlach, Mason, Shinn-Cunningham, et al. 2003; Hall et al. 2002; Kidd et al. 1998; Lutfi 1990; Neff et al. 1988; Watson and Nichols 1976). Target-masker similarity factors may include similar acoustic characteristics of the target speech and masker speech (such as intensity, spectral content, and timing, etc.) and the linguistic content and intelligibility of the target and masker speech (Brungart et al. 2001; Calandruccio et al. 2013; Cooke et al. 2008; Culling 2013; Freyman et al. 2007; Lutfi et al. 2003). Stimulus uncertainty can be related to factors such as the predictability of the presence of the target speech (whether someone is speaking) and consistency of the timing of the target relative to the background noise, both of which affect the ability to maintain focus on the target rather than become distracted by the background (Arbogast et al. 2002; Brungart et al. 2001; Freyman et al. 2004; Hoen et al. 2007; Leibold et al. 2010; Watson, Kelly, et al. 1976).

The concepts of similarity and uncertainty are not completely independent factors. For example, the more similar the target and masker on some acoustic dimension, the more uncertainty the listener may experience about whether the target is present within the masker(s). However, as theorized, these two important aspects of the target and masker relationship form the basis of experimental manipulations to better understand contributions to informational masking. It is the interaction of multiple factors that contributes to the perceived informational masking phenomenon. While behavioral research has described several aspects of target and masker speech that contribute to the perceived informational masking deficits and decreased speech recognition scores, the complex interaction, specifically the additive effects of similarity and uncertainty factors are not fully understood. In particular, how similarity and uncertainty affect physiological encoding in the auditory neural pathway and the relationship between this neural processing of speech-in-noise and resulting behavioral performance are not known.

1.2.1 Target-Masker Similarity

When target speech and background speech share similar acoustic features, speech recognition performance decreases (Brungart et al. 2001; Cooke et al. 2008; Culling 2013). Behavioral studies have shown a release from masking by *decreasing* the similarity between target and masker in a variety of acoustic dimensions including intensity, spectral content, and temporal pattern (Bertoli et al. 2005; Brungart et al. 2004; Cusack et al. 2004; Rosen et al. 2013; Sharma et al. 2014). As with energetic noise maskers, the similarity in intensity level between target and masker as reflected by the SNR plays a large role in the amount of perceived masking when speech maskers are used, with more challenging SNRs of 0 dB or poorer resulting in reduced speech recognition compared to more favorable SNRs (Billings et al. 2009; Brungart et al. 2001; Scott et al. 2004). These results are likely due to both energetic masking effects at the periphery and increased competition for the listener's attention between target and masker speech (Brungart et al. 2001; Cooke et al. 2008). However, as stated above, similarity in intensity or SNR alone do not fully account for informational masking effects as purely energetic maskers do not degrade speech recognition performance as much as speech babble masking with the same SNR (Brungart et al. 2001; Rosen et al. 2013).

Spectral similarity between target and masker speech also affects the amount of perceived informational masking in behavioral paradigms when speech maskers are used (Brungart et al. 2001; Helfer and Freyman 2008). Studies examining spectral differences between target and masker speech, such as dissimilar sex speakers or spectral vocoded speech (with differences in fundamental frequency), have shown that changes in the frequency similarity of the target and masker voices have a substantial impact on listener performance in multi-talker listening tasks (Helfer and Freyman 2008; Rosen et al. 2013). For example, Helfer and Freyman (2008) found significantly poorer sentence recognition in young normal-hearing participants when target and masker were presented in the same-sex compared to opposite-sex speakers. As previously discussed, however, it is not solely the spectral similarity that increases the masking effect, as purely energetic random noise maskers with similar spectral content do not cause the same decrease in performance (e.g. Brungart et al. (2001). Even artificial vocoded speech with the same spectral content as natural speech does not result in the same reduction in listener performance. Rosen et al. (2013) found that spectral vocoded (filtering speech into reduced freqeuncy bands) speech yielded significantly better speech recognition as compared to speech babble with matched spectral content and equal numbers of talkers (vocoded 4-talker vs. naturally spoken 4-talker). Thus, results show that spectral content of the masker alone cannot fully explain the informational masking deficit. There must be something inherent to *speech* that is contributing to this informational masking phenomena.

An aspect unique to speech is the linguistic and phonetic content, and similarity in these features between target and masker speech have been shown to contribute to informational masking in behavioral studies (Calandruccio, Buss, et al. 2014; Rhebergen et al. 2005; Van Engen et al. 2007). Linguistic similarity can be defined as how similar a target and masker language are to each other on a continuum of linguistic components, such as phonology, morphology, or semantic meaning (Calandruccio, Bradlow, et al. 2014). Phonetic similarity is one fundamental component of linguistic similarity and can be defined as how *similar permissible sound combinations* between target and masker language result in difficultly perceiving target from masker. As previously reviewed, it has been shown that speech-shaped energetic noise or even vocoded speech are less detrimental to speech recognition as compared to actual spoken language (Brungart et al. 2001; Culling 2013; Freyman et al. 2001; Rosen et al. 2013). The fact that natural speech babble maskers are more detrimental to speech understanding than other maskers suggests that the linguistic and phonetic content of the masker has a potential impact on informational masking.

Linguistic and phonetic similarity between target and masker speech have typically been studied by using a native or foreign language as the target and/or masker, as well as the use of accented speech (Burnett et al. 1996; Culling et al. 2005; Scharenborg et al. 2016). Use of a listener's native language, such as English, as target *and* masker would theoretically provide the highest level of informational masking (poorest behavioral speech recognition scores) due to highly similar linguistic and phonetic properties between the target speech signal and the background talkers (Brouwer et al. 2012; Calandruccio et al. 2013). Behavioral speech recognition studies have supported this hypothesis. For example, Rogers et al. (2004) found that listeners had significantly reduced performance when the target and background speech babble were both in the listener's native language (English) and spoken by native English speakers compared to when the background maskers spoke English with a Chinese accent. This finding suggests that listeners process speech signals differently when the phonetic structure of the target is dissimilar to the masker regardless of linguistic content. Lecumberri et al. (2006) showed that monolingual English listeners performed better on a speech perception of English consonants when the language of a competing speaker was Spanish rather than English. Van Engen and Bradlow (2007) demonstrated that for native English listeners, English sentence speech perception was better when the noise consisted of two-talker Mandarin Chinese babble than when it was composed of two-talker English babble. These results suggest that babble noise in a native language increases informational masking, decreasing behavioral speech recognition performance relative to an unknown foreign language.

The *linguistic similarity hypothesis* proposed by Brouwer et al. (2012) states that the more similar the target and masker speech language, the harder it is to segregate the two streams effectively. This definition is analogous to Durlach's definition of similarity of informational masking, but the linguistic similarity hypothesis further defines listener related factors, such as knowledge/experience with the target speech and the linguistic and phonetic content which ultimately affect behavioral speech recognition. For example, intelligible maskers with similar linguistic and phonetic content will be more detrimental to target speech recognition than unintelligible maskers (Van Engen and Bradlow 2007). Thus, the target-masker linguistic similarity hypothesis claims that a significant predictor of speech-on-speech recognition accuracy is target-masker similarity along linguistically defined dimensions. This hypothesis was tested by Calandruccio et al. (2013) using English, Dutch (linguistically dissimilar, but has similar permissible phonemes related to English), and Mandarin (linguistically and phonetically dissimilar to English) language maskers in combination with an English sentence recognition task. By using Dutch and Mandarin maskers, Calandruccio et al. (2013) with native English speakers, created a hierarchy of linguistic and phonetic effects defined by linguistic distance from English (Dutch similar in linguistic/phonetic content and Mandarin being dissimilar in linguistic/phonetic content). Linguistic distance was defined by differences in sound structure of the language at the level of the phoneme inventories, syllable- and phrase-level phonetic structures, and rhythmic structure (Calandruccio et al. 2013). Subsequently, they found that sentence recognition improved as the target-to-masker linguistic distance increased with the English-on-English masker yielding the poorest speech processing performance and Mandarin-on-English yielding the best speech performance of the study. Based on their results, they concluded that linguistic similarity operates on a continuum in the degree of linguistic distance from the target speech. In addition, Calandruccio et al. (2013) recognized the spectral and temporal differences between language maskers that could have contributed to masker effects and thus controlled for them by matching either to the long-term average speech spectra (LTASS) or to the temporal modulations of the individual language maskers. Spectral differences between the maskers accounted for some, but not all of the variation in behavioral performance between masker; however, temporal differences between language maskers were not significant. In summary, the behavioral results suggest that listeners experience a greater release from informational masking release when the target and masker speech are more linguistically and phonetically dissimilar and have greater deficits in performance when the target and masker are the most linguistically/phonetically similar, regardless of differences in spectral and temporal structure of the language. These findings highlight that the linguistic properties of speech on similar speech contributes to informational masking.

1.2.2 Stimulus Uncertainty

A second and related contribution to the amount of informational masking listeners experience is stimulus uncertainty. Uncertainty has been defined as the influence on listener's a-priori knowledge of the timing or content of the target speech and/or interfering speech has on their ability to understand the target (Brungart and Simpson 2004; Durlach, Mason, Shinn-Cunningham, et al. 2003; Watson, Kelly, et al. 1976; Watson et al. 2007). Psychophysical definitions of stimulus uncertainty have been based on the amount of trial-by-trial variability in the stimulus across a single or limited number of dimensions, such as its spatial location or timing relative to the background maskers (Dollezal et al. 2017; Durlach, Mason, Shinn-Cunningham, et al. 2003; Pollack 1975; Watson, Kelly, et al. 1976).

Behavioral informational masking studies have manipulated stimulus uncertainty using various paradigms such as dichotic listening tasks, varying spatial location, and varying the timing of stimuli (Freyman et al. 2001; Gallun et al. 2008; Kidd et al. 1998; Oxenham et al. 2003; Watson, Kelly, et al. 1976; Watson and Kidd 2007; Watson and Nichols 1976), many of which used non-speech paradigms. Uncertainty, in terms of informational masking was first examined through the work of Charles Watson and his colleagues using a novel experimental technique in which the discriminability of an alteration in some acoustic aspect of an element of a sequence of tones was measured as a function of uncertainty. In one of the first studies of stimulus uncertainty, Watson and Nichols (1976) asked listeners to detect a change in target tone frequency or intensity with trial-by-trial variation in timing of the target. Listeners performed a target tone detection task in a background sequence of tones that was either fixed or randomly varying in frequency. They found that ability to detect target tones was significantly reduced when the target tone sequence was randomly varying (more uncertainty) than when the sequence was predictable (less uncertainty). Others have shown similar results using comparable psychophysical non-speech paradigms (Kidd et al. 1998; Neff 1995; Watson and Kidd 2007). In addition, previous research has shown that presenting a cue, or preview of the stimulus or masker prior to a detection task can reduce uncertainty. For example, Richards et al. (2004) showed that when listeners received a pre-trial cue to either the target tone or masker signal, they experienced a significant release from masking as compared to not receiving a cue. Overall, these studies show that when the target tone and/or masker presentation are *more predictable* over time to the listener, the amount of informational masking decreases.

Relatively few studies have examined the influence of stimulus uncertainty on speech-on-speech masking and these results have been mixed. For example, Brungart and Simpson (2004) varied the degree of uncertainty in an informational masking dichotic listening paradigm. They used a closed-set, forced-choice, speech identification task (Coordinate Response Measure, CRM) in which participants were asked to extract information from a target phrase (“Ready <call sign> go to <color> <number> now.”) that was presented in their right ear while ignoring masking phrases that were also color-number coordinates presented in the same ear, opposite ear or both ears. In comparison to a randomized CRM task varying talker and content of the masking phrase, Brungart and Simpson (2004) manipulated masker uncertainty by either freezing the masker talker, where the talker was always the same for all trials, or by freezing the masker content, where the content of the masking phrase was always the same for all trials. Task performance was not improved when uncertainty was decreased by freezing the talker of the masker phrase, in either the target ear or the contralateral ear. Freezing the content of the masker also did not improve performance when presented to the contralateral ear. It was only when the target phrase and masker phrase were in the same ear and the content of the masker phrase was fixed (predictable) that performance improved by ~20%. That is, only uncertainty in the semantic content of the masker phrase in the same ear as the target speech had an effect on performance.

Freyman et al. (2007) also investigated the role of masker uncertainty but used open set nonsense sentences (syntactically but not semantically correct) for both target and maskers. The target sentences were a single female talker and the masker sentences were recorded by 10 different female talkers, combined into five two-talker pairs. Participants' sentence recognition performance was measured in each of these five masker conditions by in both a non-spatial condition (target and masker presented from a front loudspeaker), and under a spatially separated condition (target and masker loudspeakers separated by 60°). Listeners in their study showed larger variation in target recognition performance across the five different maskers in the non-spatial condition compared to the spatial. These differences in performance between the spatial and non-spatial conditions for each masker were interpreted as due to informational masking because in the spatial condition the target and masker were more easily separated and distinguished due to decreased uncertainty of masker spatial separation. In a second experiment, they increased or decreased masker uncertainty from trial to trial by manipulating masker content, masker talkers, and SNR in fixed and random conditions. Results showed very little effect of masker uncertainty in the spatially separated condition, and surprisingly only a small influence of uncertainty in the non-spatial condition. Recognition of the target did not improve when uncertainty was reduced by the participants knowing the masker content was the same for each trial.

While both experiments outlined above manipulated masker uncertainty, a-priori knowledge of the target was not manipulated as each experiment utilized a predictable source designation (e.g. the callsign for the CRM test and the "Ready" for the nonsense sentences) that was constant for the entirety of each study. This predictability of the target source may have been sufficient to overcome the uncertainty caused by the masker variation in both studies. In another study using the CRM task, Kidd et al. (2005) varied uncertainty and task complexity by manipulating the probability of the location of the target sentence among the three speaker locations, and by either presenting the callsign before or after the color-number coordinates. Kidd et al. (2005) presented three sentences simultaneously on each trial from three different loudspeakers (0° and ± 60°). One of the sentences was designated the target by telling the listener the callsign to listen for, and the other two sentences with different callsigns were considered maskers. Their results showed that when the listener did not know where the target sentence would be located and when the callsign was after the target color/number coordinate, performance was poor. Performance improved significantly when the callsign was before the target and when the listener was provided probability about the expected location of the target. While this study and the previously reviewed studies have varied spatial location of target and masker, spatial segregation itself decreases difficulty in separating target and masker. By increasing the listener's uncertainty about where the target may be located on a given trial, and particularly by limiting the predictability of the target sentence/content until after the coordinates are presented, informational masking and thus task difficulty is increased. Under conditions of decreased predictability of the target, more interference from the distractor sentences is likely.

As reviewed, the effects of stimulus uncertainty in informational masking have only been examined by a few studies using speech targets and speech maskers. In general, a-priori knowledge about when or where the target and/or maskers will occur, can reduce the interference of other distracting talkers. Conversely, if the target speech from trail to trial is less predictable, making the task more challenging, behavioral performance in informational masking tasks can be greatly reduced. As discussed in the next section, early cortical processing of acoustic features, as well as later auditory cognitive processing of target-masker uncertainty are likely involved in the listener's ability to separate out target and masker in challenging speech-in-noise situations.

1.2.2 Levels of processing and target-masker segregation

Understanding the relative influence of informational vs. energetic masking is important in the context of complex listening environments, where the challenge for speech understanding is not discriminating specific speech sounds but focusing on and understanding a specific speaker in a background of one or more competing talkers. As described by Brungart et al. (2001), "Higher-level informational masking occurs when the signal and masker are both audible, but the listener is unable to disentangle the elements of the target signal from a similar sounding distracter." It is this ability to separate or group sound sources, and to appropriately focus on one that is affected by informational but not energetic masking. While challenging for all listeners, understanding a target speech stream in noise is more difficult for certain populations, such as older adults, those with hearing loss, and those with acquired and developmental central auditory dysfunction (Bertoli et al. 2005; Desjardins et al. 2013; Helfer and Freyman 2008; Putter-Katz et al. 2008; Russo et al. 2008). Therefore, understanding more about the influence of informational masking on the ability to separate out talkers has important clinical applications.

In many natural settings, humans are able to perceive and attend to a specific person talking even in a complex background of multiple sound sources such as other voices and music (i.e. cocktail party effect). That is, a listener is able to separate a distinct auditory stream from other auditory objects. Both bottom-up and top-down processing influence the listener's streaming or grouping ability. Bottom-up factors, such as the physical/acoustic attributes of the sounds are processed at early peripheral and binaural sensory processing stages, for example, voice fundamental frequencies or spatial location of the talker based on inter-aural timing and intensity differences. Top-down factors may include linguistic content and familiarity of the timing of the target and masker speech, as well as the listener's memory and attentional capacity.

In the context of speech-on-speech competition, informational masking challenges the listener's ability to separate and attend to a target speech stream because multi-talker babble contains linguistic information. One way this is demonstrated is by evidence from behavioral studies that informational masking effects are highest (i.e. decreased speech understanding performance) when the masker is composed of 2-4 background talkers and decrease as the number of talkers increases beyond four (Carhart et al. 1975; Hall et al. 2002; Rosen et al. 2013). Once the number of talkers exceeds four, the temporal dips and amount of distinguishable words in the masker is decreased, thus resulting in less intelligible babble with little significant linguistic content (Hall et al. 2002). When listening to 2-talker babble, the individual streams of each talker are intelligible with relevant linguistic information that may easily divide auditory attention (Hoen et al. 2007). For example, Brungart et al. (2001) asked listeners to repeat a color and number spoken by a specific target talker and ignore a second background talker simultaneously saying color and number pairs. They found that incorrect responses made by the listeners were more likely to be the color-number pair from the masker speech than they were to be unrelated color-number pairs not spoken by either target or masker (or a combination of the target and masker pairs). Even when the listener is confident of which object is the target, auditory object selection may fail when a competing object is inherently more salient than the target due to any multitude of factors: SNR, spatial separation, linguistic or phonetic similarity, or even clarity of speech (Calandruccio et al. 2010; Freyman et al. 2001; Fritz et al. 2007). In these cases, the top-down bias of attention may be insufficient to override bottom-up salience and win the biased competition of that specific auditory scene.

How the specific effects of target-masker similarity and uncertainty in combination, as well as individually contribute to informational masking and the ability to separate target from masker streams is not fully known. This is especially true for some of the higher-level influences of the predictability of target and masker speech when considering uncertainty and the influence of the linguistic information in the target and masker when considering similarity. For example, linguistic similarity between target and masker may increase early-attentive salience of the simultaneous auditory sources. It is also possible that acoustic timing cues may serve to enable listeners to focus auditory attention, which can then impact perceptual segregation via top-down mechanisms. These higher-level factors beyond the acoustic features and spatial location of the sound sources are likely to separately and in combination challenge the listener's ability to attend to the target and ignore the interfering masker. There is evidence that even the more complex processing required when lexical, syntactic or semantic information is present can take place early as well as at later stages of auditory attention (see Bronkhorst (2000) review). There is a need for additional research to help clarify how similarity and uncertainty contribute to informational masking, where in the stages of processing these factors might influence speech recognition, and to help explain why some individuals find listening in noise particularly problematic (e.g. those with hearing impairment, older adults, auditory processing disorder, etc.). One approach toward better understanding of these issues is to use neural responses in combination with behavioral performance to help identify the time course and stages of processing where disruption in encoding the target speech in the presence of background noise may occur. Relating differences in neural activity within and across individuals in response to specific manipulations of target-masker similarity and uncertainty at cortical levels and relating these to listener performance is an important step towards understanding the informational masking phenomenon.

1.3 Neural Correlates of Informational Masking Effects

While behavioral studies have confirmed several factors related to target-masker similarity and uncertainty contributing to informational masking, the effect of these factors on underlying neuro-physiologic processes are not yet established. Neural correlates of informational masking can provide evidence to aid in further understanding the timing and level of the auditory pathway where masking interferes with speech understanding, and how this varies based on specific target and masker factors. Auditory evoked potentials (AEPs) have been used to explore many aspects of speech understanding, including speech-in-noise processing providing sensitive measures of timing and activity of postsynaptic potentials within the auditory pathway (Davis et al. 1939; Hillyard et al. 1971; Martin et al. 2008; Naatanen et al. 1987). AEPs provide information regarding the timing (latency) and salience (amplitude) of sound processing, and also a general cortical activation map (scalp topography) (Martin et al. 2008; Stapells 2008). AEPs also provide information regarding the size of the neural population indexing processing (amplitude) and can be used to infer the difficulty of stimulus detection and discrimination. The underlying assumption is that speech perception is dependent on the neural processing of the frequency, amplitude, and timing cues contained within the speech signal. While recording with non-invasive electrodes, passive and active evoked responses can be measured and manipulated by the recording paradigm. While AEPs have helped examine how speech processing in the auditory system is affected by background noise at both early (Billings et al. 2011; Kaplan-Neeman et al. 2006; Niemczak et al. 2019; Whiting et al. 1998) and late (Bennett et al. 2012;

Koerner et al. 2017; Wong et al. 2008; Zhang et al. 2016) top-down auditory levels, no current research study has combined results of multiple levels of auditory cortical processing that represents perception of spoken words in the presence of informational masking with a behavioral paradigm. In order to fill this gap in knowledge, this study seeks to examine how target-masker similarity and stimulus uncertainty relate to the informational masking deficit in normal-hearing young adults using neural correlates of auditory processing.

There are a family of AEPs, from early to late potentials that can be elicited in response to auditory stimuli, and several methods to classify them. One method of classification divides AEP into two major categories: pre-attentive "exogenous" evoked potentials and cognitive "endogenous" evoked potentials (Naatanen and Picton 1987; Stapells 2008). Exogenous AEP responses are primarily elicited by the specific acoustic properties of the stimuli (Martin et al. 2008; Naatanen et al. 1992). This type of AEP reflects activation of the auditory pathways, from the cochlea to the cortex. They occur as early as stimulus onset and extend to as late as 250ms post stimulus (Crowley et al. 2004; Naatanen and Picton 1987). In general, these AEPs are often called exogenous because they are thought to represent obligatory auditory processing occurring *before* conscious auditory processing (Naatanen and Picton 1987). In comparison, cognitive "endogenous" evoked potentials are measured in response to active conscious auditory processing that occur later than 250ms post stimulus onset. Cognitive AEP responses represent a transitional component between early attentive auditory cortical processing and the behavioral response. However, Stapells (2008) and other auditory electrophysiology researchers have highlighted that most cortical AEPs reflect a combination of exogenous and endogenous dynamics that can be affected by acoustic characteristics of stimuli in addition to top-down factors such as arousal state and attention. Thus, the reason why "before" is italicized above is because the N1 component is affected by attention, but not driven by conscious active auditory processing. For example, the N1 component of the cortical auditory evoked potential, which is of interest in the current study, is primarily considered to be exogenous, but is nonetheless affected by processes such as attention (Billings et al. 2011; Holcomb 1988; Woldorff et al. 1993; Zhang et al. 2016). Thus, the classification from exogenous to endogenous is a spectrum rather than a dichotomy. For the purpose of this project, two levels of auditory processing were examined that focus on *early* and *late* auditory top-down cortical components.

1.3.1 Early Cortical Processing Component

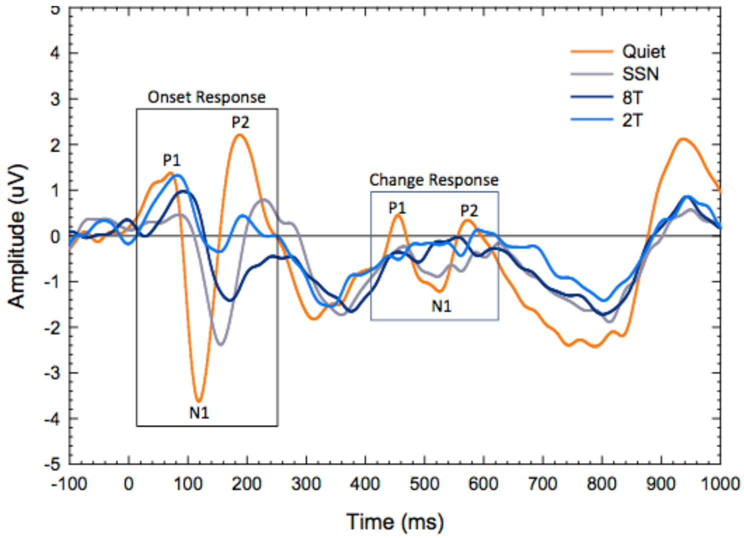
At the early level of cortical processing, the cortical auditory evoked potential (CAEP) consists of a complex of three peaks, the P1-N1-P2, can be recorded without active participation of the listener and is a transient response evoked by an acoustic change, typically from silence to sound onset (Martin et al. 2008; Naatanen and Picton 1987; Ostroff et al. 1998; Sharma et al. 2014; Whiting et al. 1998). The P1-N1-P2 generally reflects synchronous neural activity of structures in the thalamo-cortical connections to the central auditory system (Naatanen and Picton 1987) and reflects encoding of sound that underlies perceptual events. Particularly, the N1 component is present when sounds such as tones or speech are audible, but are not necessarily discriminated from other sounds (Osterhammel et al. 1973; Stapells 2008). The N1 occurs as an obligatory response in a passive listening condition, reflecting primarily sensory processing of stimulus features up to the level of the cortex. However, the response is influenced by early attention processes (Billings et al. 2013; Naatanen and Picton 1987; Pereira et al. 2014; Stapells 2008; Zhang et al. 2016). As described by Naatanen et al. (1992), the N1 wave of the CAEP has at least three generators giving rise to this cortical component. One generator of the N1 wave is thought to be most sensitive to acoustic changes within the auditory environment, the second and third generators of the N1 wave are thought to represent processes of early attention and an auditory orienting response (Naatanen et al. 1992; Roth et al. 1976). The amplitude of N1 is decreased under conditions of drowsiness and appears to be enhanced with overt attention to the stimulus (Hillyard et al. 1971; Naatanen and Picton 1987; Squires et al. 1973). P1-N1-P2, therefore, provides a method to examine the influence of informational masking on neural encoding that includes early acoustic and attentional auditory top-down processing, but prior to conscious auditory discrimination, memory, or decision making.

Although several studies have used speech syllables to elicit the P1-N1-P2, few have used naturally spoken words, which have both phonetic and linguistic structure. More complex speech stimuli with consonant and vowel changes within the stimulus can elicit multiple overlapping P1–N1–P2 responses, resulting in distinct morphologies for different phoneme sequences (Martin et al. 1999, 2000; Ostroff et al. 1998). Wagner et al. (2016) demonstrated that P1-N1-P2 morphology to spoken words that approximate the natural variability of a single speaker, reflecting the spectral and temporal features within the words. Importantly, these studies and others have demonstrated that the P1-N1-P2 waveform patterns to complex speech stimuli are sufficiently robust and reliable for use in research on the encoding of acoustic speech features from an early level of processing of speech perception in noise (Martin et al. 2008; Parbery-Clark et al. 2012; Tremblay, Friesen, et al. 2003a; Tremblay et al. 2006).

Presenting target stimuli in competing background noise affects the morphology of P1-N1-P2 responses when compared to quiet conditions, specifically by decreasing amplitude and increasing latency of component peaks. For example, degraded N1 morphology to target tones and syllables has been reported in varying energetic noise conditions, such as white noise (Billings et al. 2009; Kaplan-Neeman et al. 2006). Martin and Stapells (2005) recorded AEPs in response to /ba/ and /da/ target speech stimuli in a background of continuous masking noises filtered at various low-pass cutoff frequencies along with broadband noise and found the greatest masking effects, (i.e. greatest changes in amplitude and latency) when the noise bands directly overlapped the frequency region containing the primary acoustic cues differentiating the target signals /ba/ from /da/. This result provided evidence of energetic masking effects, specifically overlapping spectral content on the P1-N1-P2. Furthermore, Billings et al. (2011) recorded P1-N1-P2 responses to both tonal (1000 Hz) and speech stimuli (/ba/) in three types of background noise conditions (i.e. interrupted, continuous, and 4-talker babble). Results showed decreased amplitudes and increased latencies specifically for the N1 component in all background noise conditions compared to the quiet condition for the speech stimulus. The informational masking 4-talker babble condition resulted in the longest latencies and smallest amplitudes of all three conditions. This is consistent with a differential effect of background noise that is speech compared to random continuous noise, and with the idea of informational masking effects at the N1 level of auditory processing. Studies such as these suggest that changes in the amplitude and latency of the P1-N1-P2 evoked to speech stimuli may be sensitive to informational masking in a background of competing speech.

In two recent studies conducted in our lab, P1-N1-P2 morphology has been shown to be sensitive to different types of background noise that vary in the amount of informational masking (Niemczak and Vander Werff 2019; Vander Werff et al. 2016). Similar to the previous studies, neither white noise and nor continuous speech-shaped noise reduced P1-N1-P2 amplitudes and prolonged latencies as much as multitalker speech babble noise (Vander Werff et al. 2011). In a follow-up study to examine whether temporal differences between speech babble noise and continuous noise might be responsible for these effects, envelope-modulated noises were created

Figure 1- P1-N1-P2 recorded to /u-i/ stimuli in four background masking conditions representing an increase in informational masking from quiet to energetic masking (SSN), then two levels of informational masking noise (8- and 2-talker).



to temporally match the babble with different numbers of talkers. (Vander Werff et al. 2016). Results showed that even though energetic maskers that were both temporally and spectrally matched, the envelope noises did not impact amplitudes and latencies as much as babble noise. In another study (Niemczak and Vander Werff 2019) we examined the effects of both two-talker (2T) and eight-talker (8T) babble compared to continuous speech-shaped noise (SSN) on the P1-N1-P2 recorded to more complex speech stimulus with a linguistic vowel change from /u/ to /i/. We hypothesized that due to informational masking, the speech maskers would have a greater effect on waveform morphology for both the onset and change responses, and that the number of talkers would also significantly affect amplitudes and latencies. As shown in Figure 1, trends in results supported our hypotheses, with a large reduction in N1 amplitude for the 8T and 2T maskers compared to SSN at the onset in particular. All noise conditions significantly reduced onset N1 and P2 amplitudes, onset N1-P2 peak to peak amplitudes, as well as both onset and change response area compared with quiet conditions. Further, all amplitude and area measures were significantly reduced for the two babble conditions compared with continuous SSN, which is consistent with informational masking. However, the differences between 2T and 8T didn't reach statistical significance which may be due to the small size of the response for this particular stimulus paradigm.

This lack of difference by number of talkers, which is known to affect informational masking, between babble maskers in previous behavioral studies could have been due to the acoustics of the stimuli, or decreased sensitivity of the P1-N1-P2 to relative levels of informational masking. Although the P1-N1-P2 is generally considered pre-attentive, Billings et al. (2011) found that, compared to an energetic masker, a four-talker speech masker caused a larger N1 masking effect for a spoken syllable (/ba/) when listeners' attention was drawn away from the acoustic signals (passive paradigm), but not when listeners paid attention to the acoustic signals (active oddball paradigm). To further examine whether attention affected the P1–N1–P2 complex under masking conditions, Billings et al. (2011) collapsed the AEP waveforms across the three masking conditions (continuous energetic, interrupted noise, 4-talker speech) and found that the N1 amplitude was significantly larger under the active paradigm compared to the passive paradigm, indicating a facilitating effect of attention on the N1 component. There is evidence, therefore, that some aspects of informational masking have influence on these earlier processing components. Because the P1-N1-P2 represents both the encoding of stimulus features and obligatory processing, but is also sensitive to some aspects of at least early attention, this response may provide important information about whether similarity and uncertainty have an influence on this earlier stage of processing

1.3.2 Late Cortical Processing Component

Later AEP components following the P2 have been associated with active, attention-dependent, top-down cognitive processing of auditory stimuli, beyond the physical properties of the stimulus (Stapells 2008). Of these, the auditory P300 (P3) has been used extensively to study conscious processing of auditory stimuli. The P3 occurs approximately 300ms post-stimulus onset, representing the first transitional component between auditory cortical processing and behavioral responses. The P3 has been utilized in various modality paradigms, but in terms of acoustic stimulation, the P3 indexes conscious acoustic discrimination, being present when audible speech sounds are discriminable, but only when the subject is actively attending (Katayama et al. 1996; Lew et al. 1993; Martin et al. 2008; Picton 1992; Polich 2007; Polich et al. 1994). The P3 potential is generally thought to consist of two subcomponents, P3b and P3a, the presence of which vary depending on the evoking paradigm. The earlier fronto-central P3a, is primarily elicited by novel task irrelevant stimuli in a paradigm, and arises from variation in fast-acting attention mechanisms engaged to evaluate incoming stimuli. The late, more parietal P3b is thought to be proportional to the amount of attentional resources engaged in discriminating a given auditory stimulus and its peak latency is related to auditory evaluation time of that stimulus (Kutas et al. 1977b; Polich and McIsaac 1994; Schochat et al. 2012). For the paradigm used in this project, the P3b, which indexes active stimulus evaluation related to auditory context updating operations, is the main elicited sub-component and is referred to simply as the P3.

The P3 can be elicited by various auditory stimuli, and studies provide evidence of involvement of distinct mechanisms in the processing of speech targets compared to tones in an active attentional paradigm. In addition to reported differences in latency and amplitude, P3 to speech targets may also differ in the scalp distribution, generally indicating left hemisphere advantage for phonemes, syllables (Kayser et al. 2001) and words (Henkin et al. 2002; Novick et al. 1985).

Significant reductions in P3 amplitude and increases in latency have been demonstrated when target stimuli are presented in background noise, (Bennett et al. 2012; Kaplan-Neeman et al. 2006; Koerner et al. 2017; Polich 2007). Kaplan-Neeman et al. (2006) found that white noise resulted in prolonged latency for both the P3 and N1 components under various SNR conditions, but that P3 showed longer latency increases than N1 as the SNR became less favorable (Kaplan-Neeman et al. 2006). They attributed this finding to noise affecting the later stages of top-down processing involving discrimination and speed of conscious processing more than the initial stages of early attentive stimulus detection. Studies have also shown that masking noises that are presumed to contain informational masking, such as speech babble, increased P3 latency compared to purely energetic noise maskers (Bennett et al. 2012; Kaplan-Neeman et al. 2006; Krishnamurti 2001). Bennett et al. (2012) found that 4-talker babble resulted in prolonged P3 latencies compared to speech-shaped energetic noise conditions in the normal-hearing adult listeners. Bennett et al. (2012) also found that better behavioral sentence intelligibility scores were correlated with decreased P3 peak latency for a phonemic target contrast (/ba/ vs. /da/) and that slower behavioral reaction times were correlated with prolonged P3 latencies. Thus, the P3 response for a phonetic discrimination task appears to be a potential neurophysiological marker for speech-in-noise perception.

