CAN YOU HEAR ME NOW? THE RELATIONSHIP BETWEEN STATISTICAL
LEARNING AND SPEECH PERCEPTION UNDER DEGRADED LISTENING
CONDITIONS

by

Alexandria L. Muise-Hennessey

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

at

Dalhousie University
Halifax, Nova Scotia
August 2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The present study investigated whether there was a relationship between statistical learning and the ability to use top-down processing to predict incoming speech using electroencephalography (EEG). Statistical learning abilities were measured via an artificial grammar learning (AGL) task, where assimilation of transitional probabilities of stimuli were indexed using learning scores and the P600, an event-related potential (ERP) component that responds to syntactic violations. Top-down processing was indexed using the N400 — an ERP component that responds to semantic violations — in response to a speech perception task with two conditions: with and without noise. It was hypothesized that without noise, an N400 would be seen when the final word of a sentence was semantically incorrect, and that noise should attenuate this effect. Without noise, N400 and P600 amplitudes were expected to correlate, supporting evidence for a relationship between these neurocognitive processes. In the presence of noise, people who were better at the statistical learning task should have a reduced N400-mismatch effect, as they would rely on top-down processing to fill in the missing information. This should not be observed in people who were worse at the AGL task. Based on the median of the AGL learning scores, people were split into two groups: learners and non-learners. The AGL task did not elicit any significant effects in non-learners. Learners had an N400-like effect in the central parietal scalp and a frontal positivity. An N400 in response to the speech perception task was found for both quiet and noise conditions. Furthermore, there was a relationship between statistical learning and speech perception. Non-learners had a positive correlation between the N400 and AGL grammaticality effect regardless of the listening condition. In contrast, learners had a negative correlation in the absence of noise; this relationship reversed in the presence of noise, coinciding with their reduction in N400 amplitude. This reduction in N400 amplitude in noise suggests that learners may have strong expectations of what the final word should be. When hearing is impaired, learners may perceive the final word as a match rather than a mismatch. The results suggest that people who are more sensitive to the underlying statistical frequencies of stimuli may rely more on top-down processing to fill in missing information when engaged in a noisy environment.

LIST OF ABBREVIATIONS USED

SNR   Signal-to-Noise Ratio

EEG   Electroencephalography

ERP   Event-Related Potential

SPIN   Speech Perception in Noise

AGL   Artificial Grammar Learning

ICA   Individual Component Analysis

ROI   Region of Interest

GAM   Generalized Additive Mixed-Effects Model

REML   Restricted Maximum Likelihood

M   Mean

SD   Standard Deviation

CI   Confidence Interval

SNW   Slow Negative Wave

ACKNOWLEDGEMENTS

CHAPTER 1 INTRODUCTION

Speech perception is the ability to both perceive and process incoming auditory information in a variety of conditions, including quiet and noisy situations. Some individuals are better at perceiving speech than others, especially when in engaged in an environment like a noisy party. Speech perception may be impaired by a variety of factors, including hearing loss, reading abilities, short-term memory, and other cognitive factors (Lorenzi, Gilbert, Carn, Garnier, & Brian, 2006; McBride-Chang, 1995).

Statistical learning is the ability to implicitly pick up on statistical regularities of events in the environment (Kaufman et al., 2010). Statistical learning is distinct from explicit learning because it is an innate and unconscious process (Kaufman et al., 2010). Research has focused on determining whether there is a relationship between statistical learning, considered a domain-general ability, and the domain-specific abilities involved in language learning (Frost, Siegelman, Narkiss, & Afek, 2013). Research has shown that even infants appear to rely on statistical learning abilities to learn how to segment speech sounds (Aslin, Saffran, & Newport, 1998).

Statistical learning may explain why some individuals are better than others when perceiving speech: those who are better at statistical learning tasks may be better at picking up on the statistical regularities of events that they experience in general. For example, some words occur more frequently in language than other words. Based on this, statistical learning may be a key ability in understanding speech perception, and this may in part explain why some people are better at perceiving speech, especially in a noisy environment.

1.1 Speech Perception

The Mechanics of Speech Perception

Accurately perceiving spoken language relies on the perception of auditory information and the ability to process this input in a meaningful way. This perception may be impaired either at the beginning of the process, the sound signal and audition (e.g., hearing loss), or at the central processing system (e.g., aphasia). Hearing loss typically involves mechanical issues in the detection and/or transduction of the sound to action potential in the auditory nerve. Sound waves are transmitted through a medium, such as air, and are conducted through the ear canal before eventually causing the inner hair cells of the cochlea to oscillate (Yost, 2003). These vibrations are transduced into an electrical neural signal, transmitting a train of action potentials through the auditory nerve to the cochlear nucleus, located in the brainstem. From the cochlear nucleus, auditory information is eventually relayed through the thalamus before reaching the primary auditory cortex (Yost, 2003). Hearing loss can occur at multiple sites throughout this pathway, including the ossification of the ear bones (otosclerosis), loss of hair cells (either due to aging, trauma, etc), or auditory nerve damage. If the auditory signal cannot be conducted, then speech perception will be significantly, if not totally, impaired. Hearing impairment can range from mild to severe, for example, in a quiet environment, an individual with hearing loss may have a difficult time hearing a conversation and this impairment can be worsened when in a noisy environment such as a restaurant (Lorenzi et al., 2006).

The cognitive processing of auditory input, such as mapping auditory signals onto the mental lexicon, may also be impaired. Processing impairment will affect how well people

perceive speech, and processing can be affected through various ways. For example, aphasia is an acquired language deficit, typically caused by brain damage such as a stroke (A. R. Damasio, 1991, 1992; H. Damasio, 1991). Broadly, people with aphasia often have deficits in either producing or comprehending speech (Caramazza & Zurif, 1976; Goodglass, Kaplan, & Barresi, 2001; Goodglass, 1993; Hickok, 2009). Another example of a processing impairment that negatively affects speech perception is central auditory processing disorder (Bamiou, Musiek, & Luxon, 2001). Central auditory processing disorder results in impaired speech processing, especially in the presence of noise; the auditory nervous system is intact, yet there is discordance between what signal was received and how it is processed (Bamiou, Musiek, & Luxon, 2001; Rintelmann, 1985).

Processing Speech

Speech perception deficits can result from impairment at any point in the pathway. Even without impairment, spoken language is difficult to process because the auditory signal is extended through time (McClelland & Elman, 1986). The temporal nature of spoken language makes its processing complex, because speech is not processed as whole words, but rather in meaningful segments such as phonemes. Furthermore, in order to derive meaning, a record of what has been previously processed must also be maintained (McClelland & Elman, 1986).

After the auditory signal has been received in the brain, the acoustic information must be processed and mapped in a manner that leads to recognizable and meaningful stimuli such as words. McClelland and Elman (1986) developed the TRACE model to explain the mechanics behind speech processing. Spoken language is theorized to first be processed based on the acoustic features of the input, and these features provide the basic

inventory for categorizing different language sounds (Hickok & Poeppel, 2007). Spectrograms are often used to visualize the frequency spectrum for spoken language across time. Feature processing for spoken language is important because it can discriminate between the beginnings and endings of words (McClelland & Elman, 1986). After feature processing, speech signals are theorized to be processed based on phonemes. For example, the word *talk* has an initial phoneme of */t/*. As the initial phoneme */t/* is processed, words that begin with the same initial phoneme, are activated in the mental lexicon (Marslen-Wilson & Tyler, 1980). So, processing the initial phoneme */t/* may lead to activation of words such as *tree*, *trip*, *take*, *tale*, *tall*, and *talc*. Phoneme processing in the TRACE model overlaps with the COHORT model of speech perception by Marslen-Wilson and Tyler (1980). As the auditory input is further processed, words that are dissimilar (e.g., *tree* and *trip*) to the signal are eliminated from the model and those that remain similar become more highly activated (e.g., *take*, *tale*, *tall*, and *talc*) (Luce & Pisoni, 1998). As further processing occurs, potential candidates that are dissimilar or are not compatible with top-down processing are eliminated until eventually a single word is left: *talk* (Luce, 1986; McClelland & Elman, 1986). Unlike the COHORT model, the TRACE model is able to account for different errors that may occur in the speech signal, due to the fact that the TRACE model accounts for feature level processing. For example, if the initial phoneme of a word is mispronounced, the COHORT model cannot account for why people can understand these mispronunciations but the TRACE model can. This is representative of real speech signals because there are instances in which speakers accidentally mispronounce words, but listeners are still typically able to process and understand the word. In summary, the TRACE model

theorizes that auditory signal is processed based on the acoustic features of the input, which are eventually mapped onto known phonemes. As the input is further processed, more information becomes available so that the phonemes can be linked to eventually create a meaningful stimulus: a word (McClelland & Elman, 1986).

Many factors influence how spoken language is processed: some phonemes, such as */m/* and */n/,* share similar acoustic features which can result in words that contain these phonemes to be mistaken for one another (Jay, 2003). For example, the words *mice* and *nice* may be confused with one another. Phonemes are also often restructured based on the preceding and proceeding words, for example, the consonant */p/* changes depending on whether it is used in the word *pig* or *spiral* (McClelland & Elman, 1986). Speech rate and dialectical variations can also affect the pronunciation of phonemes. Another challenge with recognizing speech is that while we perceive clear boundaries between words in our native language, spectrograms show a different story. Often there is a distinct lack of boundaries between words and sounds frequently overlap. Errors in speech perception may arise due to incorrect word segmentation, and these errors can be influenced by context (McClelland & Elman, 1986). An additional factor that can influence how spoken language is processed is context (McClelland & Elman, 1986). Sentence context influences how both subsequent and prior information is processed (McClelland & Elman, 1986). Clearly, speech perception is influenced by a myriad of factors.

The Effect of Noise on Speech Perception

Speech perception also depends on the hearer's ability to detect and encode the speech signal in the presence of noise. Noise can be defined as any acoustic information

unrelated to the speech signal. Under ideal conditions (e.g., a quiet room with a slow, careful speaker), speech perception is generally highly accurate. In the absence of interfering noise, speech perception can be viewed as a largely "bottom-up" process in which auditory input is mapped to linguistic representations (Sörqvist & Rönnberg, 2012). However, speech often occurs outside of this ideal scenario, such as in a busy coffee shop or at a restaurant. These situations introduce background noise that may be louder than the target speech and occupy similar frequency ranges, which makes accurately perceiving speech difficult, even with normal hearing (Anderson, Skoe, Chandrasekaran, & Kraus, 2010). In order to make sense of the incoming input, people need to be able to focus on the target signal while ignoring extraneous auditory input (Bregman, 1994). In a noisy environment, bottom-up processing is more likely to result in a noisy target speech signal, due to the extraneous auditory signals in the environment. Thus, there may be a shift to increasingly rely on top-down processing in combination with bottom-up processing to derive meaning from the target signal.

Research on how speech is processed in a noisy environment has found several factors that are necessary to accurately perceive speech. People with normal hearing rely on environmental cues such as temporal and spectral dips to aid with processing (Peters, Moore, & Baer, 1998). Temporal dips are moments when the background noise is lower than the target signal, resulting in a higher signal-to-noise ratio (SNR) thus improving accuracy (Stuart, Givens, Walker, & Elangovan, 2006; Stuart, 2008). Spectral dips are when the target spectrum becomes more noticeable than the background input, which also improves speech perception accuracy (Peters et al., 1998; Stuart, 2008). Other cues can also be used to aid with speech perception under degraded listening conditions, such

as visual cues (Cooke, Barker, Cunningham, & Shao, 2006; Summerfield, 1992). For example, reading lips can help listeners discriminate whether the phoneme */m/* or */n/* was said, which in the absence of visual input can be difficult to discriminate (Cooke et al., 2006; Summerfield, 1992).

Temporal cues, such as pauses in words or transitions between syllables (Nourski et al., 2009), are another important factor for speech perception because they are necessary for segregating auditory input, for example, phoneme discrimination to know when one word ends and a new one begins (Anderson et al., 2010). However, in a noisy room the boundaries between words may become less clear and words may be inaccurately segmented, resulting in confusion. While inaccuracies do occur, people generally are adept at filling in missing information to determine what was actually said. Warren (1970) examined whether people used top-down processing to fill in missing information when speech was either masked by noise or segments of the speech were replaced with a non-speech sound such as a tone. Participants listened to recorded sentences and a portion of the target word was masked by a cough, tone, or silence. After listening to the recordings, participants were given a piece of paper that contained the sentence and were asked to circle the area in which the mask or silence occurred. Warren (1970) found that when a portion of the target word was silent, participants were able to accurately identify where the silence occurred. However, when the target word was masked with a tone or a cough, people filled in the missing information via sentence context to perceive the masked word. This effect, the phonemic restoration effect, demonstrates the effects of top-down processing and how it can be used to fill in missing auditory input (Warren, 1970).

The effect of noise on speech perception has been studied using electroencephalography (EEG), which records the continuous neural activity of the brain that occurs naturally. In order to investigate the relationship between EEG and specific cognitive, motor, or sensory events, EEG is time-locked to experimentally controlled events of interest, to elicit event-related potential (ERPs). ERPs are calculated by averaging the recorded activity over numerous trials (Bressler & Ding, 2006; Kaan, Harris, Gibson, & Holcomb, 2000). ERPs can reveal information about processes that may be activated by experimental procedures (Bressler & Ding, 2006). ERPs can either be positively or negatively going waveforms, and conventionally are named after the polarity and time period of the waveform (Bressler & Ding, 2006). In this way, changes in ERP amplitude can be used to investigate neurocognitive processes surrounding speech perception and the effect of noise.

One method to investigate the neurocognitive processes involved in speech perception is through the speech perception in noise (SPIN) task. In this task, people listen to sentences that either end with a predictable or unpredictable word. Predictability of the final word is measured by the Cloze probability, which is based on the context of the sentence (Block & Baldwin, 2010). To measure Cloze probability, people read sentences that are missing a word and are asked to supply what they think the correct word should be. The proportion of people who supply the same word can be measured, and this measurement is the Cloze probability. Therefore, the higher the Cloze probability of a sentence, the more predictable the final word of the sentence is. By changing the highly predictable final word of a sentence to have 0% Cloze probability, researchers can

measure the electrical activity differences between sentences that have a highly expected final word, and those that deviate from expectations (Block & Baldwin, 2010).

Kutas and Hillyard (1980) found a negative going ERP that began 400 ms post-stimulus onset during a sentence-reading task (Kutas & Federmeier, 2011). The authors used sentences that either had semantically congruous or incongruous final words. An example of an incongruous sentence would be "*He takes his coffee with sugar and horse*", where *horse* is an unexpected final word. The authors expected that the difference in activity between semantically incongruous and congruous sentences would elicit a positivity at 300 ms but instead found the N400 (Kutas & Hillyard, 1980). The N400 is related to semantic processing during language comprehension tasks (Brown & Hagoort, 1993) and is elicited by both visual and auditory words (Kutas & Federmeier, 2011). Auditory stimuli tend to elicit an N400 effect that begins approximately 200 – 600 ms post-stimulus onset (earlier than visual N400 effects) and tends to be maximal at midline electrodes around the vertex of the head (Kutas & Federmeier, 2011). Research has demonstrated that the amplitude of the N400 effect is inversely proportional to the expectancy of a target word: the smaller the Cloze probability, the higher the amplitude (Brown & Hagoort, 1993; Kutas & Federmeier, 2011). Studies have shown that the addition of background noise reduces the N400 effect (Aydelott, Dick, & Mills, 2006; Strauß, Kotz, & Obleser, 2013). It has also been shown that different levels of background noise can either enhance (high SNRs) or abolish (low SNRs) the N400 effect (Daltrozzo, Wioland, & Kotchoubey, 2012). Specifically, high SNRs increase the dB level of the target speech signal relative to the background noise, thus making it easier to hear the target speech signal. In contrast, low SNRs increase the dB level of the noise

signal relative to the target speech signal, making perception more difficult. In a low SNR, people may not hear the mismatch final word of the sentence and thus an N400 effect may not be elicited.

Other Factors that Affect Speech Perception

Research has shown that aside from the systems involved in speech perception, cognitive abilities such as executive function, reading abilities, and working memory are also important for accurate speech perception. Executive functioning is the ability to coordinate and control our behaviour and thoughts (Luria, 1966; Shallice, 1982). Executive functioning is involved in abilities such as decision-making, selective attention, and response inhibition (Blakemore & Choudhury, 2006). Selective attention is an important ability because it allows people to ignore unimportant information, which is especially important for speech perception because listeners may have to attend to one speaker and ignore others. Mesgarani and Chang (2012) had people selectively listen to one speaker while listening to a recording with two speakers. The authors reconstructed spectrograms based on the cortical responses of participant as they selectively attended to one listener. The spectrograms demonstrated salient features of the attended speaker, as if the participants had attended to that speaker in isolation. This study demonstrated that the cortical responses to speech are not solely based on the acoustic input, but also reflect relevant information regarding the intentions of the listener (Mesgarani & Chang, 2012).

Reading abilities are also implicated in speech perception: children with reading disabilities are less accurate in speech perception tasks including discrimination and identification tasks (Werker & Tees, 1987). Children with dyslexia show deficits in speech perception when performing speech perception in noise tasks (Ziegler, Pech-

Georgel, George, & Lorenzi, 2009). Other studies have concluded that decreased

performance in speech perception in reading-disabled children may either be due to

speech perception difficulties, decreased short-term memory, or a combination of both

(De Weirdt, 1988). Phonological working memory is also known to be an important

factor in speech perception. Working memory capacity is important for speech perception

because of the temporal aspects of auditory input: people need to be able to remember

what was previously said in order to contextualize and understand new input (McClelland

& Elman, 1986). Sörqvist and Rönnberg (2012) investigated the relationship between

working memory capacity and speech perception abilities. In this experiment, participants

compared item sizes, completed a reading span test, as well as listened to stories that

were either normal speech or distorted and then answered questions about the stories.

Sörqvist and Rönnberg (2012) determined that people with large working memory

capacities are less adversely affected by distorted listening conditions. It is likely that

larger working memory capacities are advantageous in this situation because a larger

memory span likely allows people to hold more information at once, which can be used

to understand sentence context, and may also have more cognitive resources available to

filter noise and perceive speech (Pichora-Fuller, Schneider, & Daneman, 1995). Pichora-

Fuller et al. (1995) examined how well older and young adults performed on a SPIN task,

as well as tested the working memory capacity of each group by asking participants to

recall final words from the SPIN task. Older adults had lower recall compared to young

adults on the working memory task, and working memory capacity was adversely

affected by noise for both groups. The authors hypothesized that when noise is present,

11

working memory capacity decreases in both groups because the energetic resources used for working memory are reallocated for auditory processing (Pichora-Fuller et al., 1995).

Statistical learning may be a reason why some people are better than others at accurately perceiving speech under degraded listening conditions. Statistical learning involves the implicit knowledge about the underlying frequencies of stimuli in the natural environment. For example, to determine the final word in the following sentence, *The doctor's suitcase was worn and obviously very ...*, people can use the context of the sentence to determine that the final word should be *old*, because if something is worn out it is likely not new. People who are able to accurately predict incoming information based on prior context are examples of individuals relying on statistical learning abilities, which the current study hypothesizes is an important ability involved in accurately perceiving speech under degraded listening conditions.

1.2 Statistical Learning

Statistical Learning Abilities

Statistical learning has been defined as the fundamental ability to implicitly learn the underlying statistical frequencies of stimuli within the environment (Kaufman et al., 2010). Aslin et al. (1998) provided infants with a continuous auditory stream containing speech syllables that had no prosodic or acoustic (e.g., pauses) cues for word boundaries. The presentation of the stimuli followed specific patterns so that some syllables were more likely than others to co-occur. The authors found that after a short time period, infants were able to segment these syllables into word-like units (Aslin et al., 1998). This study is an example of how statistical learning can be used to extract the statistical information characterizing stimuli and how expectations are formed based on these

probabilities. Statistical learning is a useful process because it allows people to unconsciously segment continuous information in a meaningful way (Frost, Armstrong, Siegelman, & Christiansen, 2015).

Statistical learning has been implicated in many domains. Saffran, Aslin, and Newport (1996) exposed 8-month old infants to a stream of nonsense syllables for a two-minute period. These nonsense syllables were presented in a probabilistic fashion, so some syllables were more likely to co-occur together than with other syllables. The speech was presented without any acoustic (e.g., pauses) or prosodic cues, as to not alert the infants to any word boundaries. Afterwards, the infants listened to non-words composed of three nonsense syllables that had a higher probability of co-occurrence and non-words of syllables that did not co-occur together. The authors found that infants listened longer to non-words in which the syllables did not co-occur in the original speech stream, suggesting that the infants found these types of non-words to be novel. Thus, infants were found to be able to use statistical learning to extract meaningful information about the underlying statistical regularities of the nonsense syllables (Saffran, Aslin, & Newport, 1996).

Statistical learning has been implicated outside of the auditory domain and in other sensory domains such as vision and touch (Conway & Christiansen, 2005). For example, in a visual statistical learning task, participants were exposed to sequences of different shapes that followed specific patterns during a learning phase (Fiser & Aslin, 2002). After the learning phase, participants were asked to judge how familiar sequences were to them. Participants were shown novel sequences that followed the shape patterns from the learning phase as well as novel sequences that did not follow the pre-established patterns.

The results demonstrated that participants were able to implicitly learn the patterns of shapes without specific instructions within a relatively short time period (Fiser & Aslin, 2002). Participants are also able to detect statistical regularities in other visual stimuli such as colours (Conway, Bauernschmidt, Huang, & Pisoni, 2010).

Statistical learning has also been demonstrated in a variety of language tasks. For example, infants who were exposed to auditory phonotactic regularities of non-English syllables for a brief period listened longer to violation syllables (Chambers, Onishi, & Fisher, 2003). Statistical learning has also been used to study how well people can detect long-distance statistical frequencies between words (Gómez, 2002; Onnis, Christiansen, Chater, & Gómez, 2003). Misyak, Christiansen, and Tomblin (2010) found that people's ability to detect non-adjacent dependencies in a statistical learning paradigm correlated with performance in accurately processing written sentences that contained non-adjacent dependencies.

Assessing Statistical Learning

A popular way to assess statistical learning is through a serial reaction time task, which involves the presentation of visual stimuli in four different locations on a computer monitor (Kaufman et al., 2010; Robertson, 2007). Participants are asked to respond as quickly as possible to the visual stimuli by pressing the corresponding key on the keyboard (Robertson, 2007). Unbeknownst to participants, the presentation order of the visual stimuli have an underlying transitional probability, so that certain stimuli sequentially appear more often than others. For example, there is a 50% probability that the next item shown after the letter A is either B or C, but there is a 0% chance of the letter D following A (Reber, 1989). Control sequences, in which the visual stimuli no

longer follow the pre-established grammar, are intermixed throughout the task (Robertson, 2007). The reaction times in response to the visual stimuli are recorded. Through repeated exposures, participants implicitly learn the statistical frequencies of the sequences and become faster at responding to the sequences that follow the pre-established grammar (Robertson, 2007).

Another method of assessing statistical learning involves the use of an artificial grammar learning (AGL) task, which involves exposing participants to items with certain transitional probabilities. After a "learning" period of repeated exposure to these statistical frequencies (grammatical sequences), a "testing" period occurs in which participants are shown a mixture of sequences that follow the grammar and sequences that do not follow this grammar (Frost et al., 2015). In some experiments, participants are explicitly aware about the existence of patterns in the sequences, while in others participants are naïve. Generally, in these types of experiments, participants are asked to either repeat back each sequence immediately after being exposed to it (Conway et al., 2010; Conway, Ellefson, & Christiansen, 2003; Conway, Karpicke, & Pisoni, 2007; Conway, Pisoni, Anaya, Karpicke, & Henning, 2011) or to judge the familiarity of sequences (Conway & Christiansen, 2005, 2006; Conway et al., 2003).

People who are better at acquiring these statistical regularities are assumed to have formed expectations based on the transitional probabilities they were exposed to (Frost et al., 2015). Depending on the AGL paradigm, learning is assessed in two different ways. If people were asked to judge the grammaticality of the testing sequences, people who perform above chance level are assumed to have assimilated the statistical regularities of the task (Frost et al., 2015). If people are asked to repeat back the sequences, a "learning

score" can be generated: the sum of correctly repeated ungrammatical sequences (multiplied by the length of each sequence) are subtracted from the sum of correctly repeated grammatical sequences (multiplied by the length of each sequence) (see Conway et al., 2010). The use of a learning score as a way to measure statistical learning has been argued to be superior over judgments based on the grammaticality of sequences because learning scores indirectly measure an individual's learning (Conway et al., 2010).

Generalizability of Statistical Learning

Over the past decade, research on statistical learning has focused on determining whether statistical learning is a comprehensive general cognitive ability that stretches across domains and explains how all cognitive systems learn (Frost et al., 2015). Statistical learning was initially believed to be generalizable so that if someone excelled at statistical learning in one domain (e.g., visual AGL task), they would also excel for similar tasks in other domains (e.g., auditory AGL task). However, research has consistently shown that statistical learning is not generalizable across domains as previously believed. Instead, research demonstrates that statistical learning is subject to both stimulus and modality specificity (Redington & Chater, 1996; Tunney & Altmann, 1999).

As an example, when people participate in a visual AGL task they may extract the information about the transitional probabilities of the sequences (Redington & Chater, 1996). However, when they participate in a subsequent auditory AGL task that uses sequences with the same transitional probabilities but different stimuli, the theory of the generalizability of statistical learning would predict that these participants would apply the previous rules from the visual task to the auditory task. This would be an example of

rule transfer across modalities. However, research has shown that this does not happen; people do not apply their previous knowledge about statistical frequencies from one task to another (Redington & Chater, 1996).

Instead, it appears that statistical learning is only generalizable within the same domain (Frost et al., 2015). For example, if people participate in two visual AGL tasks that use colours as the stimuli, people apply their knowledge about the statistical frequencies from the first task to the next. Sometimes though, this generalizability may not occur even within the same modality depending on whether the tasks use stimuli with perceptually different characteristics, for example, letters versus shapes (Conway & Christiansen, 2006). Conway and Christiansen (2006) performed three separate experiments investigating the generalizability of statistical learning across and within domains. The authors examined whether knowledge about the statistical frequencies of stimuli would be transferred across different modalities (visual versus auditory), across different dimensions in the same modality (shapes versus colours), and across dimensions of the same modality (different sets of shapes). Participants received training on grammatical sequences from both AGL tasks. For example, when comparing generalizability across visual and auditory domains, the training phase involved exposing participants to both visual and auditory sequences that each had a separate pre-established grammar. After training, participants were exposed to sequences in only one modality; half of the sequences followed the correct grammar for the modality while the other half followed the grammar from the other modality. For example, if testing was done in the visual domain, participants only saw visual stimuli but half of the sequences followed the same grammar as the visual sequences from the training phase and the other half

followed the grammar from the auditory sequences. The authors theorized that if statistical learning generalized, participants would classify all sequences as grammatical because all sequences followed the training grammar. However, if statistical learning was not generalizable across domains, participants should only judge sequences that corresponded to the same modality as grammatical (Conway & Christiansen, 2006).

Conway and Christiansen (2006) found that participants only judged sequences to be grammatical when the sequences corresponded to the modality that the testing phase occurred in. For example, if the testing phase was in the visual modality, sequences that followed the visual grammar from the training phase were identified as grammatical while sequences that followed the auditory grammar from the training phase were identified as ungrammatical. This suggests that participants did not generalize the transitional probabilities of the auditory grammar to the visual modality. Instead, participants discriminated between different modalities and did not apply previous rules to the new sequences. Similarly, across different dimensions of the same modality, participants were able to simultaneously learn two sets of grammar and did not apply previous knowledge of statistical frequencies to the new sequences. This suggests that rule transfer did not occur across different dimensions of the same modality (regardless of the modality). However, when participants performed the tasks across the same dimension of the same modality, participants were not able to learn both grammars and instead could only learn one. Thus, statistical learning is impaired when dealing with perceptually similar input and is stimulus-specific rather than generalizable (Conway & Christiansen, 2006).

Statistical Learning and its Relation to Language

Numerous studies have found that statistical learning abilities are related to language abilities in some manner. For example, individuals who are better at picking up on the non-adjacent dependencies (e.g., X-a-Y, in which X and Y are non-adjacent to one another and are more likely to appear non-adjacent to one another than Z-a-Y) of a statistical learning task are also better at reading object relative clauses (Misyak et al., 2010). Object relative clauses contain an embedded clause that pertains to the noun at the beginning of the sentence, and are generally more difficult to read. An example of an object relative clause is, *The cat that ate the fish was missing* (Misyak et al., 2010). There is much debate on whether the mechanisms involved in language learning are specific to only language, or if these mechanisms perhaps rely on general abilities such as statistical learning (Frost et al., 2013).

Frost et al. (2013) examined whether statistical learning could predict how well a native English speaker would learn to read Hebrew as a second language. Hebrew was chosen as the second-language because the statistical properties of each language differ from one another and are both different branches of language (English is an Indo-European language and Hebrew is Semitic). It was predicted that the ability to learn how to read Hebrew might be supported by statistical learning. Thus a visual AGL task was used because the ability to learn statistical frequencies of random shapes may be predictive of both success and speed for reading a new language (Frost et al., 2013). Frost et al. (2013) found that people who performed well on the visual statistical learning task had higher proficiency for reading Hebrew. The authors theorized that reading relies on the extraction of the statistical regularities of the stimuli, similar to statistical learning.

Other studies have found similar results: Misyak et al. (2010) studied how performance between a statistical learning task and reading differed between individuals. The authors used a combination of an auditory AGL and serial-reaction time task that tested how well people could extract non-adjacent dependencies from three non-word strings. For example, the first word of the triplet would always be matched to the same final word (e.g., *pel* and *dek* always go together) but the middle word could be any other word (e.g., *pel*, *wiffle*, and *dek* versus *pel*, *nilbo*, *dek*). After training, participants judged the grammaticality of the sequences. Afterwards, participants completed a self-paced reading task that contained both subject-relative and object-relative clauses; participants were scored based on their comprehension of the sentence. Comprehension for the object-relative clauses was lower than for the subject-relative clauses. Misyak et al. (2010) divided individuals into learners or non-learners based on their reaction times for the statistical learning task, and the authors found that learners were significantly faster at reading the critical verb region of object-relative clauses. This result is interesting because it suggests that there is a relationship between how well people can process non-adjacent dependencies in language and in a statistical learning paradigm (Misyak et al., 2010).

Given the evidence described above, statistical learning appears to be important for reading abilities (Frost et al., 2013; Misyak et al., 2010). Researchers are also interested in the relationship between spoken language and statistical learning. Conway et al. (2010) were interested in whether successful acquisition of the rules from an AGL task would correlate with speech perception in degraded listening conditions. Previous studies have shown that knowledge about the statistical regularities of language can help a listener

better predict what word should come next (Miller, Heise, & Lichten, 1951; Onnis, Farmer, Baroni, Christiansen, & Spivey, 2008). This is especially true when listening conditions are degraded, such as a noisy room (Elliott, 1995; Kalikow, Stevens, & Elliott, 1977; McClelland, Mirman, & Holt, 2006). Conway et al. (2010) hypothesized that statistical learning ability would correlate with better speech perception in a degraded listening condition. Speech perception ability was measured as the difference in how well participants perceived the final word of sentences that were either highly predictable or unpredictable. Statistical learning was assessed via various AGL tasks: a visual AGL task that used colours as stimuli and an auditory AGL that used spoken non-words. The authors found that there was a high correlation between speech perception in noise abilities and statistical learning and this correlation was significant for the visual AGL task but not the auditory AGL task. Based on these results, it appears that statistical learning plays an important role for language processing (Conway et al., 2010).

Another study by Conway et al. (2011) investigated statistical learning abilities in deaf children who received cochlear implants. In this study, deaf children who received cochlear implants were tested on their statistical learning abilities in comparison to normally hearing children. Statistical learning was assessed via a visual AGL task. To assess how well participants assimilated the transitional probabilities of the training sequences, learning scores were calculated based on the methods previously explained. The authors found that half of the normally hearing children had learning scores above zero, but only a third of the deaf children with cochlear implants had learning scores above zero (Conway et al., 2011). The authors then controlled for age and computed partial correlations on the duration of implant use and age of implantation for the deaf

children and found that the later a child received their implant, the lower their learning

score. They also found that children who had longer durations of cochlear implant use

had higher learning scores. The authors additionally reported a link between language

abilities and statistical learning for deaf children with cochlear implants. The results of

the study suggest that statistical learning abilities are important for language outcomes in

deaf children with cochlear implants. The authors speculated that this relationship

between statistical learning abilities and language may explain why some individuals fare

better with a cochlear implant than others: those who are better at statistical learning may

be better at picking up on the statistical regularities of spoken language (Conway et al.,

2011).

Many studies have found correlations between language and statistical learning

abilities (Conway et al., 2010, 2011; Frost et al., 2013; Misyak et al., 2010). It seems

highly likely that language learning may in part rely on a general ability such as statistical

learning (Conway et al., 2011). Logically, it would be redundant for two separate abilities

to exist in the human brain that are used for the extraction of the statistical frequencies of

events, though may be possible. Tallal, Stark, Kallman, and Mellits (1981) investigated

temporal processing disorders in language-impaired children who had both visual and

auditory deficits. The authors found that the visual deficits resolved, but not the auditory

deficits, due to age. This study suggests that the auditory and visual deficits associated

with temporal processing rely on different processing systems (Tallal et al., 1981).