Another benefit of the P3 is the ability to manipulate task complexity and active top-down listener factors such as attention allocation and working memory during an auditory discrimination task. The classic two-stimulus oddball paradigm is a common experimental design to evoke the P3 and has proven particularly useful for investigating timing of stimulus-evaluation processes. In this paradigm, rare target stimuli (occurrence ~20%) are inserted in series of much more frequent standard stimuli (occurrence ~80%) of the same modality (Polich et al. 1988). The task given to the subject is to perceive the target stimulus and to react to it, typically by pressing a button, or just by mental counting. P3 responses with a similar topography can also be generated in a single stimulus task where a single target is randomly presented as in the oddball paradigm, but with the standard stimuli replaced by silence (Polich

2007; Wronka et al. 2008). The P3 appears to reflect processing time of attending to and discriminating auditory stimuli and the updating of working memory with sequential presentations of stimuli (Polich 2007; Steiner et al. 2014). A three-stimulus variant is an example of a more complex paradigm that can elicit the P3, in which an additional infrequent non-target stimulus is inserted into a sequence of rare target and frequent standard stimuli (Katayama et al. 1998, 1999). The resulting P3 wave is elicited by both target and non-target stimuli, however the waveform for the stimulus which the subject is instructed to respond (target) is typically larger in amplitude as the subject is actively listening for and attending to the target stimulus (Polich 2007). Using this more complex stimulus paradigm can provide information about listener attention allocation and effects of task complexity.

Another way to vary task complexity relevant is to manipulate the trial-to-trial variability in the presentation of the target and/or non-target stimuli. In a three-stimulus auditory P3 paradigm, this has been referred to as varying the target-to-target interval (TTI) (Gonsalvez et al. 1995). In general, P3 studies have shown an inverse relationship between target probability and P3 amplitude under a wide range of experimental conditions, with optimal infrequent target probability around 20% of total stimulus presentations (Duncanjohnson et al. 1977; Kutas et al. 1977a; Sqires et al. 1977). Probability can also be considered in terms of not only overall rate of occurrence but in time between evoking stimuli. For example, P3 amplitude increases linearly when targets (T) follow a longer rather than a shorter string of non-targets (N) (Gonsalvez et al. 2007). The effects of this stimulus presentation structure have been termed "sequence effects," and several studies have demonstrated that P3 amplitude to targets is significantly affected by sequential structure, such as match-versus-mismatch (TT vs. NT) and repetitions-versus-alternation sequences (TTTT vs. NTNT) (Johnson et al. 1982a, 1982b; Squires et al. 1977). TTI is defined as the time between a given target and the preceding target and is manipulated by varying the number/timing of non-targets between target presentations. TTI provides a direct way to manipulate stimulus uncertainty in an informational masking paradigm. For example, for a fixed TTI, uncertainty would be decreased as listeners would recognize the pattern in which target stimuli occur, with little trial-to-trial variability of the target speech occurrence relative to the background speech noise. However, for a random TTI, trial-to-trial variability and therefore stimulus uncertainty is increased. Listeners would presumably need to invest increased auditory attention to discriminate targets from non-target words and separate these from the distracting background maskers, which may be varied in terms of target-masker similarity. Currently, the P3 auditory evoked potential has not been studied by means of linguistic and phonetic similarity. The distinctive effect of linguistic similarity and stimulus uncertainty effects on the P3 may provide insight into how active attentional neural processing effects of sound discrimination in the presence of informational masking.

Furthermore, studies have used AEPs to better understand the difficulties of speech-in-noise processing and shown that they are sensitive to several aspects of informational masking. Few studies have examined two levels of multiple levels auditory neural processing by means of an informational masking paradigm. In addition, no studies to date utilizing AEPs have directly manipulated similarity in terms of the linguistic content of the maskers vs. targets, as well as stimulus uncertainty in a complex auditory task. Results of this study will aid in understanding how earlier and later stages of auditory neural encoding are related to the processes of informational masking effects.

1.4 Specific Aims and Secondary Objectives

While informational masking research has identified factors of both similarity and uncertainty that contribute to reduced behavioral performance in noise, critical gaps in knowledge regarding the informational masking phenomenon, including the underlying neural-perceptual processes remain. In order to further understand informational masking deficits and provide converging evidence with behavioral findings, the current project examined neural correlates of informational masking by systematically manipulating both target-masker similarity and stimulus uncertainty in an auditory evoked potential (AEP) paradigm. The goals of this study were to provide objective evidence of how both increased target-masker similarity and stimulus uncertainty increase the amount of informational masking experienced by the listener and examine two levels of top-down auditory processing using temporally sensitive auditory electrophysiology measures.

By their definitional structure these aspects of informational masking are not completely dichotomous. However, by systematically manipulating aspects of similarity and uncertainty, the current study proposed to examine the time course and objectively quantify these informational masking effects at both early and late stages of auditory processing using auditory evoked potentials in a two-factor repeated measures paradigm. Specifically, linguistic similarity between target and masker were manipulated by varying language of the talkers in the noise maskers. Stimulus uncertainty was manipulated by task complexity, specifically target-to-target interval (TTI). This design allowed for simultaneous recording of multiple AEP peaks, including analysis of amplitude, area, and latency characteristics, which were used to determine the relative influence of target-masker similarity and stimulus uncertainty, known to affect behavioral speech recognition performance, on neural indices of both early (focusing on the N1) and late (focusing on the P3) levels of top-down auditory processing. Finally, AEP measurers were compared to performance on a behavioral speech-in-noise task to further explore if AEP components of amplitude/area and latency are correlated to behavioral outcomes across proposed maskers.

***Specific Aim I*: To establish AEP correlates of increased informational masking through manipulations of 1) linguistic similarity between target and masker, 2) uncertainty in target speech timing in the presence of these maskers, and 3) interaction between target-masker similarity and target uncertainty.** To accomplish this aim, an oddball paradigm was used to evoke AEP components to target speech consisting of spoken words in English (listener's native language) in the presence of background maskers.

Objective 1: Objectively quantify the effects of linguistic similarity on neural indices of cortical auditory processing.

Target-masker similarity was manipulated through the use of multi-talker maskers (2-talker) using three different languages and a quiet control condition. AEP responses were recorded to the target stimuli, which will consist of spoken words in English, the native language of the listener (discussed in objective 2). Masking conditions will set up a hierarchy of target-masker similarity from low to high using *Mandarin*, *Dutch*, and *English* 4-talker maskers.

Table 1 - Proposed linguistic/phonetic hierarchy

Condition	*Target-masker similarity*	*Masker*
1	None	No masker (quiet condition)
2	Low	Mandarin – linguistically and phonetically dissimilar
3	Medium	Dutch – linguistically dissimilar, but phonetically similar
4	High	English – linguistically and phonetically the same

Based on existing behavioral and electrophysiological literature, the hypothesis was that the English babble masker with the highest linguistic similarity to the target English words would result in the largest reduction in amplitude/area and increase in latency for the N1 and P3 AEP components. The Mandarin masker was expected to provide the least amount of informational masking because of the largest linguistic differences to the target English words, and therefore have the smallest effect on amplitudes/areas and latencies. The Dutch babble condition was hypothesized to yield shorter latencies and larger amplitudes compared to Mandarin, but smaller amplitudes compared to the English masker due to a similar linguistic structure and similar permissible phonemes to the English language.

Because the P3 component reflects active top-down stages of processing that are more affected by active discrimination, the effects of target-masker similarity were expected to be most prominent for this component. Based on previous research, however, including previous studies in our lab (Figure 1.), we expected informational masking to also have an effect at the earlier level on the N1 component, but robust effects were expected at the P3 level of auditory processing.

Objective 2: Objectively quantify the effects of stimulus uncertainty on neural indices of cortical auditory processing.

Stimulus uncertainty were manipulated by systematically varying the expected timing of target presentation, specifically by varying target-to-target interval (TTI) in an oddball paradigm. TTI will either be *fixed* or *random*, thus increasing task difficulty in discriminating target vs. non-target stimuli in the presence of each multi-talker masker described in Objective 1. The two levels of experimental task were as follows:

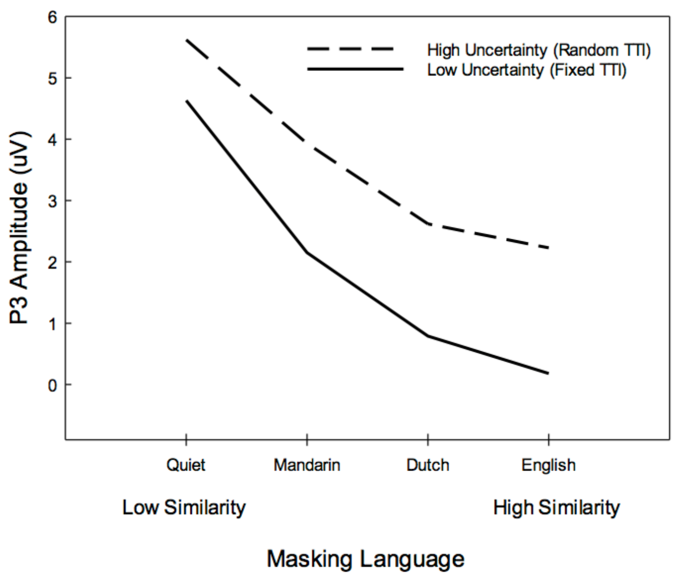
Experimental Task	*Stimulus Uncertainty*	*Presentation Paradigm**	*TTI*
1	Low	Target /pæt/, Standard /bæt/	Fixed 6 seconds
2	High	Target /pæt/, Standard /bæt/	Random 2-10 seconds

Table 2 - *Targets and Standards were switched systematically such that half of the subjects receive /pæt/ as the target and /bæt/ as the target to control for one-way stimulus detection.

We hypothesized, based on previous literature, that because informational masking is increased when stimulus uncertainty relative to the masker is increased, AEP amplitudes/areas would be increased, and latencies prolonged as uncertainty in timing of the target stimulus is increased from low to high due to random variation in TTI. Amplitudes/areas, at least for the P3, were expected to increase when stimulus uncertainty is high due to increased attention required for the task. It was expected that the effects on the earlier N1 would be more complex, but as active attention is known to have some effect, similar but less robust effects were also expected on this peak due to later top-down effects.

Objective 3: Determine the extent of interaction between target-masker similarity and target uncertainty on neural indices of cortical auditory processing

Figure 2 - Proposed masking/task interaction for P3 amplitude. High uncertainty and low uncertainty task conditions are indicated in dashed and solid lines respectively. An additive effect of uncertainty and similarity is hypothesized to occur at the highest level of similarity (English masking).



The third objective was to determine the interaction of stimulus uncertainty and target-masker similarity on AEP outcomes resulting in a comprehensive examination of two informational masking factors. A significant interaction term would indicate that the similarity and uncertainty contributions to informational masking are additive. If increasing the uncertainty specifically increased the cognitive demand of the task, then the listener would be likely to have more difficulty under conditions where the target and masker are the most similar (i.e. the most challenging to separate). Because the P1-N1-P2 is early attentive level of top-down processing, no significant interaction effect between similarity and uncertainty was expected for the amplitude or latency of the N1 component. However, if a significant interaction was found, this would be consistent with previous work (e.g. Billings et al. (2011) suggesting that auditory attention does have an effect at this level of processing for speech-in-noise tasks.

A hypothesized interaction effect for P3 amplitude is shown in Figure 2, in which increasing target-masker similarity is shown from left to right on the x-axis and uncertainty conditions are shown by the solid (fixed TTI, low uncertainty) and dashed (random TTI, high uncertainty) lines. The additive effect of target uncertainty and target-masker similarity is seen in the separation of the two lines as target-masker similarity increases. We hypothesized that increased uncertainty would modulate active top-down auditory neural processing, such that overall P3 amplitudes increased relative to the low uncertainty conditions, more so for (the most difficult) highest target-masker similarity conditions.

***Specific Aim II*: To characterize the relationship between physiological AEP outcomes of informational masking and behavioral performance on a task of speech recognition in noise.** To accomplish this aim, behavioral word recognition was measured under the same linguistic masking conditions as Aim I. The underlying theory of this objective was that accurate speech recognition in noise is dependent on the neural encoding of the auditory stimulus. Therefore, we proposed to answer two questions: 1.) Are individual changes in AEP component amplitude and latency across masking conditions related to changes in speech recognition scores across masking conditions? 2.) Which AEP components better explain the variability in behavioral performance between masker conditions?

In order to answer the proposed questions, we measured reaction time, accuracy, and d-prime (d') performance during the AEP tasks and behavioral word recognition-in-noise scores for all masker conditions. We analyzed behavioral performance at multiple SNRs, including the same SNR as the AEP conditions as well as more and less favorable SNRs. We hypothesized that AEP outcome variables would be related to behavioral speech recognition-in-noise performance such that larger amplitudes and decreased latencies would be correlated with better speech recognition abilities, specifically on the P3 due to active top-down auditory processing.

2.0 Design and Methodology

2.1 Design

A factorial (4x2) repeated measures experimental design was implemented to examine the effects of target-masker similarity and stimulus uncertainty on AEP morphology. The independent variable of target-masker similarity consisted of four different background speech babble masker conditions including quiet, Mandarin, Dutch, and English. (Aim I - objective 1). The independent variable of stimulus uncertainty consisted of two levels of target-to target interval (TTI), fixed and random (Aim I - objective 2). Primary dependent variables included mean amplitude, area, and peak latency measures for the N1 and P3 components. Eight total AEP waveforms were recorded per person (4x2). Masking language and TTI conditions were randomized and counter-balanced across participants.

2.2 Methodology

2.2.1 Participants

In order to establish the relationships between target-masker similarity and stimulus uncertainty on neural indices of informational masking in the normal auditory system, participants for the current study were young adults with clinically normal audiometric thresholds. To sample this population, individuals between the ages of 18–30 years (Helfer and Freyman 2008; Polich et al. 1985) with no history of auditory pathology were eligible to participate in this study. The sample size was based on *a-priori* power analysis for two-way repeated measures ANOVA (RM-ANOVA) conducted using estimated effect sizes based on data from previous electrophysiology studies related to speech-in-noise and aspects of informational masking (Bennett et al. 2012; Billings et al. 2013; Koerner et al. 2017). Using a significance level of $\alpha = 0.05$, power level of 0.9, and medium effect size (eta squared (η^2) of 0.09 and Cohen's f^2 0.25), produced a required sample size of approximately 31 total participants. Participants were recruited from the general public from Syracuse University via approved flyers, and approved college email list-serves, including the SU-News. To minimize the confounding factor of peripheral hearing loss, participants were required to demonstrate clinically normal hearing, defined as bilateral pure tone air conduction thresholds ≤ 15 dB HL at octave frequencies from 0.25 - 8 kHz. Normal tympanometric compliance (226 Hz tympanogram, peak compliance ≥ .3 mmho) were required for inclusion. In order to further account for possible variability across individuals in peripheral hearing that could confound results, both distortion-product otoacoustic emissions (DPOAEs: Mimosa Acoustics HearID Software: f2/f1 ratio = 1.22, 55/65 dB SPL, 1.0-6.0 kHz) and extended high frequency audiometric thresholds (9 – 16 kHz) were measured for all participants. Neither measure was exclusionary.

Exclusionary criteria included non-native English speakers, extended exposure to Dutch or Mandarin languages (dual citizenship, study abroad, etc. see appendix for screening survey), left handedness (Hoffman et al. 1999; Polich et al. 1998), regular nicotine smokers (Friedman et al. 1980; Polich et al. 2004), history of head injury/concussion (Hall et al. 1982; Rugg et al. 1993; Vander Werff et al. 2019), history of diagnosed learning, speech/language (Ferguson et al. 2011), psychiatric or neurological disorders, specifically schizophrenia and bipolar disorder (Blackwood et al. 1987; Souza et al. 1995). Additionally, participants were excluded if they have had ≥1 year of professional musical training to rule out any enhancement of speech-in-noise processing (Anderson et al. 2013; Oxenham et al. 2003; Parbery-Clark et al. 2012). Use of categories of medications (e.g. benzodiazepines, prescription sedatives, anticholinergics, antipsychotics) known to affect AEP responses were also exclusionary (Polich et al. 1995; Qidwai et al. 2002). Handedness was assessed in the screening survey using questions from the Edinburgh handedness inventory (Veale 2014).

Measures of working memory and attention were also implemented as a control for cognitive functioning using two subtests of the NIH Toolbox Cognition Battery for the Assessment of Neurological and Behavioral Function (Weintraub et al. 2013; Weintraub et al. 2014). Specifically, working memory was be measured by the List Sorting Task and attention scores on the Flanker Attentional Task. Both tests were administered on an iPad. The List Sorting Task is a series of stimuli presented on the iPad screen visually (object) and orally (spoken name), one at a time. Participants are instructed to repeat the stimuli to the examiner in order of size, from smallest to largest. In one condition, all stimuli come from 1 category. In the second, stimuli are presented from two categories, following which the participant must report first all stimuli from the first category, then from the other, in order of size within each. The number of items in each series increases from one trial to the next and the test is discontinued when two trials of the same length are failed. The List Sorting task takes approximately 7 minutes to administer (Weintraub et al. 2014). Test scores consist of total items correct across all trials.

The Flanker Attentional Task, version of the Eriksen Flanker Task, tests the ability to inhibit visual attention to irrelevant task dimensions. On each trial, a central directional target (using arrows) is flanked by similar stimuli on the left and right. The task is to indicate the direction of the central stimulus. On congruent trials, the flankers face the same direction as the target. On incongruent trials, they face the opposite direction. A scoring algorithm integrates accuracy and reaction time, yielding scores from 0 to 10. There are 40 trials and the average time to complete the task is 4 minutes (Weintraub et al. 2014).

Both instruments were validated in English, in a sample of 476 participants ranging in age from 3 to 85 years, with representation from both sexes, 3 racial/ethnic categories, and 3 levels of education (Weintraub et al. 2013). Both the List Sorting Task and Flanker Attentional Task have normative values accessible within the iPad app (Weintraub et al. 2014).

2.2.2 AEP Stimuli

Target and standard stimuli for all experimental tasks consisted of two English words spoken by a female native speaker of English. These experimental *target* and *standard* stimuli for all experimental tasks consisted of the consonant-vowel-consonant (CVC) English words /bæt/ (“bat”) and /pæt/ (“pat”). Recording of stimuli took place in a sound treated booth using a Senhiesser mke 600 microphone, Behringer umc22 amplifier, and Praat software (Boersma and Weenink, 2012). Praat software was utilized to create equal duration of 500ms for each word, as well as equal duration of plosive lead, formant transition, steady-state portion of the vowel, and plosive stop to minimize secondary acoustic length cues for identification. Amplitude of the stimuli were standardized by normalizing the burst, formant transition, and stead-state vowel amplitude in Audacity.

All stimuli were presented binaurally using Etymotic ER-3A insert earphones at 70 dB SPL in a double walled sound attenuated booth. This intensity was chosen due to its similarity to normal conversational speech levels and is also sufficiently loud enough to elicit a reliable AEP response. In order to control for acoustic differences between stimuli in a one-way oddball (/bæt/ as target and /pæt/ as standard), stimuli were be randomized such that half of recruited subjects received /bæt/ as the target with /pæt/ at the standard and the other half received /pæt/ as the target and /bæt/ as the standard.

Both /bæt/ and /pæt/ translate to /knuppel/ and /tijke/ respectively in Dutch. In addition, both /bæt/ and /pæt/ do not have an English translation to Dutch or Mandarin (i.e. those specific phonetic combinations are meaningless). The primary reason this word-word contrast was chosen was due to voice onset time (VOT), which is a relevant acoustic precept necessary for accurate speech perception (Oden et al. 1978). The /b/ and /p/ voiced and unvoiced bilabial stops, which represents a significant phonetic contrast that needs to be accurately coded in order to accurately perceive speech. In addition, this paradigm more directly connects with the masking hierarchy making interpretation clearer, such that the masker language can be interpreted in comparison to only English words (i.e. reduces the difference between interpretation a semantic effect between word and non-word.)

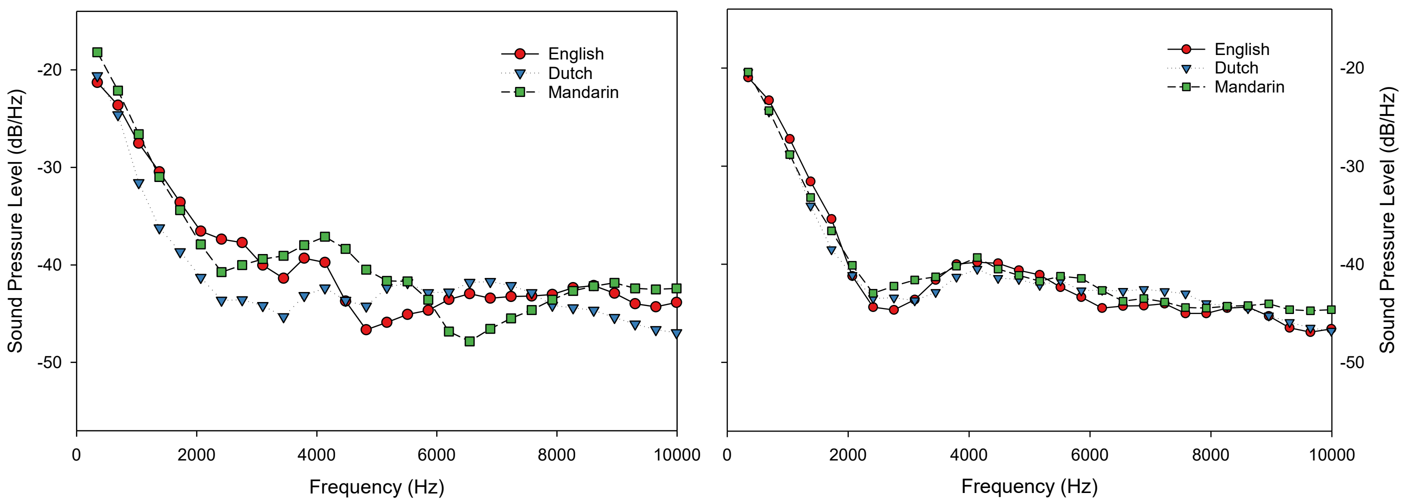
2.2.3 Masking Conditions

Following the masking hierarchy from Calandruccio et al. (2013), target and masker similarity were manipulated by using multi-talker babble maskers in three languages; one native to the listener, (English), and two foreign languages, (Dutch and Mandarin). A control quiet condition was also recorded (no masker present). Each babble masker was composed of 2-talkers spoken by female native speakers. Female target and maskers were chosen to eliminate sex differences between target and masker. 2-talker babble has been shown to provide the highest amount of informational masking (Brungart et al. 2001). The English masker consisted of syntactically correct, meaningful sentences spoken in English taken from the Harvard/Institute of Electrical and Electronics Engineers (IEEE) sentence lists (IEEE, 1969). An example of a sentence from these lists is, "A white silk jacket goes with any shoes." The Dutch sentences used during testing were direct translations of the IEEE sentences that are syntactically correct but semantically anomalous. An example of these sentences is, "The great car met the milk." An example of the same sentence translated into Dutch is "De geweldige auto ontmoette de melk." The Mandarin sentences, originally used in Van Engen and Bradlow (2007), are also syntactically correct but semantically anomalous materials. It should be noted that although the English competing sentences were meaningful and the Dutch and Mandarin competing sentences were semantically anomalous, all listeners were monolingual speakers of English and had no knowledge of either Dutch or Mandarin. Brouwer et al. (2012) reported data for monolingual English listeners in the presence of meaningful and anomalous Dutch maskers. Results indicated no significant differences between the masker conditions; therefore, we would expect that because the listeners in the present study were all monolingual English speakers, the fact that Dutch and Mandarin maskers were anomalous should not matter.

The two-talker maskers were created by concatenating sentences spoken by each talker with no silent intervals between sentences to eliminate the potential for glimpsing. Though each of the two talkers spoke the same sentences in each language, the order of concatenation differed between the talkers in each masker condition. The sentences were equalized to the same root-mean-square (RMS) pressure level using Praat prior to concatenation. The two strings of sentences were combined into a single audio file using Audacity. The final audio files (one for each masker condition) were RMS equalized to the same overall pressure.

In order to account for the spectral differences between maskers, we manipulated the long-term average speech spectrum (LTASS) of all two-talker babble tracks as a means of reducing unequal amounts of energetic masking between conditions. Figure 3 plots the LTASS of all three maskers before (left panel) and after (right panel) the normalization.

Figure 3 - Long term average speech spectrum (LTASS) for each language masker before (left panel) and after (right panel) normalization.



The left panel of figure 3 shows substantial spectral differences in the higher frequencies (6-8 kHz), specifically for the Mandarin masker. LTASS normalization eliminated these difference by adjusting each masker LTASS to match the average LTASS. This normalization procedure was implemented in Praat (scripting acquired from Dr. Susanne Brouwer and originally created by Dr. Chun Liang Chan) and involved first computing the LTASS separately for each masker speech wave file. The LTASS for a given wave file was then computed by breaking up the file into windows of 2048 samples. The fast Fourier transformation was then taken of each window and the mean was subsequently taken across all windows. After that, the average LTASS across all masker files was computed and each masker file LTASS was adjusted to the average LTASS. Following this manipulation, informal listening tests with native English and native Dutch listeners on the original and the spectrally-transformed sound files to ensure that the stimuli maintained their naturalness after signal processing. The results of these tests showed that normal-hearing listeners could not reliably distinguish between the original and normalized sound files. This was not surprising since the amount of spectral manipulation was relatively small.

To examine effects of whether the target stimuli are differently “glimpsed” across time for the various maskers (heard in the temporal gaps, affecting the amount of energetic masking), masking sentences were concatenated and ideal time-frequency segregation (ITFS) was implemented to quantify the spectral overlap of maskers and target stimuli across time. First, each 2-talker masker was created such that no sentence began or ended at that same time creating an interleaved pattern with no visible temporal gaps in each masking waveform. Second, ITF, a signal processing technique implemented through Matlab, was used to identify both temporal areas within a combined target-masker acoustic waveform where the target would be undetectable in the presence of the masker due to spectral overlap and temporal areas where the target would be detectable or glimpsed above the masker. The ITFS scripting that was used by Brungart and colleagues (Brungart et al. 2006, 2009) for separating energetic and informational masking components in psychophysical and speech perception tasks was implemented in this study and adapted from a Matlab toolbox created by Dr. DeLiang Wang. However, ITFS in this experiment was simply implemented to measure and compare the energetic masking effects and evaluate whether these affects are similar across the maskers and ensure that the relationship between the amount of masking and glimpsing of the targets are approximately equalized across maskers.

SNR for each target-masker condition was held constant at +3 dB, which has been shown to not only be representative of real world difficulty listeners experience, but produce reliable AEP morphology (Bennett et al. 2012; Koerner et al. 2017). Previous behavioral results have also showed that this SNR with a same sex target and masker, yields approximately 50-75% correct detection (Words-In-Noise test, and CRM task), limiting celling and floor effects (Brungart et al. 2001; Wilson, Carnell, et al. 2007).

2.2.4 Stimulus Paradigm/Listener Tasks

Target uncertainty was manipulated by the target stimulus paradigm, specifically the target-to-target interval (TTI), in two task conditions – fixed and random. Both tasks used a three-stimulus auditory paradigm, with AEPs recorded to naturally spoken words in each phonetic/linguistic masking condition, as well as quiet. As stated above, *target* (T) and *non-target standard* (N) stimuli were either /bæt/ or /pæt/ depending on subject randomization. In addition to target and standard words, periods of *quiet* (Q) (500ms duration, identical to word stimuli) were also presented as additional non-target standard stimuli to reduce habituation effects of perceiving between two CVC words (Kok 2001; Polich and McIsaac 1994). In this active paradigm, subjects attended to the stimuli and pressed buttons corresponding to either target or standard stimuli as quickly as possible while refraining from pressing a button during silent blocks.

Table 3- Depicts presentation sequence for fixed TTI (low uncertainty - top panel) and random TTI (high uncertainty - bottom panel). T stands for target, S standard, and Q quiet block.

Fixed TTI – Low Uncertainty

	5s						5s						5s				
T	S	Q	S	Q	S	T	Q	Q	S	Q	S	T	S	S	Q	S	Q
6s TTI						6s TTI						6s TTI					
Time																	

Random TTI – High Uncertainty

	5s						2s			8s							
T	S	Q	S	Q	S	T	Q	Q	T	Q	S	Q	S	S	Q	S	Q
6s TTI						3s TTI			9s TTI								
Time																	

For the fixed task, as shown in the top part of Figure 4, the TTI was held constant at 6 seconds. Probability of target words in the experiential paradigm was 20% as compared to 60% non-target standard words and 20% silent blocks (Gonsalvez et al. 1999; Gonsalvez et al. 2002; Katayama and Polich 1996). While the TTI was fixed at 6 seconds, the order of standard stimuli between targets was random and counterbalanced occurrence of silence and standards. A TTI of 6 seconds has been shown to result in measurable P3 morphology, and this pattern of results has been linked with working memory-updating processes (Steiner et al 2016). The interstimulus interval (ISI) between any N, T, and Q interval was 1100ms with a 100ms jitter from trial to trial. 150 target presentations were recorded for each noise condition for a total test time of approximately one hour per task (~12 min per noise condition – four noise conditions).

For the random task, as shown in the bottom part of Figure 4, the same three-stimuli and four masker conditions were used but with pseudo-randomization of TTI. TTI was randomized from 2-10 seconds (2, 4, 6, 8, 10). In order to control for TTI effects on AEP morphology, TTI was averaged to 6 seconds, similar to Task 1, to compare across tasks. Probability of target words, ISI, button press response, and noise randomization was identical to task 1. This manipulation of TTI provided a high level of uncertainty of target presentation, in that the participant was unsure of the timing of the target across trials under each noise condition.

2.2.5 *Electrophysiological Methodology*

Auditory neural responses were recorded using the Neuroscan Synamps[2] recording system using a 64-channel electrode cap. Responses were referenced to the mastoids in the online and offline analysis. Cap position from nasion to inion was measured for each subject to ensure consistent cap placement between testing sessions. Blink artifact rejection using a vertical eye channel was utilized. Inter-stimulus interval (ISI) from 1200-1000ms (jittered 100ms) was employed to ensure no overlapping of AEP activity and that temporal and spectral content of stimuli and noise for each condition was random. Responses were analyzed over a 100ms pre- and 1200ms post stimulus window. Within each task/noise condition, there was approximately 750 stimulus presentations (150 target, 450 non-target, and 150 quiet). Participants were seated in a comfortable chair within a double-walled sound-treated booth. Subjects were given an indicator with two buttons, in which they responded with a button press using two fingers of the same hand to each target and standard word stimulus to reduce motor artifact differences (Luck 2014a). The subjects were instructed to remain awake and attended, pressing the corresponding button to the stimuli as quickly as possible whenever they hear the stimulus. Prior to testing, a practice condition, in which all stimuli are played in quiet, was presented to acclimate all subjects to the experimental paradigm. Performance measures of reaction time, accuracy, and stimulus discrimination indexed by measures of d' were recorded simultaneously via Neuroscan software and button response box. Using signal detection theory (Macmillan, and Creelman, 1991) as shown in Table 4, the d' statistic was calculated as the difference between z-transforms for hit rate and false alarm rate (d' = z(H) - z(F)). Subjects were given a 2-5 minute break between conditions if needed.

	Response: Yes (button click)	Response: No (no button click)
Stimuli: Present	**Hit**	**Miss**
Stimuli: Absent	**False alarm**	**Correct rejection**

Table 4 - d' stimulus discrimination response example used for calculation.

Evoked responses were analyzed offline using Matlab ERPLab toolbox (Lopez-Calderon and Luck 2014), an open-source Matlab package for analyzing ERP data. Offline artifact rejection was performed using Matlab ERPLab moving-window technique with a window width of 200ms and step size of 100ms. Trials in which the amplitude exceeded the threshold value of 100μV were marked for rejection. After artifact rejection, remaining sweeps were averaged and filtered from 0.1-30 Hz for standard and 0.1-20 Hz for target responses (12 dB/octave). Response analysis focused on waveforms recorded from the three midline locations (F_z, C_z, P_z), but overall scalp distribution and global field power (GFP) were used to aid in component identification. Outcome measures included peak and mean amplitudes and local peak latencies within specified time windows of the primary peaks N1 and P3, as well as total rectified area of the N1 and P3 (Luck 2014b). All amplitude and latency measurements were made using the ERPLab measurement tool and verified by two judges. The judges were blinded to the conditions and used surrounding electrodes sites to verify peaks. Absolute peak amplitudes were calculated relative to baseline and absolute peak latencies were calculated relative to stimulus onset (0ms). Time windows for peak identification were based on Martin et al. (2008) and confirmed across all noise conditions as to not exclude any peaks due to increased latency, specified as N1: 50-

150ms, and P3: 250-550ms from stimulus onset. Absolute peak amplitude and peak latency were also measured to compare across conditions and to previous electrophysiological studies. In addition, exploratory analysis of scalp distribution of informational masking effects were conducted (e.g. left vs. right hemisphere, frontal vs. parietal).

2.2.6 *Behavioral Methodology*

Behavioral speech-in-noise performance was assessed using a word recognition task in quiet and each of the three language masker conditions described in Aim I. Northwestern University Auditory Test No. 6.4 (NU-6) monosyllabic word lists were the stimuli, chosen to provide an open-set test free of contextual cues and to match AEP stimuli (Tillman et al. 1963). Testing was administered under insert earphones and participants were be asked to repeat the last word following each carrier phrase ("Say the word…"). In addition to testing behavioral speech recognition at a fixed +3 dB SNR (identical to AEP conditions), SNR at which 50%-word recognition was also found using an adaptive procedure in order to better characterize individual differences in performance. Methodology was adapted from the Words-in-Noise (WIN) Test, which evaluates word recognition in multitalker babble at seven signal-to-noise ratios and uses the 50% correct point (in dB SNR) calculated with the Spearman-Kärber equation as the primary metric (Wilson and McArdle 2007). SNR-50 was calculated at 12, 8, 4, and 0 dB SNR, by taking accuracy of the first five words in each SNR (+3 dB SNR was left out of the equation due to mathematical constraints). The Spearman-Kärber equation which is expressed as:

$$T_{50\%} = i + \tfrac{1}{2}(d) - (d)\left(\frac{\#\text{ words correct}}{\#\text{ words per step}}\right)$$

in which *i* is the highest presentation level (12 dB SNR), *d* is the step size (4 dB steps), and five words were presented per step. This procedure is similar to the procedure used in the Words in Noise Test (WIN).

2.2.7 Statistical Analysis

There were six primary AEP outcomes analyzed as dependent variables as shown in Table 3: N1 amplitude, N1 latency, P1-N1-P2 response area, P3 amplitude, P3 latency, and P3 response area. A two-factor repeated measures analysis of variance (4x2 RM-ANOVA) was used to analyze the main effects of target-masker similarity (4 levels: quiet, Mandarin, Dutch, and English language maskers) and target-masker uncertainty (2 levels: fixed TTI and random TTI task) and the interaction of these main effects. The three N1 primary outcome variables were measured at C_z and the three P3 primary outcome variables were measured at P_z, resulting in total of six separate RM-ANOVAs (Table 3). Tests of normality and sphericity were conducted, with Bonferroni correction for multiple comparisons. Accuracy and reaction time to target stimuli obtained from button press responses was analyzed across masking and task conditions.

Behavioral word recognition in noise performance (% correct and SNR50) scores were used as outcome variables for Aim II. A 5 x 4 RM-ANOVA was used to evaluate the effects of SNR and masker similarity condition on behavioral performance outcomes. Stepwise linear regression and correlation analyses were conducted to assess relationship among behavioral and AEP outcomes for each language masker condition masking conditions. This statistical procedure was used to probe the relationship between AEP outcomes and behavioral responses to better understand what specific AEP components (or combination of components) are correlated to behavioral outcomes. Accuracy and reaction time to target stimuli obtained from button press responses, were also be analyzed across masking and task conditions and utilized in the stepwise linear regression.

Paired t-tests were used to test for significant differences between left and right ears for peripheral measures include pure tone audiometric thresholds (0.25-16 kHz) and DPOAE amplitudes.

Primary Outcome Variables

	Early N1 Response			Late P3 Response		
4x2 RM-ANOVA	N1 Mean Amplitude	N1 Peak Latency	N1 Rectified Area	P3 Mean Amplitude	P3 Peak Latency	P3 Rectified Area

Table 5 – Primary outcome variables for both early N1 and late P3 responses.

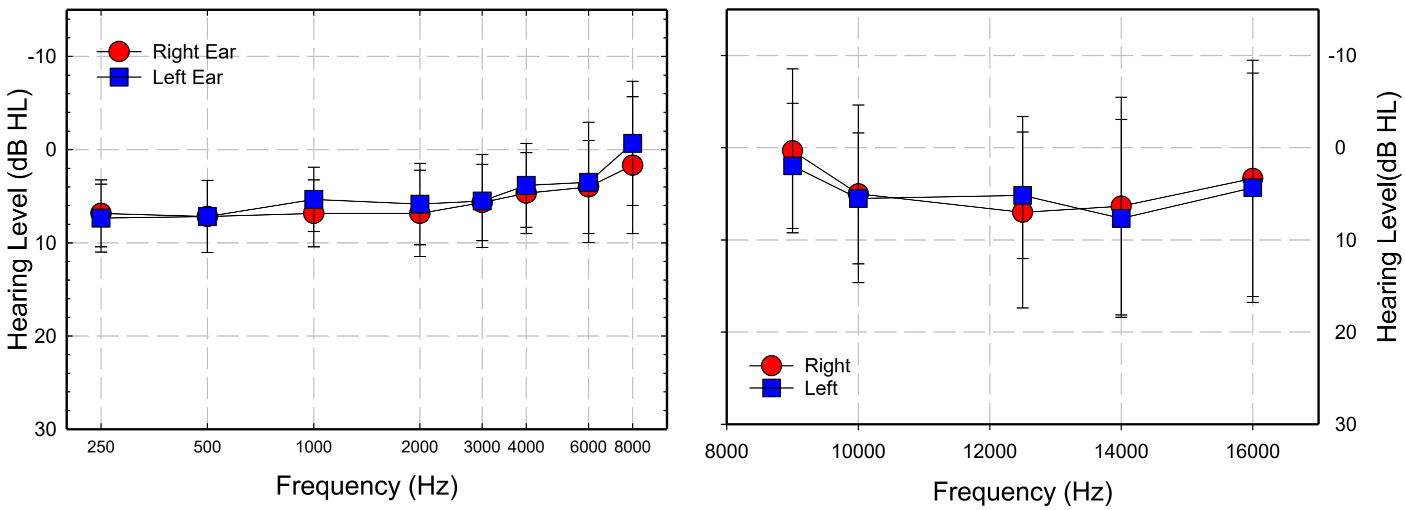
3.0 Results

3.1 Participant characteristics and peripheral hearing status

Thirty-four individuals consented and were enrolled in the study, but three individuals were excluded due to pure tone thresholds exceeding criteria and one subject withdrew due to the electrode preparation materials irritating the scalp. Therefore, a total of 30 individuals completed the study, 25 female and 5 male. Participants were 18-30 years of age, with an average age of 23.4 years (SD=3.32). All participants met the inclusion criteria as listed in the methodology. All participants were native English speakers as determined by self-report, none of the participants reported any exposure to Dutch or Mandarin, and all were dominantly right-handed. No subject reported being diagnosed with a speech/language impairment (or ever had speech therapy), learning, psychiatric, or neurological disorder. No history of ear disease was reported, but one subject did report occasional mild tinnitus in the right ear.

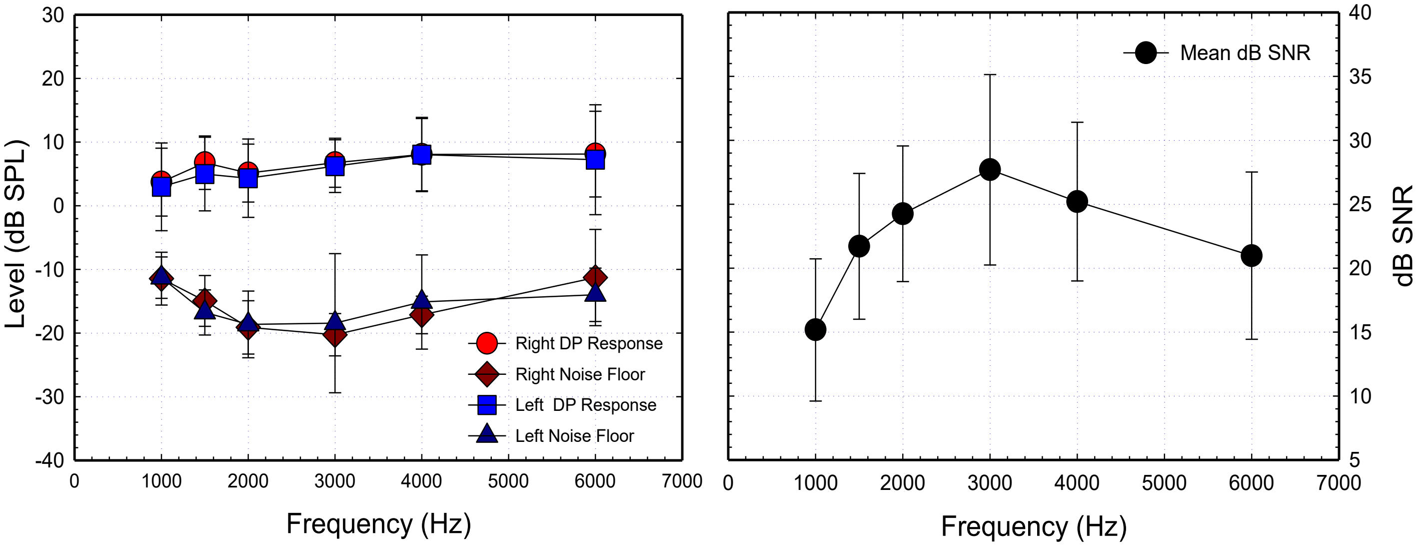
As determined by inclusion criteria, all participants had pure-tone thresholds of 15 dB HL or better in both ears for standard audiometric frequencies from 0.25-8 kHz. Figure 4 shows mean pure tone thresholds for standard and extended high frequencies for the right and left ears for all participants. There were no significant differences in thresholds between right and left ears for any single frequency from 0.25-16 kHz by paired t-test (p>.05).

Figure 4 - Mean (+/- 1 SD) pure tone behavioral thresholds for standard (left panel) and extended high-frequencies (right panel).



Distortion product otoacoustic emissions were also measured to further establish peripheral hearing status. All subjects had present DPOAE responses across all tested frequencies, with the exception of one subject with present, but abnormally low DPOAE responses at 1 and 1.5 kHz in the left ear (4.6 and 5.9 dB SNR respectively). Figure 5 shows mean DPOAE signal and noise levels for the left and right ears of all subjects. Of note, the lowest mean DPOAE SNR was found for 1000 Hz (15.17 ±5.57 and 14.27 ±6.57 for right and left ears, respectively). There were no significant differences between right and left DPOAE SNRs by paired t-test (p>.05).

Figure 5 - DPOAE response levels and noise floor (left panel) and mean dB SNR across participants' ears (right panel) error bars are +/-1 SD.



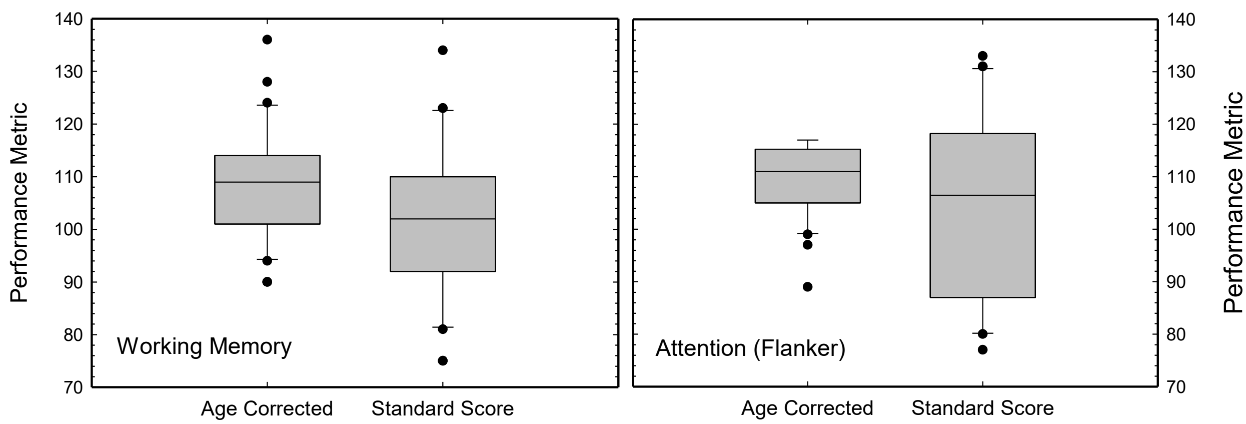
Overall, pure tone threshold and DPOAE results suggest that the participants had little peripheral hearing damage. As a group, they also reported relatively low levels of noise exposure. Self-reported noise exposure history was assessed using the Noise Exposure Questionnaire (NEQ). Scores for each question on the NEQ represent a rating from 0 to 4 of how often they were exposed to loud noises in each category over the past year (0 – never, 1 – every few months, 2 – monthly, 3 – weekly, and 4 – daily). Summed (up to 120 possible) and mean ratings by question as shown in Table 6. Results of the NEQ demonstrate that participants had relatively low amounts of noise exposure, except for listening to music under headphones and by speakers. Interestingly, every subject reported listening to music on a daily basis, whether it be at home, in a car, or on a device through headphones. A screening score based on the average rating for either the first three or six questions has been suggested by Johnson et al. (2017) to identify individuals at risk for noise induced hearing loss (last two columns of Table 6). For a more conservative estimate, the six question screen was used to assess noise exposure risk. Johnson et al. (2017) stated that a score of 3 or 4 should be used as a possible criterion for noise exposure risk, and 5 or 6 as a more lenient criterion to reduce false positives. The average screening scores were 1.70 (SD 1.24) and 2.54 (SD 1.81) for the three and six item screening scores, respectively, which are less than either cutoff criteria. Only three individuals had screening scores that exceeded 4, indicating that their noise exposure over the past year was reported to be higher risk. Interestingly, all three of these scores were highly driven by attending Syracuse University sporting events, but none of these individuals had exposure to heavy machinery or firearms.

Table 5- NEQ ratings per question and summed screening scores.

	Power tools	Heavy Equipment	Sporting events	Motorized vehicles	Small aircraft	Firearms	Play a musical instrument	Listen to music using headphones	Listen to music from speakers	3 Item screen score	6 Item screen score
Sum of ratings	7	5	39	15	6	4	28	100	115	51	76
Average rating	0.23	0.16	1.33	0.50	0.20	0.13	0.93	3.33	3.83	1.70	2.53
SD	0.56	0.59	0.95	0.86	0.44	0.34	1.41	1.12	0.59	1.23	1.81

Because it is possible that general cognitive function would influence AEP results, particularly in an auditory oddball task, tests of working memory and selective attention were administered to all subjects. Working memory was measured by the List Sorting Task and attention was measured on the Flanker Selective Attentional Task from the NIH Toolbox Cognition Battery (Weintraub et al. 2014). Figure 6 shows the range of age-corrected scores and uncorrected standard scores, with a normal performance metric mean of 100 and SD of 15 for each. The age-corrected scores account for age and other demographic characteristics (education, gender, and race/ethnicity) that may affect the performance of individuals in the general population. A cut-off score of 70 (two standard deviations below the mean) was used as a criterion for normal cognitive function to be enrolled in the study. All participants met this criteria for both the working memory and attention tests, and only two subjects scored below a 90 for the standard/age-corrected score. While no subjects scored below the normal cut-off, one subject scored just above two SD above the mean on the working memory and another subject scored two SD above the mean on the attentional task. The results are consistent with normal working memory and attentional skills across participants, with minimal across-subject variation, particularly for the age-corrected scores.

Figure 6 - Box and whisker plots for age corrected and standard scores for both working memory (left panel) and attentional tasks (right panel). The box represents the 25th and 75th percentile, the middle line the median, and the whiskers the 10th and 90th percentiles. Filled circles are outliers.



3.3 AEP grand means and individual variability

Informational masking effects were examined using an AEP oddball paradigm in which similarity between target and masker was manipulated by four background noise masker conditions varying in linguistic/phonetic content (Quiet, Mandarin, Dutch, and English), and uncertainty was manipulated by varying the TTI of target stimulus presentations in a predictable (low uncertainty) and random (high uncertainty) manner. For each participant, therefore, a total of eight average waveforms to the standard stimulus and eight waveforms to the target stimulus were obtained. The P1-N1-P2 was analyzed in the standard waveform and the P3 was analyzed in the target average waveform. Waveforms were averaged across all button press accuracy responses including hits, misses, false alarms, and correct rejections. The number of misses and false alarms as determined per participant was low, and these epochs containing performance errors were not removed from the average waveform for each condition (see section 3.6.1 for further discussion).

Figure 7 shows standard and target grand mean averaged time waveforms by electrode location across the scalp (shown for the quiet low uncertainty condition). Consistent with the literature (Hillyard et al. 1971; Martin et al. 2008; Naatanen and Picton 1987; Picton 1992), response amplitudes were generally largest at midline locations and decreased at sites moving away from the midline. Note that electrodes were re-referenced to mastoids, resulting in flat waveforms for the M1 and M2 electrode locations (pictured in the bottom right corner of Figure 7). In addition, heat maps for the time windows of the N1 (25-300ms) and P3 (250-550ms) generally verify the expected distribution of maximum voltages at the vertex for the N1 and a wider pattern with more parietal distribution for the P3. Neither N1 (amplitude averaged from 25-300ms) nor P3 (averaged from 250-550ms) average peak amplitude were significantly different across the left and right hemispheres as determined by comparing averaged electrodes sites of Fc6, C6, Cp6, and P6 on the right and Fc5 C5, Cp5, and P5 on the left hemisphere ($p>.05$ for both N1 and P3). Further analyses were therefore conducted only for the Cz site for N1 and Pz site for P3, consistent with the literature and maximum amplitude distributions.

Figure 7 - Scalp topography for quiet low uncertainty target and standard waveforms across the entire electrode cap. A heat map of N1 (amplitude averaged from 25-300ms) and P3 (averaged from 250-550ms) voltages are shown in the bottom panels. Voltage scales are in µV for both grand mean waveforms and heat maps.

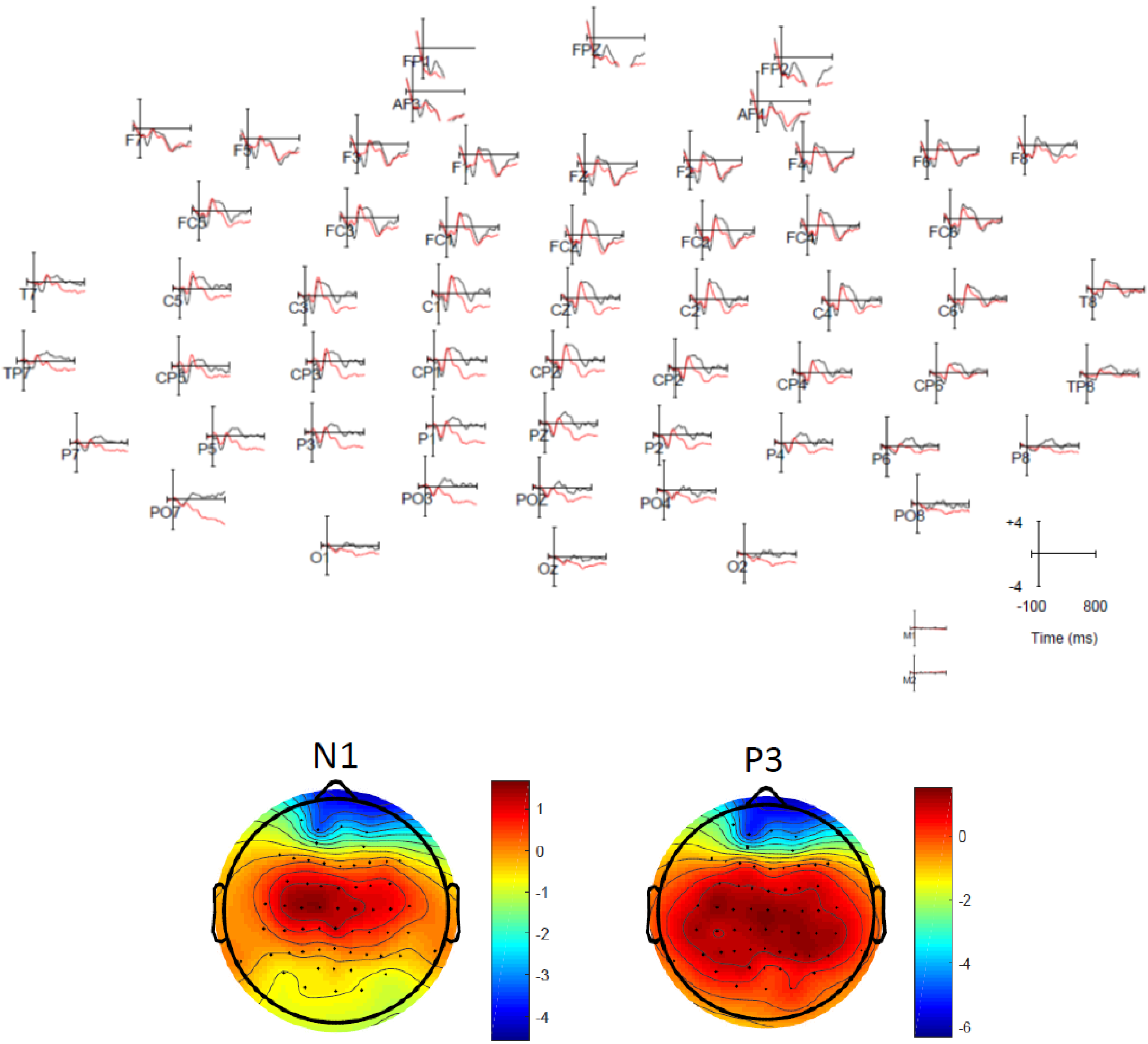
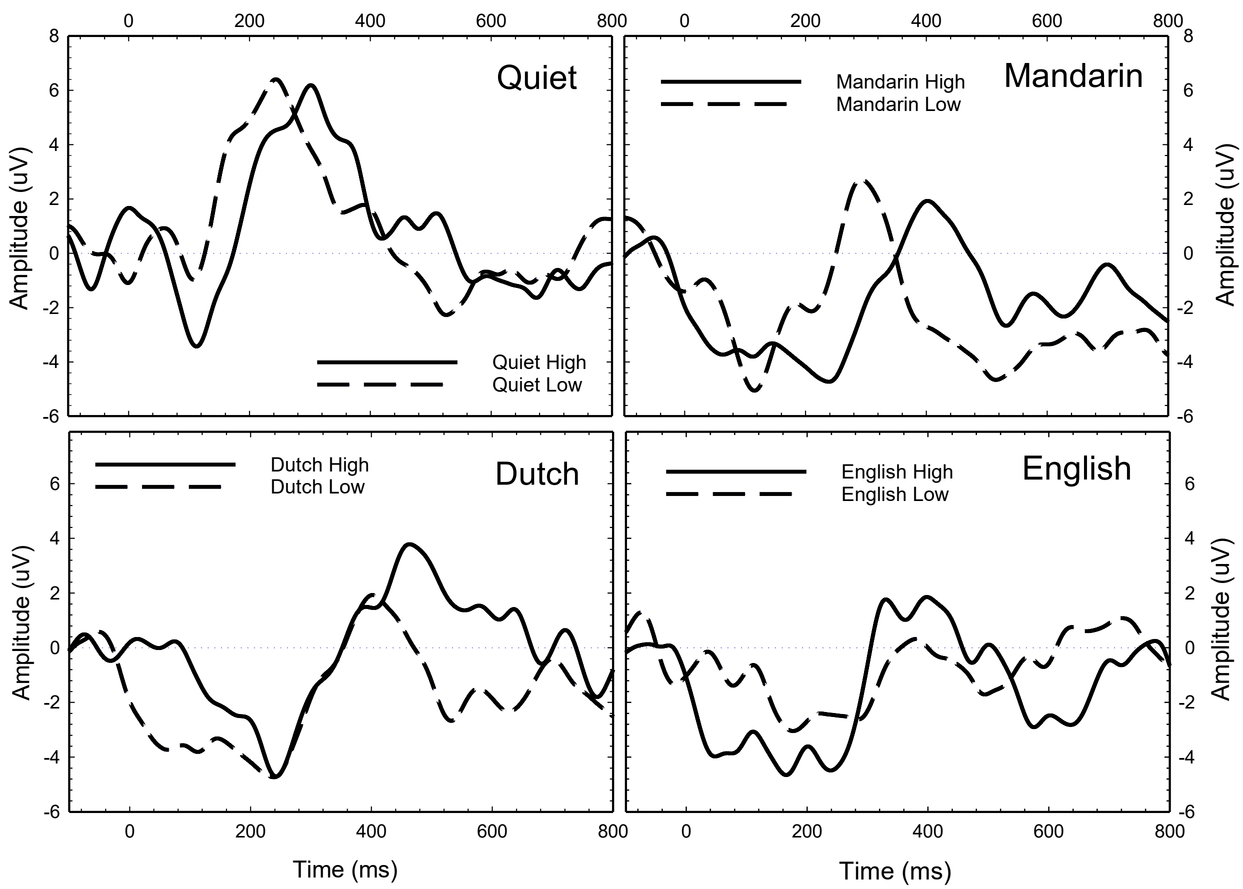


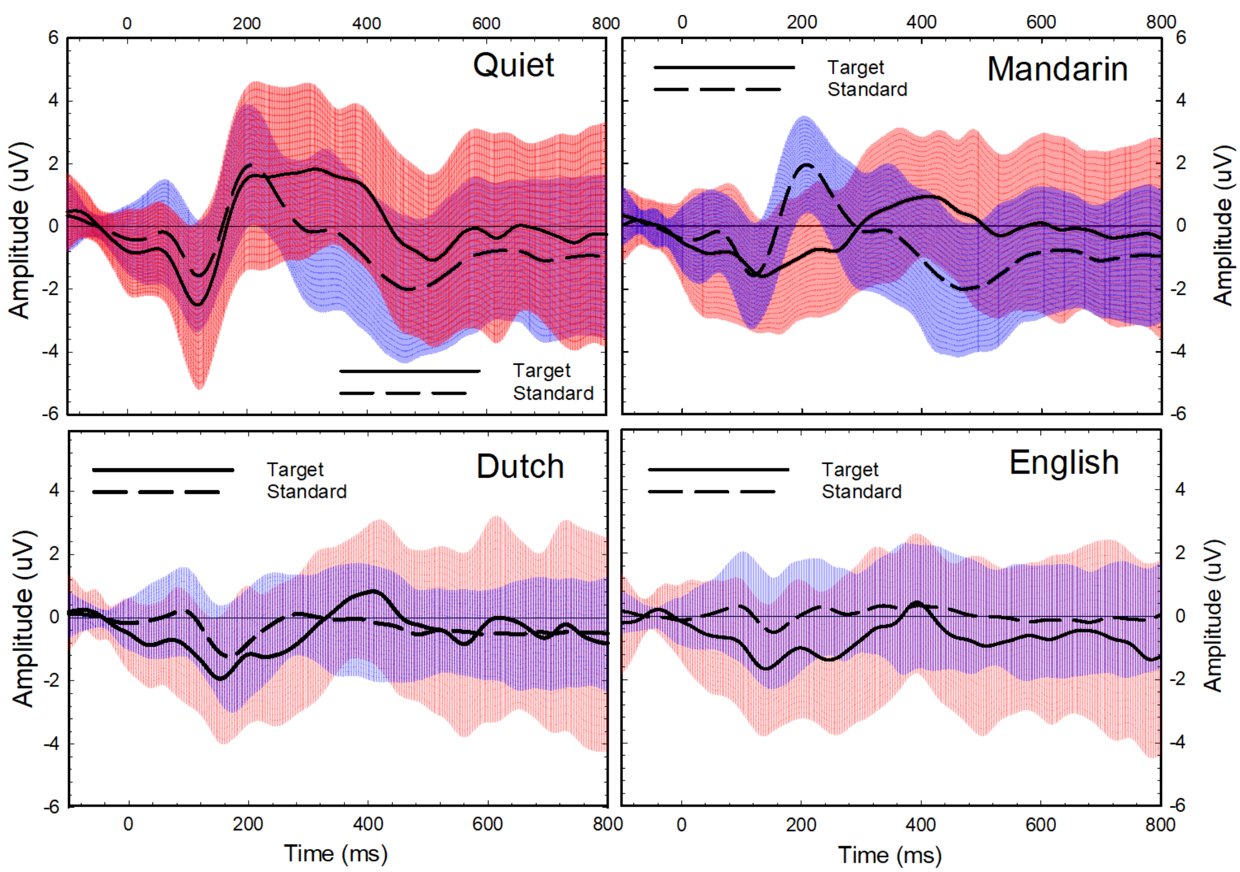
Figure 8 shows individual waveforms from one participant recorded to target stimuli in the infrequent waveform at Pz for all similarity and uncertainty conditions. As demonstrated by this individual, general trends in the individual data are consistent with smaller amplitudes, longer latencies and poorer morphology for the language masking conditions compared to quiet. Waveforms generally showed the smallest amplitudes and poorest morphology in the English condition. Individual differences in waveforms between the two uncertainty conditions were generally consistent with longer peak latencies for the high uncertainty condition compared to low uncertainty.

Figure 8 - Example individual waveforms from a single subject for all conditions to target stimuli (infrequent) at the Pz electrode Similarity conditions are separated across each of the four panels and uncertainty conditions are indicated with either a solid or dashed lines.



Variability across individual waveforms is demonstrated in Figure 9, which shows the grand mean and a shaded range of ±1 standard deviation shading around the mean for target and standard stimuli in across all language maskers. Responses are shown for the high uncertainty condition only. In this figure, the blue shading shows variability around the mean for the standard stimulus and red shading the variability around the waveform for the target stimulus. Though there is relatively large inter-individual variability observed, the range is consistent across conditions. There is a slightly greater variability for target responses (red) compared to standard, consistent with the smaller number of stimulus presentations for the target compared to standard. Variability is also greater for later time periods compared to earlier.

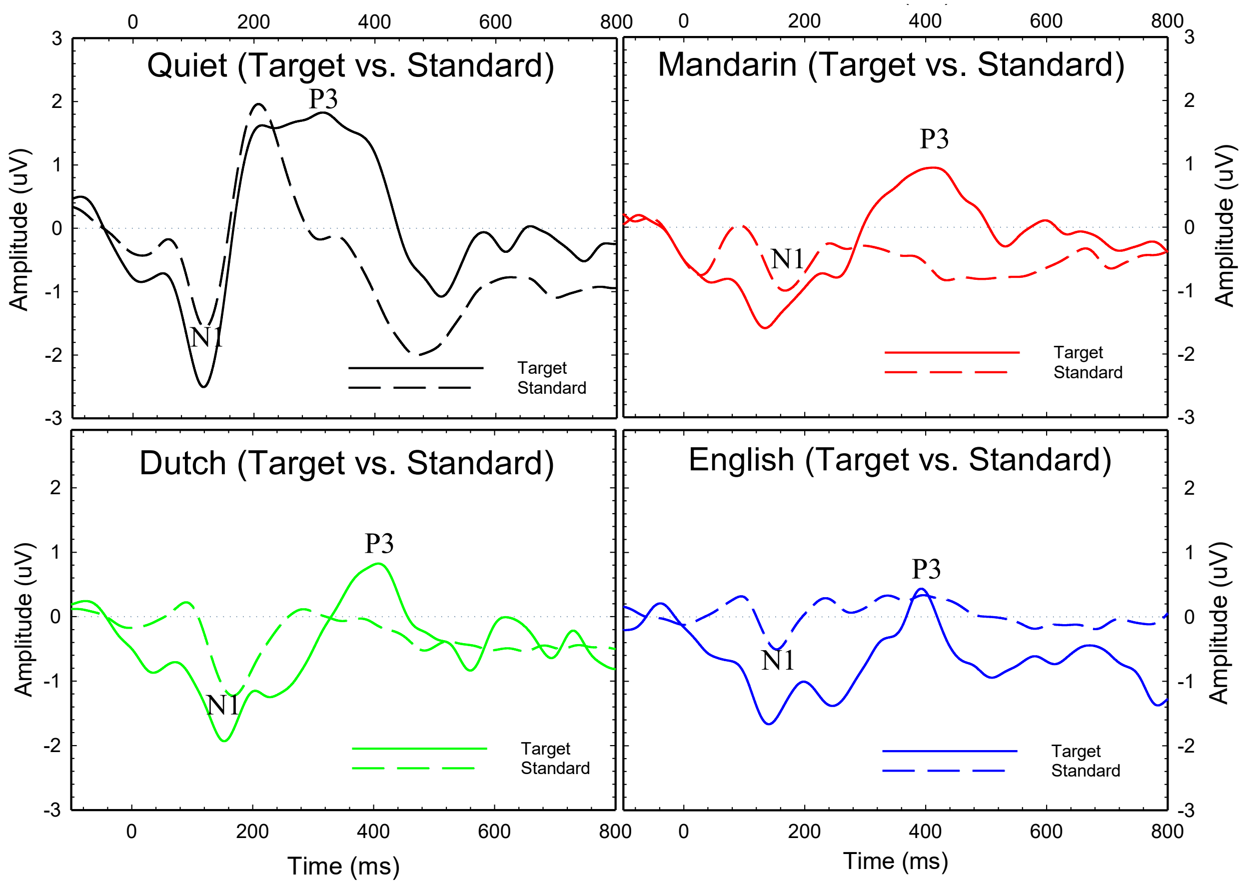
Figure 9 - Grand mean and standard deviation of grand mean waveforms target (at Pz) and standard (at Cz) stimuli across language masking conditions for high uncertainty conditions. Standard deviation of ±1 SD is shown in red for the target and blue for the standard.



Figures from here will show only the grand mean waveforms, but peak measures (amplitude, latency, and area) were obtained from the individual waveforms for each participant in each condition, therefore the inter-subject variability is accounted for in the statistical analyses. All further time waveform figures display the grand means only.

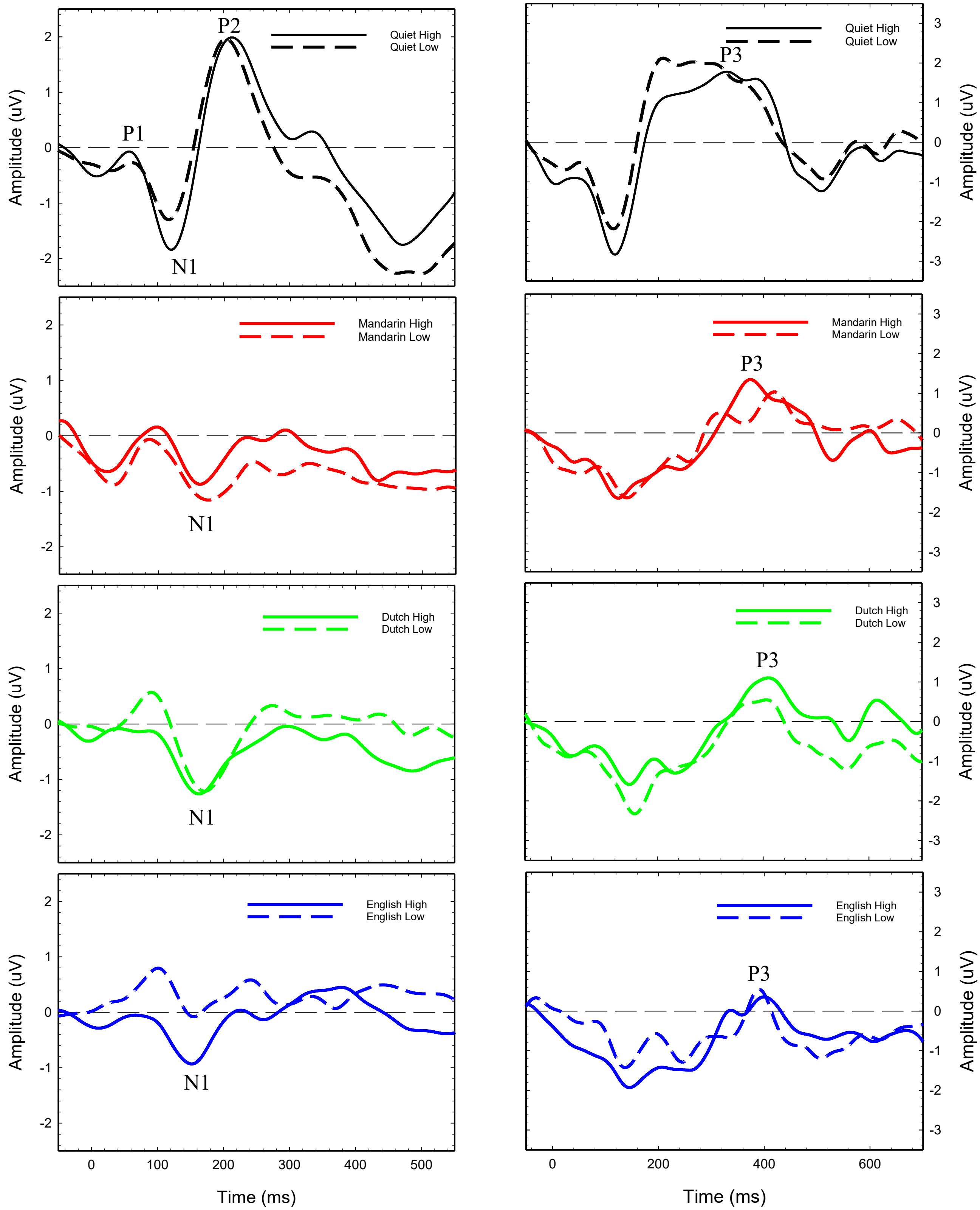
Figure 10 shows overall grand means for the similarity conditions. As in the previous figure (but without the range of variability), target and standard grand average responses for each of the language masking conditions are shown only for the high uncertainty. N1 and P3 peaks, as analyzed for each individual in the standard and target waveforms respectively, are visible and labeled in the grand means. In quiet, a large P1-N1-P2 predominates the standard waveform, and is clearly visible in the target waveform. The P3 response is identified as a broad and large peak with a maximum in the 300ms range. Within each language masking conditions there is an evident N1 around 150ms and a clear difference in amplitude around 300-400ms between the standard and target responses consistent with the P3 response. General changes in the morphology of these grand average waveforms are apparent across the noise conditions, with a decreasing amplitude N1 and P3 from Mandarin to Dutch to English, roughly consistent with the linguistic/phonetic similarity hypothesis

Figure 10 - Grand mean waveforms for target (P3) and standard (N1) stimuli in high uncertainty conditions. Target waveforms are from the Pz electrode, where P3 responses were analyzed, and standard waveforms from the Cz electrode where N1 responses were analyzed.