Christiansen, Conway, and Onnis (2012) investigated the relationship between statistical

learning and language processing using an EEG paradigm. The authors investigated

whether language and statistical learning processing relied on the same processes to

process syntactic violations. After training on an AGL task, participants were asked to judge whether the testing sequences were grammatical or not. Afterwards, participants read sentences and decided whether the sentence contained a grammatical violation or not. The authors found evidence that suggested that statistical learning utilized a similar system as language to process the presence of a syntactic violation. The authors found an ERP that was common to both tasks — the P600 — and it was maximal over the central parietal regions of the scalp (Christiansen et al., 2012).

The P600 is a positive deflection that begins 600 ms post-stimulus onset (Brouwer, Fitz, & Hoeks, 2012). The P600 was discovered by Osterhout and Holcomb (1992) when participants read sentences that contained different grammatical errors (phrase structure violations and sub-categorization constraint deviations). For example, the word *to* in the following sentence, *The broker persuaded to sell the stock was tall*, elicited a P600 effect. The authors found that these types of syntactic violations resulted in an P600 effect over the right anterior scalp that was positive deflecting and slow (Osterhout & Holcomb, 1992). The P600 is elicited by syntactic anomalies including grammar violations and syntactically correct sentences that result in inaccurate parsing (garden path sentences) and commonly found in the central parietal region of the scalp (Kaan et al., 2000; Osterhout & Holcomb, 1992). The P600 is generally interpreted to reflect revisions in syntactic processing (Coulson, King, & Kutas, 1998). The P600 is elicited by both visual and auditory words, similar to the N400 (Hagoort & Brown, 2000). Interestingly, studies have shown that the P600 may be frontally distributed rather than posteriorly, when exposed to sentences that are syntactically correct but are not preferred (Friederici, Hahne, & Saddy, 2002; Kaan & Swaab, 2003a, 2003b; Osterhout & Holcomb, 1992).

These differences in scalp distribution may reflect costs that are associated with reading non-preferred sentences when the P600 is frontally distributed versus parsing failure when a more posterior distribution is seen (Friederici et al., 2002; Hagoort & Brown, 2000; Kaan & Swaab, 2003a).

1.3 The Current Study

The objective of the current study was to determine whether there was a relationship between statistical learning and speech perception in noise abilities for normally hearing listeners. More specifically, whether implicit statistical learning in a visual AGL task — indexed via the P600 effect — would correlate with the ability to predict a final word on a SPIN task, as indexed via the N400 effect.

The visual AGL task involved exposing participants to sequences of letters. Unbeknownst to participants, the letters followed pre-established transitional probabilities so that some letters were more likely to sequentially occur, while other letters would never sequentially occur. For example, the letters B and C had a 50% chance to occur after the letter A, but the letter D would never follow the letter A. Participants received extensive training in which they were asked to reproduce the sequences immediately after they saw the sequence. After training, participants were shown sequences of letters that either followed the grammar from the training phase or saw sequences that violated the grammar at either the third, fourth, or fifth letter of the sequence. It was expected that participants would extract the statistical frequencies of the letters during the training phase. Learning was assessed via the learning scores and through the P600 effect, which has been previously shown to be elicited by syntactic violations on an AGL task (Christiansen et al., 2012).

The SPIN task was used to measure people's ability to apply prior knowledge about the context of a sentence in order to predict incoming information. People listened to pre-recorded sentences and were asked to repeat the final word. However, half of the sentences had a final word that had a high Cloze probability and the other half had 0% Cloze probability. The SPIN task was presented under two different conditions: in quiet and with background noise (-1 dB SNR). The background noise was multi-talker babble and the SNR was chosen in order to simulate real-life environments in which speech perception accuracy decreases and people must increase their reliance on top-down processing in order to accurately perceive speech.

1.4 Hypotheses

Based on the experiment by Christiansen et al. (2012), I hypothesized that the syntactic violations of the ungrammatical sequences would elicit a central parietal P600 effect in people who were sensitive to the underlying statistical regularities of the stimuli. Additionally, learning scores were expected to vary between participants. People with higher learning scores were expected to have an enhanced P600 effect in contrast to those with lower learning scores.

The SPIN task was predicted to elicit an N400 effect in both quiet and with background noise conditions, but I expected that behavioural performance would be negatively affected by background noise. Additionally, it was expected that the low SNR chosen for the background noise condition would diminish the N400 mismatch effect as previous studies have demonstrated that background noise reduces the N400 mismatch effect (Aydelott et al., 2006; Strauß et al., 2013). This was because the purpose of the

noise condition was to mimic a natural environment, so speech perception should be slightly impaired, and mismatch words may erroneously be heard as a match.

Finally, it was hypothesized that there would be a significant positive correlation between the amplitude of the N400 and P600 effects in the SPIN and AGL tasks, respectively. In other words, as the amplitude of the N400 effect became more negative in the SPIN task, the amplitude of the P600 effect was predicted to become more positive for the AGL task. Thus, people who were more aware of the underlying statistical frequencies of language would have a greater N400 effect, and this would correlate with higher awareness for the transitional probabilities of the AGL sequences, resulting in an enhanced P600. It was expected that when background noise was added to the SPIN task, there would be a reduced correlation in people who were more sensitive to the transitional probabilities of the AGL task, because these people were predicted to be better at using the context of the sentence to supply a correct match word, thus showing a reduced N400. The correlation between N400 and P600 amplitude was expected to be maintained in people who were worse at the AGL task, regardless of the SPIN condition, because they would be worse at generating a correct match word.

CHAPTER 2 METHODS

2.1 Participants

Fourteen adults (5 males, mean age = 25, SD = 8, range = 18 – 50; mean years of

education = 18, SD = 3, range = 13 – 23) with normal hearing (defined as hearing

threshold of 30 dB or lower in the worse ear at 1000 Hz and confirmed through pure-tone

audiometry at 1000 Hz during the study) were recruited for this study. All subjects were

native English speakers, right-handed, and were not taking medication known to affect

EEG or attention. Participants provided informed consent to participate in this study and

were compensated $50. The study was approved by the Dalhousie University ethics

board.

Four of the fourteen participants had previously participated (range = 1 – 2 years) in a

behavioural study that used an AGL task with the same stimuli as the one used here.

However, the AGL task for the current study used unique sequences that were not

previously used in the behavioural AGL task. Additionally, the ungrammatical sequences

for the current AGL task were otherwise grammatical but contained a single

ungrammaticality in the sequence, after which the sequence became grammatical again.

In contrast, the AGL task of the previous study did not follow this procedure.

2.2 Experimental Design

The study took place at the NeuroCognitive Imaging Lab at Dalhousie University

over two study sessions: the first session involved acquiring informed consent, and

completing a variety of questionnaires and cognitive measures which are outside of the

scope of the present study and are thus not mentioned further. The final session of the

study involved the collection of neuroimaging data. Each session took place inside of the same quiet, isolated room.

2.3 EEG Tasks

The EEG tasks and data collection took place inside of the same quiet, isolated room that was used for collecting the cognitive measures. Both EEG tasks were programmed using DirectRT version 2014.1.114 (Jarvis, 2012). Both tasks were delivered on a computer monitor that was specialized for enhanced graphics (BenQ, Taiwan) with a refresh rate of 120 Hz.

2.3.1   Artificial Grammar Learning Task

The AGL task was used to assess people's ability to implicitly learn arbitrary sequences. Previous studies have determined that there is a relationship between speech comprehension and performance on the AGL task in both children with cochlear implants (Conway et al., 2007) and in adults (Conway et al., 2011).

Stimuli

Four letters were chosen as the stimuli for the AGL task: Q, X, N, and D. These four letters were chosen based on the fact that their names are not phonetically similar, and they are visually distinct from one another. One-hundred and thirty-three unique sequences of 5 – 7 letters in length were generated for the AGL task and were based on a Markovian finite-state grammar (Conway et al., 2010).

From the perspective of the experimental design, the AGL task was divided into seven blocks: a practice block (seven sequences), two learning bocks (21 sequences per block), and four testing blocks (21 sequences per block). However, during the study participants were only alerted to the practice versus "real" blocks; they were not told that

28

there were sequences to learn nor that there would be a testing phase. The practice block was used to familiarize participants with the AGL task and used different letters as stimuli for the task (L, K, W, and B). The sequences in the practice block did not use the same Markovian finite-state grammar as the rest of the experiment. Participants started with three letter sequences and progressed to seven letter sequences. The learning blocks each consisted of 21 'grammatical' sequences: sequences that followed the pre-established Markovian final-state grammar. Based on the recommendations from , who recommend that participants are trained on shorter grammatical sequences first, the first learning block started with five letter sequences and sequentially increased to seven letter sequences in the final learning block.

The testing blocks were used to determine how well participants picked up on statistical frequencies of the grammatical sequences. Of the 84 sequences in the learning blocks, half followed the Markovian finite-state grammar used in the learning blocks. The remaining 42 sequences were 'ungrammatical': the sequences followed the pre-established Markovian finite-state grammar, but contained a violation of the grammar at either the third, fourth, or fifth letter of the sequence. An example of a grammatical sequence is, "DX<u>N</u>QD", in which the third stimulus (N) is the grammaticality of interest. For the ungrammatical sequence, "XN<u>X</u>QD", the sequence followed the pre-established grammar until there is a syntactic violation at the third stimulus (X), the sequence again follows the pre-established grammar after the ungrammaticality. The grammatical and ungrammatical sequences were pseudo-randomly assigned so that there was an equal distribution of grammatical and ungrammatical sequences in each block. Constraints were also placed on sequence length and the position of a violation in an ungrammatical

29

sequence for each block. For example, each testing block contained sequences that ranged from 5 – 7 letters.

Procedure

At the start of the AGL task, participants received both verbal and written instructions on how to complete the task. Written instructions were provided after every rest break. Participants were informed that they would see a series of letters that they then had to repeat back using a keypad. The keypad (a USB numeric keypad) was labeled with the four letters used in the AGL task. This was done so that the response keys would be located in a single row, rather than their typical positions on a QWERTY keyboard. Participants completed the practice block, then after ensuring that they understood the task, participants were provided with a rest break before beginning the experiment.

The sequence of events in a trial is shown schematically in Figure 1. At the start of each trial, a fixation cross (1.9 degrees viewing angle) was presented in the centre of the computer screen for a random interval of 500 – 1500 ms. After fixation, an empty box (9.5 degrees viewing angle) appeared in the centre of the screen and remained on for the duration of the trial. This box cued the participant that the sequence was starting, and participants were asked to try to not blink while the box was shown. The first letter in the sequence appeared 500 ms after the box appeared. During sequence presentation, each letter (1.9 degree viewing angle) was displayed in the centre of the box for 500 ms, followed by an empty box for 500 ms, after which the next letter in the sequence was shown. After the trial finished, there was a 1500 ms delay period in which participants saw an ellipsis in the centre of the screen. After this delay period, a response prompt appeared on the screen asking participants to type back the sequence that they had just

observed via the keypad. Participants received six rest breaks throughout the duration of the experiment. After completing the AGL task, participants answered a questionnaire on how they remembered the sequences and whether they noticed any patterns in the sequences.
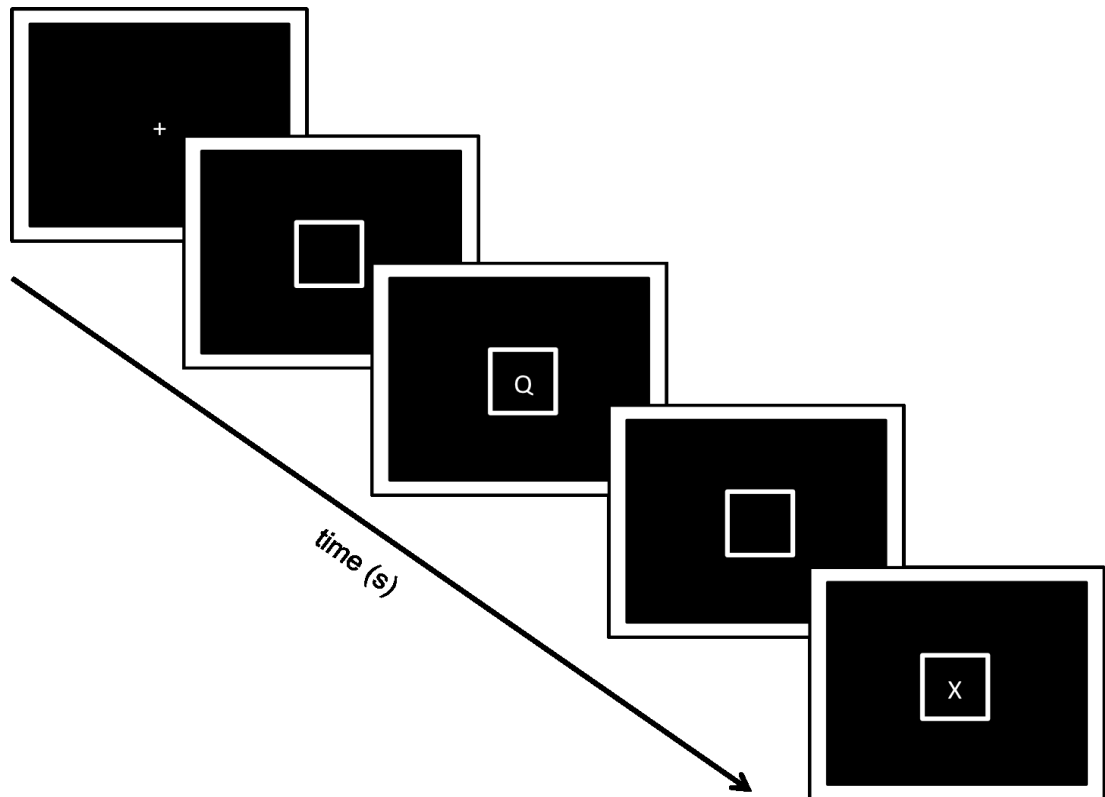
Figure 1. Schematic of the procedure for the AGL task.

For two participants, technical problems occurred during the tasks in which the visual stimuli of a sequence stopped appearing, but the response prompt still appeared. This problem occurred on random trials: once within the learning block and another time in the final experimental block. The AGL task was restarted at the last sequence prior to the problem's occurrence. These technical difficulties were accounted for in both participants

by removing both the behavioural and EEG data for the trials where sequences did not display.

Behavioural Analysis

An AGL learning score was calculated using the following equation:

$$\text{AGL Learning Score} = \sum(g_c * L) - \sum(ug_c * L)$$

Where g_c is the number of correctly answered grammatical sequences, ug_c is the number of correctly answered ungrammatical sequences, and L is length of the sentence. Positive learning scores thus indicate superior performance on grammatical than ungrammatical sequences.

Generalized linear mixed-effects modelling with a binomial family was used to analyze the AGL accuracy for all trials as a function of fixed effects including grammaticality and learning status, as well as random effects of subject and sequences.

2.3.2   Speech Perception in Noise Task

Stimuli

A SPIN task was used to investigate sentence-processing abilities. Short, declarative sentences that were normed (Block & Baldwin, 2010) based on the predictability of the final word (Cloze probability) were chosen for this task. The Cloze probability of the 176 chosen sentences ranged from 83% – 99%. None of the chosen sentences ended with the same final word as another chosen sentence in order to avoid practice effects.

The Corpus of Contemporary American English database was used to generate alternate final words (mismatches) for each of the 176 sentences. Mismatches were generated based on the following criteria: the natural log of the frequency of occurrences

of the mismatch word in English was matched to the original final word (matches), mismatches contained the same number of letters as matches; mismatches were of the same class as the matches; mismatches did not begin with the same letter as matches (to avoid priming effects); mismatch words did not rhyme with match words; and, while the mismatches were semantically incongruent, they were grammatically correct. For example, *In the heat of his performance, Sean broke a guitar string / clinic*, in which *string* is the expected match (Cloze probability of 87%) and *clinic* is the mismatch. A two-tailed *t*-test was performed to ensure that the natural log of the frequencies for match and mismatch words were not significantly different from one another, $t(350) = -0.27$, $p = 0.79$.

The sentences used in this task were recorded by a male native Canadian English speaker in a neutral tone. Each of the 176 sentences with the expected match word were recorded as separate audio files; each mismatch word was recorded as a separate audio file. Sentences were normalized so that there would be no volume (dB) differences between sentences. Sentences were presented at 65 dB within an isolated room. Each audio file was modified to include 500 ms of silence at the beginning and at the end of each sentence. Each audio file was also clipped so that the final word in the sentence was a separate audio file from the rest of the sentence.