For a better visual comparison within each language masking condition and between uncertainty conditions, Figure 11 displays grand mean waveforms for the P1-N1-P2 time window for the standard waveform (left column) and the P3 time window (right column) for the target waveform across language masking conditions (rows) and uncertainty conditions (solid and dashed lines). Several trends consistent with the hypotheses for the effects of linguistic/phonetic similarity and uncertainty factors on informational masking are observed in these grand averages. In quiet, morphology of each component is strong with distinct component peaks and expected latencies. Across language masking conditions, the amplitude of N1 and P3 is markedly reduced compared to quiet. The latencies of the peaks can also be observed to generally shift later from top to bottom. Between the uncertainty conditions, there are differences in overall morphology, amplitude and latency of the peaks within each quiet or noise condition, but there are not obvious consistencies. Of note, the low uncertainty condition for the P1-N1-P2 waveform standard response in English masking displayed an overall higher baseline amplitude compared to high uncertainty. A review of the individual data indicated that this was a relatively consistent trend across subjects. Specific analysis of the peak amplitude, latency and area measures are presented in the next section.

Figure 11 - Grand mean waveforms for the P1-N1-P2 (standard) recorded at Cz (left panels) and for the P3 (target) at Pz (right panels) with low uncertainty and high uncertainty indicated with solid and dashed lines. Grand means in quiet are shown in the top row, with the three language masker conditions in the rows below. The color scheme is consistent with the previous figure and throughout the following figures, with Mandarin in red, Dutch in green, and English in blue.



3.4 Similarity and uncertainty effects on peak measures

Peak amplitude, mean amplitude, peak latency, and area were measured on the individual waveforms for all participants as described in the methods. For each of these primary outcomes for N1 and P3, 4x2 RM-ANOVA analyses were conducted to analyze main effects of similarity and uncertainty, as well as the interaction. Results are presented below for the P1-N1-P2 first, as analyzed in the standard waveform, and then for the P3, with the effects of similarity (Aim 1-objective 1) presented first and the effects of uncertainty (Aim 1 – objective 2) presented second.

3.4.1 N1 Similarity

Table 7 shows the ANOVA results for the primary outcome variables related to P1-N1-P2 responses. Significant results based on corrected p-values ≤ 0.05 are highlighted in gray. Mauchly's test of sphericity was performed for the similarity effect (4 repeated conditions) for all ANOVAS. Results of this test are reported only in cases where the assumption of sphericity was violated. If sphericity was violated, Greenhouse-Geisser corrections were implemented.

Table 6- Results of the overall 4x2 RMANOVA for N1 and P1-N1-P2 primary outcome measures, values, F-values, and effect sizes (partial eta squared) are shown for each of the main effects and the interaction term. Degrees of freedom for the F statistic were (3,87), (1,29), and (3,87) for similarity, uncertainty, and interaction terms respectively. Significant results (p<.05) are highlighted in gray.

	N1												**P1-N1-P2**		
	N1 Peak Amplitude			N1 Mean Amplitude			N1 Latency			N1 Area			P1-N1-P2 Area		
	p	F	η_p^2	p	F	η_p^2	p	F	η_p^2	p	F	η_p^2	p	F	η_p^2
Main Effect of Similarity	**.003**	**4.95**	**.146**	.617	.600	.020	**.005**	**4.56**	**.136**	.165	1.75	.057	**.001**	**5.76**	**.166**
Main Effect of Uncertainty	.181	1.87	.061	.253	1.36	.045	**.025**	**5.61**	**.162**	.692	.160	.005	.307	1.082	.036
Interaction	.217	1.51	.050	.477	.837	.028	.851	.264	.009	.764	.410	.014	.537	.730	.025

First, it is noted in Table 7 that the interaction effect between similarity and uncertainty was not significant for any of the outcome measures for the P1-N1-P2 response. Effect sizes were small based on classification partial eta squared (η_p^2), based on a small effect $\leq$.01, moderate ~ .09, and strong $\geq$.24 (Kotrlik, and Williams, 2003; Tabachnick, and Fidell, 2001). In the left panel of Figure 12, the hypothesized experimental interaction (figure 2 in the objectives section) is shown, although it was hypothesized this would be an additive or a non-parallel effect would be seen for the P3 more than for N1. The panel on the right shows the pattern of the actual N1 amplitude data. Consistent with the hypothesis, there is not an apparent interaction effect observed for this outcome measure, or any of the P1-N1-P2 outcomes. Because none of the interactions were significant, the ANOVA results for the simple main effects can be considered separately for each of the variables for the P1-N1-P2, and the results for similarity and uncertainty and their relationship to the hypotheses will be presented below

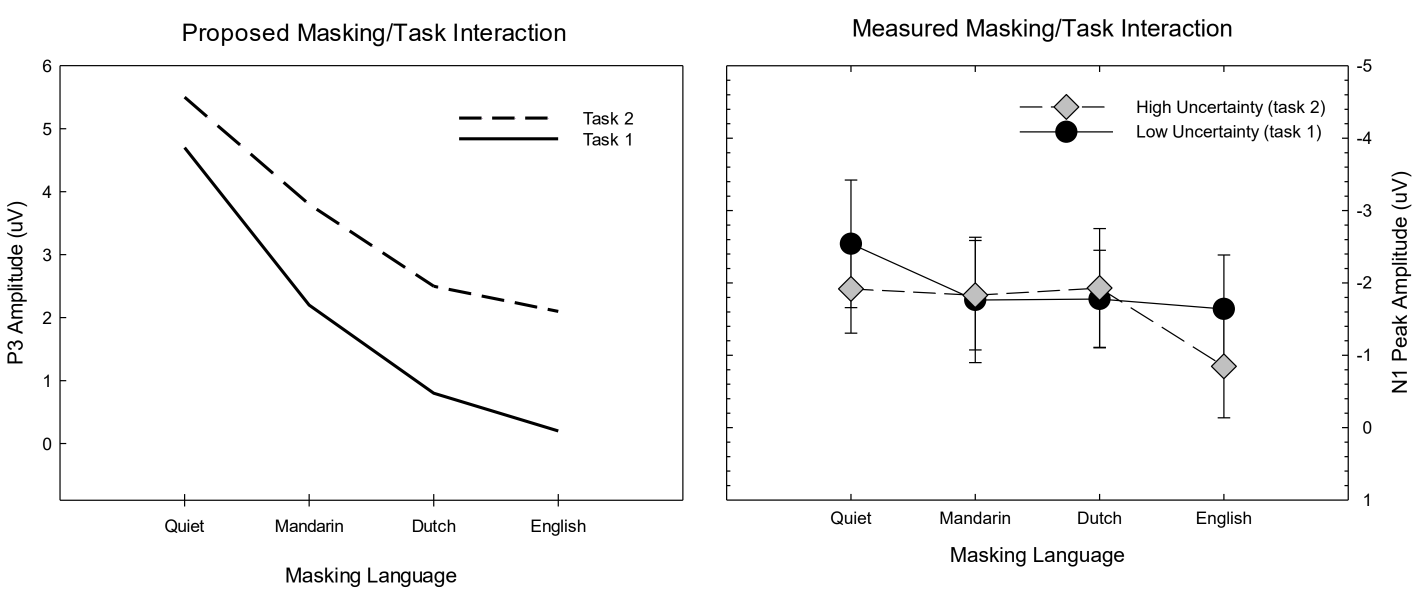


Figure 12 - Hypothesized interaction effect between similarity and uncertainty compared to actual N1 amplitude data.

To better illustrate the effects of just one factor at a time, the grand means for each similarity condition as evaluated by the effect of the four conditions, including Quiet and three different 2-talker language maskers, combined across the two uncertainty conditions for the P1-N1-P2 and within each uncertainty condition are shown in Figure 13. While these grand means are useful in visualizing some of the trends and statistical outcomes for similarity to follow, it should be remembered that the results for the outcome measures below are obtained from the individual waveforms for each separate condition (e.g. English masker, high uncertainty) and analyzed in the overall 4x2 ANOVA shown in Table 1 above.

Figure 13 - Grand mean P1-N1-P2 standard waveforms at Cz for similarity conditions averaged across combined high and low uncertainty conditions (top panel) and for each uncertainty condition separately in the bottom left (low uncertainty conditions) and right (high uncertainty conditions) panels.

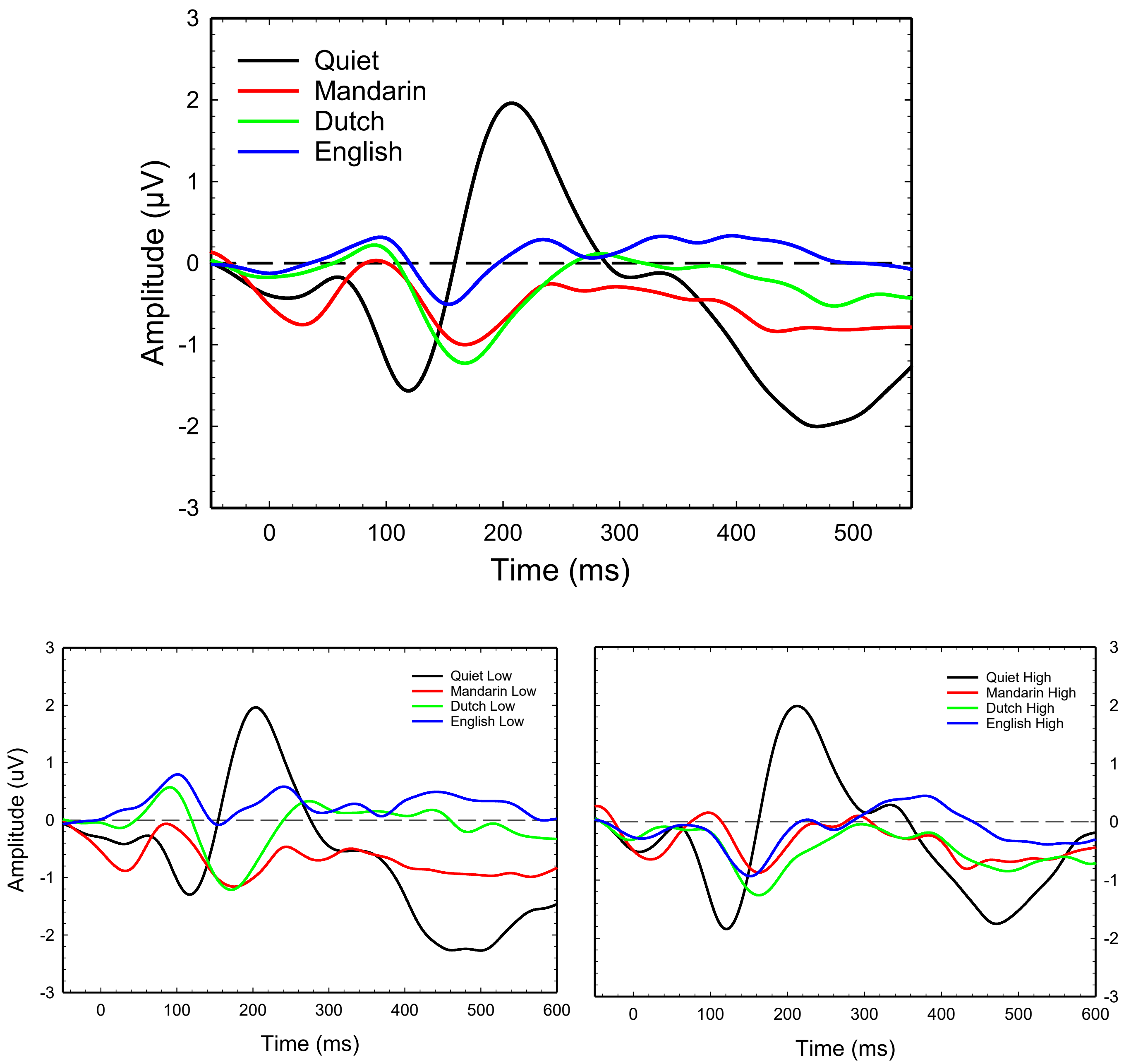
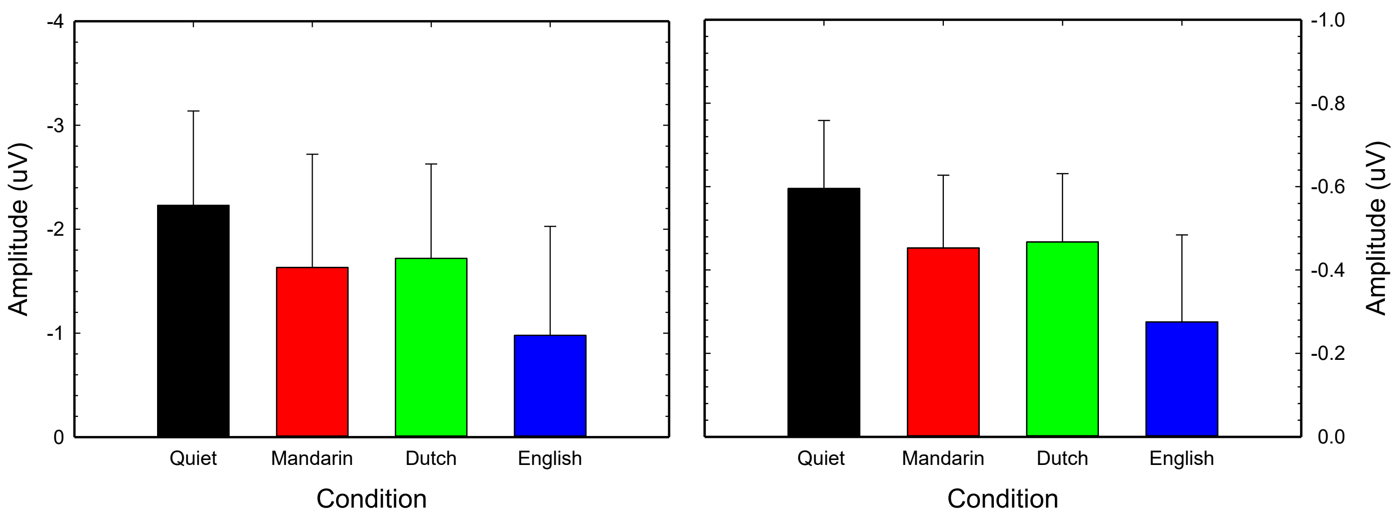


Figure 14 displays average peak and mean amplitude measures across all participants for the similarity conditions. As a reminder, peak and mean amplitude were both measured because peak measures can be easily compared to previous studies, while mean amplitude gives a better understanding of the entire amplitude of the underlying component. For either method of measuring amplitude, the magnitude was largest for quiet and decreased in the presence of any of the language maskers. In both cases, the average amplitude was the smallest for the English masker condition, and appear similar between the Mandarin and Dutch conditions.

Figure 14 - Average N1 peak amplitude (left panel) and mean amplitude (left and right panels) for all similarity conditions averaged across uncertainty conditions. Error bars represent +1 SD.



Although the trends appear similar, as shown in Table 7, the main effect of similarity was significant for N1 peak amplitude, but did not reach significance for the mean amplitude measure. Pairwise comparisons with adjustment for multiple comparisons (Bonferroni) revealed that peak amplitudes were larger in the quiet condition compared to English (p=.002). Although they appear larger on average, peak amplitudes in quiet were not significantly different compared to either Dutch (p = .083) or Mandarin (p=.059), although they approached significance in comparison to Mandarin. Peak amplitudes were also not significantly different between Dutch and Mandarin ($p>.05$). However, peak amplitude in the Dutch condition was significantly larger than the English condition ($p=.027$), while Mandarin and English were not significantly different ($p = .124$).

Figure 15 - Average N1 peak latency for all similarity conditions averaged across uncertainty conditions. Error bars represent +1SD.

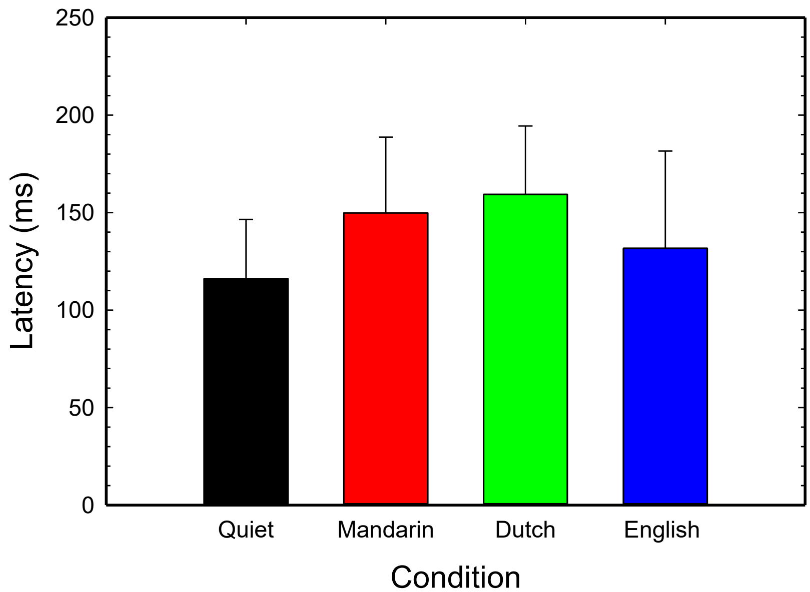


Figure 15 shows mean N1 peak latencies across language maskers. As shown in Table 7, there was a significant overall main effect of similarity for N1 peak latency. Pairwise comparisons revealed that the N1 in quiet was significantly earlier than for Dutch ($p=.002$), but not earlier than the Mandarin condition ($p=.140$). N1 latency was not significantly different between the Dutch and Mandarin conditions ($p>.05$). Contrary to the hypothesized delay in latency for the most linguistic/phonetic similar masker, N1 latency for the English condition appeared to be shorter than the other language maskers. This resulted in N1 latency in English masking being not significantly different to either Dutch, Mandarin, or quiet ($p>.05$).

Because both peak amplitude and peak latency can be affected by small waveform variations and do not provide a encompassing measure of the entire component, area measures were also calculated for a comparison of changes in size and morphology of neural responses less dominated by the absolute peaks of the waveform. The rectified area under the curve (negative area in the case of N1 region, and total rectified area for the P1-N1-P2) was calculated as described in the methods for the N1 time window as well as for the entire P1-N1-P2 response area (25-300ms). Figure 16 shows mean N1 area across similarity conditions. While N1 area for the English condition had the smallest mean, the main effect of similarity did not reach significance for N1 area (p>.05).

Figure 16 - Average N1 area measures for similarity conditions averaged across uncertainty conditions. Error bars represent +1SD.

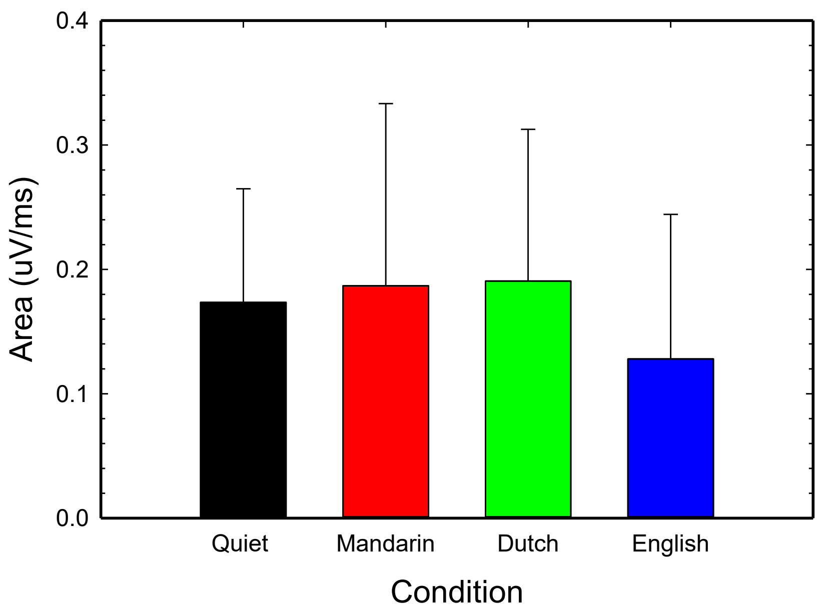


Figure 17 - Average P1-N1-P2 total area across similarity conditions averaged across uncertainty. Error bars represent +1SD.

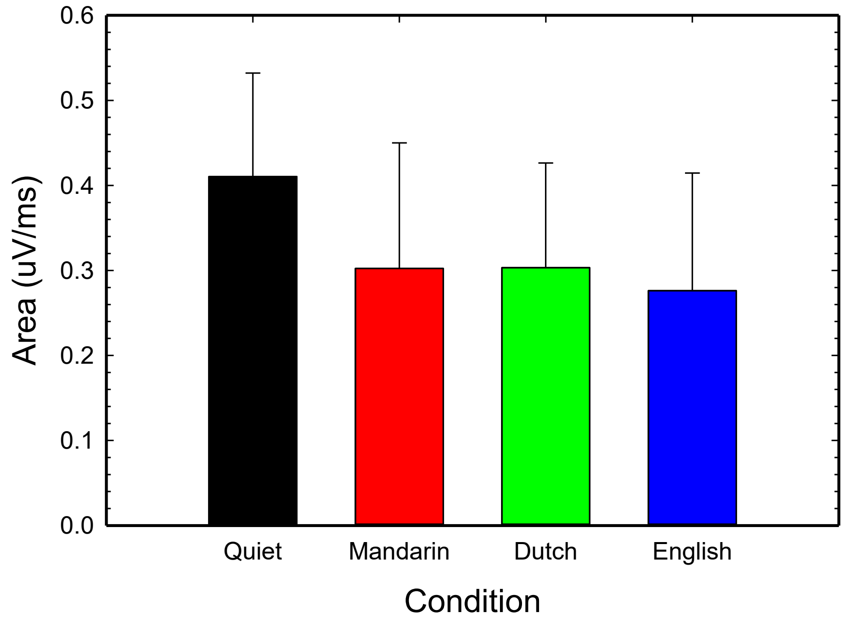


Figure 17 shows area measures across the entire P1-N1-P2 time window (25-300ms) for all similarity conditions averaged across uncertainty conditions. The effect of similarity was significant for P1-N1-P2 area. Pairwise comparisons showed that response are in quiet was significantly larger than all other conditions (p=.015, .017, and .002 for Mandarin, Dutch and English comparisons respectively). However, there were no other significant differences in overall area among the three language masking conditions (p>.05).

3.4.2 P3 Similarity

Similar peak outcome measures were analyzed for the later processing P3 component for the similarity main effect, graphed and summarized in the same way below. Table 8 shows the results from the 4x2 RMANOVA for the primary outcome variables related to the P3 responses. Significant results based on corrected p-values ≤ .05 are highlighted in gray. Mauchly's test of sphericity was performed for the similarity effect (4 repeated conditions) for all ANOVAS. Results of this test are reported only in cases where the assumption of sphericity was violated, which only happened in the P3 area measure, and Greenhouse-Geisser corrections were implemented in this case.

Table 7 –Results of the overall 4x2 RMANOVA for the P3 primary outcome measures, values, F-values, and effect sizes (partial eta squared) are shown for each of the main effects and the interaction term. Degrees of freedom for the F statistic were (3,87), (1,29), and (3,87) for similarity, uncertainty, and interaction terms respectively. Significant results (p<.05) are highlighted in gray.

	P3											
	P3 Peak Amplitude			P3 Mean Amplitude			P3 Latency			P3 Area		
	Sig	F	η_p^2	Sig	F	η_p^2	Sig	F	η_p^2	Sig	F	η_p^2
Main effect of Similarity	**.000**	**7.23**	**.199**	**.006**	**4.39**	**.132**	**.000**	**7.51**	**.206**	**.000**	**12.4**	**.300**
Main Effect of Uncertainty	.203	1.68	.055	.771	.143	.005	**.005**	**9.37**	**.244**	**.031**	**5.11**	**.150**
Interaction	.866	.244	.008	.918	.167	.006	.917	.169	.007	.989	.042	.001

As was the case for all N1 outcomes, table8 shows that there were also no significant interaction effect between similarity and uncertainty for any of the P3 outcome measures for the P3. Effect sizes were small based on the classifications for partial eta squared presented previously. Figure 18 shows the hypothesized experimental interaction (based on hypothetical data) and actual measured data for P3 amplitude, similar to the figure presented in the previous section for N1. In the case of P3 amplitude, the trend across similarity conditions was more similar to the hypothesized pattern, but the difference between high and low uncertainty conditions small and largely parallel, indicating a lack of interaction between the two factors. For the P3, the hypothesis was that for the English masking (the highest level of similarity), uncertainty would have the largest effect (i.e. there would be an additive effect at this level of processing). The statistical outcome indicate and the graphed data verify that this was not the case.

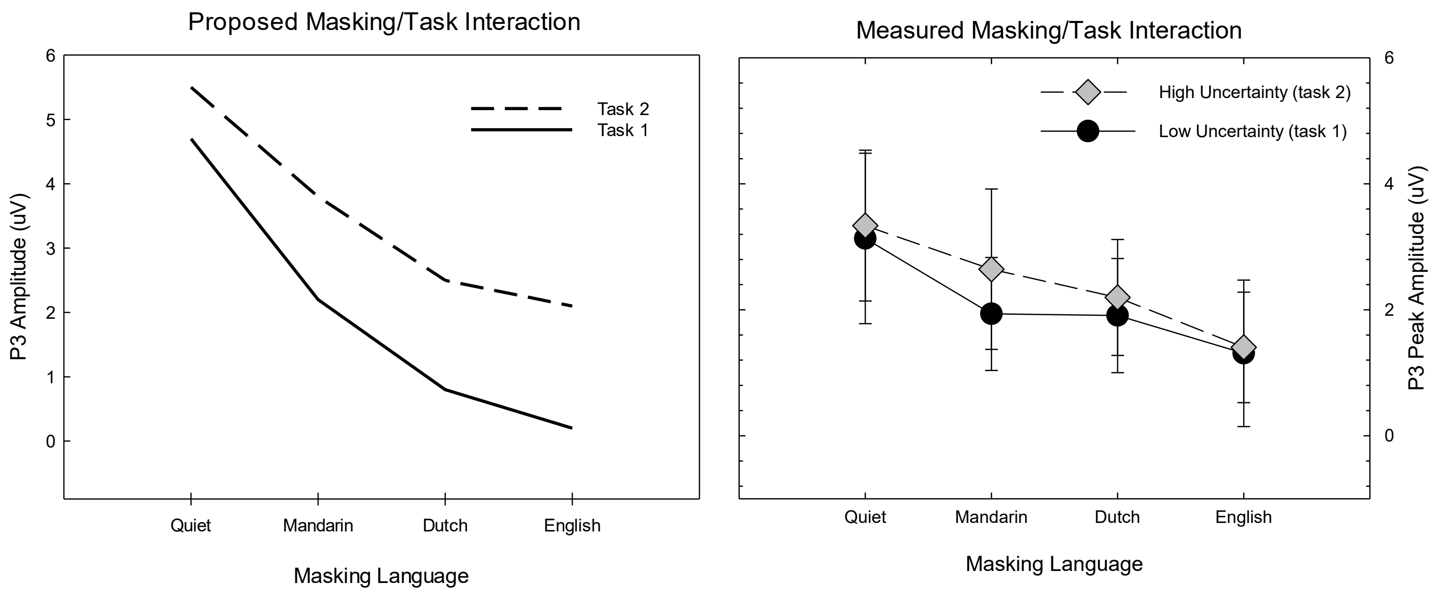
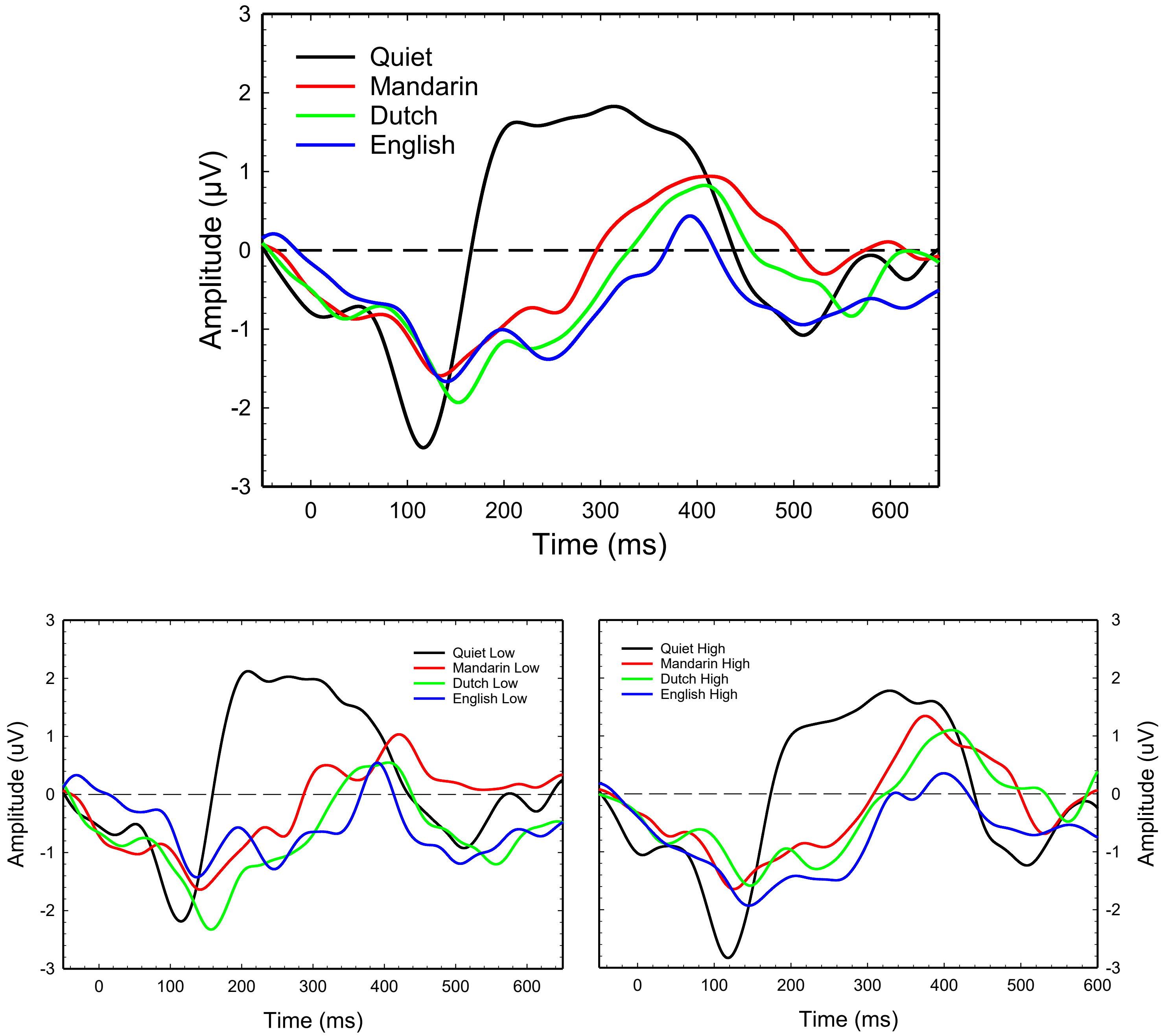


Figure 18 - Hypothesized interaction effect between similarity and uncertainty compared to actual P3 amplitude data.