For the second condition (with background noise), sentences were masked with a multi-talker babble audio file that began 500 ms prior to stimulus presentation and continued after stimulus presentation ended. The original multi-talker babble audio file from the Words-in Noise test (NIH Toolbox, Illinois) was normalized so that the multi-talker babble audio file was not louder than the pre-recorded sentences. The multi-talker

babble audio file was randomly sampled to generate 500 unique audio files that were 6860 ms long. This was to ensure that participants were not inadvertently prompted to the beginning of the stimulus. The multi-talker babble audio files were presented at a -1 dB SNR relative to the sentence audio files.

The 176 sentences were randomly divided into two lists of 88 sentences per list. Each stimulus list was divided between the two SPIN conditions so that only one list was completed per condition. List order was counterbalanced between participants. Within each list, 44 sentences were randomly selected to end with a mismatch word rather than the expected match word. Each list was constrained so that there would be no more than four consecutive sentences that had a match or mismatch final word. Additionally, the first eight sentences within each list were used as a practice block. The remaining 80 sentences were then divided into four testing blocks of twenty sentences per testing block.

Procedure

Each participant began the SPIN task in the no background noise condition then began the with background noise condition. Only one stimulus list was completed per condition. List order was counterbalanced between participants. Prior to beginning the experiment, participants received written and verbal instructions on how to complete the task. Participants also received written instructions at the start of every block.

At the beginning of each trial, an outline of a white box (9.5 degrees visual angle) appeared in the centre of the screen to direct the participant's attention to the screen and to inform participants when a new trial began. The box was displayed for a random interval of 250 – 1000 ms prior to auditory stimulus presentation and remained on screen

until 1000 ms post auditory stimulus presentation. After the box disappeared, a response prompt appeared on the screen that asked participants to repeat the final word of the sentence that they had just heard into the microphone. If participants did not know what the final word was, they were instructed to guess or to say "pass", to begin the next trial. The reaction time of the spoken utterance was coded automatically through DirectRT using an algorithm that determined voice onset time according to a loudness threshold (Jarvis, 2012). The next trial began after a two second delay. Participants did not receive any feedback regarding the identity of the final word nor about their accuracy. Participants received rest breaks after completing 25%, 50%, and 75% of the experiment.

Behavioural Analysis

Any trials that did not contain a response or were labeled as missing were removed from the data set. Linear mixed-effects modelling (Baayen, Davidson, & Bates, 2008) was used to analyze reaction times for all trials, as a function of fixed effects including condition, target word category (match or mismatch), as well as random effects of subject and the target word. Outliers were defined as data points that fell more than 2.5 standard deviations outside of the mean and were removed from analysis. Afterwards, the model was refitted to the trimmed data.

Generalized linear mixed-effects modelling using a binomial family was used to analyze accuracy for all trials as a function of fixed effects including condition, target word category, as well as random effects of subject and target word category. Additionally, a linear mixed-effects model was used to correlate the SPIN accuracy scores with the AGL learning scores, with a random by-subject effect.

2.4 EEG Data Acquisition

EEG data were collected using a 64 channel QuickAmp (Advanced Neuro Technology, Enschede, Netherlands) amplifier that was connected to an Acticap (Brain Vision, Morrisville) electrode system. This system had active preamplifiers located on each electrode in order to improve signal quality. The electrodes were attached to an elastic cap.

Participants were prepared for EEG recording by cleaning areas of exposed skin (including behind the ears, forehead, and cheeks) where the electrodes sat with a hypoallergenic NuPrep cream. After participants were prepared, the electrode cap was placed on the head and the electrodes were positioned. In order to ensure a good electrode connection, each electrode was filled with hypoallergenic electrolyte gel (SuperVisc) through a blunt-tipped syringe. The gel was gently rubbed into the scalp. Impedance was lowered below a threshold of 30 kΩ at each electrode. To monitor eye movement and blinks, bipolar, self-adhesive electrodes were also placed above and below the left eye and on the outer canthi lateral to each eye. This allowed me to correct for artefacts in the EEG recordings. Data were on-line filtered at 138 Hz, were averaged-reference, digitized at a sampling rate of 512 Hz via ASALAB software (Advanced Neuro Technology, Enschede, Netherlands), and stored on a computer for later analysis.

2.5 EEG Data Analysis

Data were processed offline through EEGLAB 13.4.3b (Delorme & Makeig, 2004).The EEG data were band-pass filtered in the $0.1 - 30$ Hz range for the AGL task and in the $0.3 - 30$ Hz range for the SPIN task. After filtering, trials that contained excessive noise (e.g., channel drift, head movement) were identified and were removed

from each individual participant's dataset. Excessively noisy channels were also eliminated from each individual participant's dataset. Ocular artefacts as well as other artefacts such as channel drift were identified and corrected using independent component analysis (ICA) with the FastICA 2.5 MATLAB package (Hyvärinen & Oja, 2000). The ICA components were visually inspected to determine the components that contained ocular artefacts as well as other type of noise. These components were removed from the dataset.

After ICA, noisy channels that had been removed prior to ICA were interpolated using data from the surrounding channels with spherical spline interpolation. Continuous EEG data were segmented into discrete epochs (200 ms pre-stimulus onset and 1000 ms post-stimulus onset) that were time-locked onto the events of interest. For the AGL task, events that were of interest were stimuli that were either grammatical or ungrammatical at the third, fourth, or fifth position of the sequence. For the SPIN task, events of interest were the target word category (match or mismatch). Epochs were baseline corrected by subtracting the mean of the 200 ms pre-stimulus period. After epoching, EEG data were re-referenced to an average of the left and right mastoids.

To investigate the effects of the experimental conditions on ERP amplitude, the ERP amplitude, timing, and scalp distribution, were exported from EEGLAB to R version 3.2.1 (R Core Team, 2013). The mean amplitude at each electrode for each trial was computed and exported for the time windows of interest: 300 – 500, and 600 – 800. These time windows were selected based on the time range that the N400 and P600 ERPs were expected to occur (300 – 500 ms for the N400, 600 – 800 ms for the P600) for the SPIN and AGL task, respectively. Electrode positions on the scalp were categorized into

seven regions of interest (ROI) based on areas where the ERPs of interest appear. Electrodes in the midline central (electrodes C1, Cz, C2, CP1, CPz, CP2, P1, Pz, and P2) were chosen as an ROI for the N400 and for the P600. As an exploratory measure, other ROIs were examined and included the left anterior (electrodes AF7, AF3, F7, F5, and F3), middle anterior (electrodes Fp1, Fp2, F1, Fz, F2, FC1, and FC2), right anterior (electrodes AF4, AF8, F4, F6, and F8), left central (electrodes FT9, FT7, FC5, FC3, T7, C5, C3, TP7, CP5, and CP3), right central (electrodes FC4, FC6, FT8, FT10, C4, C6, T8, CP4, CP6, and TP8), and the posterior (electrodes P7, P5, P3, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, PO9, and PO10).

Outliers were defined as data points that fell more than 2.5 standard deviations outside of the mean and were removed from analysis. A generalized additive mixed-effects model (GAM) was used to analyze the amplitude of the EEG data as a function of the variables of interest, as well as random effects such as trial and subject. GAM is an extension of the general linear model and is well suited for EEG analysis because it is able to minimize the amount of time required for data processing and thus is optimal for processing large data sets (Wood, Shaw, & Goude, 2015). Additionally, GAM is able to include non-linear factors and non-linear interactions in the model (Meulman, Wieling, Sprenger, Stowe, & Schmid, 2014). Since GAM is able to detect non-linear effects, GAM is an optimal method for analyzing EEG data because there may be non-linear effects (Meulman et al., 2014). For example, the mean amplitude of all electrodes at each trial may vary in a non-linear way.

It was important that an optimal GAM was identified for data analysis. Iterative tests that compared more complex models to simpler models were performed for each

dependent measure in order to determine the optimal model (Tremblay & Ransijn, 2013; Tremblay & Tucker, 2011). By comparing simpler models to more complex models, I was able to remove variables and interactions that did not explain significant amounts of data variation (Baayen, Davidson, & Bates, 2008). The optimal model was defined as the model that accounted for the most variance with the fewest interactions and factors, and was evaluated using the restricted maximum likelihood (REML) (Dodge, 2006). REML removes effects of nuisance parameters by using the likelihood function calculated from the dataset (Dodge, 2006).

Artificial Grammar Learning

GAM was used to analyze the amplitude of the EEG data at the time window of interest (600 – 800 ms), as a function of grammaticality, learning (non-learners or learners), ROIs, as well as a non-linear effect of trial, random intercepts for each subject, and a random non-linear effect of trial for each combination of subject and ROI. Outliers were identified and removed from analysis. Afterwards, the model was refitted to the trimmed data. The model explained 5.69% of the deviance in the data.

Speech Perception in Noise

GAM was used to analyze the amplitude of the EEG data at the time window of interest (300 – 500 ms), as a function of condition (no background noise or noise), target word category, ROIs, as well as a non-linear effect of trial, random intercepts for each subject, and a random non-linear effect of trial for each combination of subject and ROI. Outliers were identified and removed from analysis. Afterwards, the model was refitted to the trimmed data. The model explained 5.79% of the deviance.

Speech Perception in Noise Amplitude Correlated with Artificial Grammar Learning

  Amplitude

To determine whether there was a relationship between the amplitudes for the SPIN

and AGL tasks, the ERP data from each task were correlated using GAM. To simplify

this analysis and isolate the effect of interest in each experiment, the differences in

amplitude between mismatch and match conditions (mismatch – match sentences for the

SPIN task; ungrammatical – grammatical sequences for the AGL task) for each subject,

ROI, and SPIN condition (without or with background noise) were calculated. The mean

amplitude at each electrode was calculated independently, but those within an ROI were

not differentiated from one another during modelling. This reduced the dataset to $n = 98$

per SPIN condition. When there was no background noise, the model explained 75.1% of

the deviance. With background noise, the model explained 69.6% of the deviance.

CHAPTER 3 RESULTS

3.1 Behavioural Results

3.1.1   Artificial Grammar Learning

Learning Scores

The mean (*M*) AGL learning score was 6.64 (median = 4). The descriptive statistics

for the learning scores can be seen in Table 1.

Table 1. Descriptive statistics for the grammatical span, ungrammatical span, and
learning score for the AGL task.

Measure	Mean	Standard deviation (SD)	Minimum score	Maximum score
Grammatical span	55.43	22.55	19	89
Ungrammatical span	48.79	12.94	25	72
Learning score	6.64	18.25	-26	36

While the range of the learning scores was similar to other studies that used a visual

AGL task (Conway et al., 2010, 2011), it was a concern that ERP effects in response to

grammaticality might not be seen at the group level if there were participants who did not

successfully acquire the statistical probabilities of the sequences. Participants were

divided into two groups, learners and non-learners, based on the median of the AGL

learning scores, as shown in Figure 2.
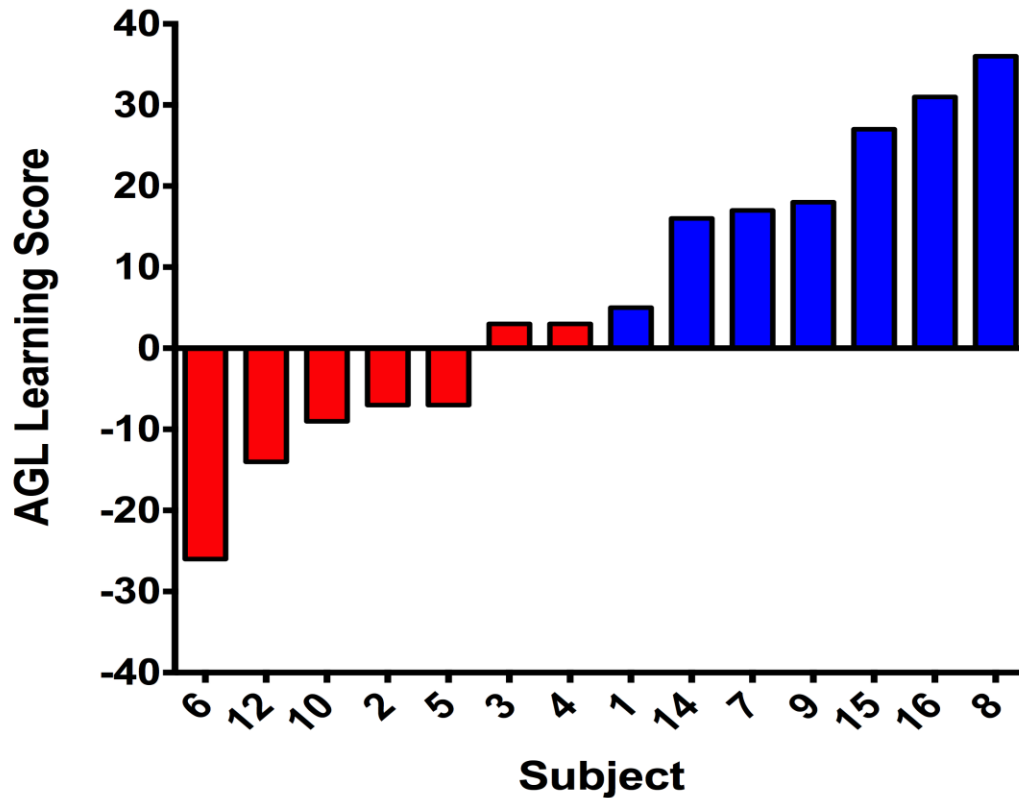
Figure 2. Distribution of the AGL learning scores, by subject. Red indicates participants defined as non-learners and blue indicates participants defined as learners, based on a median split of learning scores.

There were seven learners and seven non-learners on the AGL task. The average learning score for learners was 21.43 ± 95% confidence interval (CI) of 3.98 (range = 5 – 36). The average learning score for non-learners was -8.14 ± 5.04 (range = -26 – 3). A two-tailed *t*-test revealed that learners had a significantly higher learning score than non-learners on the AGL task, $t(12) = 3.45$, $p = 0.005$.

Accuracy

The average accuracy on the AGL task demonstrated that learners were more accurate for grammatical ($M = 66\%$, $SD = 48\%$) than ungrammatical sequences ($M = 55\%$, $SD = 50\%$). Non-learners were less accurate for grammatical ($M = 43\%$, $SD = 50\%$) than

ungrammatical sequences ($M = 47\%$, $SD = 50\%$). A generalized linear mixed-effects

model was fitted in which the logits of accuracy were regressed onto a two-way

interaction including grammaticality and learning status, with random effects of subject

and sequences. There were no significant main effects of grammaticality, $F(1, 1074) =$

0.2, $p = 0.67$, or learning status, $F(1, 1074) = 0.4$, $p = 0.53$. However, there was a

significant interaction between grammaticality and learning status, $F(1, 1074) = 10.2$, $p =$

0.002. Post-hoc analysis was performed to determine the nature of the interaction;

probability values were two-tailed and were Bonferroni corrected for three comparisons.

Figure 3 demonstrates that learners were not significantly more accurate for

grammatical ($M = 72\% \pm 27\%$) than ungrammatical sequences ($M = 57\% \pm 31\%$;

difference $= 14\%$, $z = 1.53$, corrected $p = 0.38$). Non-learners were not significantly more

accurate for grammatical ($M = 46\% \pm 31\%$) than ungrammatical sequences ($M = 53\% \pm$

31%; difference $= -7\%$, $z = -0.79$, uncorrected $p = 0.43$). Finally, while the accuracy

differences between ungrammatical and grammatical sequences did significantly differ

between learners ($M = -14\%$) and non-learners ($M = 7\%$), this difference did not survive

multiple comparison correction (difference $= 28\%$, $z = -3.1$, corrected $p = 0.054$).
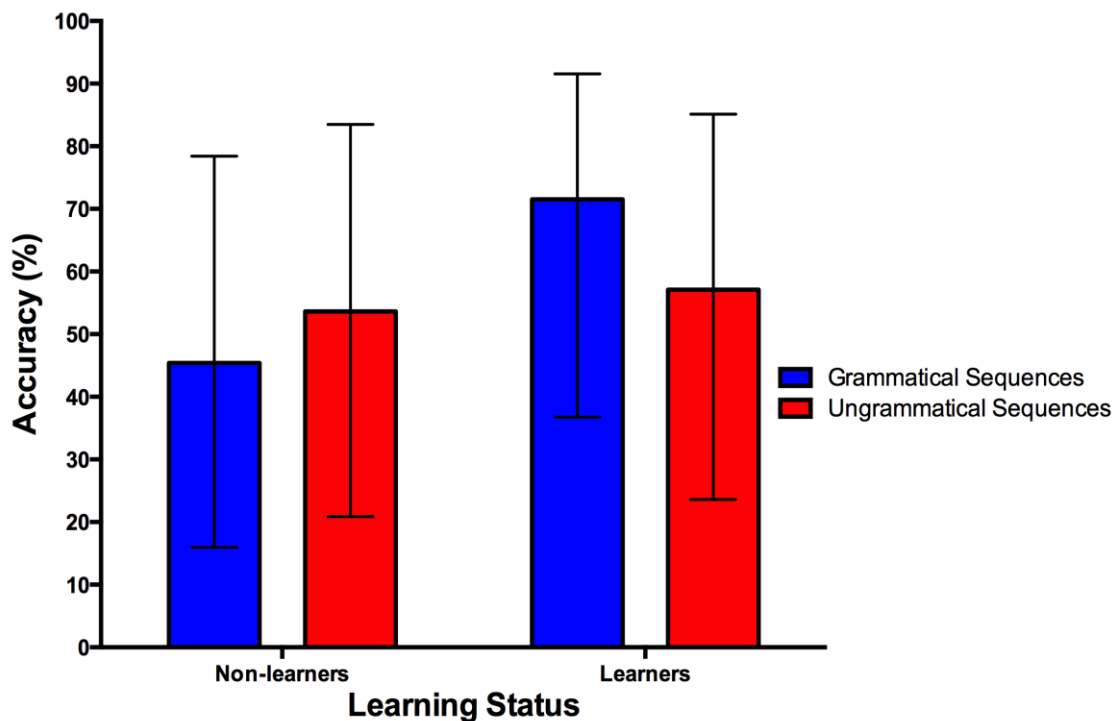
Figure 3. Model predicted accuracy (%) for grammatical (blue) and ungrammatical (red) sequences for non-learners and learners.

3.1.2 Speech Perception in Noise

Reaction Times

When there was no background noise, the average reaction time for participants to repeat the final word of the sentence was 794.73 ms ($SD = 1.51$ ms) and 981.42 ms ($SD = 1.46$ ms) when there was background noise. On average, people took approximately 866.10 ms, ($SD = 1.47$ ms) to repeat the final word when it was a match and 908.69 ms ($SD = 1.54$ ms) when the final word was a mismatch. A linear mixed-effects model was fitted where the log of the reaction times were regressed on a two-way interaction including the condition and target word category. The model also included by-subject and by-item random intercepts. Outliers were removed from analysis and represented 4.13% of the data (90 trials of 2179 trials were removed). There were main effects of condition, $F(1, 1912) = 211.4$, $p < 0.001$, and target word category, $F(1, 1912) = 7.2$, $p = 0.007$, on

reaction times. There was no significant interaction between condition and target word category. Post-hoc testing was performed to determine the significant differences condition and target identity; probability values were two-tailed and were Bonferroni corrected for two comparisons.

Post-hoc testing revealed that participants were significantly faster at repeating the target word in the no-noise condition ($M = 794.61$ ms ± 116.15 ms) in comparison to the with noise condition ($M = 960.61$ ms ± 140.28 ms; difference $= 165.70$ ms, $t(1912) = 0.01$, corrected $p < 0.001$), as seen in Figure 4. Participants were also significantly faster at repeating the target word when it was a match ($M = 794.61$ ms ± 116.15 ms) rather than a mismatch ($M = 832.54 \pm 121.74$ ms; difference $= -37.93$ ms, $t(1912) = 0.02$, corrected $p = 0.02$), as shown in Figure 5.
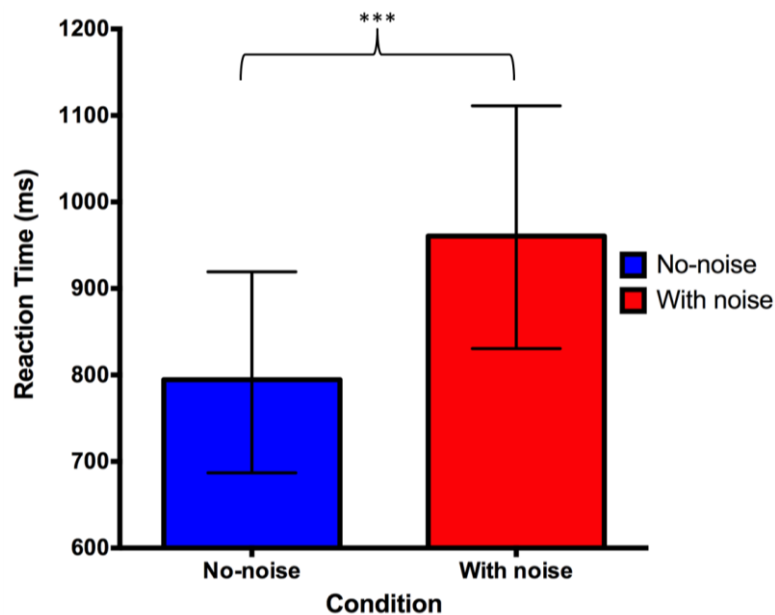
Figure 4. Model predicted reaction time (ms) of responses on the SPIN task for the no-noise (blue) and with noise (red) conditions. Error bars represent the 95% confidence interval. *** indicates a $p < 0.001$.
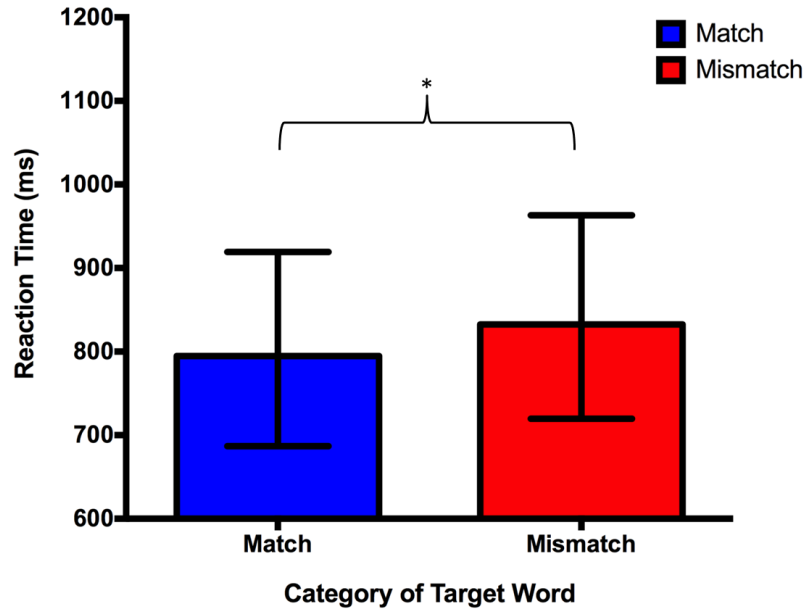
Figure 5. Model predicted reaction time (ms) for responses on the SPIN task for the match (blue) and mismatch (red) sentences. Error bars represent the 95% confidence interval. * indicates a $p < 0.05$.

Accuracy

When there was no background noise, the mean accuracy for participants correctly identifying the final word of the sentence was 99% ($SD = 8\%$) and accuracy decreased to 82% ($SD = 38\%$) when there was background noise. When the final word of the sentence was a match, participants accurately identified the final word 96% ($SD = 21\%$) of the time. When the final word was a mismatch, accuracy decreased to 85% ($SD = 36\%$). A generalized linear mixed-effects model was fitted where the logits of accuracy were regressed onto a two-way interaction including the target word category and condition. The model also included by-subject and by-item random intercepts. There was a significant interaction between target word category and condition on accuracy, $F(1, 1911) = 4.07$, $p = 0.044$. There were main effects of condition, $F(1, 1911) = 71.14$, $p < 0.001$, and target word category, $F(1, 1911) = 28.71$, $p < 0.001$, on accuracy.

To elucidate the nature of the differences, post-hoc analysis was performed. Probability values were two-tailed and Bonferroni corrected for three comparisons. Figure 6 shows that in the with noise condition, participants were more accurate at identifying the final word when it was a match ($M = 97\% \pm 2.19\%$) rather than a mismatch ($M = 81\% \pm 8.1\%$; difference = 15%, $z = -5.17$, corrected $p < 0.001$). In the no-noise condition, accuracy was not significantly different between match ($M = 99\% \pm 0.31\%$) and mismatch words ($M = 99\% \pm 0.31\%$; difference = 0.01%, $z = 0.07$, uncorrected $p = 0.94$). Accuracy did not significantly differ between both conditions, $z = -2.21$, corrected $p = 0.08$.
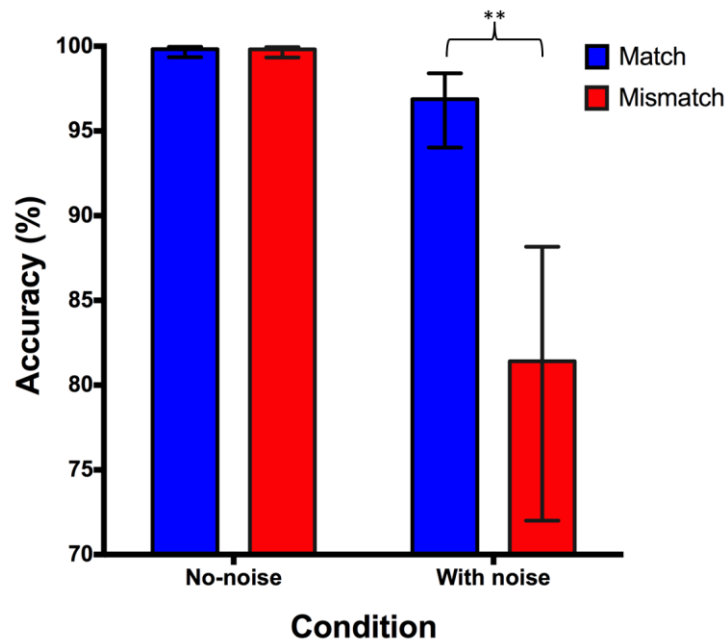
Figure 6. Model predicted accuracy (%) of the responses for each condition (no-noise and with noise) and for the target word category (match or mismatch) for the SPIN task. Error bars represent the 95% confidence interval. ** indicates a significance of $p < 0.01$.

3.1.3   Behavioural Correlations

A linear mixed-effects model was fitted in which the behavioural performances on the

SPIN task were regressed as a function of the AGL learning scores, with a by-subject

random effect. Figure 7 demonstrates that there was no correlation between accuracy on

the SPIN task and the AGL learning scores, $\beta_{weight} = 0.02$, $t(12) = 0.18$, $p = 0.86$.
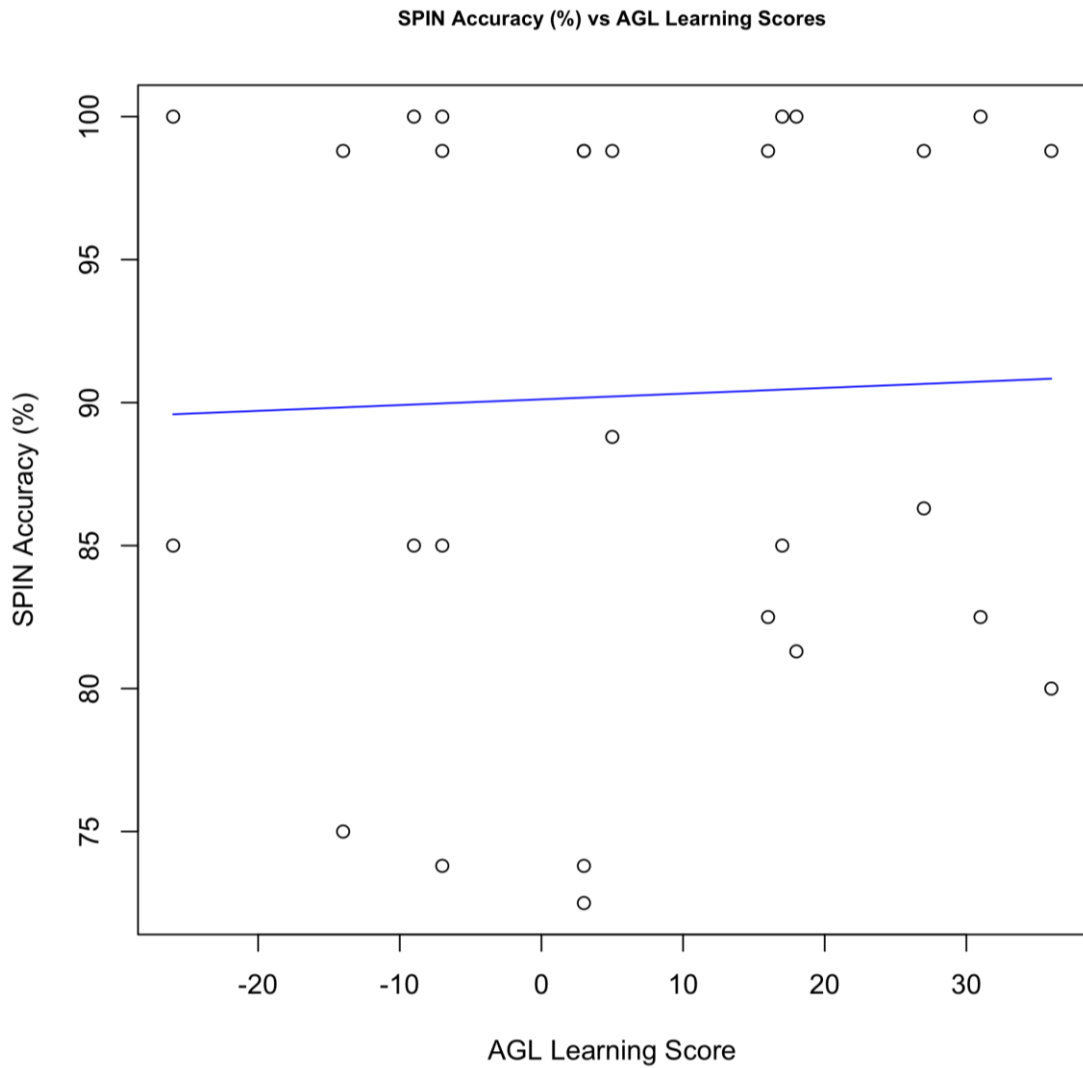
Figure 7. The correlation between SPIN Accuracy (%) and AGL learning scores. Line
represents the trend line.

3.2 EEG Results

For the purposes of ERP visualization only, EEGLAB (Delorme & Makeig, 2004) was used to smooth the grand average and difference waveforms with a 15 Hz low pass filter.

3.2.1 Artificial Grammar Learning

Event-Related Potential Data

Visual inspection of the ERP waveforms of learners and non-learners suggested that there were clear differences in how learners and non-learners processed the stimuli for the AGL task. The grand averaged waveforms for learners are shown in Figure 8. Over the central parietal and anterior electrodes, the waveforms for both grammatical and ungrammatical stimuli had an initial positive deflection at approximately 50 ms, followed by a negative deflection at 100 ms, after which there was a positive deflection at approximately 150 ms then at approximately 200 ms, a second negative deflection. These patterns are representative of the P1-N1-P2-N2 complex that typically occurs in response to visual stimuli.

Figure 8. The grand average waveforms across time for learners in response to ungrammatical (red) and grammatical (blue) stimuli on the AGL task, across all electrodes. Negative is plotted upward.

Figure 9 shows that learners had larger apparent differences between the grammatical and ungrammatical waveforms at both the central parietal and anterior electrodes. At the central parietal electrodes, the ungrammatical waveform was less positive than the grammatical waveform beginning approximately 250 ms post-stimulus onset, with maximal positivity at 600 ms. While there was no evidence of a P600 difference at the central parietal electrodes, the ungrammatical waveform at the most anterior electrode sites had an increased positivity relative to grammatical that began at 600 ms and peaked

around 700 ms post-stimulus onset, as seen in Figure 9. The central parietal negativity

and anterior positivity can also be observed in the scalp topography maps in Figure 13.

Figure 9. The grand average waveforms across time for learners in response to ungrammatical (red) and grammatical (blue), at the midline frontal and central ROIs. Negative is plotted upward. Scale is -8 to 8 µV.