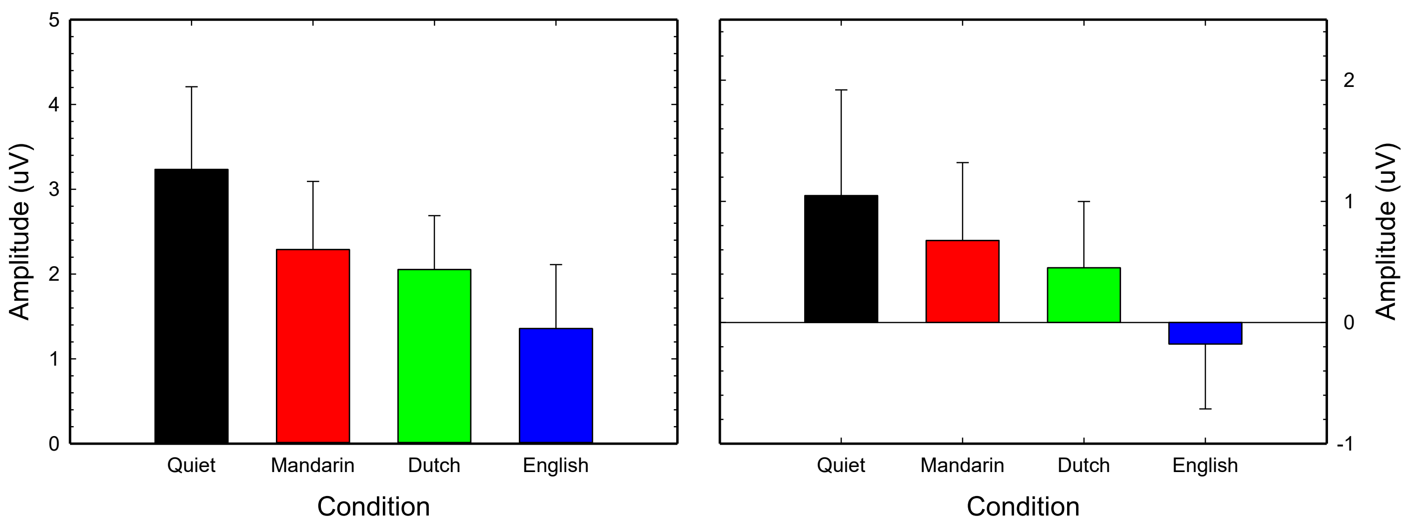
Because none of the interactions were significant, the ANOVA results for the simple main effects of uncertainty can be considered separately for each of P3 outcome variables. Figure 19 shows the averaged similarity grand mean waveforms for the P3 responses in the target waveform for all language masking conditions averaged across both uncertainty conditions. As with the P1-N1-P2 grand means, low and high uncertainty are also show in the left and right bottom panels respectively. Overall results show a relatively large width of the P3 component response (verified across electrode locations) and a clear decrease in morphology of the P3 in the presence of all language maskers. The trends in these grand means show a reduction in P3 amplitude from Mandarin to Dutch to English. In addition, the reduction in area under the curve of the P3 component appears to follow the hypothesized increase in informational masking from least to most similar to the English target words. This trend is more apparent in the high uncertainty condition than for low uncertainty.

Figure 19 - Grand mean P3 target waveforms at Pz for the similarity conditions averaged across combined high and low uncertainty conditions (top panel) and for each uncertainty condition separately in the bottom left (low uncertainty conditions) and right (high uncertainty conditions) panels.



As previously described, P3 outcomes across conditions similarity were analyzed by the 4x2 RM-ANOVA, but are shown graphically averaged across uncertainty conditions. Figure 20 shows average P3 amplitude for both peak and mean amplitude measures across similarity conditions. Trends in the mean individual data displayed the same effects observed in the grand mean waveform, in that amplitudes were largest for the quiet condition and decreased for all masker conditions. As shown in Table 8, the main effect of similarity was significant for both peak and mean amplitude. Results of the pairwise comparisons revealed that peak amplitude measures were significantly larger in quiet compared to English ($p=.001$), no significant differences between Mandarin and Dutch ($p>.05$), and English had the smallest peak amplitude compared to quiet, Mandarin, or Dutch ($p=.001$, $p=.012$, $p= .027$ respectively). P3 mean amplitudes were also significantly larger in quiet compared to English ($p=.017$), but were not significantly different from either Mandarin or Dutch ($p>.05$). The difference between English and Mandarin approached significance ($p=.067$). It is important to note that because mean amplitude is measured across a specified time window, rather at the maximum peak, the average voltage around the P3 component in the English background noise was below baseline, resulting

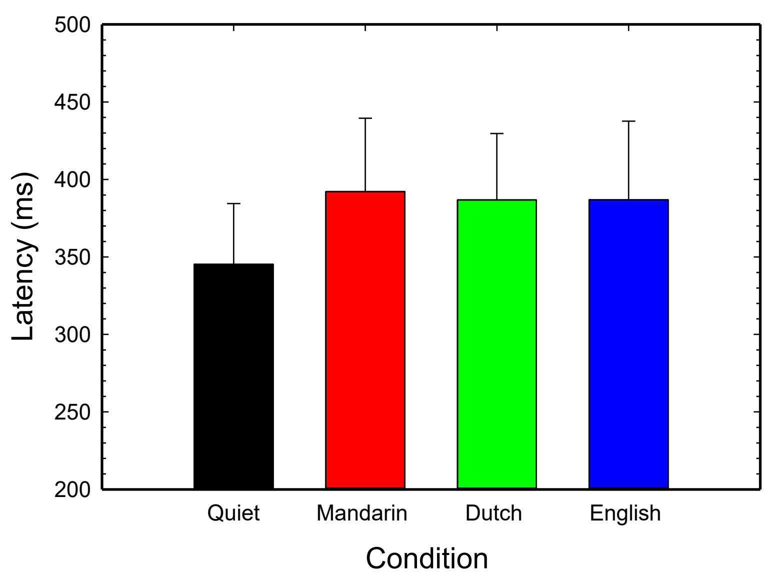
Figure 20 - Average P3 peak and mean amplitude (left and right panels respectively) for all language masking conditions averaged across uncertainty conditions. Error bars represent +1SD.



in a negative value for this measure. Both P3 peak and mean amplitude show a trend of decreasing amplitude consistent with the proposed informational masking hierarchy, although few of the amplitude differences reached statistical significance.

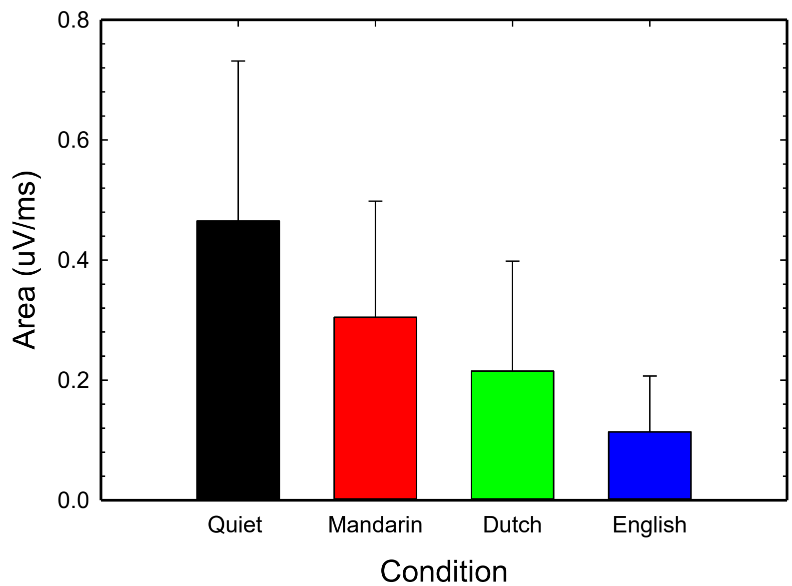
Figure 21 shows P3 peak latency across similarity conditions. Latency results for the P3 were less consistent with the proposed increase in latency as linguistic/phonetic similarity increased across maskers. As shown in Table 8, there was a significant overall main effect of similarity on P3 latency. However, as shown in the figure and revealed by pairwise comparisons, the only differences were between quiet and the noise conditions. P3 peak latency was earlier in quiet compared to Mandarin, Dutch, or English ($p=.003$, $p= .005$, $p=.001$ respectively) but none of the latency differences between language masker were significant ($p>.05$).

Figure 21 - Average P3 peak latency for similarity conditions averaged across uncertainty conditions. Error bars represent +1SD.



As observed in the grand means, the overall width of the P3 responses was wide in quiet and became smaller and narrower across language masker. To better capture this change in waveform morphology, total rectified area of the P3 response was calculated across the time window from 275-550ms. Figure 22 shows average P3 area for each similarity condition averaged across uncertainty. There was a significant overall effect of similarity on P3 area, with the largest effect size of any of the AEP outcome measure ($\eta_p^2 = .300$, interpreted as a large effect). Pairwise comparisons showed that P3 areas for the quiet condition were significantly larger than those for either Dutch (p=.005) or English (p<.001). P3 area for the English condition, with the highest similarity, was significantly smaller than all other conditions (quiet p<.001, Mandarin p<.001, Dutch p=.015). P3 areas fell between quiet and English for the two non-native maskers, but were not significantly different between Mandarin and Dutch (p=.315).

Figure 22 - Average P3 total area across similarity conditions averaged across uncertainty. Error bars represent +1SD.



In summary, the overall effect of similarity was significant for N1 peak amplitude, latency and total P1-N1-P2 area, as well as P3 amplitude, latency, and area outcomes measures. Differences among the amplitude and area measures outcomes for N1 and P3 showed that in general quiet was significantly different from all other background noise conditions. Across the different language masker conditions there were trends consistent with reductions across the proposed hierarchy of increasing linguistic/phonetic similarity from Mandarin to Dutch to English, but only a few of the comparisons reached statistical significance including P3 peak amplitude and area. Results for latency also showed that quiet was significantly different from other background noise conditions, but none of the differences within language maskers were significant for N1 or P3 latency.

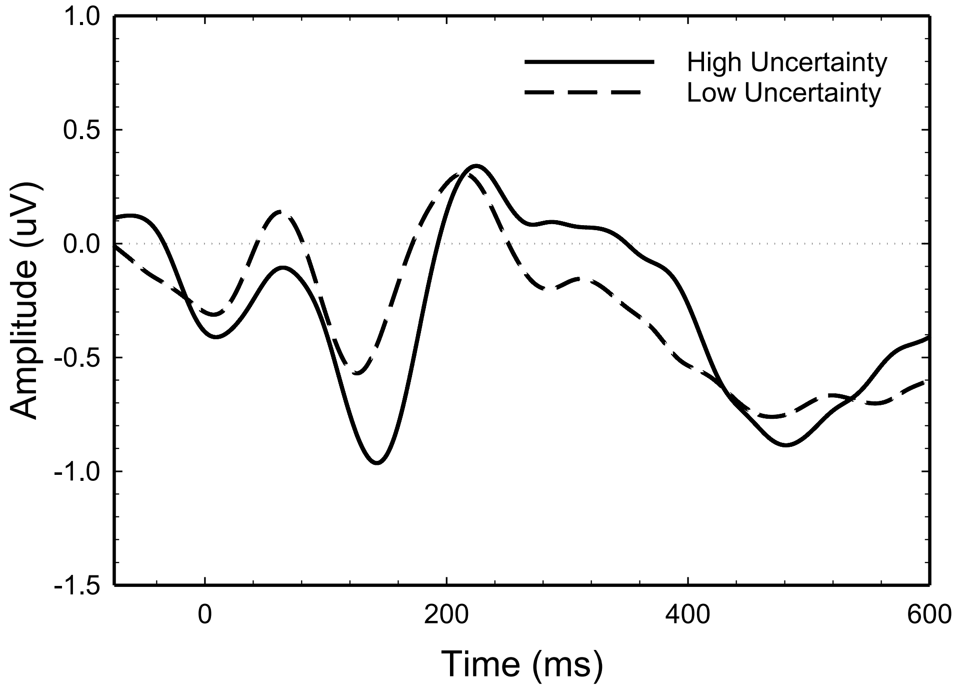
3.5 Uncertainty

3.5.1 N1 Uncertainty

The effect of stimulus uncertainty was manipulated through fixed (low uncertainty) vs. random (high uncertainty) time-to-target interval (TTI) conditions in the auditory oddball paradigm. Results for the main effect of uncertainty on P1-N1-P2 outcomes from the overall ANOVA results Table 7 (in section 3.4.1) indicated a significant effect only for N1 latency. N1 amplitude (peak or mean) and area of either N1 or P1-N1-P2 did not show an overall main effect of the uncertainty manipulation, and the effect sizes were small. These results are graphed and discussed further below.

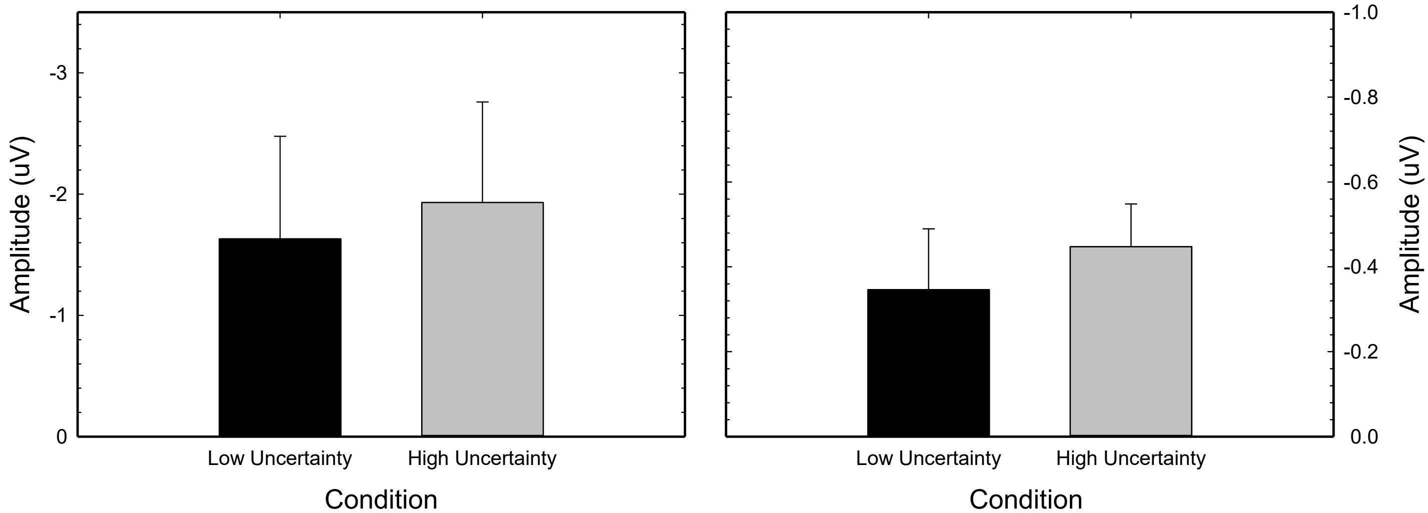
Grand means for high and low uncertainty combined across all similarity conditions for the P1-N1-P2 response to the standard stimulus are shown in Figure 23. Overall, the grand mean trends show shows a smaller and earlier N1 response for the low uncertainty condition (easier task, dashed lines compared to the high uncertainty condition (more difficult task, solid line).

Figure 23 - Grand mean P1-N1-P2 waveforms for the standard stimuli at Cz for high and low uncertainty conditions averaged across all similarity conditions. compared to low uncertainty.



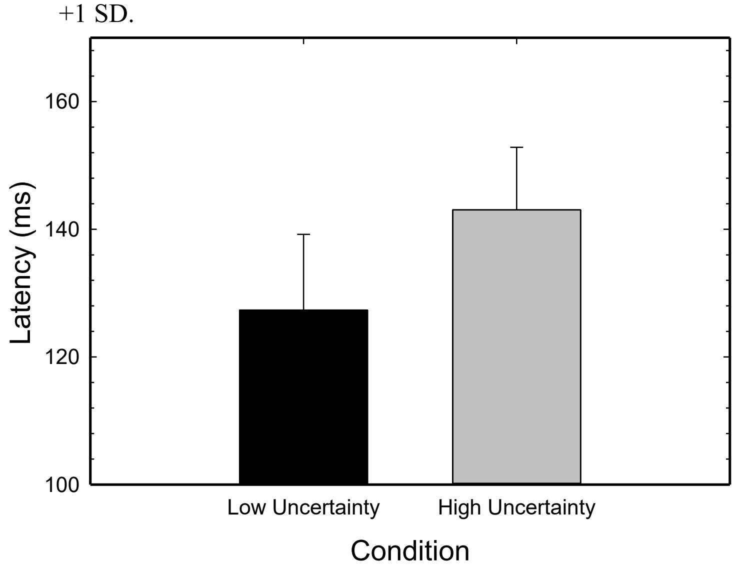
Average N1 amplitudes across individuals for both peak and mean amplitude measures across uncertainty conditions are shown in Figure 24. While the mean peak measures also show a trend for increasing amplitude from low to high uncertainty, as previously mentioned, the main effect of uncertainty did not reach significance for either amplitude measure (p=.181 and p=.253 respectively).

Figure 24 - Average N1 peak and mean amplitude (left and right respectively) for both low and high uncertainty conditions averaged across similarity conditions. Error bars represent +1 SD.



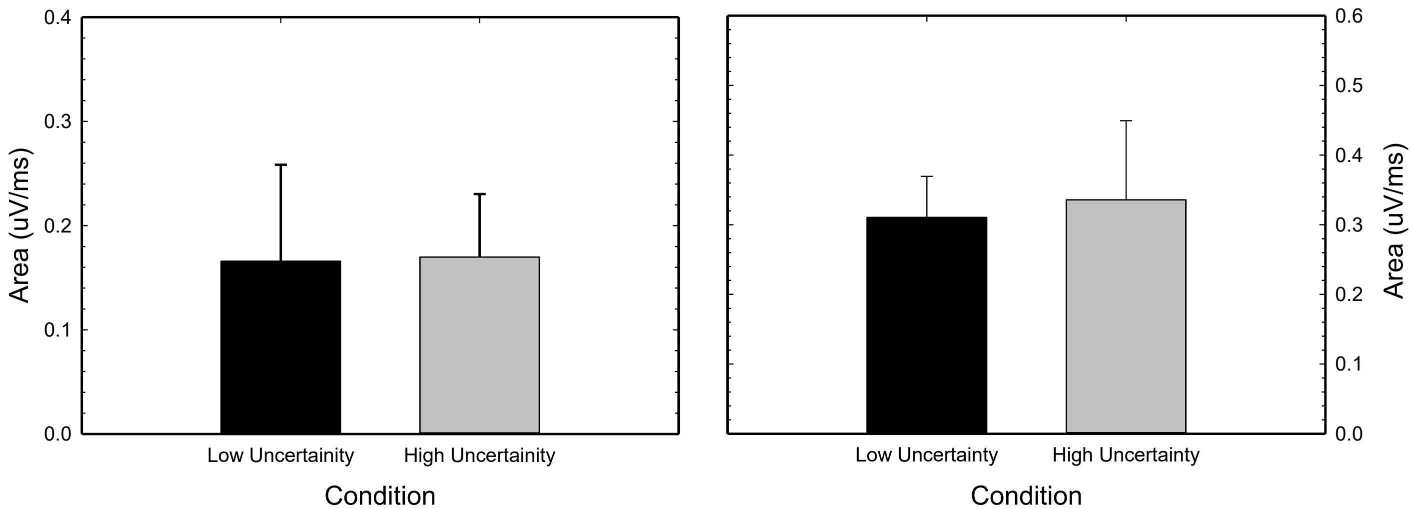
N1 peak latency, on the other hand, did show an overall significant effect of uncertainty (p=.025). This effect can be seen in Figure 25, which shows that N1 peak latency increased from low to high uncertainty. This was consistent with the hypothesis that an increased difficulty on an auditory task manipulated by TTI, would increase the stimulus processing time even at the earlier level of the N1, as well as increasing the attentional resources required for the task.

Figure 25 - Average N1 peak latency for low and high uncertainty conditions averaged across similarity conditions. Error bars represent +1 SD.



Area outcomes for the P1-N1-P2 are shown in Figure 26, with the N1 peak area and overall P1-N1-P2 area for each uncertainty condition averaged across all similarity conditions. The main effect of uncertainty on N1 or P1-N1-P2 area did not reach significance (p= .692 and p=.307 respectively).

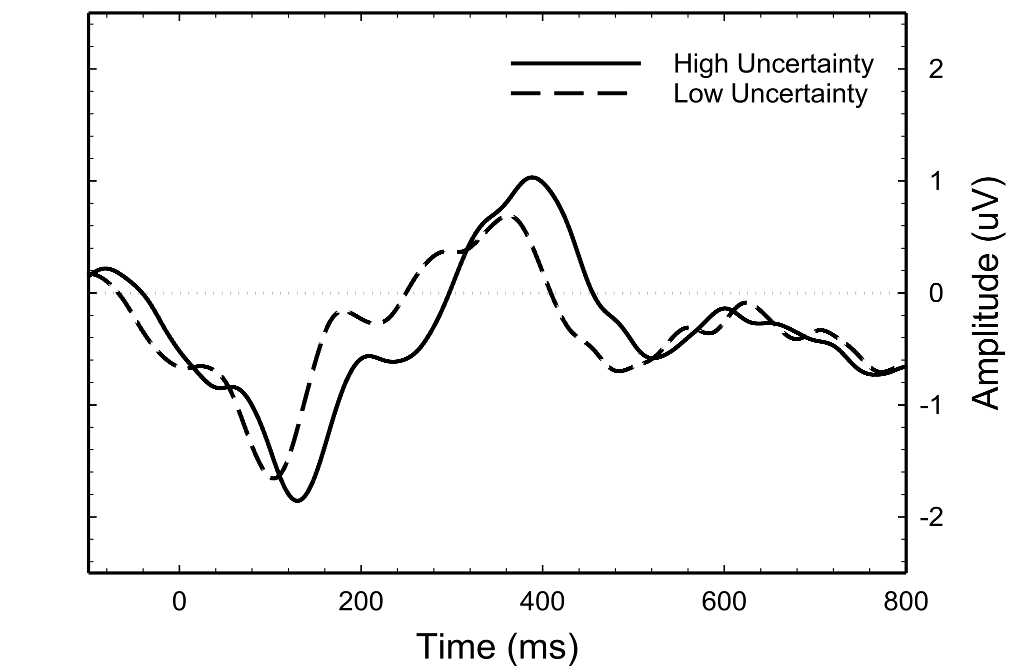
Figure 26 - Average N1 (left panel) and P1-N1-P2 area (right panel) measures for low and high conditions averaged across similarity conditions. Error bars represent +1 SD.



3.5.2 P3 Uncertainty

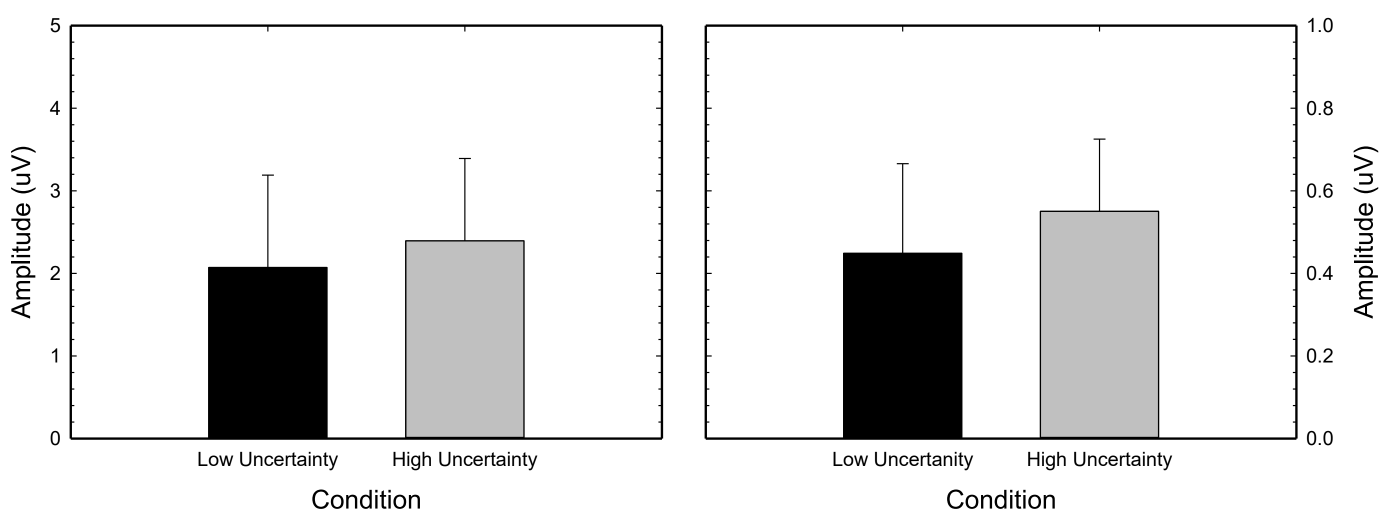
Effects of uncertainty on the later P3 component were summarized previously as part of the overall ANOVA results in Table 8. There were significant main effects of uncertainty on P3 latency and area overall, but not for P3 amplitude. Figure 27 shows the grand mean waveforms for each uncertainty condition averaged across all similarity conditions to visualize these results. The significant increase in latency between low and high uncertainty is observed in these grand means.

Figure 27 - Grand mean waveforms for the P3 response to target stimuli at Pz for high and low uncertainty averaged across all similarity conditions.



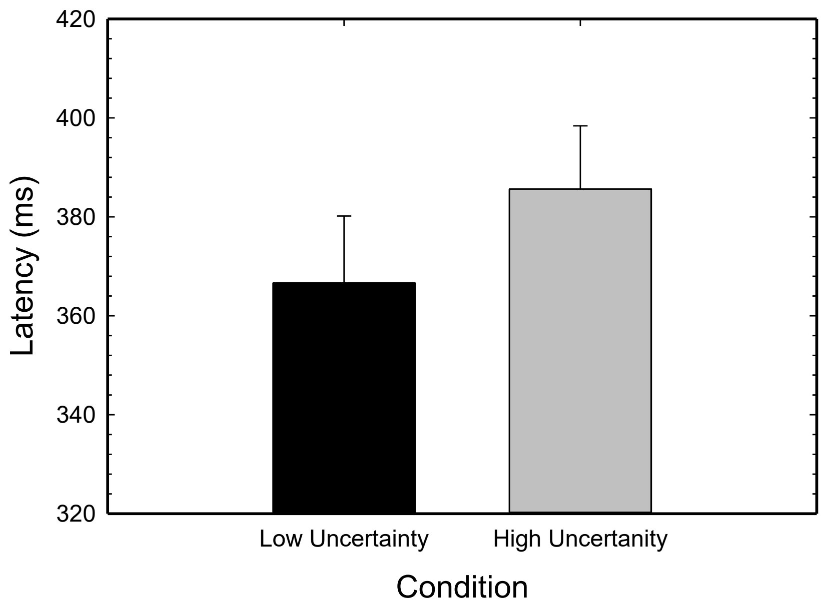
Average P3 amplitudes across individuals for both peak and mean measures across uncertainty conditions are shown in Figure 28. Although the mean P3 amplitude increases between low and high uncertainty for both measures, this effect did not reach significance (p=.203 and p=.771 respectively).

Figure 28 - Average P3 amplitude for peak (left panel) and mean (right panel) measures for both uncertainty conditions averaged across similarity conditions. Error bars represent +1 SD.



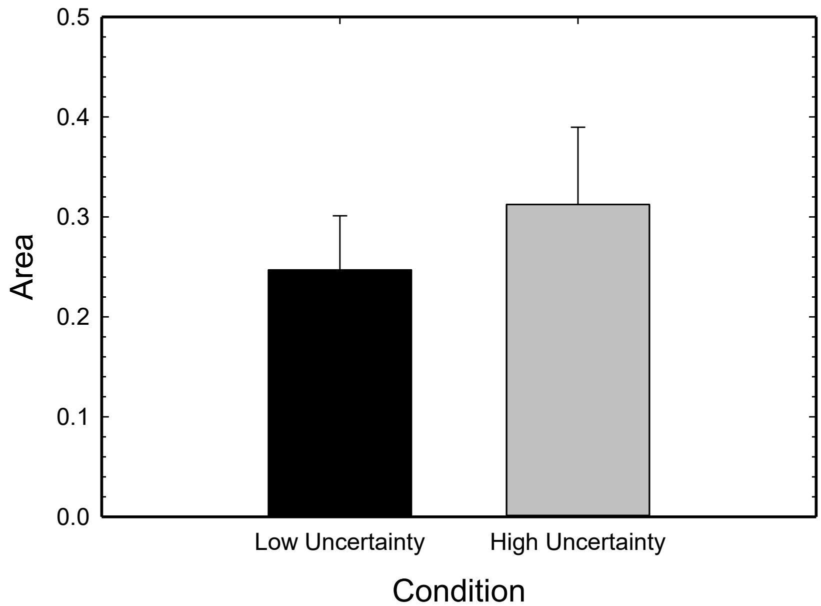
P3 latency, alternatively, did show a significant effect of uncertainty as shown in Table 8 (p=.005). Mean individual peak latency data in Figure 29 show this significant increase in latency from the low uncertainty to the high uncertainty condition. This significant effect for P3 latency is consistent with the hypothesis that a harder auditory task would increase latency of the P3 response due to increased time in categorizing and discriminating stimuli

Figure 29 - Displays N1 peak latency measures for uncertainty conditions averaged across similarity. Error bars represent +1SD.



Finally, the effect of uncertainty on P3 area is shown in Figure 29, with increasing average response area from low to high uncertainty. This effect was significant for P3 area (p=.031) is consistent with our hypothesis that a harder auditory task would yield larger neural responses, combined with increases in latency.

Figure 30 - Average P3 area for low and high uncertainty conditions averaged across similarity. Error bars represent +1 SD.



3.6 Comparison between AEP and behavioral outcomes

3.6.1 Reaction Time, Accuracy, and d'

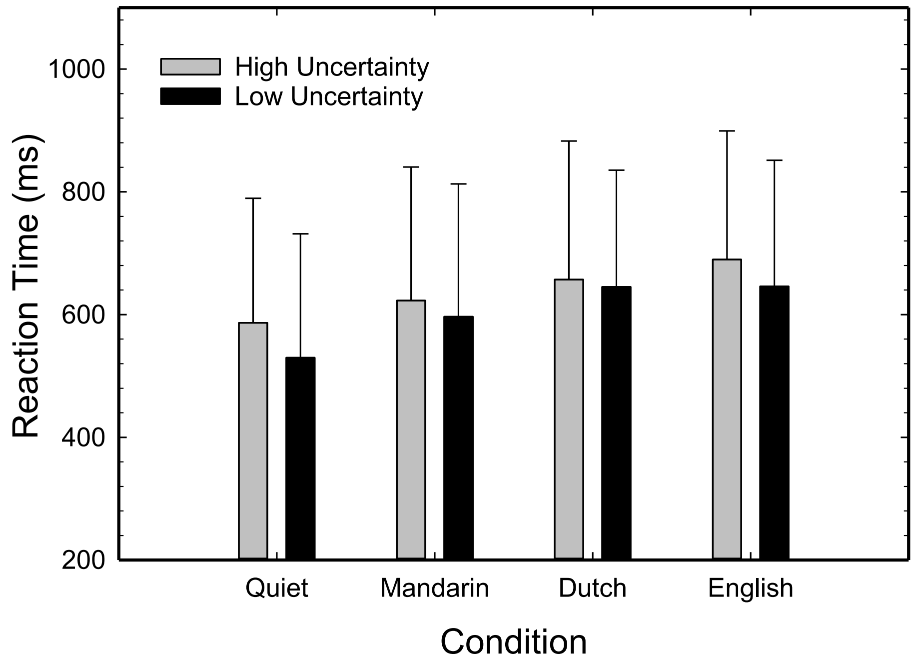
For the AEP oddball task, participants actively attended to and pressed separate buttons for each presentation of both the standard and target stimuli. The reaction time for each button press following stimulus presentation and whether it was a correct or incorrect response were recorded. These behavioral outcomes were recorded for all eight conditions of similarity and uncertainty for 26 of the 30 participants. For 4 of the participants, button press data was not accurate for all stimulus presentations because of insufficient pressure applied to the center of the response button. Once this problem was determined, subjects were instructed on how to press the button correctly so that all responses would be recorded correctly. Even though a full set of data were not available for these four subjects, there were at least 20-60 recordable responses in each condition that were accurately recorded, and the data was evaluated and considered reliable enough across this more limited number of responses to include in the overall analysis. Table 9 displays results of reaction time, accuracy, and d' of the button press response. Of note, large effect sizes in accuracy and d' across both effects of similarity and uncertainty were observed.

Table 8- Results of the overall 4x2 RMANOVA for reaction time, accuracy and d' measures. p- values, F-values, and effect sizes (partial eta squared) are shown for each of the main effects and the interaction term. Degrees of freedom for the F statistic were (3,87), (1,29), and (3,87) for similarity, uncertainty, and interaction terms respectively. Significant results ($p<.05$) are highlighted in gray.

	Reaction Time			Accuracy			d'		
	p	F	η_p^2	p	F	η_p^2	p	F	η_p^2
Main Effect of Similarity	**.001**	**6.16**	**.175**	**.000**	**40.9**	**.585**	**.000**	**8.71**	**.231**
Main Effect of Uncertainty	**.037**	**4.72**	**.141**	**.000**	**69.2**	**.732**	**.000**	**14.5**	**.334**
Interaction	.797	0.34	.012	.620	.595	.020	.447	.896	.030

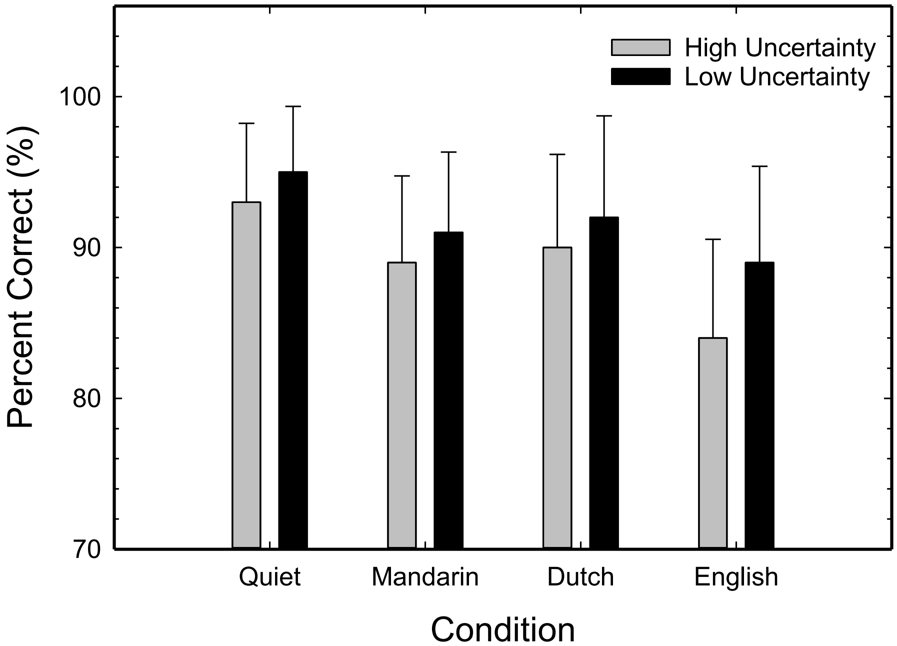
Mean reaction times to target stimuli in each language masker and uncertainty condition are shown in Figure 31. A factorial 4x2 RMANOVA on reaction time revealed a significant main effects of similarity (p=.001) and uncertainty (p=.037) but the interaction was not significant (p=.797). Pairwise comparisons revealed that reaction times were significantly shorter overall for quiet compared to either the Dutch (p=.012) or English (p=.002) masker conditions, but not significantly different from the Mandarin condition (p=.291). Reaction times were not significantly different among any of the three language conditions (p>.05). Reaction times were faster overall for the low uncertainty task compared to the high uncertainty task (p=.037), which can also be observed within each similarity condition.