Figure 10 shows that for non-learners, the grammatical and ungrammatical

waveforms were similar across all electrodes with the exception of the anterior

electrodes. At the anterior electrodes, the grammatical waveform had a larger negativity

relative to the ungrammatical waveform. There was no clear suggestion of a typical P600

difference between the waveforms.
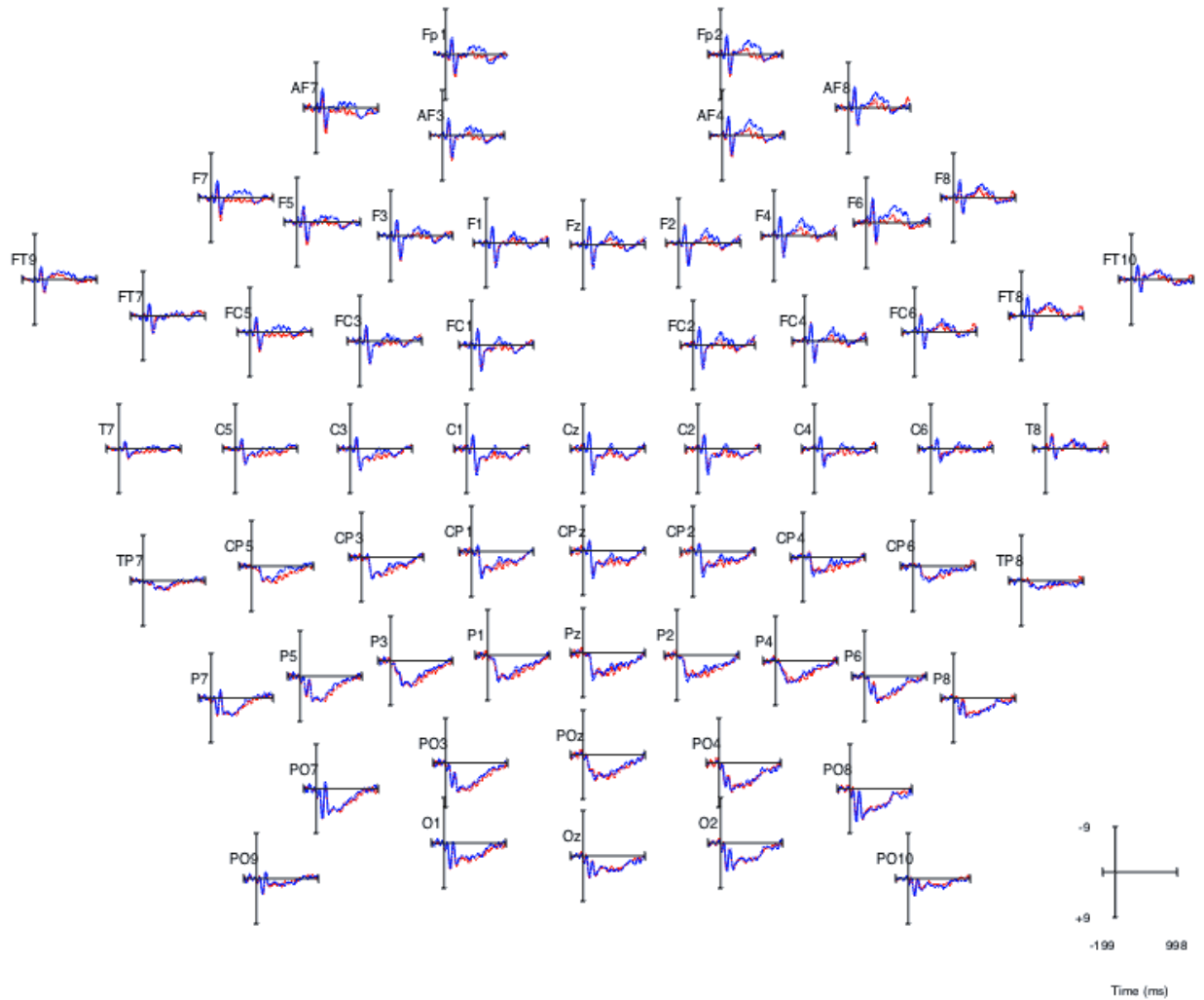
Figure 10. The grand average waveforms across time for ungrammatical (red) and grammatical (blue) sequences on the AGL task for non-learners, across all electrodes. Negative is plotted upward.

Figure 11 shows that similar to learners, non-learners had a typical visual P1-N1-P2-N2 complex. The ungrammatical and grammatical waveforms were similar to one another at central parietal electrode sites. In contrast to learners, the most apparent differences between grammatical and ungrammatical waveforms for non-learners was an enhanced negativity for grammatical items over anterior electrode sites, particularly over the right hemisphere, that lasted from approximately 300 – 800 ms. In contrast, the

ungrammatical waveforms at the anterior electrode sites were more positive-going than the grammatical waveforms, and this can also be seen in Figure 13.
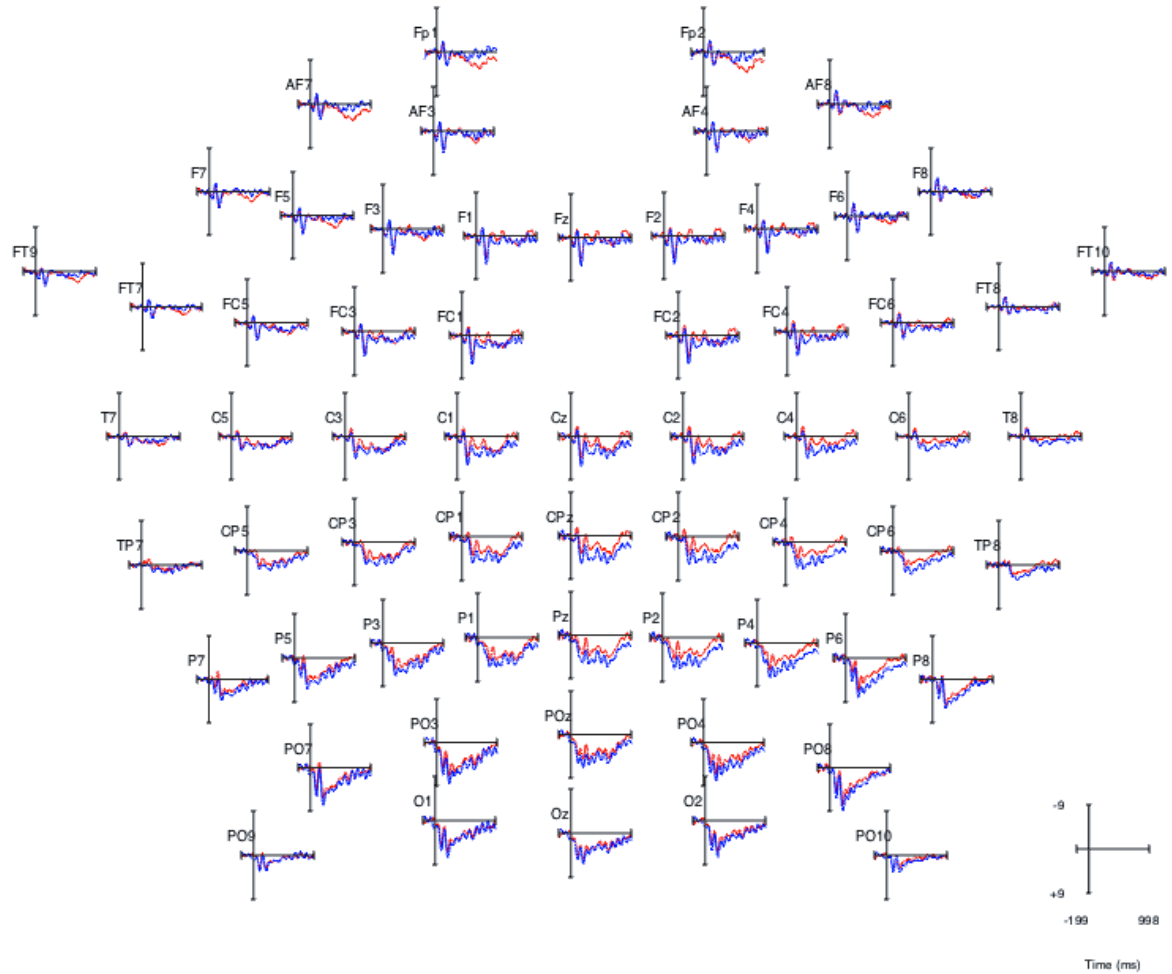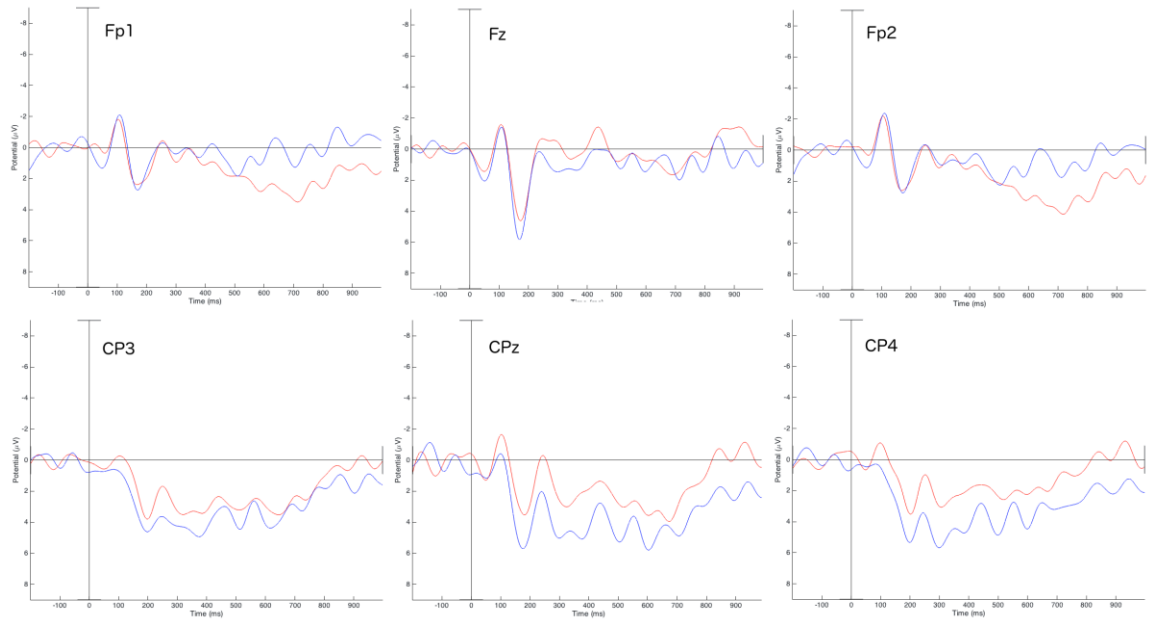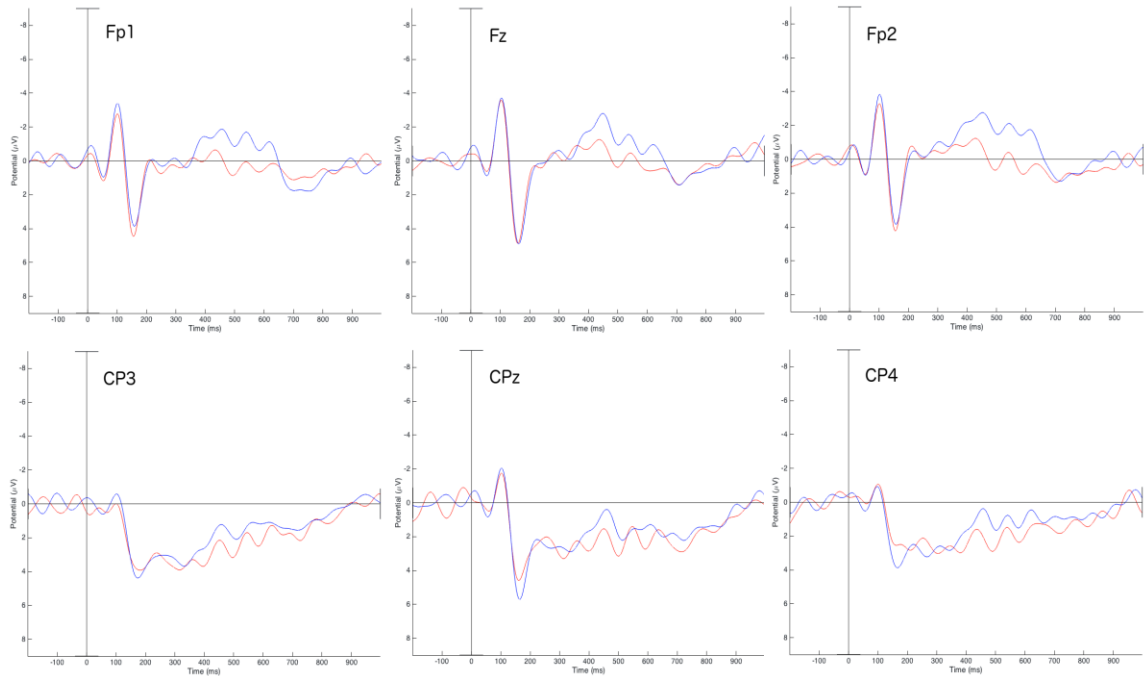

Figure 11. The grand average waveforms across time for non-learners in response to ungrammatical (red) and grammatical (blue), at the midline frontal and central ROIs. Negative is plotted upward. Scale is -8 to 8 µV.

The difference waveforms (ungrammatical – grammatical) at the midline frontal and central ROIs are shown in Figure 12 and allow a direct visual comparison of the effects elicited in each group. The central parietal negativity and later-onset anterior positivity in learners appear distinct from the difference waveforms for non-learners. Although both groups showed an anterior positivity for ungrammatical sequences, this was larger in amplitude and later in onset, as well as longer-lasting, in learners.
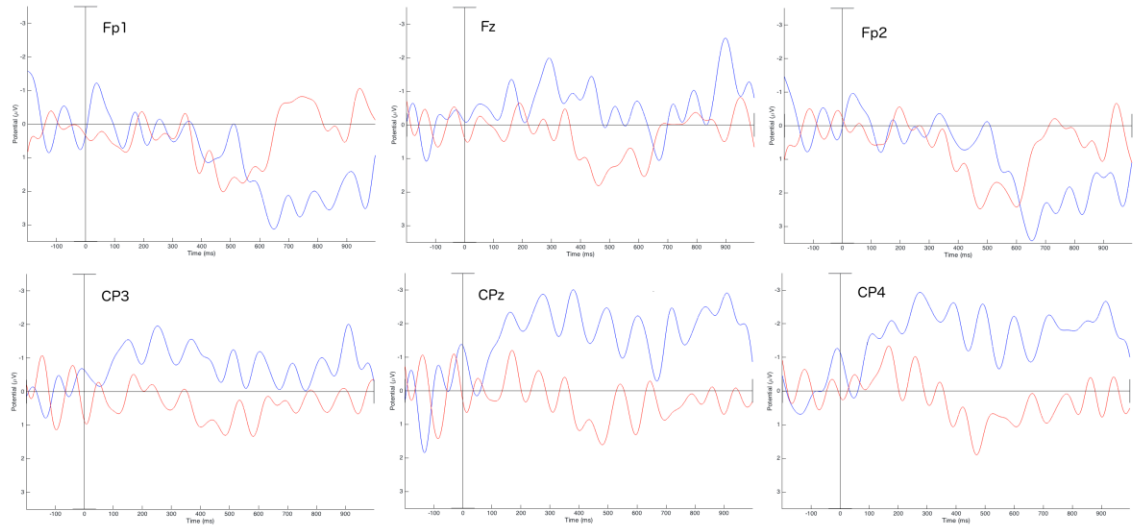
Figure 12. The difference waveforms between learners (blue) and non-learners (red) for the AGL task, at midline frontal and central ROIs. Negative is plotted upward. Scale is -8 to 8 µV.

Figure 13 shows the scalp distributions for the difference waves, averaged over three consecutive time windows for learners and non-learners. Learners had a sustained negativity, which was maximal over the central-parietal region (midline central ROI) and began in the 200 – 400 ms time window. This negativity was sustained for the duration of the trial, though gradually dissipated over time. Additionally, there was maximal positivity over the anterior regions approximately 600 – 800 ms post-stimulus onset, suggestive of a frontal P600. In contrast, non-learners had little activity until the 400 – 600 ms time window, in which there was a positivity that was maximal over the anterior ROIs. After 600 ms, this positivity disappeared.
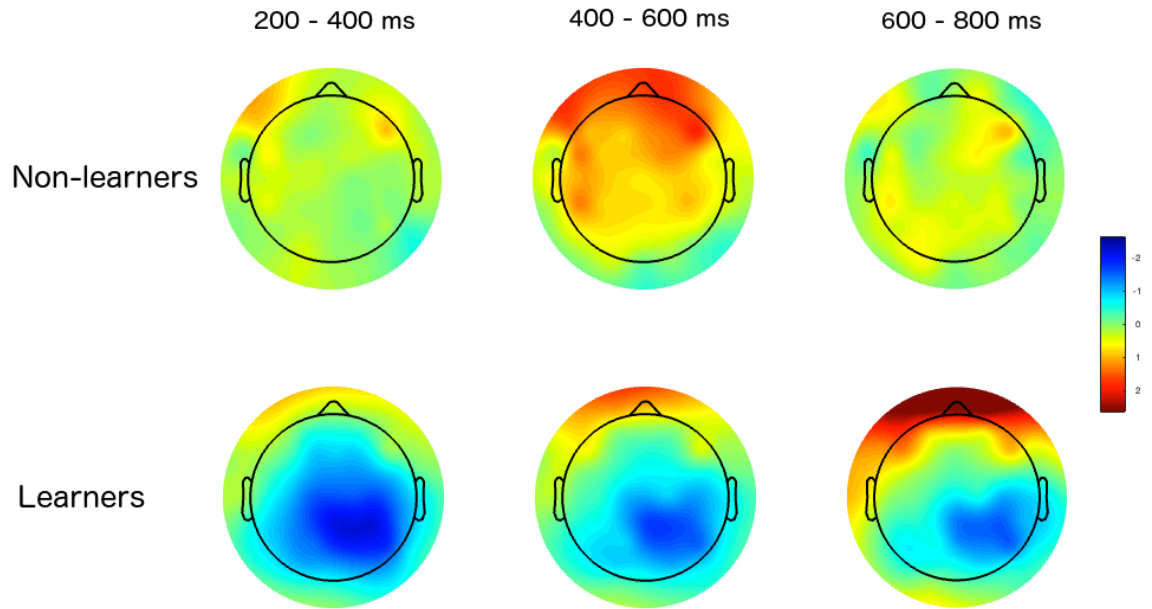
Figure 13. Scalp voltage maps for the AGL task showing the differences in voltage distribution between ungrammatical and grammatical sequences, averaged over selected time windows, for the non-learners and learners. Scale is -2.5 μV (blue) to 2.5 μV (red).

Statistical Analysis

A GAM was fitted in which the amplitude of the EEG data from the 600 – 800 ms time window were regressed as a function of grammaticality interacting with learning status, ROI, as well as error terms including a non-linear effect of trial, random intercepts for each subject, and a random non-linear effect of trial for each combination of subject and ROI. There were main effects of ROI, learning status, and grammaticality. Outliers were removed and represented approximately 2.11% of the data. There was a significant three-way interaction between grammaticality, learning status, and ROI. There were significant two-way interactions between learning status and ROI, as well as between grammaticality and learning status. There was no significant interaction between grammaticality and ROI. The main effects and interactions are shown in Table 2.

Table 2. Summary table for the AGL GAM in the 600 – 800 ms time window
(Denominator Lower-bound df = 68743.4).

Coefficient	df	F	p-value
Grammaticality	1	4.94	0.026
Learning Status	1	7.81	0.005
ROI	6	2.10	0.05
Grammaticality x Learning Status	1	20.72	0.001
Grammaticality x ROI	6	0.433	0.858
Learning Status x ROI	6	9.13	< 0.001
Grammaticality x Learning Status x ROI	6	4.95	< 0.001

Post-hoc testing was performed to determine where the significant differences for each variable originated, the summary for each post-hoc comparisons are shown in Table 3, along with the means for each condition and learner group, within each ROI. The threshold probability values (α) were Bonferroni corrected for 21 comparisons (3 comparisons per ROI).

Table 3. AGL post-hoc comparisons in the 600 – 800 ms time window (Denominator lower-bound df = 68743.4). U = ungrammatical, G = grammatical, n.s. = non-significant.

Comparisons	Estimate (μV)	Standard error	t-value	Corrected p-value
Amplitude differences (U – G) for learners at each ROI				
Left anterior	0.99	0.28	3.56	0.008
Midline anterior	0.47	0.24	1.98	n.s.
Right anterior	0.90	0.28	3.22	0.027
Left central	0.10	0.20	0.51	n.s.
Midline central	-0.88	0.21	-4.21	< 0.001
Right central	-0.33	0.20	-1.67	n.s.
Posterior	-0.60	0.16	-3.85	0.002
Amplitude differences (U – G) for non-learners at each ROI				
Left anterior	0.18	0.29	0.62	n.s.
Midline anterior	0.38	0.24	1.56	n.s.
Right anterior	0.38	0.29	1.34	n.s.
Left central	0.45	0.20	2.24	n.s.
Midline central	0.47	0.21	2.22	n.s.
Right central	0.20	0.20	1.00	n.s.
Posterior	0.16	0.16	1.02	n.s.
The ERP effect differences (U – G) between learners and non-learners at each ROI				
Left anterior	0.81	0.40	2.03	n.s.
Midline anterior	0.09	0.34	0.26	n.s.
Right anterior	0.51	0.40	1.28	n.s.
Left central	-0.35	0.28	-1.26	n.s.
Midline central	-1.35	0.30	-4.55	< 0.001
Right central	-0.53	0.28	-1.89	n.s.
Posterior	-0.76	0.22	-3.45	0.011

There was a significant difference in the ERP amplitude (ungrammatical –

grammatical) at several different ROIs for learners, as shown in Figure 14. This

difference was significantly positive at the left and right anterior ROIs for learners. In the

midline central and posterior ROIs, learners had a significant sustained negativity. In

contrast, there were no differences in the amplitudes between ungrammatical and

grammatical stimuli for non-learners. There were also significant differences in the ERP

amplitudes between learners and non-learners: at the midline central and posterior ROIs,

learners had a significantly more negative ERP amplitude than non-learners.
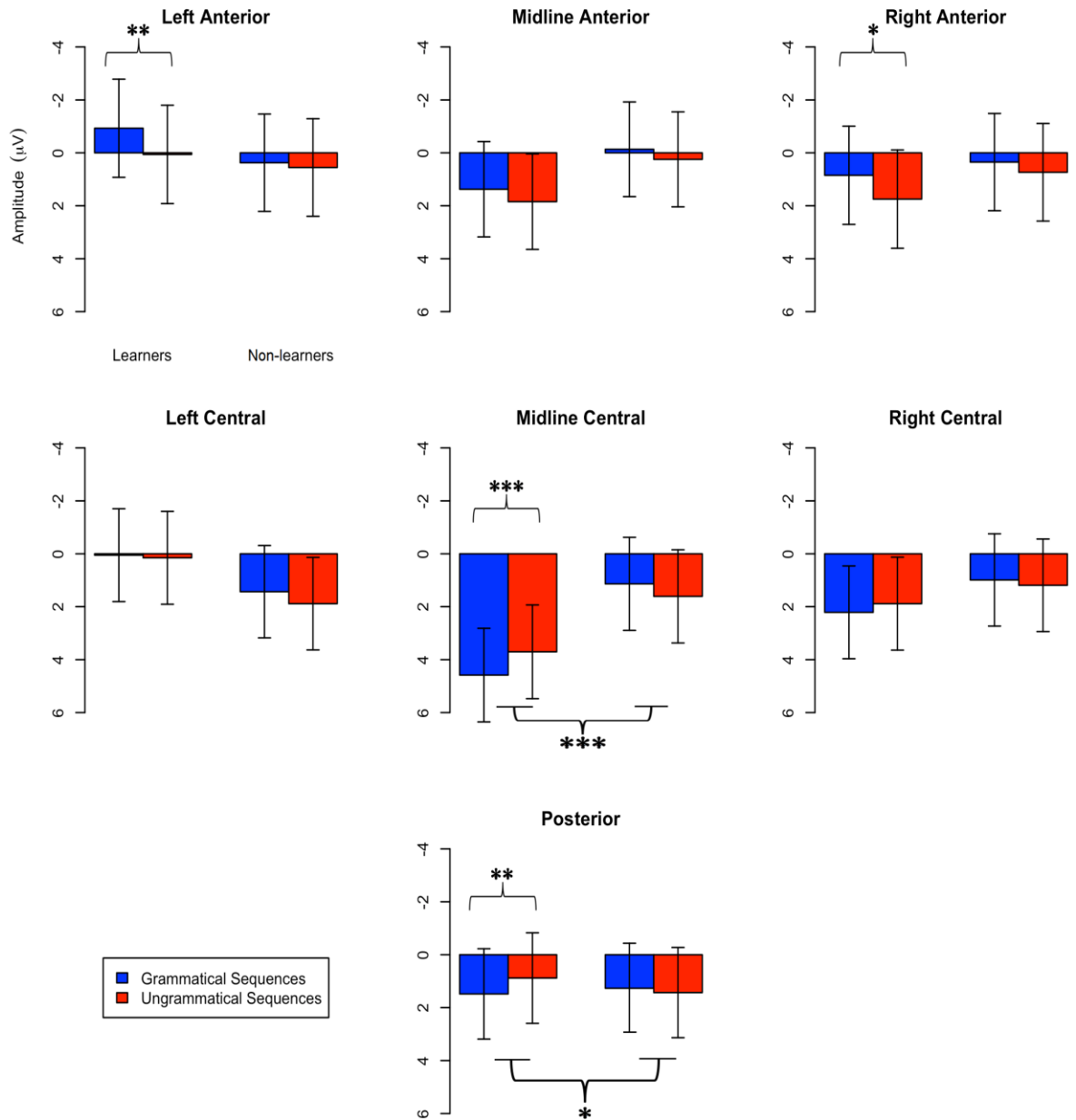
Figure 14. The amplitude (μV) for the AGL task in response to grammatical and ungrammatical sequences, across the seven different ROIs. Negative amplitude is plotted in the upward direction. Error bars represent the 95% confidence interval. * indicates a significance of $p < 0.05$, ** indicates a significance of $p < 0.01$, and *** indicates a significance of $p < 0.001$.

3.2.2 Speech Perception in Noise

Event-Related Potential Waveforms

Visual inspection of the ERP waveforms for the SPIN task suggested that there were N400 effects in each condition. Figure 15 shows the grand averaged waveforms for mismatch and match stimuli in the no-noise SPIN condition across all electrodes. The waveform in response to match items was generally positive-going; the mismatch waveform negative-going.
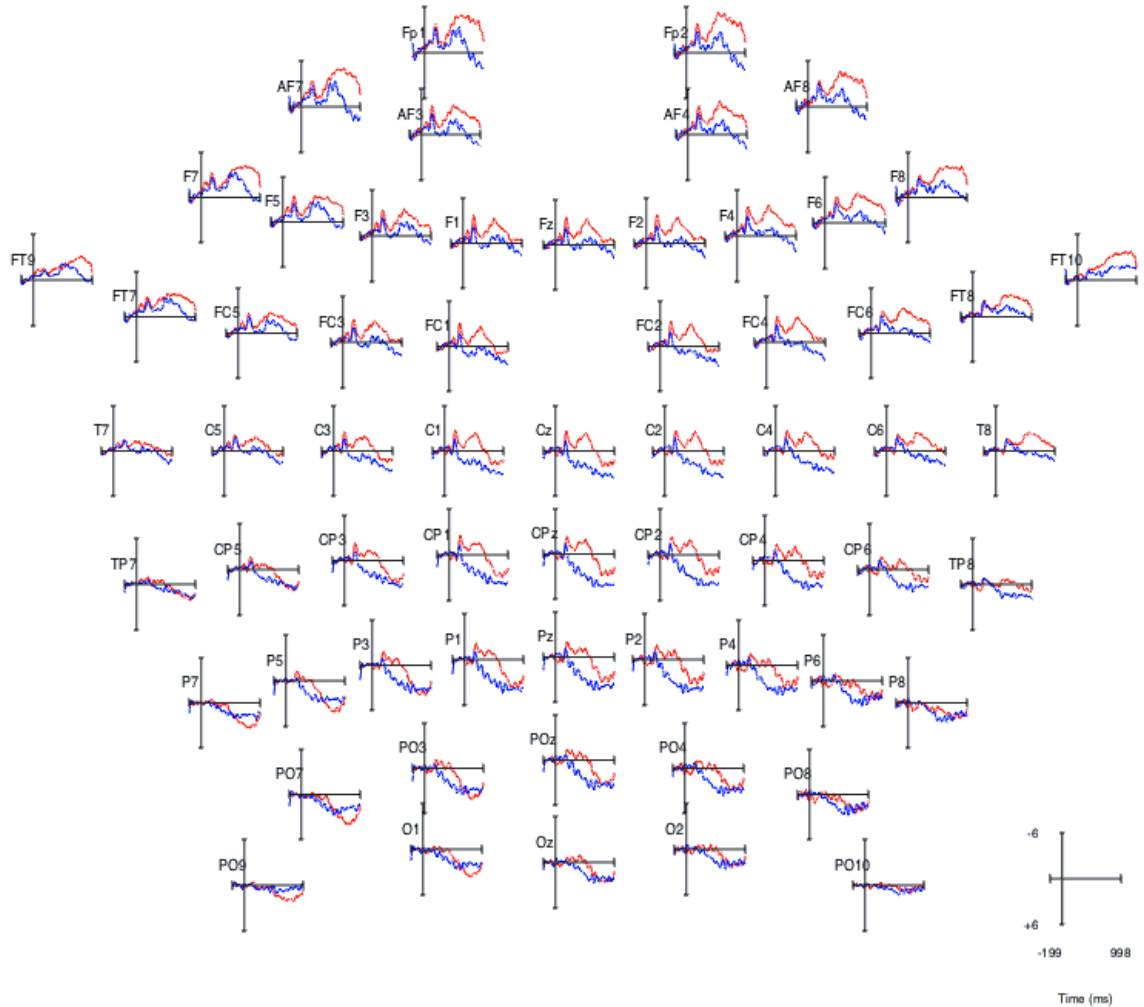
Figure 15. The grand average waveforms across time for mismatch (red) and match (blue) sentences in the no-noise SPIN task, across all electrodes. Negative is plotted upward.

Figure 16 shows that at the central parietal and anterior electrodes, there was an initial P1-N1-P2-N2 complex for both mismatch and match words. For mismatch words, an enhanced negativity began during the N1 peak and was sustained throughout the epoch, relative to match words. The peak difference occurred at around 500 ms.
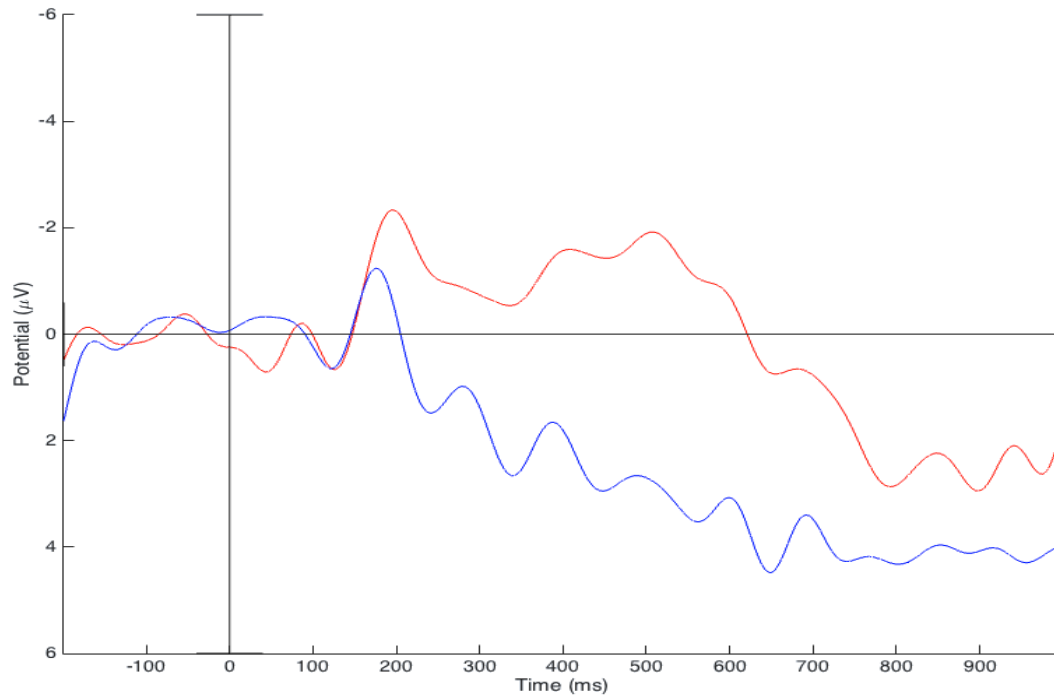
Figure 16. The grand average waveforms across time for mismatch (red) and match (blue) sentences in the no-noise SPIN task, at electrode CPz. Negative is plotted upward.