Figure 31 - Average reaction time in for high and low uncertainty conditions in each similarity. Error bars represent +1 SD.



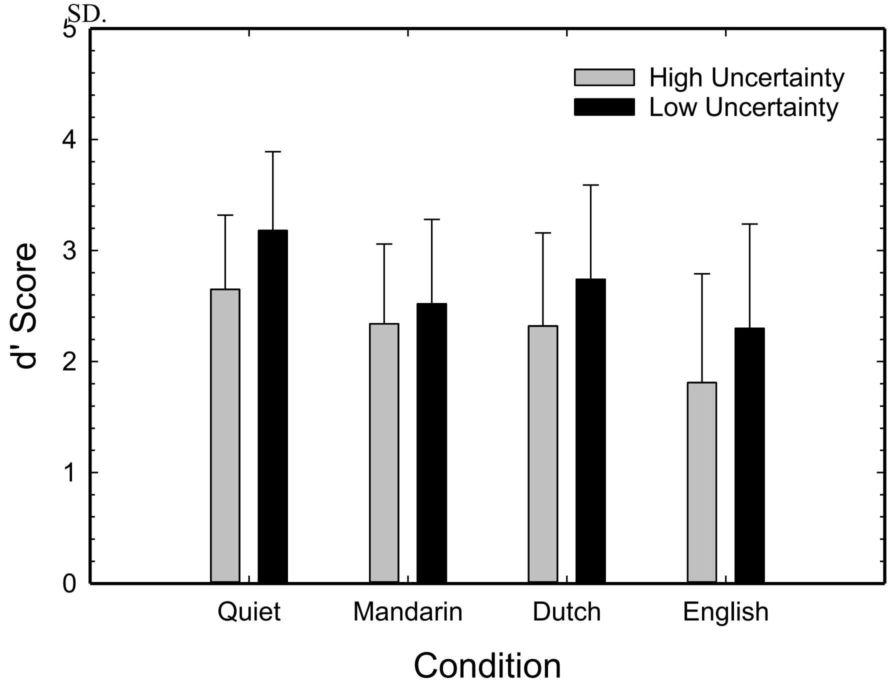
The accuracy of button press identification of target stimulus measured in percent correct is shown in Figure 32. As shown in Table 9, percent correct scores, revealed a significant main effects of similarity ($p<.001$) and uncertainty ($p<.001$) but not a significant interaction ($p=.602$). Mauchly's test of sphericity was significant, indicating that the assumption of sphericity was violated, and Greenhouse-Geisser corrections were applied. Pairwise comparisons revealed that accuracy was highest in the quiet condition compared to all three background noise conditions ($p\leq.004$). Accuracy of target identification was not significantly different between Mandarin and Dutch conditions ($p=.150$), but was significantly worse in English compared to either of the other language masker conditions ($p\leq.001$). Overall accuracy was also significantly higher in the low uncertainty conditions compared to high uncertainty conditions ($p=.001$).

Figure 32- Average accuracy scores (% correct) for button press identification of the target stimulus for high and low uncertainty conditions within each similarity condition. Error bars represent +1 SD.



As described in the methods, because percent correct scores may have a ceiling effect and not accurately represent the participants difficulty in identifying standard and target stimuli in each condition, d' was calculated as a measure of behavioral discrimination. Figure 33 shows mean d' scores for each similarity and uncertainty condition. In general, average behavioral discrimination between target and standard was better for low compared to high uncertainty across all similarity conditions. As displayed in Table 9, d' scores showed significant main effects of similarity ($p<.001$) and uncertainty ($p<.001$), but the interaction was again not significant ($p=.447$). Pairwise comparisons revealed that d' scores were significantly higher for the quiet condition compared to either the Mandarin ($p=.011$) or English ($p=.003$) conditions, but were not significantly different from d' scores for the Dutch condition ($p=.219$). However, d' scores for the Dutch condition were significantly higher than those for the English condition ($p=.024$). The difference between d' scores for Mandarin and Dutch conditions was not statistically significant (p>.05). Comparisons of uncertainty showed that d' in low uncertainty conditions yielded significantly higher d' compared to high uncertainty. In general, therefore, it was most difficult for participants to perform the task of detecting target stimuli in the English language background masker, easiest in quiet, and no different between the two non-native language maskers. In addition, across all language maskers, it was easier to detect target stimuli in the low uncertainty conditions compared to the high.

Figure 33 - Average d' behavioral discrimination scores for high and low uncertainty with each similarity and uncertainty condition. Error bars represent +1 SD.

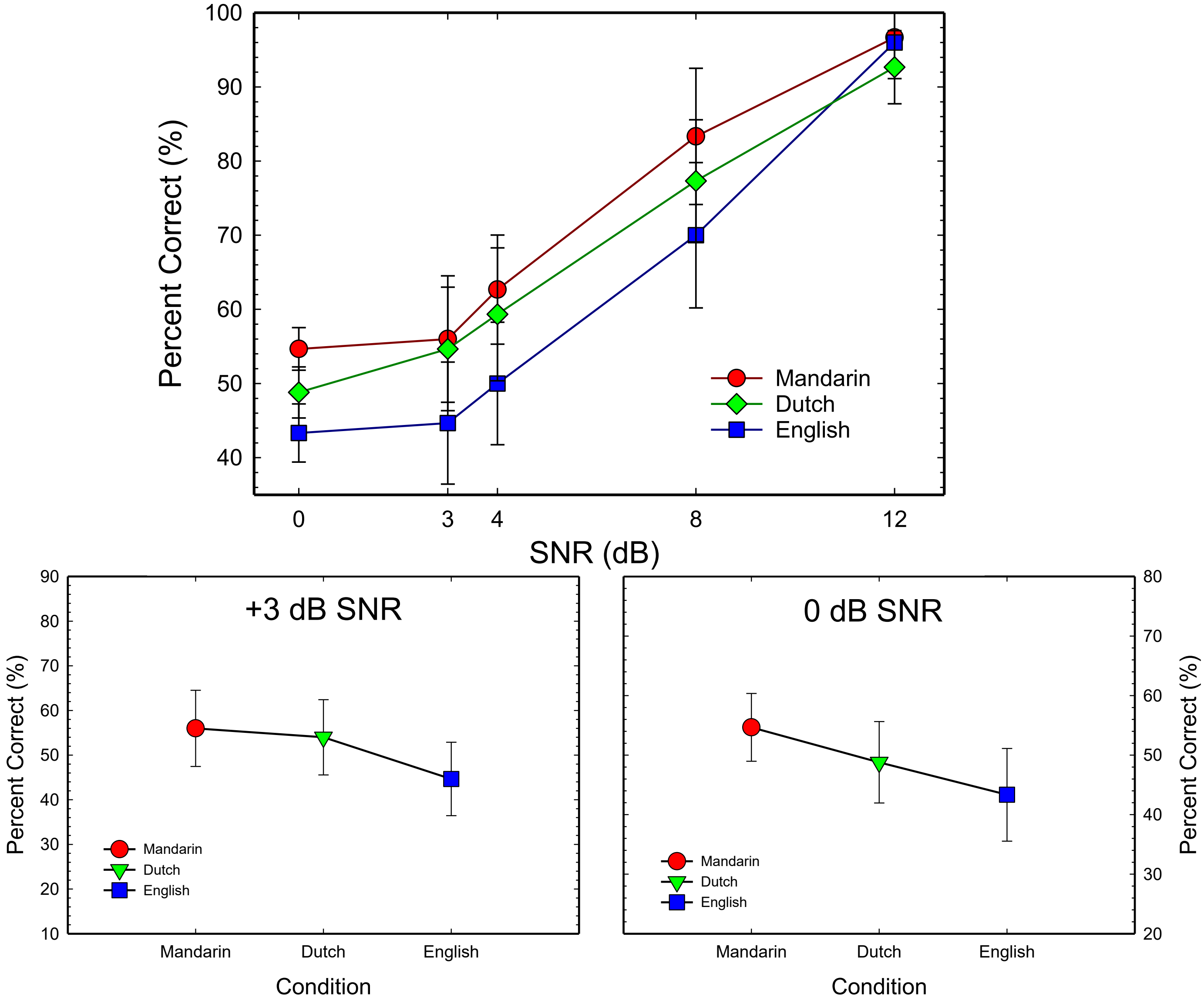


3.6.2 Word recognition in noise

The effect of informational masking on a behavioral speech recognition performance task was also assessed for the effect linguistic-phonetic similarity between target and masker. While the same Mandarin, Dutch and English language masker noises were used, the task differed from the AEP task in a number of ways. Target stimuli were NU-6 words (in English) and the SNR of the masker varying from more favorable (+12) to less favorable (0) compared to the +3 dB SNR condition used for AEPs. Five NU-6 words were presented in SNRs of +12, 8, 4, and 3 dB, in addition to 25 words in 0 dB SNR. As discussed in the methods, multiple SNRs were presented to calculate the SNR50. There was no manipulation of uncertainty for this task.

Results for percent correct word recognition performance in each of the language masker conditions by SNR are shown in Figure 34. In general, performance improved for SNRs better than 4 dB for all maskers. Results for +3, the same SNR as the AEP task, and the more difficult 0 dB SNR condition are shown in the bottom panel of the figure. Performance for these two and all but the 12 dB SNR condition, followed a trend of higher percent correct in the presence of the Mandarin masker, followed by the Dutch masker condition, and the poorest performance in the English masker condition.

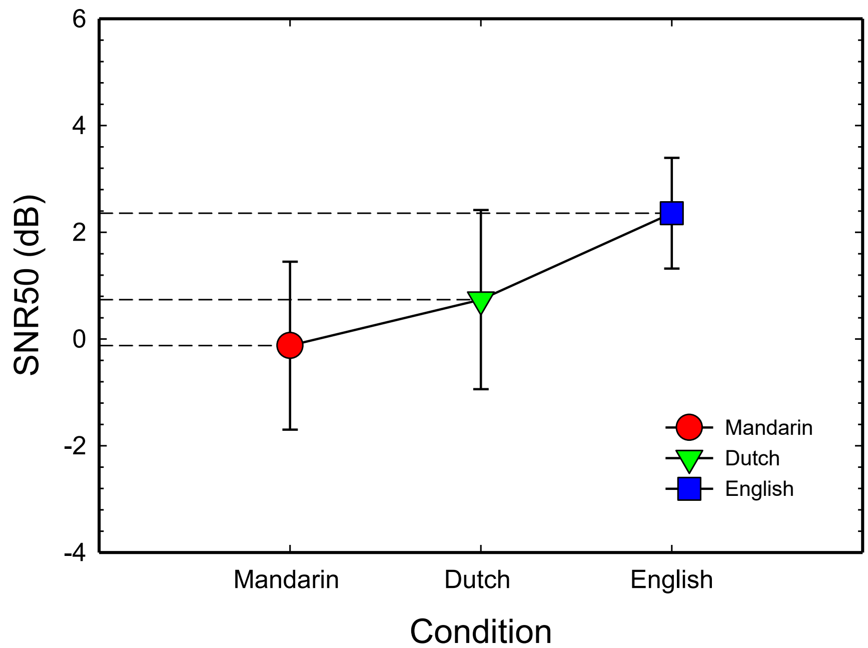
Figure 34 - Average percent correct word recognition for each language masking condition for all SNR levels (top panel). The bottom panels highlight results for the +3 (bottom left) and 0 dB SNR (bottom right) conditions in each language masker condition. Error bars represent $\pm$1 SD.



A factorial 5x3 (SNR by similarity condition) RM-ANOVA was conducted on these behavioral percent correct scores. Because Mauchly's test of sphericity indicated a significant violation of sphericity, Greenhouse-Geisser corrections were used. The main effects of SNR ($p=<.001$, $F(2.76, 80.2)=201.4$, $\eta_p^2=0.874$) and language masker ($p=<.001$, $F(0.186, 57.3)=18.27$, $\eta p^2=0.387$) were significant with large effect sizes, but there was not a significant interaction between these two factors ($p=.186$, $F(4.99, 144.7)=1.43$, $\eta_p^2=.047$). Pairwise comparisons for the effect of SNR revealed a significant decrease performance between +12 and +4 dB SNR, with all comparisons significant ($p\leq.001$). The overall difference between +4 to +3 ($p=.139$) and between +3 to 0 dB SNR ($p=.753$) was not significant. Pairwise comparisons for the effect of similarity condition revealed that overall performance for Mandarin and Dutch were not significantly different ($p=.069$), but overall performance in the English condition was significantly worse than either Dutch or Mandarin ($p=.001$, each). SNR50 was calculated for each individual for each language masker to estimate the level at which subjects could repeat back 50% of the words correctly. SNR50 has been used in previous studies of linguistic/phonetic masking and was used in this study to as a comparison and to equalize performance rather than choose a fixed SNR over which performance may vary. Figure 35 shows calculated SNR50 for each noise condition (error bars = $\pm$1 SD).

One-way RM-ANOVA (3 language masker conditions) revealed a significant effect of masker similarity on SNR50 ($p<.001$, $F(2,56)=22.0$, $\eta_p^2=0.441$) with a large effect size. Post hoc analysis revealed that SNR50s were not significantly different between Mandarin and Dutch conditions ($p=.076$), but SNR50 was significantly higher for English (a more favorable SNR was required to achieve 50% correct) compared to either Mandarin ($p<.001$) or Dutch ($p=.001$).

Figure 35 - Average SNR50 behavioral word recognition across language maskers. Lower scores indicate better performance (50% correct achieved in a less favorable SNR). Error bars represent +1 SD.



To establish whether there were significant relationships between AEP outcomes and behavioral word recognition, six stepwise linear regressions were conducted. Each of the measured AEP variables (N1 and P3 peak and mean amplitude, N1 and P3 peak latency, N1 and P3 area, P1-N1-P2 area, reaction time, accuracy and d') were added into the model to establish which significantly accounted for the variance in behavioral performance outcomes within each of the language masker conditions separately for SNR50 and at the same +3 dB SNR as used for the AEP task. Figure 36 shows Pearson correlation coefficients for all AEP variables entered into the model and the behavioral outcomes of SNR50 and % correct at +3 dB SNR. This figure was included to visualize all predictor variable correlation coefficients instead of individual plots. Dotted lines across -.4 and .4 are shown to indicate moderate correlations that were found. Note that Figure 36 shows Pearson correlation coefficients for all predictor variables entered into the model, but significant predictors are summarized in Table 10.

For the Mandarin masker, only button press accuracy in the high uncertainty condition was a significant predictor (p=.003), which accounted for 28% of the variance in SNR50 scores. None of the other variables significantly increased the r^2 ($p>.05$) for prediction of SNR50. For word recognition scores at +3 dB SNR, no variable resulted in a significant prediction in the model summary. For the Dutch masker, two variables significantly contributed to the prediction of SNR50, N1 peak amplitude (p=.012, r^2=.218) and N1 latency (p=.032, r^2= .134) for the Dutch high uncertainty conditions accounted for 35% of the variance. For +3 SNR, no variables were significant predictors of performance in the model summary. For the English masker, P3 area for the English high uncertainty condition (p=.007), was a significant predictor of SNR50 (p-value), accounting for 23% of the variance. None of the AEP variables were significant for +3 dB SNR for the English condition.

Table 9 - Displays significant predictor variables for stepwise linear regressions for Mandarin, Dutch, and English SNR50. p-value, r^2 value and correlation coefficients are also listed. Regressions for +3 dB SNR are not listed due to lack of statistical significance.

	Stepwise Linear Regressions			
	Mandarin	Dutch		English
	SNR50	SNR50		SNR50
Significant Predictor Variables	Accuracy for Mandarin high uncertainty	N1 Peak Amplitude for Dutch high uncertainty	P3 Latency for Dutch low uncertainty	P3 Area for English high uncertainty
p-value	.003	.012	.032	.007
r^2 value	.280	.218	.134	.233
Correlation Coefficient	-.530	-.467	.407	-.483

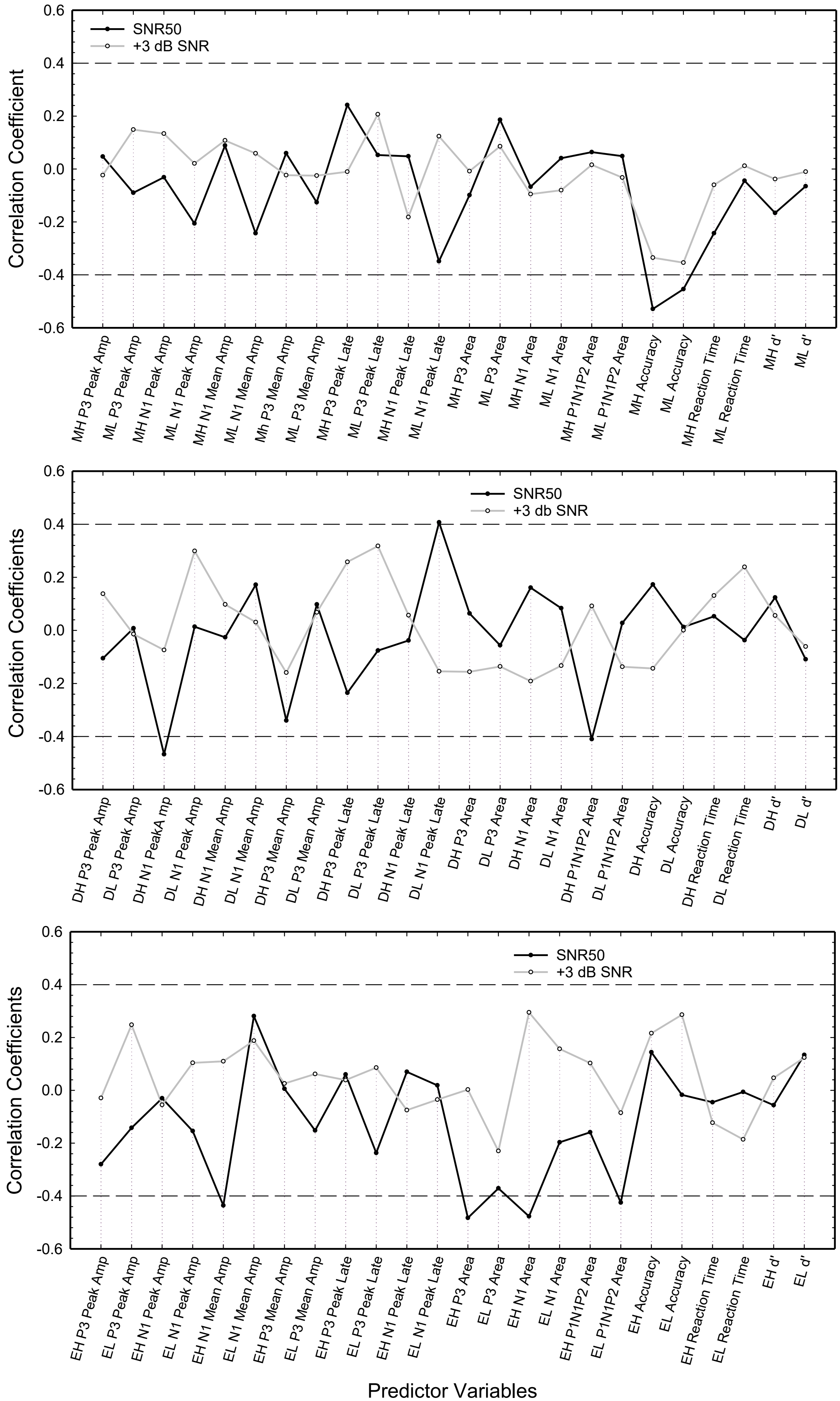
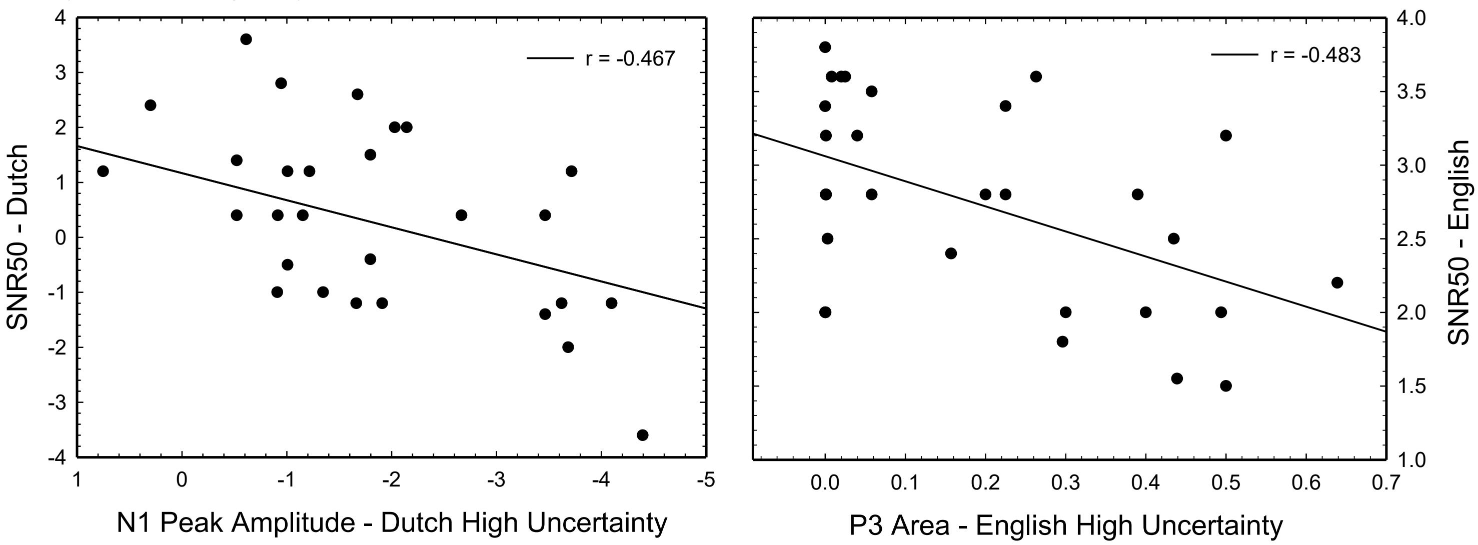


Figure 36 - Pearson correlation coefficients for all variables entered into the stepwise linear regression for Mandarin (top panel), Dutch (middle panel), and English (bottom panel). Naming convention for each of the predictor variables followed the language masker (first letter) then uncertainty condition (second letter). For example, MH stands for Mandarin high uncertainty. Correlation coefficients are depicted for SNR50 in black and +3 dB SNR in grey. Dotted lines show moderate correlations.

While a few AEP outcomes were found to serve as significant predictors of behavioral performance, however patterns of correlations were not consistent across language maskers. N1 amplitude and latency were significantly related to performance for Dutch, such that larger and earlier N1s were associated with better performance. P3 latency provided the strongest correlation with any of the aspects of behavioral performance, with shorter latencies associated with higher performance, but only for the English masker condition SNR50. Figure 37 shows correlations for N1 peak amplitude in the Dutch high uncertainty condition for Dutch SNR50 and P3 area in the English High uncertainty for English SNR50.

Figure 37 - Correlations for significant predictors variables. Left panel shows N1 peak amplitude in the Dutch high uncertainty condition(x-axis) by SNR50 in Dutch (y-axis). The right panel shows P3 area in the English high uncertainty condition (x-axis) by SNR50 in English (y-axis).



4.0 Discussion

4.1 General Discussion

The purpose of the present study was to objectively identify the effects of two factors of informational masking on neural information processing along the auditory pathway. In order to further understand informational masking deficits and provide converging evidence with behavioral findings, the goals of the current project were twofold. The first was to determine how informational masking factors of increased target-masker similarity and stimulus uncertainty alter neural processing of speech information at earlier and later stages of top-down auditory processing using temporally sensitive auditory electrophysiology measures. The second was to characterize the relationship between physiological AEP outcomes of informational masking and behavioral performance on a word recognition task. This was the first study to manipulate the aspects of linguistic/phonetic similarity between target and masker speech while also experimentally controlling stimulus uncertainty via timing of target stimuli in the same paradigm. In addition, this was the first study to use objective indices to identify the neural processing time course within the auditory nervous system of these factors related to informational masking. The main findings of the study as they are related to target-masker similarity, stimulus uncertainty, and the interaction between these factors are discussed below, as well as the relationship between these AEP outcomes and behavioral performance in the same participants. The strengths and limitations of the study, and directions for future research are presented.

The present results support an effect of linguistic/phonetic similarity and stimulus uncertainty on AEP responses and linguistic/phonetic similarity on behavioral word recognition. Several trends in AEP and behavioral outcomes were consistent with the hypothesized hierarchy of increasing linguistic/phonetic similarity from Mandarin to Dutch to English, but not all differences were significant. The most supported findings for this factor were that all babble maskers significantly affected outcomes compared to quiet, and that the native language English masker had the largest effect on outcomes in the AEP paradigm, including N1 amplitude, P3 amplitude and area, as well as decreased reaction time, accuracy, and d' behavioral discrimination to target word responses. AEP outcomes for the Mandarin and Dutch maskers, however, were not significantly different. Outcomes for AEP latencies for both N1 and P3 also supported an effect stimulus uncertainty, consistent with the hypothesized increase in processing time related to increased task complexity when target stimulus timing was randomized. An unanticipated result was the absence of the expected additive effect between linguistic/phonetic similarity and stimulus uncertainty. None of the AEP outcome measures demonstrated a statistically significant interaction effect in the RMANOVA analyses. Finally, trends in behavioral word recognition performance were generally consistent with those observed for AEP component measures. Behavioral word recognition at +3 dB SNR, the same SNR used in the AEP paradigm, was no different in the Dutch and Mandarin maskers, but the English masker yielded the lowest percent correct scores. Furthermore, correlations between behavioral word recognition and AEP component measures yielded some moderate correlations, but no common AEP components accounted for a majority of variance for behavioral word recognition.

4.2 Effect of linguistic/phonetic similarity on neural indices of cortical auditory processing

The presence of a masker and the degree of target-masker similarity affected auditory neural processing. Specifically, significant effects of similarity were found on AEP outcomes for

N1 and P3 where the target was an English word (native-language) and the maskers were linguistically and phonetically similar (English talkers), linguistically dissimilar but phonetically similar (Dutch talkers), and linguistically and phonetically dissimilar (Mandarin talkers). That is, the findings of this study supported a target-masker similarity effect at both earlier and later levels of top-down auditory processing.

4.2.1 N1 Similarity

At the earlier level of processing, there were large effects on the morphology of the response in the noise conditions compared to the quiet condition. N1 amplitudes were significantly decreased and N1 latencies significantly prolonged by the presence of any of the three language maskers. This result is consistent with a body of previous literature and consistent with the N1 indexing an early orienting response of acoustic change in the stimulus (Billings et al. 2011; Culling 2013; Koerner et al. 2017; Niemczak and Vander Werff 2019; Sharma et al. 2014; Stapells 2008; Wong et al. 2008). When noise of any kind is present, but particularly speech babble noise, the N1 response is known to be decreased and delayed (Billings et al. 2011; Niemczak and Vander Werff 2019; Vander Werff and Arduini 2011; Vander Werff et al. 2016).

One possible interpretation of the current results is that Mandarin and Dutch speech babble maskers provided less informational masking, and functioned as energetic maskers, because they were not understandable to the listeners. Several of the previously mentioned studies have directly compared the effects of speech babble noise to energetic maskers such as continuous white noise or speech-shaped noise. These results are consistent with speech babble masking having a larger effect on P1-N1-P2 responses compared to continuous noise maskers. Unlike continuous white noise or speech-shaped noise, Mandarin and Dutch babble have temporal envelopes similar to the English masker, which could comparable to interrupted or modulated noise maskers. The effects of non-speech noise with temporal gaps are not consistent with the same informational masking effects as speech babble in either behavioral or electrophysiological research. For example, Cooke (2006) demonstrated improved behavioral speech recognition scores in various temporally modulated energetic masking conditions compared to continuous noise maskers, and that the proportion of the glimpses in a noise source is a good predictor of intelligibility. In contrast, Brungart et al. (2001) found poorer sentence recognition in normal-hearing young adults in the presence of babble masking compared to energetic speech-shaped noise modulated to mimic the temporal variations of the babble. That is, noise with the same spectral shape and temporal fluctuation of babble maskers provided the listeners an advantage (masking release) over continuous noise, but speech babble degraded performance through informational masking. In AEP studies, noise that is interrupted or has temporal envelopes similar to speech babble has not been found to have the same effects as speech babble on P1-N1-P2 amplitudes and latencies. Billings et al. (2011), for example, found decreased N1 amplitudes for the response to both tone and speech (/ba/) stimuli in a 4-talker babble condition compared to either continuous or interrupted noise. Differences in amplitude or latency between the interrupted and continuous noise conditions were not significant in either an active or passive recording paradigm. Similarly, Zhang et al. (2014) showed reduced amplitude of N1 to a speech target /bi/ in 2-talker babble compared to either modulated or steady state energetic noise in an active paradigm. Vander Werff et al. (2016) recorded the P1-N1-P2 to /ba/ in speech babble with two to eight talkers, continuous SSN, and SSN modulated using the same temporal envelope as each of the multi-talker babble maskers. Results showed greater effects of speech babble on the entire P1-N1-P2 responses, including significantly greater reductions in amplitude and latency of the N1 and P2 peaks compared to the enveloped noise conditions.

Babble noise degraded the overall response morphology considerably, while the envelope noises resulted in reduction in amplitude and increase in latency, but typical P1-N1-P2 morphology. P1-N1-P2 responses in all three language maskers in the current study are morphologically similar to the responses in babble in the previous experiment, rather than to the energetic modulated masker conditions, which argues against Dutch and Mandarin acting as purely energetic modulated noise maskers. While Mandarin and Dutch babble maskers therefore appear to have contributed some informational masking, though to a lesser extent than English, a purely energetic masker condition was not implemented in the study to verify this conclusion. In a future study, a direct comparison of spectrally and temporally matched enveloped speech noise and the foreign language maskers would be needed to determine whether they provide additional informational masking beyond the energetic content. The current study adds to the existing literature on auditory neural processing of speech in noise at the level of the P1-N1-P2 by using 2-talker babble and three different language maskers and recording responses to word stimuli rather than syllables in an active oddball paradigm. The convergence of evidence across studies, therefore, demonstrates that N1 amplitude is sensitive to informational masking in that it is reduced when the masker is composed of multi-talker babble.

The presence of any background babble also delayed N1 latencies suggesting an influence of the masking noise on stimulus processing time at this earlier level of top-down auditory processing. These results are also consistent with multiple studies (Bennett et al. 2012; Billings et al. 2011; Niemczak and Vander Werff 2019; Zhang et al. 2016). For example in our previous study using the acoustic change complex (ACC) in a passive listening paradigm to /u-i/ stimuli in SSN (energetic masker), 8-talker, and 2-talker babble, we found that all background noise conditions significantly delayed latency of N1, but in the babble conditions more than the energetic maskers (Niemczak and Vander Werff 2019). In addition, Zhang et al. (2016) found that in both a passive and active paradigm recorded to speech stimuli, N1 latency was delayed more for 2-talker babble compared to continuous noise at all SNRs tested (4, 0, -4, -8 dB SNR). That is, in both passive and active listening paradigms, N1 latency is generally affected by the presence of babble noise compared to energetic masking regardless of linguistic/phonetic similarity.

Area measures also supported the overall effect of informational masking provided by babble noise. The main result of decreased P1-N1-P2 area across all language maskers is most likely the result of not only decreased N1, but the large change in the P2 component between quiet and noise conditions. P2 in the quiet condition dominated the area measure, while P2 amplitudes in language masking conditions were greatly reduced and had increased variability. While the P2 component was not a primary outcome in this study, the area measure was used to account for this effect. Decreased P2 amplitude in speech babble masking, in addition to reduction in N1 has been reported in numerous studies (Billings et al. 2011; Billings et al. 2009; Niemczak and Vander Werff 2019; Vander Werff et al. 2016; Zhang et al. 2016). Area measures taken from our previous study (Niemczak and Vander Werff 2019) showed that both levels of informational masking (8- and 2-talker) resulted in decreased area compared to continuous speech-shaped noise, but there were no area differences by number of talkers in the babble. The previous study was a passive listening paradigm compared to the active oddball paradigm for the current study. Nevertheless, these results are consistent with the previous literature suggesting that speech babble maskers provide more masking than continuous noise makers, which is consistent with an informational masking effect.

While supporting a general informational masking effect of babble noise, outcomes for the hierarchy of linguistic/phonetic similarity across languages at the N1 level were only partially consistent with hypotheses. N1 peak amplitudes in English were the smallest compared to all other conditions. That is, the native language masker caused the largest reduction in the neural response for preconscious detection of stimulus features and orienting to the English words. N1 amplitude between the non-native language masker conditions, however, was not significantly different, suggesting that the similarity contrasts between these non-native languages did not have an effect at this level.

Although there was a significant overall effect of similarity, the results for N1 latency were also not clearly consistent with the hypothesized hierarchy of linguistic/phonetic similarity between target and masker. The overall effect was only quiet vs. babble, as across the different language maskers there no significant differences in latency. Only N1 latency for the Dutch masking condition was significantly longer in latency compared to quiet. This result does not follow expectations that the most similar English condition would result in the longest latency and Mandarin, with the least similarity to the target English words, would result in the earliest latency. In fact, in the grand mean waveforms and mean bar graphs (e.g. Figure 13 and 15) the N1 peak for the English masker condition is actually slightly earlier, than for either Mandarin or Dutch. This non-significant result may relate to individual variability and generally poorer waveform morphology in the noise conditions, particularly those for the English masker which had the smallest amplitudes. It is likely absolute peak latency measures were not sensitive to differences between noise conditions due to this. There is also a possibility that the active paradigm used in this experiment had an effect on N1 latency measures such that the English condition required more attention, or that the non-native languages were more distracting. While

N1 amplitude increases in an active compared to an purely passive listening condition have been demonstrated fairly consistently, the effects of active attention on N1 latency are inconclusive in the literature, such that latency is unchanged between active and passive conditions (Michalewski et al. 2009; Picton et al. 1974; Woldorff et al. 1993; Zhang et al. 2016; Zhang et al. 2014). However, Zhang et al. (2014) found conflicting results on N1 latency in their study using an active and passive paradigm comparing between modulated noise to 2-talker babble. They found significantly longer latencies for babble compared to the modulated noise, but only in the active condition. The results suggest that the attentional effects on the cortical representation of the speech stimulus was masker specific. However, in the current study, the latency results for N1 also supported a difference in processing time at this earlier level of processing for all 2-talker language maskers, but not between the maskers as an effect of linguistic/phonetic similarity. Addressing some of the limitations in absolute peak measures, overall P1-N1-P2 showed a decrease in area of the entire response complex in the presence of language maskers compared to quiet, but there were no significant differences in N1 or overall P1-N1-P2 area among the three language masker conditions. Therefore, only the results for N1 amplitude in English partially supported this hypothesis at the early cortical level.

Amplitudes in Mandarin and Dutch conditions did not differ, which may suggest that because they are not understandable to the listener, they are more similar to modulated noise in terms of the effects on neural responses. However, the general similarity in N1 component morphology across all language maskers that was not observed in previous published comparisons between babble and modulated noise maskers suggests that theses speech maskers were still influencing the neural encoding of the targets in a different way from purely energetic maskers. Future experimental designs should include purely energetic maskers for comparisons of energetic and informational masking, as well as examine the effects of “understandability” of the masker. Overall, these changes in the N1 response to all language maskers suggest at least some decrease in the ability to encode stimulus features and perhaps selectively orient to standard stimuli in the presence of multi-talker babble masking consistent with an informational masking effect.

4.2.2 P3 Similarity

At later level of auditory processing, all P3 outcomes (amplitude, latency, and area) were significantly affected by the overall manipulation of target-masker similarity in this paradigm. In addition to responses in all language masking conditions being significantly different from quiet, some of the findings were consistent with the English language masker having a largest effect compared to the non-native language maskers.

P3 peak and mean amplitude as well as area outcomes were consistent with reduced P3 amplitudes/area in babble noise compared to quiet. P3 latency was also significantly increased for the all masking conditions compared to quiet. These results are overall consistent with previous studies that have recorded P3 in active paradigms under speech noise conditions. Several studies have demonstrated that P3 amplitude is reduced in speech masking compared to quiet and compared to non-speech masking conditions (Bennett et al. 2012; Kaplan-Neeman et al. 2006; Koerner et al. 2017; Whiting et al. 1998). Both Bennett et al. (2012) and Koerner et al. (2017), used 4-talker babble to mask speech syllable targets in active paradigms. Both studies found significantly decreased P3 amplitude and increased P3 latencies in those conditions compared to quiet. In addition, Bennett et al. (2012) found that 4-talker babble compared to continuous noise maskers resulted in further reductions in amplitude and increases in latency compared to continuous noise.

These results and the current study are consistent with an informational masking effect of babble noise in general, regardless of language. P3 amplitude indexes attentional resources and the cognitive demands of the task, while latency is usually interpreted as the speed of processing the difference between stimuli. In general, P3 amplitude would be expected to increase as more attentional resources are required, and P3 latency would be expected to increase as the time to process and categorize/discriminate stimuli increase (Hillyard et al. 1971; Picton 1992; Polich 2007). While the latency prolongation in noise is consistent with this expected effect, P3 amplitude decreased rather than increased. This result is likely consistent with difficulty in detecting and discriminating stimuli in noise associated with a reduction in neural synchrony under babble maskers. Stimuli near threshold or with small differences near detection thresholds (like small frequency and intensity differences) evoke smaller and later P3s (Gonsalvez et al. 1999; Hillyard et al. 1971; Hink et al. 1976; Polich 1987). Therefore, this effect may be dominated by difficulty discriminating the stimuli in all the babble noises versus the overall attentional demands of the task. In quiet, the task of discriminating between target and standard words is relatively undemanding resulting in a larger and earlier P3 peak. In the presence of the informational masking of various language babble, therefore, it appears the latency is prolonged due to slower time to evaluate and increased difficulty discriminating the target from the standard overall.

Differences in P3 amplitude and latency outcomes across the hierarchy of increasing similarity from Mandarin to Dutch to English followed the hypothesized trends, but were not statistically significant. Findings for P3 peak amplitude and area, however, had the largest effect sizes of outcomes in the study, and results most consistent with the proposed hierarchy of informational masking by linguistic/phonetic similarity. P3 amplitude and area was the largest in quiet, the smallest for the English masker condition, and the outcomes for Dutch and Mandarin fell in between. The differences between quiet and both English and Dutch were significant. There was not a significant difference in P3 peak amplitude and area between the two non-native language conditions.

These results for P3 are therefore at least partially consistent with a high similarity or native language informational masking effect. It may be that the combined effect of changes in attentional resources between native versus non-native maskers and the difficulty of the discrimination task at the +3 dB SNR converge making between-language differences smaller or more variable. Percent correct detection of targets showed that the discrimination task was more challenging in English compared to the other languages. This could have resulted in a larger reduction in amplitude for that masker, but counteracted by an increased attention paid to the task because the masker was understandable and required more cognitive resources to separate from target stimuli. Previous studies have also found similar decreases in P3 amplitude for active speech tasks in babble noise conditions suggesting that even with an increase in task difficulty, P3 amplitudes are decreased partiality in the native language babble. Both Koerner et al. (2017) and Bennett et al. (2012) found that P3 amplitude decreased along with button press accuracy to a target speech stimulus in 4-talker babble masking.

There was also not a significant between-language masker difference in P3 latency, and in fact mean latencies were very similar between conditions (as seen in the bar graph in Figure 21). This result was contrary to our hypothesis that there would be an effect of linguistic/phonetic similarity, specifically at this later top-down level of auditory processing. The current results could be interpreted as a language effect on speed of processing at this level, no matter what language is comprises the auditory background.

The morphology of the peaks may have contributed to this result, in that peak latency might not be the most sensitive measure to between condition differences. The advantage of P3 area measures in capturing this overall morphology change may have contributed to why measures of P3 area were more consistent with the effect of a linguistic/phonetic similarity hierarchy (i.e. a native vs. non-native language masking effect). Visually (e.g. Figure 22), the change in area between masker conditions is apparent, while the absolute peaks show less change. English had the smallest area, quiet the largest area, and area for Mandarin and Dutch masker conditions fell between but were not significantly different from each other. P3 area measures had the largest effect size of all AEP component measures (η_p^2=0.300) for the target-masker similarity condition, which would be interpreted as a large effect size. This supports that an active cognitive top-down auditory response was at least partially sensitive to informational masking effects across the linguistic/phonetic similarity hierarchy, and possibly more so than the earlier N1 component based on the effect size.

General results for the linguistic/phonetic similarity manipulation across both N1 and P3 auditory neural responses were consistent with a general effect of informational masking effect provided by language maskers compared to quiet at both earlier and later stages of auditory processing. Specifically, there was an overall significant effect on measures of amplitude, area, and latency, where response in babble noise were smaller and later than those in quiet. There was some indication that the English masker, with the highest level of linguistic/phonetic similarity had at least a partial effect on the N1 and P3 amplitudes and the strongest effect on P3 area consistent with the stated hypotheses, but these were not consistent across all measures. There were no significant differences between AEP outcomes in the English masker compared to the non-native language maskers, except for P3 amplitude and area. Because of the differences in magnitude of absolute latencies and amplitudes of the N1 and P3, it is difficult to directly compare in the same analysis whether there was a larger effect of informational masking at one level compared to the other. Effect sizes were largest for the P3 amplitude and area outcomes compared to any other variables in the study, which partially supports that the linguistic/phonetic similarity contrast had a larger effect at this level of processing. However even for these variables of amplitude and area there were no difference between the two non-native language maskers as Dutch and Mandarin did not reach statistical significance. The maskers used in the current study were based on previous behavioral research demonstrating significant effects of linguistic and phonetic similarity on behavioral performance. The AEP outcomes did not completely agree with the findings from these behavioral studies. These comparisons are discussed further in section 4.4 below.

4.3 Effect of stimulus uncertainty on neural indices of cortical auditory processing

The other classic factor theorized to contribute to the amount of informational masking listeners experience is stimulus uncertainty. Behaviorally, uncertainty has been defined as the influence on listener's a-priori knowledge of the timing or content of the target speech and/or interfering speech has on their ability to understand the target (Brungart and Simpson 2004; Durlach, Mason, Shinn-Cunningham, et al. 2003; Watson, Kelly, et al. 1976; Watson and Kidd 2007). Uncertainty in this AEP experimental paradigm manipulated the timing between presentation of the target stimuli relative to the standard and quiet intervals. An advantage of using AEPs as an index of uncertainty effects is that the timing of the presentation of the stimuli and therefore task difficulty, which has known effects on the timing of the neural response (Gonsalvez and Polich 2002; Kutas et al. 1977b; Picton 1992; Polich et al. 1985). For the paradigm used in the current study, uncertainty was manipulated by changing the TTI or timing between target stimuli. When the timing of the target was consistent and predictable (i.e. every 6 seconds) uncertainty of target presentation was coming was low. When target to target timing was random (i.e. 2, 4, 6, 8, or 10 seconds), uncertainty of target presentation was high.

4.3.1 N1 Uncertainty

At the earlier stage of processing, the effect of uncertainty was only significant for N1 latency, specifically that N1 peak latency was longer in the high uncertainty condition compared to low uncertainty. Results of N1 peak and mean amplitude as well as area measures across both levels of uncertainty are consistent with no effect of uncertainty on the strength of the neural response. Although visual inspection of both bar graphs and grand mean waveforms trend toward higher amplitude for high uncertainty conditions, none of the mentioned measures were statistically different between low and high uncertainty conditions. N1 latency, on the other hand, was significantly longer for the high uncertainty condition consistent with hypotheses, suggesting that the more difficult task increased thalamo-cortical processing time for identifying and orienting to the standard stimuli.

Few studies of the P1-N1-P2 on its own, which is typically recorded in a passive paradigm, directly manipulate task difficulty. Between passive and active tasks, studies have generally shown that amplitudes are larger for the active paradigm, but have not generally shown significant differences in latency between passive and active (Billings et al. 2011; Pereira et al. 2014; C. Zhang et al. 2016; C. X. Zhang et al. 2014). In relation to task difficulty, AEP studies have shown that the size of attention effects on early AEP increases with increasing task difficulty, as manipulated for example by lower stimulus intensity (Schwent et al. 1976) or lower target discriminability (Alho 1992; Squires et al. 1973). Mulert et al. (2007) found increased N1 amplitude when task difficulty increased in an active auditory oddball paradigm recorded to standard stimuli. On the contrary, Cranford et al. (2004) found significant task related decreases in P2 amplitude, but no effects on N1 amplitudes using a stimulus oddball procedure to tone stimuli. Task difficulty was manipulated by decreasing the size of the frequency differences and/or adding competing speech babble to the non-test ear. This finding of selective effects on components of the P1-N1-P2 provides objective evidence that there is at least some effect of task difficulty at this level of the central auditory system.

4.3.2 P3 Uncertainty

Results of stimulus uncertainty on are also consistent with an effect of delayed auditory discrimination on later auditory top-down processing as evidenced by increased latency and area of P3 in high uncertainty conditions. Effect sizes for the uncertainty outcomes for these P3 latency and area measures were stronger than observed for the N1 outcomes. This finding supports our hypothesis that a later cognitive AEP that indexes auditory discrimination would yield larger effects between experimental conditions, due to the higher level auditory processing that is present during informational masking tasks. Results of stimulus uncertainty on P3 latency and area are consistent with a delay in top-down evaluation time of the target stimulus during high uncertainty conditions. Specifically, results showed that across all language maskers, high uncertainty yielded longer latencies and smaller areas compared to low uncertainty. Therefore, it could reasonably be interpreted that participants had to pay more attention to stimuli under this uncertainty condition, due to the complexity of the task.

In general, P3 studies have shown an inverse relationship between target probability and P3 amplitude under a wide range of experimental conditions, with optimal infrequent target probability around 20% of total stimulus presentations (Duncan-Johnson et al. 1977; Kutas et al. 1977a; Sqires et al. 1977). For this study uncertainty was defined in terms of the time between target stimuli being fixed or random and not the percentage of presentations (targets equaled 20% of presentations). Given that attention allocation reflects by P3 size and timing, the nature of the timing between target events appears to affect P3 values of latency across all background noise conditions. Results from this study are consistent with previous studies that have shown increased P3 latency when task difficulty is increased (Comerchero et al. 1999; Gonsalvez et al. 2007; Hoffman and Polich 1999; Katayama and Polich 1996). For example, Polich (1987) employed a binaural listening task to target tones presented at 40 (standard) and 60 (target) dB SPL for an easy listening condition and at 40 (standard) and 45 (target) dB SPL as a hard condition. They found significantly longer P3 latency in the hard condition compared to the easy condition. Another example, Comerchero and Polich (1999) showed that P3 latency was increased in a difficult auditory perceptual discrimination task versus an easy tone frequency discrimination task. Interestingly, Comerchero et al. (1999) used a three stimulus auditory oddball and found that latencies to the target were delayed, but also to the infrequent non-target tone. That is, instead of using a silent block as a third stimulus to control for attention, as was implement in the current study, they introduced a tone that had a large frequency difference and also measured the P3 and found increased latencies.

Even though the peak amplitude difference between conditions did not reach significance, P3 area reflected a strong overall change in morphology that occurred across conditions. This is again consistent with effects of increased task difficulty seen in other studies (Bennett et al. 2012; Blackwood et al. 1987; Kok 2001; Polich 1987; Wronka et al. 2008). As area is a more encompassing measure of an auditory neural component that combines amplitude and latency, it was unexpected that uncertainty yielded significant effects due to the absence of statistical significance of P3 peak and mean amplitude. However, when visualizing grand means particularly between P3 uncertainty conditions (e.g. Figure 28), area appears larger in higher uncertainty conditions compared to lower uncertainty. This effect was found across all noise conditions and this is consistent with auditory task difficulty affecting both amplitude and latency aspects of the complex waveform. Consistent with effects of latency mentioned above, Polich (1987) also found significant reduced amplitude of the P3 in the hard auditory discrimination task as compared to the easy task. Moreover, this effect was observed across all conditions, including quiet. This can be reasonably interpreted that auditory uncertainty has an equal effect across all tested background noise conditions.

Interestingly, for all but the quiet condition not only was the task difficulty manipulated through TTI, but also the task was more difficult because participants were listening in background noise. The presentation of noise in general can be considered a manipulation of task difficulty and uncertainty. As already reviewed, noise decreased amplitudes and decreased latencies compared to quiet in general. The effects of similarity and uncertainty are therefore not completely separate in this study, nor can they be considered completely separately in informational masking theory in general. This is part of the reason it was hypothesized there would be an additive or interactive effect of the two factors on AEP outcomes. This is discussed further in the next section.

4.4 Interaction between linguistic/phonetic similarity and stimulus uncertainty

One of the most unanticipated results of this study was the absence of the interaction between linguistic/phonetic similarity and stimulus uncertainty. A significant interaction was hypothesized to indicate that the similarity and uncertainty contributions to informational masking were likely to be additive in nature, such that the highest target-masker similarity in the most uncertain condition would have the largest effect on AEP outcomes. That is, if increasing stimulus timing uncertainty specifically increases the cognitive demand of the task, then the listener was hypothesized to have more difficulty under conditions where the target and masker were the most linguistically/phonetically similar (i.e. English).

We hypothesized that this interaction would be true for the higher level of auditory processing, but not necessarily at the lower level of earlier auditory stimulus processing. This hypothesis was partially supported in that there was no significant interaction effect between similarity and uncertainty for the amplitude or latency of the N1 component. Although previous work demonstrates that auditory attention significantly affects this earlier level of auditory processing, specifically N1 amplitude and not latency (Billings et al. 2011; Pereira et al. 2014; C. Zhang et al. 2016; C. X. Zhang et al. 2014), the combined effects of language of the masker and unpredictability of target stimuli did not yield an additive effect. Both similarity and uncertainty affected N1 latency separately, but the effects of stimulus uncertainty were the same across the English, Dutch and Mandarin maskers. It is possible that the amplitude was already so small in noise conditions that further reductions were not reliably measurable, contributing to a potential floor effect. It is also important to note that N1 did not yield strong effects across Mandarin, Dutch, and English and the effects of similarity was largely driven by quiet vs. noise at the level of the N1.

At the level of the P3, a later auditory processing level, we expected that there would be a significant interactive effect of the two factors related to informational masking, because this level should relate most highly to the cognitive processing that would be required in behavioral recognition tasks. As shown in the proposed and actual interaction results in Figure 18, uncertainty affected each of the masking conditions equally by reducing the amplitude approximately 0.5-1.0 uV. This linear pattern of significant main effects, but absence of interactions, as indicated by parallel lines, was generally seen for all AEP measures.

The fact that uncertainty or task difficulty had the same effect on auditory encoding of target words in noise regardless of the language of the masker is a novel finding in the informational masking literature. Few studies have manipulated aspects of similarity and uncertainty in the same paradigm, and in particularly not for speech in noise tasks. In traditional auditory psychophysical approaches that study informational masking, it is common to map perception of changes across individual acoustic dimensions. For example, measurement of the frequency difference threshold between two tones could be operationally defined as the measurement of the similarity between those tones. If the target stimulus remains fixed, the only uncertainty is due to the change in comparison tone frequency. However, in more complex designs the two factors may result in changes in opposite directions that cancel out (e.g., Kidd et al., 1994; Neff, 1995; Durlach et al., 2003). In this study, for example, the increases in amplitude for high uncertainty could have been counteracted by decreases in amplitude when similarity is increased (made harder, more similar) between target and background noise. Thus, it is comparatively difficult to truly define the interaction of experimental stimuli within the construct of complex auditory paradigms like this experiment, which is one of the reasons why knowledge about this phenomena remains incomplete. In addition, it possible that the task manipulation of uncertainty was not a robust enough effect to yield an interaction in the current study. While TTI is a valid manipulation of trial-by-trial variability, and did have significant effects on AEP outcomes, it is possible that randomly varying TTI did not sufficiently alter the perceived uncertainty in the relatively repeated simple oddball paradigm to provide additional task difficulty that depended on the masker condition. Nevertheless, in the confines of this experimental paradigm we can interpret one finding that aids in our understanding of informational masking, which is that the manipulations of similarity and uncertainty did not interact at either level of auditory cortical processing for this AEP paradigm. Because we see effects of both speech babble masking and target-to-target interval, we know these manipulations affected auditory cortical processing in their own way, but they weren't additive for this paradigm at either level.

4.5 The relationship between AEP outcomes and behavioral performance

To compare informational masking effects on AEP outcomes to behavioral performance, word recognition performance in noise was measured for English words under the same linguistic/phonetic similarity masking conditions that were used in the AEP experiment. The linguistic similarity hypothesis proposed by Brouwer et al. (2012) states that the more similar the target and masker speech language, the harder it is to segregate the two streams effectively. This definition is analogous to Durlach's definition of similarity of informational masking, but the linguistic similarity hypothesis further defines listener related factors, such as knowledge/experience with the target speech and the linguistic and phonetic content which ultimately affect behavioral speech recognition. This hypothesis was tested by Calandruccio et al. (2013) using Mandarin, Dutch, English and language maskers in combination with an English sentence recognition task. This hierarchy provided the basis for the current study. Calandruccio et al. (2013) found that sentence recognition in -5 dB SNR improved as the target-to-masker linguistic distance increased with the English-on-English masker yielding the poorest speech processing followed by Dutch, then Mandarin.

For the current study behavioral performance was measured using the same language masker conditions. Unlike the Calandruccio et al. (2013) study, however, stimuli for the current study were NU-6 single-syllable words, which are theoretically harder to discriminate compared to sentences due to lack of context. The pattern of performance across language maskers was the same, with SNR50s of 2.35 dB in English, 0.77 dB in Dutch, and -0.12 dB in Mandarin. While performance was worst for the English masker, scores were not significantly different between Dutch and Mandarin. This was also true for the fixed +3 dB SNR level, a direct comparison to the AEP results, and visualized across levels. Even at 0 dB SNR, hypothesized to provide a larger performance spread between language maskers, Dutch and Mandarin were not significantly different.

These behavioral results are generally in agreement with the AEP results, which also did not show any significant differences for any of the outcomes between the Dutch and Mandarin conditions. There could be several reasons for the lack of a hierarchy effect in behavioral performance in the current study compared to previous literature. As mentioned, the task used words rather than sentences, which made the task more difficult in terms of speech recognition but also limited the linguistic demands of the task. It may be that the effects of linguistic and phonetic similarity were not as relevant in this case. The English condition still resulted in the poorest performance, which partially supports the similarity hypothesis. Previous studies support the result of a native language masker contributing to decreased behavioral speech perception (Brouwer et al. 2012; Calandruccio, Bradlow, et al. 2014; Calandruccio et al. 2013; Cooke et al. 2008; Van Engen and Bradlow 2007). Lecumberri and Cooke (2006) showed that monolingual English listeners performed better on a speech perception of English consonants when the language of a competing speaker was Spanish rather than English. Van Engen and Bradlow

(2007) demonstrated that for native English listeners, English sentence speech perception was better when the noise consisted of two-talker Mandarin Chinese babble than when it was composed of two-talker English babble. These results suggest that babble noise in a native language increases informational masking, decreasing behavioral speech recognition performance relative to an unknown foreign language.

Although these behavioral word recognition performance outcomes only reflect the target-masker similarity portion of the informational masking question, reaction time, accuracy, and d' of button press responses during AEP testing provide behavioral outcomes that can be compared for both factors. Reaction time results showed significant effects of similarity and uncertainty similar to that of P3 latency, with increased reaction times for all language maskers as well as for high uncertainty conditions. Results of button press accuracy revealed result similar to P3 area, with the highest accuracy in the quiet condition, performance in the middle for Dutch and Mandarin in the middle, and the lowest accuracy scores for the English masker. The d' behavioral discrimination scores were higher in quiet compared to Mandarin and English, but interestingly not compared to Dutch. However, d' scores were higher in Dutch compared to English. Both % correct accuracy and d' scores were significantly higher for low vs. high uncertainty, and none of the interactions were significant for any of these behavioral measures. In general, these results were consistent with quiet being the easiest task, no difference between Mandarin and Dutch, and English being the most difficult behavioral detection task. The effect of uncertainty was consistent across all measures, with low uncertainty results indicating an easier task than high, but the effect was the same for all maskers. These results follow the same trend in AEP component measures. This result is similar to previous studies which have shown increases in button press reaction time and decreases to accuracy with the addition of the maskers (Bennett et al. 2012; Kaplan-Neeman et al. 2006; Koerner et al. 2017; Martin et al. 2008), consistent with listeners taking longer to make a distinction between the speech targets and standards and increasing the rate of errors.

Results of the stepwise linear regression and correlations showed that some AEP outcomes were significantly related to behavioral outcomes in the same individuals, but did not reveal an identifiable pattern of AEP outcomes related to behavioral performance across masking conditions. For example, N1 mean amplitude accounted for 11% of the variance on Mandarin SNR50, but <2% on Dutch and English. P3 peak amplitude and latency accounted for 35% of the variance in Dutch, but <5% in Mandarin and English. Furthermore, significant model predictors did not account for a majority of the variance in any language masking condition. There were significant correlations for each language masking condition. Area measures seemed to have the most numerous significant correlations across SNR50 for each language masking condition, but again no AEP measure was consistently correlated with SNR50, or word recognition scores across SNR. For English SNR50, P1-N1-P2 and P3 area in English high uncertainty conditions showed significant moderate correlations respectively, with increasing area associated with decreasing (better) SNR50. There were also moderate correlations for Dutch SNR50 with N1 peak amplitude in Dutch high uncertainty and P3 latency in low uncertainty and for Mandarin SNR50 withN1 latency). Interestingly, correlations between reaction time, accuracy, and d' scores and AEP components were not significant for all languages and only moderate at best. It has been previous reported to have a mixed relationship with behavioral response accuracy or reaction time (Koerner et al. 2017; McCarthy et al. 1981; Verleger 1997). This is consistent with the discrepancy of correlations of behavioral word recognition performance and accuracy, reaction time, and d' across listening conditions AEP component measures. For example, Bennett et al. (2012) and Koerner et al. (2017) found that P3 latency or amplitude did not predict sentence-level recognition across listening conditions or accuracy and reaction time of target speech stimuli. Differences between analysis methods were noted by Koerner et al. (2017), as Bennett et al. (2012) used Pearson correlations to examine brain-behavior relationships while their analysis used stepwise regression models similar to the current study.

Relationships between AEP outcomes and behavioral outcomes are complex and while there was some agreement between outcomes in the current study, no consistent trend between brain and behavior emerged. In fact, the only trend that was similar between both AEP and behavioral outcomes was the effect of similarity across language maskers. Therefore, while AEP and behavioral tasks are not entirely parallel, AEP outcomes did show the same effects of language masking. Results of this study do support the use of AEPs as windows into the levels of auditory processing that behavioral outcomes overshadow.

4.6 Limitations and Future Directions

Although this study adds to knowledge about informational masking and the effects on neural processing, there are several limitations in our ability to draw conclusions about target-masker similarity and stimulus uncertainty effects on the earlier and later stages of auditory processing.

As it has been previously noted, the AEPs recorded in this study were recorded in an active paradigm, which may have influenced results for N1 in some confounding ways, as well limit the ability to interpret this response as primarily sensory. Processing of auditory information occurs throughout the central auditory nervous system, reflected by the series of AEP components that represent different aspects of detecting and interpreting the input stimuli.

Although sometimes categorized as pre-attentive "exogenous" evoked potentials and cognitive "endogenous" evoked potentials, it is well known that AEPs reflect a combination of exogenous and endogenous dynamics that can be affected by acoustic characteristics of stimuli in addition to top-down factors such as arousal state and attention (Martin et al. 2008; Naatanen and Picton 1987; Squires et al. 1973; Stapells 2008). Due to the active paradigm used in the current study, higher level cognitive processing, such as working memory, auditory inhibition, and listening effort, could have affected AEP components. For example, all participants were actively attending to the oddball paradigm as indicated by button press responses, but variability in motivation to quickly and accurately press the correspond button to the target or standard was not measured. All participant were told to listen and press the button as soon as they discriminated between stimuli, but there was no incentive for increased accuracy or decreased reaction time. The addition of the third stimulus in the oddball paradigm (the silent interval) was also implemented to maintain attention throughout the task. Future studies could measure and account for listener attention and motivation or effort across the tasks in active auditory oddball tasks using EEG or other methods of quantifying vigilant attention and effort. Regardless, because a completely passive condition where participants did not have to respond to the target or standard words was not recorded, effects of passive versus active attention on these AEP components in general and between language masker and uncertainty conditions cannot be quantified. A true auditory obligatory effect of linguistic/phonetic similarity and stimulus uncertainty on the P1-N1-P2 cannot be interpreted from these results. While previous research and results from the current study are consistent with informational masking affecting higher-level active auditory processing, it would be interesting to record each of the conditions in a non-attended or ignored condition to evaluate this effect.

Another limitation of the current study that is present in all AEP studies is the comparison between the current AEP paradigm and behavioral word recognition. The linguistic complexity of repeatedly discriminating between /bæt/ and /pæt/ is not entirely comparable to behavioral word recognition of NU-6 open set list of words. The primary reason the /bæt/ vs. /pæt/ word-word contrast was chosen was due to voice onset time (VOT), which is a relevant acoustic precept necessary for accurate speech perception (Oden et al. 1978). The /b/ and /p/ voiced and unvoiced bilabial stops, represent a significant phonetic contrast that needs to be accurately coded in order to accurately perceive speech. While, VOT is a distinction of production of plosive consonants, it is only one of a multitude of phonetic place, manner, and voicing properties that makes up the English language. That is, in order to distinguish between /bæt/ vs. /pæt/ in the AEP paradigm could be considered relatively simple as compared to behavioral word recognition. This could be the reason for an absence of a consistent AEP predictor variable between SNR50 and AEP component measures. However, it is particularly interesting, given the difference between AEP and behavior, that the effects of linguistic/phonetic similarity are similar in trends across languages (no difference between Mandarin and Dutch, but English yields the lowest percent correct and decreased amplitudes/areas).

Additionally, word stimuli in AEP paradigms have been used sparingly and thus further investigation of more auditory complex stimuli in AEP paradigms was implemented in the current study. The results of this study support the use of word stimuli in AEP recording paradigms, but a direct comparison to behavioral word recognition, as stated above, should be interpreted with caution. While previous studies have used tone (Billings et al. 2011; Katayama and Polich 1996; Naatanen and Picton 1987; Picton and Hillyard 1974; Tremblay, Friesen, et al.

2003b) and speech tokens such as /ba/, /pa/, and /ga/ (Bennett et al. 2012; Koerner et al. 2017; Niemczak and Vander Werff 2019; Novick et al. 1985; Whiting et al. 1998), this study adds to the literature by using a simple word stimuli to evoke an N1 and P3 in an active oddball paradigm. It has been previous shown that speech as AEP speech stimuli typically results in delayed component peak latencies relative to tonal signals. (Billings et al. 2011; Dai et al. 2016; Onishi et al. 1968; Ruhm et al. 1969). The reason for the results may be due to the specific spectral characteristics of the speech token itself. A relatively simple and interesting follow-up manipulation would be to use variation of frequency tones instead of /bæt/ and /pæt/ to further investigate whether the same trends of experimental manipulations would occur. In other words, is the effect of similarity and uncertainty still present for a tone discrimination task that does not carry semantic meaning. Or, a more interesting future direction, would be to use stimuli with increased semantic meaning directly related to behavioral word or sentence discrimination, such as categorical discrimination of words with an oddball category. Regardless, future AEP investigations into informational masking employing word stimuli, or even more complex stimuli such as sentences, will be important to better relate AEP measures to real-world behavioral speech discrimination.

Limitations of experimental manipulations were also present, but should lead to future examinations of both informational masking factors. First, linguistic/phonetic similarity had a strong effect on both amplitude and latency of AEP response overall when the quiet condition and all maskers were included. There was some evidence that a native language masker is more detrimental to neural auditory perception especially at a later level of auditory processing. Language maskers consisted 2-talker IEEE sentences across language maskers, but were only understandable by the listeners in the English condition (as none of the participants had experience with either Dutch or Mandarin). It would be interesting to determine whether native-language maskers that were less understandable, or less meaningful, would have the same effects as understandable maskers. For example, would the effects of similarity to auditory neural processing be changed if the masker was constructed of anomalous sentences in English (e.g. "The walks table by cat the")? In other words does, does the correct syntactic order of the words in each sentence affect the ability to detect a speech target when the masker is essentially incomprehensible? Previous behavioral research studies have generally shown that anomalous sentences provide less masking compered to sentences that are syntactically constructed (Brouwer et al. 2012; Calandruccio, Bradlow, et al. 2014; Calandruccio et al. 2010). For example, found that accuracy in discrimination of sentences was higher in an anomalous English masker than a syntactically meaningful English masker. This result indicates that speech-on-speech masking not only involves interference at the phonetic/phonemic level, but also at the syntactic level. In addition, it was found that the syntactic content effect was not present with the Dutch maskers due to a non-native language effect. In other words, the listeners were not affected by the difference between meaningful and anomalous Dutch background speech. Nevertheless, examination of meaningfulness of speech for both the target and maskers (i.e. the semantic content of the sentences) would need to be further examined at both the auditory cortical and behavioral level.

Another limitation of the current study is that definitional structure of theoretical constructs of similarity and uncertainty as two independent manipulations. Specifically, the factor of stimulus uncertainty is difficult to pinpoint and manipulate in an auditory paradigm. For the purposes of this study, stimulus uncertainty was manipulated by varying the timing between target stimuli, but it is also possible that decreasing or increasing target-masker similarity also affected stimulus uncertainty. However, while manipulation of uncertainty did contribute to task difficulty, as evidenced by AEP latency, reaction time, and accuracy, it did not change the masker type or language of the masker. This could be interpreted in two ways. First, the manipulation of stimulus uncertainty could be viewed as not contributing to an informational masking effect, but more so a task difficulty effect. The absent interaction supports this interpretation to which presenting target stimuli in a random compared to predictable fashion just increased latency across every condition equally. Another interpretation could be that uncertainty is essential to studying informational masking causing a change in the auditory environment that cannot be classified by the acoustic properties of the target and masker in isolation. That is, without taxing the auditory system in informational masking tasks by varying the uncertainty (via timing, location, etc.) of target and masker speech, is informational masking actually being studied? It is important to understand these two factors of similarity and uncertainty in the framework of the listening task, and future research will be needed to extend the definitional structure of similarity and uncertainty and find objective correlates of processing of informational masking using more externally valid manipulations that better relate to behavioral speech perception.

The population in which this study was conducted was young normal hearing individuals. There was relatively little evidence that peripheral hearing differences or differences in cognitive function confounded results. Although most of the participants reported frequent listening to music under headphones, and a few individuals met screening criteria for higher risk of noise exposure, mean pure tone thresholds were better than 20 dB HL across all frequencies, including extended high frequencies. DPOAEs were also robust across all frequencies tested. It seems unlikely that peripheral hearing damage influenced the variability in AEP or behavioral outcomes. Similarly, there was little inter-individual variation in cognitive abilities as assessed by the working memory and attention tests.

It may be that the effects of informational masking on AEPs were not robust in these normal hearing young adults, but may be more apparent in populations more vulnerable to the effects of masking, such as older adults and those with hearing loss. It will be interesting in future studies to focus on the aging auditory system and the influence of hearing loss on how linguistic/phonetic similarity and uncertainty may contribute to increased difficulty understanding speech and how this is encoded in the pathological auditory nervous system. The overall goal in better defining the informational masking deficits is to help those with auditory difficulties, especially in background noise. The effects and possible interaction of informational masking factors could be different in these populations. Previous research has demonstrated delayed latency and increased variability of AEP responses in older adults (Bertoli et al. 2005; Helfer and Freyman 2008; Picton 1992; Polich et al. 1985; Tremblay, Piskosz, et al. 2003; Woldorff et al. 1993). One potential explanation for this age effect suggested by Tremblay, Piskosz, et al. (2003), could be age-related neural refractory differences in younger and older auditory systems. Refractory issues, evidenced by differences in latency, may affect synchronized neural activity underlying the perception of time-varying speech cues. While the effects of age on behavioral speech perception in noise have been extensively studied, limited research on auditory physiological effects of background noise and age is still remain largely undefined. One such study by Maamor et al. (2017) found age related differences for P2 latencies elicited by speech tokens in a passive oddball paradigm in various background noise where morphology of the P2 peak was delayed in latency for two older groups (one with hearing impairment) than for the younger group. In contrast, Maamor and Billings (2017) also found that

N1 peaks appeared later and with poorer morphology for younger and older normal-hearing groups relative to an older group with hearing impairment. Interestingly, they also found that the N1 wave was sensitive to the more energetic masking (continuous and modulated energetic masking) effects whereas, the P2 was more sensitive to informational masking noise (4-talker babble), whose magnitude varied as a function of age. The authors suggested that due to the P2 involving more high-level auditory processing effects of informational masking were more apparent. Comparing to the current study, both the N1 and P3 and the overall area of the responses were susceptible to informational masking, but if older individuals were recruited for this study, it is possible that the hypothesized effect of increased target-masker similarity and stimulus uncertainty would have been present at a later stage of auditory processing. While the P2 was not a primary outcome variable, all language maskers greatly reduced the amplitude of the P2 component in the current study. Area measures of the entire P1-N1-P2 encompass this effect, and support decreased strength of auditory processing in all language masking conditions. Moving farther along the auditory pathway, the P3 indexed a language masking effect and in particular a native language masking effect for P3 amplitude and area. However, this result needs to be validated in and older and/or hearing loss population. Future studies should seek to examine later, more cognitive levels of auditory processing of speech perception in informational background noise in older adults with and without hearing loss.

5.0 Conclusions

The results of this study add to our understanding of auditory perception in informational masking in four ways. First, observable effects of both similarity and uncertainty are evidenced at both top-down levels of auditory neural processing. This supports the use of AEP to better understand the informational masking deficit by providing a window into the auditory pathway. Second, larger effect sizes were found for some outcomes for the P3 response, providing some suggestion that while informational masking degradation happens at lower levels, higher level cognitive auditory processing may be more sensitive to informational masking deficits. Future studies on informational masking should focus on the entire auditory pathway, but should emphasize cognitive affects such as auditory attention and discrimination. Third, the lack of interaction of main effects leads us to a linear interpretation of similarity and uncertainty with an equal effect across listening conditions. This results of this experimental paradigm provides a better understanding to how similarity and uncertainty factors of informational masking may be processed separately and do not have a combined effect. Fourth, even though there were few and only moderate correlations to behavioral word recognition, AEP and behavioral performance data followed the same trends of linguistic/phonetic similarity. Through both auditory neural and behavioral testing, language maskers degraded AEPs and reduced word recognition, but particularly using a native-language masker. Nevertheless, it is important to be cautious of the direct comparison between AEP and behavior due to substantial differences in the auditory stimuli and tasks. The auditory neural and behavioral results from this study provide a foundation for further investigation of how the linguistic content of target and masker, as well as task difficulty contribute to difficulty understanding speech in noise.

6.0 References

Abutan, B. B., Hoes, A. W., Vandalsen, C. L., et al. (1993). Prevalence of Hearing Impairment and Hearing Complaints in Older Adults - a Study in General-Practice. *Family Practice, 10*, 391-395.

Alho, K. (1992). Selective Attention in Auditory Processing as Reflected by Event-Related Brain Potentials. *Psychophysiology, 29*, 247-263.

Anderson, S., White-Schwoch, T., Parbery-Clark, A., et al. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hear Res, 300*, 18-32.

Arbogast, T. L., Mason, C. R., Kidd, G., Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *J Acoust Soc Am, 112*, 2086-2098.

Bennett, K. O., Billings, C. J., Molis, M. R., et al. (2012). Neural encoding and perception of speech signals in informational masking. *Ear Hear, 33*, 231-238.

Bertoli, S., Smurzynski, J., Probst, R. (2005). Effects of age, age-related hearing loss, and contralateral cafeteria noise on the discrimination of small frequency changes: psychoacoustic and electrophysiological measures. *J Assoc Res Otolaryngol, 6*, 207-222.

Billings, C. J., Bennett, K. O., Molis, M. R., et al. (2011). Cortical encoding of signals in noise: effects of stimulus type and recording paradigm. *Ear Hear, 32*, 53-60.

Billings, C. J., McMillan, G. P., Penman, T. M., et al. (2013). Predicting perception in noise using cortical auditory evoked potentials. *J Assoc Res Otolaryngol, 14*, 891-903.

Billings, C. J., Tremblay, K. L., Stecker, G. C., et al. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear Res, 254*, 15-24.

Blackwood, D. H., Whalley, L. J., Christie, J. E., et al. (1987). Changes in auditory P3 event-related potential in schizophrenia and depression. *Br J Psychiatry, 150*, 154-160.

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

Brouwer, S., Van Engen, K. J., Calandruccio, L., et al. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. *J Acoust Soc Am, 131*, 1449-1464.

Brungart, D. S., Chang, P. S., Simpson, B. D., et al. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J Acoust Soc Am, 120*, 4007-4018.

Brungart, D. S., Chang, P. S., Simpson, B. D., et al. (2009). Multitalker speech perception with ideal time-frequency segregation: effects of voice characteristics and number of talkers. *J Acoust Soc Am, 125*, 4006-4022.

Brungart, D. S., Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty. *Journal of the Acoustical Society of America, 115*, 301-310.

Brungart, D. S., Simpson, B. D., Ericson, M. A., et al. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am, 110*, 2527-2538.

Burnett, I. S., Parry, J. J. (1996). On the effects of accent and language on low rate speech coders. *Icslp 96 - Fourth International Conference on Spoken Language Processing, Proceedings, Vols 1-4*, 291-294.

Calandruccio, L., Bradlow, A. R., Dhar, S. (2014). Speech-on-speech masking with variable access to the linguistic content of the masker speech for native and nonnative english speakers. *J Am Acad Audiol, 25*, 355-366.

Calandruccio, L., Brouwer, S., Van Engen, K. J., et al. (2013). Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *Am J Audiol, 22*, 157-164.

Calandruccio, L., Buss, E., Hall, J. W., 3rd. (2014). Effects of linguistic experience on the ability to benefit from temporal and spectral masker modulation. *J Acoust Soc Am, 135*, 1335-1343.

Calandruccio, L., Van Engen, K., Dhar, S., et al. (2010). The effectiveness of clear speech as a masker. *J Speech Lang Hear Res, 53*, 1458-1471.

Comerchero, M. D., Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clin Neurophysiol, 110*, 24-30.

Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America, 119*, 1562-1573.

Cooke, M., Garcia Lecumberri, M. L., Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *J Acoust Soc Am, 123*, 414-427.

Cranford, J. L., Rothermel, A. K., Walker, L., et al. (2004). Effects of discrimination task difficulty on N1 and P2 components of late auditory evoked potential. *J Am Acad Audiol, 15*, 456-461.

Crowley, K. E., Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin Neurophysiol, 115*, 732-744.

Culling, J. F. (2013). Energetic and informational masking in a simulated restaurant environment. *Adv Exp Med Biol, 787*, 511-518.

Culling, J. F., Porter, J. S. (2005). Effects of differences in the accent and gender of competing voices on speech segregation. *Auditory Signal Processingp: Physiology, Psychoacoustics, and Models*, 307-313.

Cusack, R., Deeks, J., Aikman, G., et al. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J Exp Psychol Hum Percept Perform, 30*, 643-656.

Dai, L. S., Shinn-Cunningham, B. G. (2016). Contributions of Sensory Coding and Attentional Control to Individual Differences in Performance in Spatial Auditory Selective Attention Tasks. *Frontiers in Human Neuroscience, 10*.

Davis, H., Davis, P. A., Loomis, A. L., et al. (1939). Electrical reactions of the human brain to auditory stimulation during sleep. *Journal of Neurophysiology, 2*, 500-514.

Desjardins, J. L., Doherty, K. A. (2013). Age-related changes in listening effort for various types of masker noises. *Ear Hear, 34*, 261-272.

Dollezal, L. V., Tolnai, S., Beutelmann, R., et al. (2017). Release from informational masking by auditory stream segregation: perception and its neural correlate. *Eur J Neurosci*.

Duncanjohnson, C. C., Donchin, E. (1977). Effects of a Priori and Sequential Probability of Stimuli on Event-Related Potential. *Psychophysiology, 14*, 95-95.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., et al. (2003). Note on informational masking. *J Acoust Soc Am, 113*, 2984-2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., et al. (2003). Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J Acoust Soc Am, 114*, 368-379.

Elhilali, M., Chi, T., Shamma, S. A. (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication, 41*, 331-348.

Fausti, S. A., Wilmington, D. J., Gallun, F. J., et al. (2009). Auditory and vestibular dysfunction associated with blast-related traumatic brain injury. *J Rehabil Res Dev, 46*, 797-810.

Ferguson, M. A., Hall, R. L., Riley, A., et al. (2011). Communication, listening, cognitive and speech perception skills in children with auditory processing disorder (APD) or Specific Language Impairment (SLI). *J Speech Lang Hear Res, 54*, 211-227.

Festen, J. M., Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am, 88*, 1725-1736.

Freyman, R. L., Balakrishnan, U., Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *J Acoust Soc Am, 109*, 2112-2122.

Freyman, R. L., Balakrishnan, U., Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J Acoust Soc Am, 115*, 2246-2256.

Freyman, R. L., Helfer, K. S., Balakrishnan, U. (2007). Variability and uncertainty in masking by competing speech. *J Acoust Soc Am, 121*, 1040-1046.

Friedman, J., Meares, R. (1980). Tobacco smoking and cortical evoked potentials: an opposite effect on auditory and visual systems. *Clin Exp Pharmacol Physiol, 7*, 609-615.

Fritz, J. B., Elhilali, M., David, S. V., et al. (2007). Auditory attention--focusing the searchlight on sound. *Curr Opin Neurobiol, 17*, 437-455.

Gallun, F. J., Durlach, N. I., Colburn, H. S., et al. (2008). The extent to which a position-based explanation accounts for binaural release from informational masking. *J Acoust Soc Am, 124*, 439-449.

Glasberg, B. R., Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear Res, 47*, 103-138.

Gonsalvez, C. J., Barry, R. J., Rushby, J. A., et al. (2007). Target-to-target interval, intensity, and P300 from an auditory single-stimulus task. *Psychophysiology, 44*, 245-250.

Gonsalvez, C. J., Gordon, E., Anderson, J., et al. (1995). Numbers of preceding nontargets differentially affect responses to targets in normal volunteers and patients with schizophrenia: a study of event-related potentials. *Psychiatry Res, 58*, 69-75.

Gonsalvez, C. J., Gordon, E., Grayson, S., et al. (1999). Is the target-to-target interval a critical determinant of P3 amplitude? *Psychophysiology, 36*, 643-654.

Gonsalvez, C. J., Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology, 39*, 388-396.

Hall, J. W., 3rd, Grose, J. H., Buss, E., et al. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear Hear, 23*, 159-165.

Hall, J. W., 3rd, Huang-fu, M., Gennarelli, T. A. (1982). Auditory function in acute severe head injury. *Laryngoscope, 92*, 883-890.

Helfer, K. S., Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear Hear, 29*, 87-98.

Henkin, Y., Kishon-Rabin, L., Gadoth, N., et al. (2002). Auditory event-related potentials during phonetic and semantic processing in children. *Audiology and Neuro-Otology, 7*, 228-239.

Hillyard, S. A., Squires, K. C., Bauer, J. W., et al. (1971). Evoked Potential Correlates of Auditory Signal Detection. *Science, 172*, 1357-&.

Hink, R. F., Hillyard, S. A. (1976). Auditory Evoked-Potentials during Selective Listening to Dichotic Speech Messages. *Perception & Psychophysics, 20*, 236-242.

Hoen, M., Meunier, F., Grataloup, C. L., et al. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Communication, 49*, 905-916.

Hoffman, L. D., Polich, J. (1999). P300, handedness, and corpus callosal size: gender, modality, and task. *Int J Psychophysiol, 31*, 163-174.

Holcomb, P. J. (1988). Automatic and Attentional Processing - an Event-Related Brain Potential Analysis of Semantic Priming. *Brain and Language, 35*, 66-85.

Johnson, R., Donchin, E. (1982a). 2 P300s Elicited by a Single Stimulus. *Psychophysiology, 19*, 566-567.

Johnson, R., Donchin, E. (1982b). Sequential Expectancies and Decision-Making in a Changing Environment - an Electro-Physiological Approach. *Psychophysiology, 19*, 183-200.

Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., et al. (2006). Identification of syllables in noise: electrophysiological and behavioral correlates. *J Acoust Soc Am, 120*, 926-933.

Katayama, J., Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *Int J Psychophysiol, 23*, 33-40.

Katayama, J., Polich, J. (1998). Stimulus context determines P3a and P3b. *Psychophysiology, 35*, 23-33.

Katayama, J., Polich, J. (1999). Auditory and visual P300 topography from a 3 stimulus paradigm. *Clin Neurophysiol, 110*, 463-468.

Kayser, J., Bruder, G. E., Tenke, C. E., et al. (2001). Event-related brain potentials (ERPs) in schizophrenia for tonal and phonetic oddball tasks. *Biological Psychiatry, 49*, 832-847.

Kidd, G., Jr., Arbogast, T. L., Mason, C. R., et al. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America, 118*, 3804-3815.

Kidd, G., Jr., Mason, C. R., Rohtla, T. L., et al. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J Acoust Soc Am, 104*, 422-431.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., et al. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am, 116*, 2395-2405.

Koerner, T. K., Zhang, Y., Nelson, P. B., et al. (2017). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A P3 study. *Hear Res, 350*, 58-67.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*, 557-577.

Krishnamurti, S. (2001). P300 auditory event-related potentials in binaural and competing noise conditions in adults with central auditory processing disorders. *Contemporary Issues in Communication Science and Disorders, 28*, 40-47.

Kujawa, S. G., Liberman, M. C. (2015). Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. *Hear Res, 330*, 191-199.

Kutas, M., Mccarthy, G., Donchin, E. (1977a). Augmenting Mental Chronometry - P300 as a Measure of Stimulus Evaluation Time. *Science, 197*, 792-795.

Kutas, M., McCarthy, G., Donchin, E. (1977b). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science, 197*, 792-795.

Lecumberri, M. L. G., Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America, 119*, 2445-2454.

Leibold, L. J., Hitchens, J. J., Buss, E., et al. (2010). Excitation-based and informational masking of a tonal signal in a four-tone maskera). *The Journal of the Acoustical Society of America, 127*, 2441-2450.

Lew, G. S., Polich, J. (1993). P300, habituation, and response mode. *Physiol Behav, 53*, 111-117.

Luck, S. J. (2014a). *The Design of ERP Experiments*.

Luck, S. J. (2014b). *Quantifying ERP Amplitudes and Latencies*.

Lutfi, R. A. (1990). How much masking is informational masking? *J Acoust Soc Am, 88*, 2607-2610.

Lutfi, R. A., Kistler, D. J., Oh, E. L., et al. (2003). One factor underlies individual differences in auditory informational masking within and across age groups. *Percept Psychophys, 65*, 396-406.

Maamor, N., Billings, C. J. (2017). Cortical signal-in-noise coding varies by noise type, signal-to-noise ratio, age, and hearing status. *Neuroscience Letters, 636*, 258-264.

Martin, B. A., Boothroyd, A. (1999). Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear, 20*, 33-44.

Martin, B. A., Boothroyd, A. (2000). Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *J Acoust Soc Am, 107*, 2155-2161.

Martin, B. A., Tremblay, K. L., Korczak, P. (2008). Speech evoked potentials: from the laboratory to the clinic. *Ear Hear, 29*, 285-313.

McCarthy, G., Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. *Science, 211*, 77-80.

Michalewski, H. J., Starr, A., Zeng, F. G., et al. (2009). N100 cortical potentials accompanying disrupted auditory nerve activity in auditory neuropathy (AN): Effects of signal intensity and continuous noise. *Clinical Neurophysiology, 120*, 1352-1363.

Mulert, C., Leicht, G., Pogarell, O., et al. (2007). Auditory cortex and anterior cingulate cortex sources of the early evoked gamma-band response: Relationship to task difficulty and mental effort. *Neuropsychologia, 45*, 2294-2306.

Naatanen, R., Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*, 375-425.

Naatanen, R., Teder, W., Alho, K., et al. (1992). Auditory attention and selective input modulation: a topographical ERP study. *Neuroreport, 3*, 493-496.

Neff, D. L. (1995). Signal properties that reduce masking by simultaneous, random-frequency maskers. *J Acoust Soc Am, 98*, 1909-1920.

Neff, D. L., Callaghan, B. P. (1988). Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *J Acoust Soc Am, 83*, 1833-1838.

Niemczak, C., Vander Werff, K. (2019). Informational Masking Effects on Neural Encoding of Stimulus Onset and Acoustic Change. *Ear and Hearing, 40*, 156-167.

Novick, B., Lovrich, D., Vaughan, H. G., Jr. (1985). Event-related potentials associated with the discrimination of acoustic and semantic aspects of speech. *Neuropsychologia, 23*, 87-101.

Oden, G. C., Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychol Rev, 85*, 172-191.

Onishi, S., Davis, H. (1968). Effects of duration and rise time of tone bursts on evoked V potentials. *J Acoust Soc Am, 44*, 582-591.

Osterhammel, P. A., Davis, H., Wier, C. C., et al. (1973). Adult auditory evoked vertex potentials in sleep. *Audiology, 12*, 116-128.

Ostroff, J. M., Martin, B. A., Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear Hear, 19*, 290-297.

Oticon. (2016). Poll Results Provide Snapshot of Hearing Loss Struggles. In: The Hearing Review.

Oxenham, A. J., Fligor, B. J., Mason, C. R., et al. (2003). Informational masking and musical training. *J Acoust Soc Am, 114*, 1543-1549.

Parbery-Clark, A., Tierney, A., Strait, D. L., et al. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience, 219*, 111-119.

Pereira, D. R., Cardoso, S., Ferreira-Santos, F., et al. (2014). Effects of inter-stimulus interval (ISI) duration on the N1 and P2 components of the auditory event-related potential. *International Journal of Psychophysiology, 94*, 311-318.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *J Clin Neurophysiol, 9*, 456-479.

Picton, T. W., Hillyard, S. A. (1974). Human Auditory Evoked-Potentials .2. Effects of Attention. *Electroencephalography and Clinical Neurophysiology, 36*, 191-199.

Polich, J. (1987). Task-Difficulty, Probability, and Inter-Stimulus Interval as Determinants of P300 from Auditory-Stimuli. *Electroencephalography and Clinical Neurophysiology, 68*, 311-320.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol, 118*, 2128-2148.

Polich, J., Donchin, E. (1988). P300 and the word frequency effect. *Electroencephalogr Clin Neurophysiol, 70*, 33-45.

Polich, J., Hoffman, L. D. (1998). P300 and handedness: on the possible contribution of corpus callosal size to ERPs. *Psychophysiology, 35*, 497-507.

Polich, J., Howard, L., Starr, A. (1985). Effects of age on the P300 component of the event-related potential from auditory stimuli: peak definition, variation, and measurement. *J Gerontol, 40*, 721-726.

Polich, J., Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol Psychol, 41*, 103-146.

Polich, J., McIsaac, H. K. (1994). Comparison of auditory P300 habituation from active and passive conditions. *Int J Psychophysiol, 17*, 25-34.

Polich, J., Ochoa, C. J. (2004). Alcoholism risk, tobacco smoking, and P300 event-related potential. *Clinical Neurophysiology, 115*, 1374-1383.

Pollack, I. (1975). Auditory Informational Masking. *Journal of the Acoustical Society of America, 57*, S5-S5.

Putter-Katz, H., Adi-Bensaid, L., Feldman, I., et al. (2008). Effects of speech in noise and dichotic listening intervention programs on central auditory processing disorders. *J Basic Clin Physiol Pharmacol, 19*, 301-316.

Qidwai, J. C., Watson, G. S., Weiler, J. M. (2002). Sedation, cognition, and antihistamines. *Curr Allergy Asthma Rep, 2*, 216-222.

Rhebergen, K. S., Versfeld, N. J., Dreschler, W. A. (2005). Release from informational masking by time reversal of native and non-native interfering speech. *J Acoust Soc Am, 118*, 1274-1277.

Richards, V. M., Neff, D. L. (2004). Cuing effects for informational masking. *Journal of the Acoustical Society of America, 115*, 289-300.

Rogers, C. L., Dalby, J., Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Lang Speech, 47*, 139-154.

Rosen, S., Souza, P., Ekelund, C., et al. (2013). Listening to speech in a background of other talkers: effects of talker number and noise vocoding. *J Acoust Soc Am, 133*, 2431-2443.

Roth, W. T., Krainz, P. L., Ford, J. M., et al. (1976). Parameters of temporal recovery of the human auditory evoked potential. *Electroencephalogr Clin Neurophysiol, 40*, 623-632.

Rugg, M. D., Pickles, C. D., Potter, D. D., et al. (1993). Cognitive brain potentials in a three-stimulus auditory "oddball" task after closed head injury. *Neuropsychologia, 31*, 373-393.

Ruhm, H. B., Jansen, J. W. (1969). Rate of Stimulus Change and Evoked Response .1. Signal Rise-Time. *Journal of Auditory Research, 9*, 211-216.

Russo, F. A., Pichora-Fuller, M. K. (2008). Tune in or tune out: age-related differences in listening to speech in music. *Ear Hear, 29*, 746-760.

Scharenborg, O., Kolkman, E., Kakouros, S., et al. (2016). The Effect of Sentence Accent on Non-native Speech Perception in Noise. *17th Annual Conference of the International Speech Communication Association (Interspeech 2016), Vols 1-5*, 863-867.

Schochat, E., Matas, C. G., Samelli, A. G., et al. (2012). From otoacoustic emission to late auditory potentials P300: the inhibitory effect. *Acta Neurobiol Exp (Wars), 72*, 296-308.

Schwent, V. L., Hillyard, S. A., Galambos, R. (1976). Selective Attention and Auditory Vertex Potential .1. Effects of Stimulus Delivery Rate. *Electroencephalography and Clinical Neurophysiology, 40*, 604-614.

Scott, S. K., Rosen, S., Wickham, L., et al. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J Acoust Soc Am, 115*, 813-821.

Sharma, M., Purdy, S. C., Munro, K. J., et al. (2014). Effects of broadband noise on cortical evoked auditory responses at different loudness levels in young adults. *Neuroreport, 25*, 312-319.

Souza, V. B. N., Muir, W. J., Walker, M. T., et al. (1995). Auditory P300 Event-Related Potentials and Neuropsychological Performance in Schizophrenia and Bipolar Affective-Disorder. *Biological Psychiatry, 37*, 300-310.

Sqires, K. C., Donchin, E., Herning, R. I., et al. (1977). Influence of Task Relevance and Stimulus Probability on Event-Related-Potential Components. *Electroencephalography and Clinical Neurophysiology, 42*, 1-14.

Squires, K. C., Hillyard, S. A., Lindsay, P. H. (1973). Cortical Potentials Evoked by Confirming and Disconfirming Feedback Following an Auditory Discrimination. *Perception & Psychophysics, 13*, 25-31.

Squires, K. C., Petuchowski, S., Wickens, C., et al. (1977). Sequential Effects on Event-Related Potentials - Comparison of Auditory and Visual Modalities. *Psychophysiology, 14*, 96-96.

Stapells, D. R. (2008). *Cortical Event-Related Potentials to Auditory Stimuli*. (6 ed.): Lippincoctt Williams & Wilkins.

Steiner, G. Z., Barry, R. J., Gonsalvez, C. J. (2014). Nontarget-to-nontarget interval determines the nontarget P300 in an auditory equiprobable Go/NoGo task. *Int J Psychophysiol, 92*, 113-121.

Stone, M. A., Canavan, S. (2016). The near non-existence of "pure" energetic masking release for speech: Extension to spectro-temporal modulation and glimpsing. *J Acoust Soc Am, 140*, 832.

Stone, M. A., Fullgrabe, C., Moore, B. C. J. (2012). Notionally steady background noise acts primarily as a modulation masker of speech. *Journal of the Acoustical Society of America, 132*, 317-326.

Stuart, A., Phillips, D. P., Green, W. B. (1995). Word recognition performance in continuous and interrupted broad-band noise by normal-hearing and simulated hearing-impaired listeners. *Am J Otol, 16*, 658-663.

Tillman, T. W., Carhart, R. (1963). Effect of Competing Speech on Aided Discrimination of Monosyllabic Words. *Journal of the Acoustical Society of America, 35*, 1900-&.

Tremblay, K. L., Friesen, L., Martin, B. A., et al. (2003a). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear Hear, 24*, 225-232.

Tremblay, K. L., Friesen, L., Martin, B. A., et al. (2003b). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear and Hearing, 24*, 225-232.

Tremblay, K. L., Kalstein, L., Billings, C. J., et al. (2006). The neural representation of consonant-vowel transitions in adults who wear hearing AIDS. *Trends Amplif, 10*, 155-162.

Tremblay, K. L., 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 Engen, K. J., Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *J Acoust Soc Am, 121*, 519-526.

Vander Werff, K. R., Arduini, S. (2011). Masker type and SNR effects on cortical responses to speech. In *American Auditory Society*. Scottsdale, AZ.

Vander Werff, K. R., Niemczak, C. E., O'Flaherty, S., et al. (2016). Effects of Number of Talkers and Spectrotemporal Content of Informational Masking on the CAEP. In *The American Academy of Audiology*. Phoenix, AZ. .

Vander Werff, K. R., Rieger, B. (2019). Impaired auditory processing and neural representation of speech in noise among symptomatic post-concussion adults. *Brain Injury, 33*, 1320-1331.

Veale, J. F. (2014). Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. *Laterality, 19*, 164-177.

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology, 34*, 131-156.

Wagner, M., Roychoudhury, A., Campanelli, L., et al. (2016). Representation of spectro-temporal features of spoken words within the P1-N1-P2 and T-complex of the auditory evoked potentials (AEP). *Neurosci Lett, 614*, 119-126.

Watson, C. S., Kelly, W. J., Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *J Acoust Soc Am, 60*, 1176-1186.

Watson, C. S., Kidd, G. R. (2007). Studies of tone sequence perception: effects of uncertainty, familiarity, and selective attention. *Front Biosci, 12*, 3355-3366.

Watson, C. S., Nichols, T. L. (1976). Detectability of auditory signals presented without defined observation intervals. *J Acoust Soc Am, 59*, 655-668.

Weintraub, S., Dikmen, S. S., Heaton, R. K., et al. (2013). Cognition assessment using the NIH Toolbox. *Neurology, 80*, S54-S64.

Weintraub, S., Dikmen, S. S., Heaton, R. K., et al. (2014). The Cognition Battery of the NIH Toolbox for Assessment of Neurological and Behavioral Function: Validation in an Adult Sample. *Journal of the International Neuropsychological Society, 20*, 567-578.

Whiting, K. A., Martin, B. A., Stapells, D. R. (1998). The effects of broadband noise masking on cortical event-related potentials to speech sounds /ba/ and /da. *Ear Hear, 19*, 218-231.

Wilson, R. H., Carnell, C. S., Cleghorn, A. L. (2007). The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. *J Am Acad Audiol, 18*, 522-529.

Wilson, R. H., McArdle, R. (2007). Intra- and inter-session test, retest reliability of the Words-in-Noise (WIN) test. *J Am Acad Audiol, 18*, 813-825.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc Natl Acad Sci U S A, 90*, 8722-8726.

Wong, P. C., Uppunda, A. K., Parrish, T. B., et al. (2008). Cortical mechanisms of speech perception in noise. *J Speech Lang Hear Res, 51*, 1026-1041.

Wronka, E., Kaiser, J., Coenen, A. M. (2008). The auditory P3 from passive and active three-stimulus oddball paradigm. *Acta Neurobiol Exp (Wars), 68*, 362-372.

Zhang, C., Arnott, S. R., Rabaglia, C., et al. (2016). Attentional modulation of informational masking on early cortical representations of speech signals. *Hear Res, 331*, 119-130.

Zhang, C., Lu, L. X., Wu, X. H., et al. (2014). Attentional modulation of the early cortical representation of speech signals in informational or energetic masking. *Brain and Language, 135*, 85-95.

Christopher Edward Niemczak, AuD

Education

Fall 2013-Present	Syracuse University
Degree	AuD/PhD Candidate in Audiology with a concentration in Neuroscience. AuD coursework completed May 2017, externship completed July 2018, PhD anticipated Spring 2020. GPA 3.93
Dissertation	"Cortical auditory processing of informational masking effects by target-masker similarity and stimulus uncertainty."
Advisor	Kathy R. Vander Werff, PhD
Fall 2009-May 2013	University of Vermont BS Communication Sciences and Disorders. Minor: Chemistry. GPA 3.35

Research Experience

Fall 2015-Present	Graduate Research Assistant, Communication Sciences and Disorders, Syracuse University Auditory Electrophysiology Lab, Kathy Vander Werff, PhD
Projects	Informational masking effects on neural encoding of an acoustic change
	Comparison of silent gap in noise cortical auditory evoked potentials in matched tinnitus and no-tinnitus control subjects
July 2017-July 2018	Fourth-year Research Externship at The National Center for Rehabilitative Auditory Research (NCRAR), Portland, OR, Naomi Bramhall, AuD, PhD
Projects	Evoked potentials reveal noise exposure-related central auditory changes despite normal audiograms
	Uncovering physiological markers of hidden hearing loss
Fall 2012-Spring 2013	Research Assistant, Autism Spectrum Disorder Social Cognition and language development lab, University of Vermont, Tiffany Hutchins, PhD
Project	Visual attention in Autism Spectrum Disorder eye tracking study

Publications

Niemczak, C.E. & Vander Werff, K.R. (2019). Informational masking effects on neural encoding of stimulus onset and acoustic change. *Ear and Hearing,* 40(1): 156-167. doi:10.1097/AUD.0000000000000604

Bramhall, N.F., **Niemczak, C.E.,** Kampel, S.D., Billings, C.J., McMillan, G.P. (2020). Evoked potentials reveal noise exposure-related central auditory changes despite normal audiograms. *American Journal Of Audiology*, 10(1): 1-13. doi:10.1044/2019_AJA-19-00060

Professional Presentations

Niemczak, C.E., & Vander Werff, K.R. (2020, March). Objective and behavioral indices of linguistic/phonetic similarity in informational masking. NIH Mentored Student Award poster presented at the American Auditory Society Annual Meeting, Scottsdale, AZ.

Niemczak, C.E., Kampel, S., Billings, C.J., Bramhall, N. (2019, February). Examining the effects of noise exposure on the auditory system using evoked potentials. Podium presentation presented at the Association for Research in Otolaryngology, Baltimore, MD.

Niemczak, C.E. (2018, March and April). Effect of noise exposure on auditory evoked potentials in humans. Oral presentation at the Oregon Hearing Research Center and at Oregon Health and Science University Grand Rounds at the NCRAR, Portland, OR.

Vander Werff, K.R., **Niemczak, C.E.,** O'Flaherty, S., Rekola, S., and Doci, S. (2016, April). Effects of number of talkers and spectrotemporal content of informational masking on the CAEP. Poster presented at the AudiologyNOW Annual Convention of the American Academy of Audiology of Audiology, Phoenix, AZ.

Niemczak, C.E. & Vander Werff, K.R. (2016, February). Informational masking effects on neural encoding of acoustic change. NIH Mentored Student Award poster presented at the American Auditory Society Annual Meeting, Scottsdale, AZ.

Awards and Academic Achievements

Spring 2020	Syracuse University Graduate Student Organization Travel Grant
Spring 2019	Council of Academic Programs in Communication Sciences and Disorders (CAPCSD) PhD Scholarship ($20,000)
Fall 2018	Association for Research in Otolaryngology Travel Award
Fall 2018	Founder and Director of Syracuse University Communication Sciences and Disorders Data Club
Fall 2014-Spring 2017	Syracuse University Graduate Fellowship
Spring 2016	AAA Poster Award for Excellence in Electrophysiology,
Fall 2015-Fall 2016	President, Syracuse University Chapter Student Academy of Audiology (SAA)

Fall 2015-Spring 2020	Syracuse Neuroscience Graduate Travel Award
Fall 2015-Spring 2020 (NIH/NIDCD),	Graduate Student Mentored Research Travel Award to attend the American Auditory Society Meeting, Scottsdale AZ
Fall 2014	Audiology/Hearing Science Research Travel Award (ARTA) Scholarship to attend the ASHA Convention, Orlando FL

Teaching Experience

Undergraduate Course Instructor

Fall 2018, 2019	CSD 212, Introduction to Communication Sciences and Disorders

Guest Undergraduate Lecturer

Spring 2019, 2020	CSD 212, Introduction to Communication Sciences and Disorders, *Hearing aids and assistive listening devices*
Fall 2018	CSD 315, Anatomy and Physiology of Speech and Hearing, *Phonological physiology, Articulation and resonance anatomy part I and II*
Spring 2017	CSD 212, Introduction to Communication Sciences and Disorders, *Communication and multicultural considerations, Deaf culture*

Guest Graduate Lecturer

Spring 2019,2020	CSD 663, Auditory Evoked Potentials, *Middle latency response and P1-N1-P2 basics, Advanced cortical auditory evoked potentials: ACC, MMN, & P300*
	CSD 658, Auditory Anatomy and Physiology, *Neuronal anatomy and physiology*
Fall 2018	CSD 637, Auditory Instrumentation, *Transducers, microphones, and sound level meters*

Lab Coordinator

Spring 2019	CSD 663, Auditory Evoked Potentials
Fall 2018	CSD 637, Instrumentation for Speech and Hearing
Fall 2016	CSD 637, Instrumentation for Speech and Hearing
Spring 2016	CSD 673, Hearing Aids I, CSD 663, Auditory Evoked Potentials
Fall 2015	CSD 637, Instrumentation for Speech and Hearing

Clinical Experience

Summer 2018-2019	Fourth-year Clinical Externship at Portland, OR, VA Medical Center
Spring/Fall 2014, 2016	Gebbie Clinic Hearing Aid Technician, Syracuse University
Fall 2013-Spring 2017	Syracuse University Clinical Audiology Internships: o The Chafe Center ENT, Utica, NY, Summer 2016

- Preferred Audiology (private practice), Camillus, NY, Spring 2016
- Oswego Health System, Oswego, NY, Fall 2015
- Syracuse VA Medical Center, Syracuse, NY, Summer 2015
- Westside Audiology Services (private practice), Camillus, NY, Spring 2015
- Onondaga-Cortland-Madison Board of Cooperative Educational Services, Syracuse, NY, Fall 2014
- Gebbie Clinic Graduate Student Clinician, Syracuse, NY, Fall 2013-Fall 2014

Spring 2010-Summer 2013	Emergency Medical Technician, City of Philadelphia, PA

Service Experience

Spring 2020	Syracuse neuroscience research conference planning committee, website coordination/registration, Syracuse University
Summer 2019	Graduate teaching mentor and teaching consultant for all-university teaching assistant orientation program and graduate school, Syracuse University
Spring 2019	Syracuse neuroscience research conference planning committee, registration and scheduling coordinator, Syracuse University
	Graduate admissions consultant for live admissions webinar, Communication Sciences and Disorders, Syracuse University
	Ad-Hoc Reviewer for American Journal of Audiology (AJA)
Spring 2016	Graduate admissions committee, Communication Sciences and Disorders, Syracuse University,
Fall 2015-Spring 2016	President of SAA, Syracuse University Chapter

Professional Associations

Spring 2019-Present	Nu Rho Psi, National Honor Society in Neuroscience
Fall 2014-Present	American Auditory Society (AAS) Student Member
Fall 2014-Present	National Student Speech, Language, and Hearing Association
Fall 2013-Present	Syracuse University Student Academy of Audiology