In the noise condition, the grand-averaged waveforms for mismatch and match stimuli followed a similar trend as those observed in the no-noise condition although, as seen in Figure 17, the P1-N1-P2-N2 complex was reduced. ERPs for match and mismatch words began to diverge between 100 – 200 ms, with an enhanced negativity typical of the N400 for mismatch words that peaked between 500 – 600 ms and was sustained until approximately 900 ms. This effect can be seen more clearly in Figure 18 depicting electrode CPz, where the negativity was maximal.
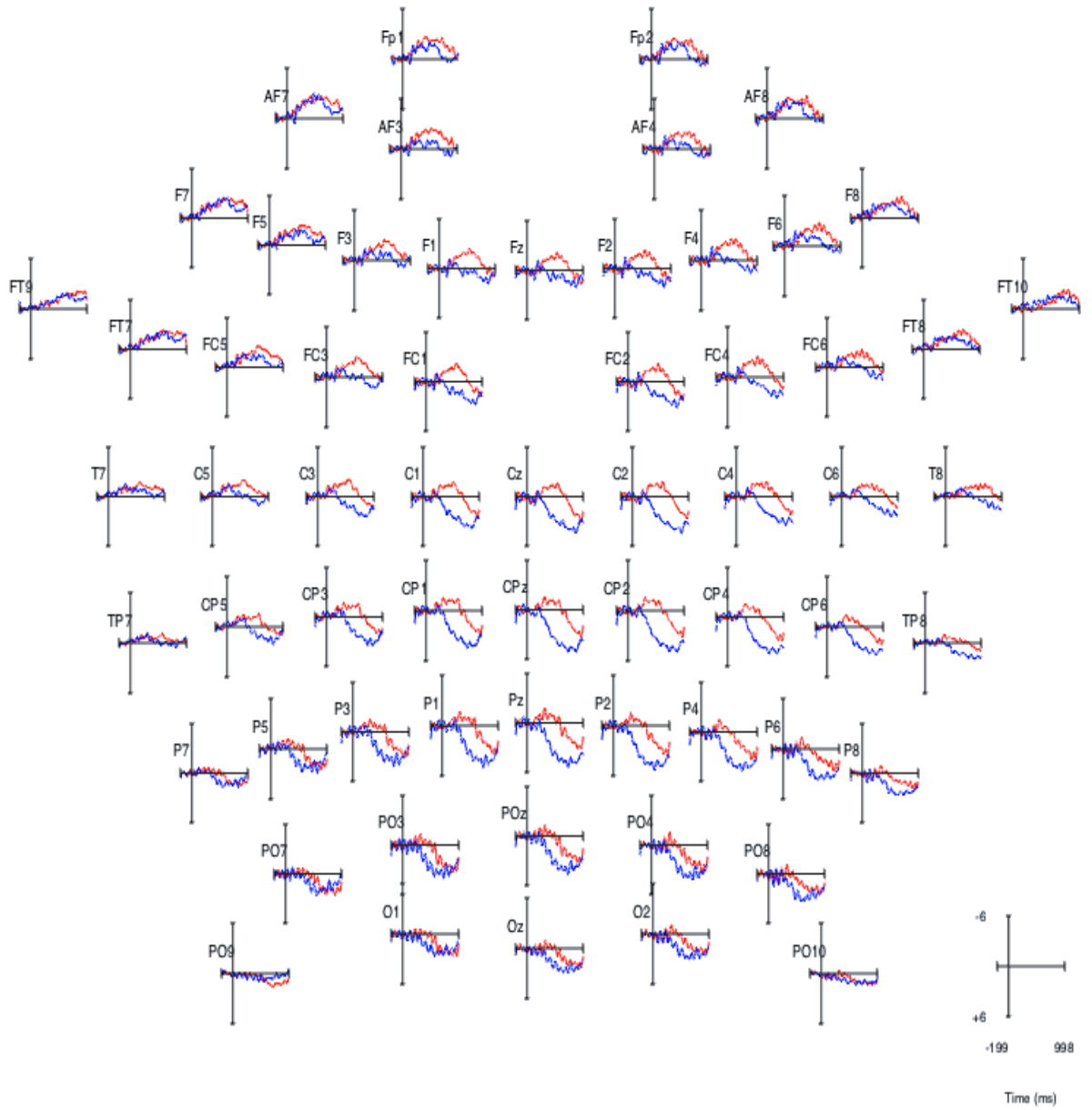
Figure 17. The grand average waveforms across time for mismatch (red) and match (blue) sentences in the with noise SPIN task, across all electrodes. Negative is plotted upward.
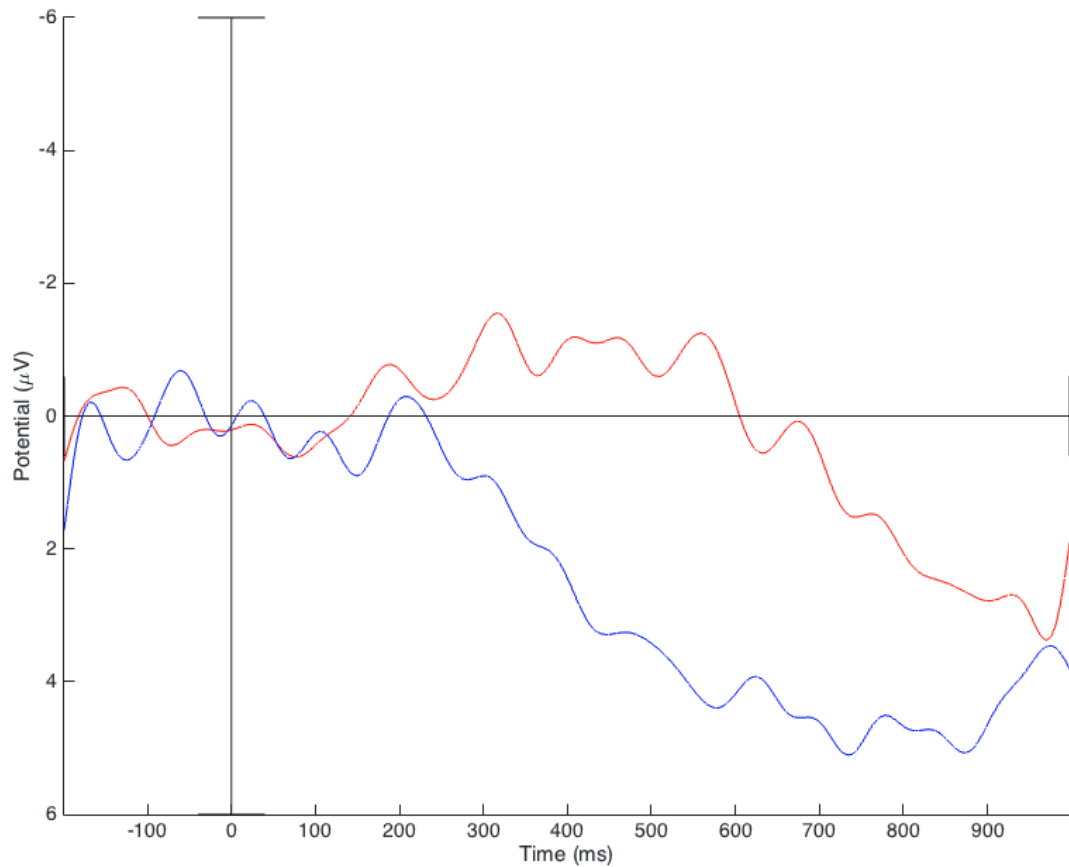
Figure 18. The grand average waveforms across time for mismatch (red) and match (blue) sentences in the with noise SPIN task, at electrode CPz. Negative is plotted upward.

Figure 19 shows the difference waveforms (mismatch – match), for both the noise and no-noise conditions at electrode CPz (where the difference waves were maximal). Overall, the difference waveforms were very similar in their amplitude and timing. Close inspection of the difference waveforms does suggest however that the effect of mismatches elicited an earlier response in the no-noise condition, whereas the effect in the with noise condition had higher amplitude later in the time window than the no-noise condition.
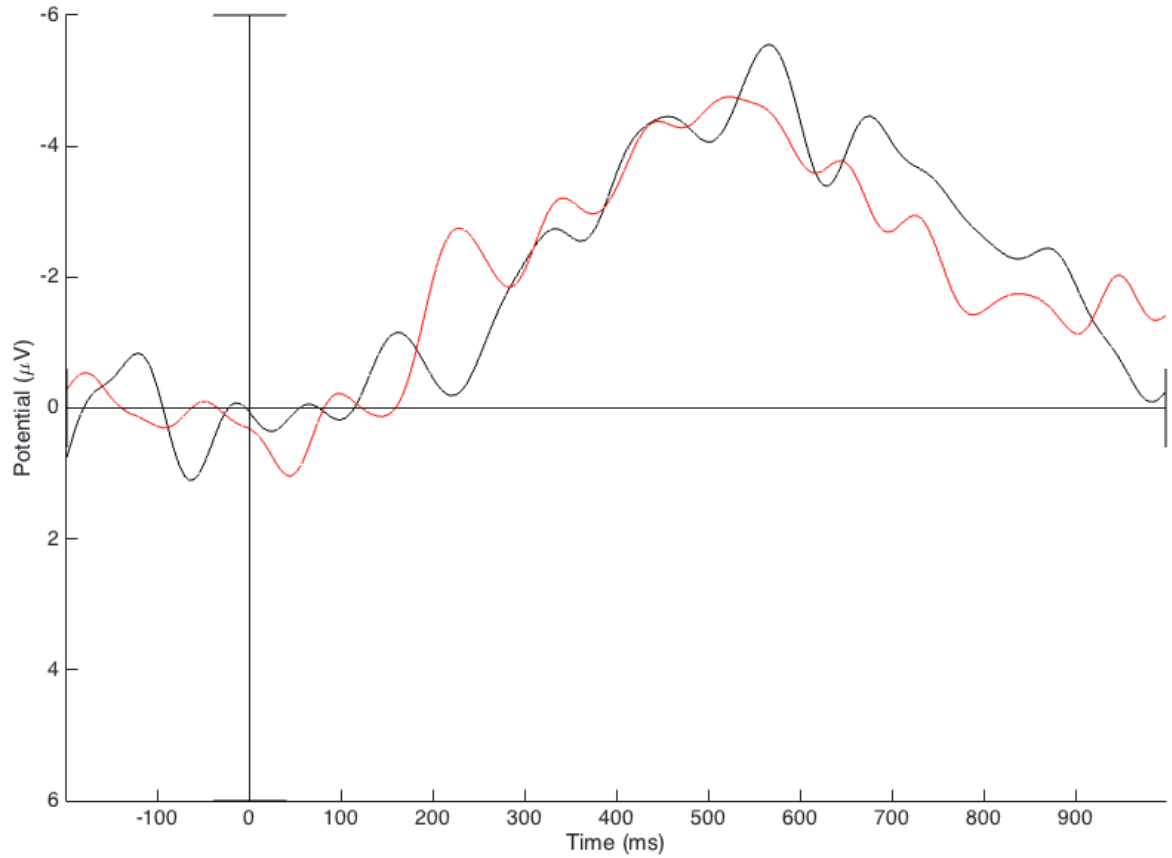
Figure 19. ERP difference waveforms (mismatch – match sentences) between the with noise (black) and no noise (red) conditions of the SPIN task, at electrode CPz. Negative is plotted upward.

Figure 20 shows the scalp distributions for the difference waves, averaged over three consecutive time windows, for each SPIN condition. The scalp distribution of the N400 was maximal over the central-parietal regions (midline central ROI) for both SPIN conditions. However, Figure 20 again illustrates the earlier onset of the N400 effect in the no-noise condition, with a more widespread negativity in the 100 – 300 ms time window compared to the with noise condition. Likewise, the N400 effect in the with noise condition is more widespread in the latest time window shown, from 500 – 800 ms.
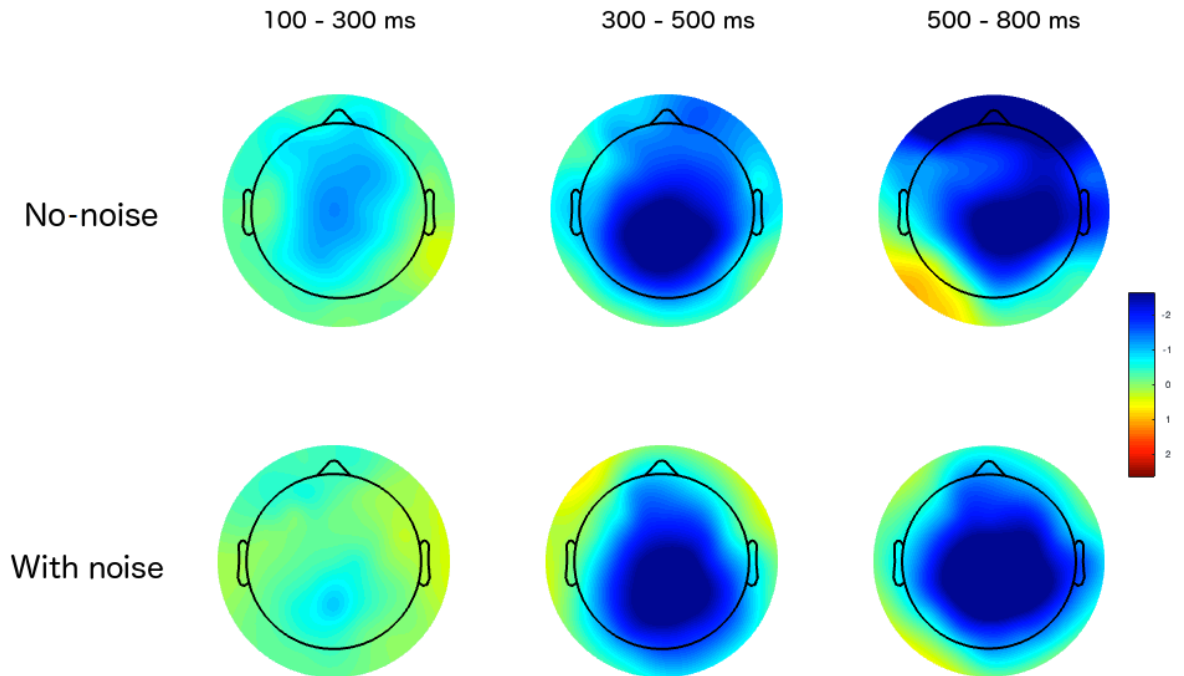
Figure 20. Scalp voltage maps for the SPIN task showing the differences in voltage distribution between mismatch and match sentences, averaged over selected time windows, for the no-noise and with noise conditions. Scale is -2.5 µV (blue) to 2.5 µV (red).

Statistical Analysis

A GAM was fitted in which the amplitude of the SPIN EEG data from the $300 - 500$ ms time window were regressed as a function of noise (no-noise or with noise) interacting with the category of the target word (match or mismatch), ROI, as well as a non-linear effect of trial, random intercepts for each subject, and random non-linear effect of trial for each combination of ROIs and subject. Outliers were removed and represented approximately 2.40% of the data. There were main effects for noise, target word category, and ROI, as can be seen in Table 4. There was a significant three-way interaction between noise, target word category, and ROI. Finally, there was a significant two-way interaction between target word category and ROI, between noise and ROI, and between noise and target word category.

Table 4. Summary table for the SPIN GAM in the 300 – 500 ms time window (Denominator Lower-bound *df* = 134044.5).

Coefficient	*df*	*F*	*p*-value
Noise	1	4.57	0.033
Target word category	1	461.36	< 0.001
ROI	6	93.51	< 0.001
Noise x Target word category	1	6.99	0.008
Noise x ROI	6	4.40	< 0.001
Target word category x ROI	6	23.96	< 0.001
Noise x Target word category x ROI	6	3.52	0.002

Post-hoc testing was performed to determine where the significant differences the interactions originated, by examining differences between mismatch and match words for each noise condition, within each individual ROI. The amplitude differences for each contrast can be seen in Table 5. The threshold probability values (α) were Bonferroni corrected for 21 comparisons (3 comparisons per ROI). The N400 mismatch effect (mismatch – match difference) was significant in each individual ROI in both the with background noise and no-noise conditions, except in the left anterior ROI in the with background noise condition. The amplitude of the N400 mismatch effect did not differ between noise conditions at any ROI except for the left central ROI where it was smaller in the with noise condition, as shown in Figure 21.

Table 5. SPIN post-hoc comparisons in the 300 – 500 ms time window (Denominator Lower-bound df = 134044.5). MS = mismatch, M = match, n.s. = non-significant.

Comparisons	Estimate (µV)	Standard error	t-value	Corrected p-value
Amplitude differences between MS and M stimuli at each ROI (No-noise)				
Left anterior	-0.74	0.18	-3.98	0.001
Midline anterior	-1.65	0.16	-10.56	< 0.001
Right anterior	-1.14	0.19	-6.16	< 0.001
Left central	-1.09	0.13	-8.54	< 0.001
Midline central	-2.90	0.14	-21.48	< 0.001
Right central	-1.44	0.13	-11.31	< 0.001
Posterior	-1.31	0.10	-13.06	< 0.001
Amplitude differences between MS and M stimuli at each ROI (With noise)				
Left anterior	0.04	0.18	0.20	n.s.
Midline anterior	-1.34	0.15	-8.68	< 0.001
Right anterior	-0.65	0.18	-3.54	0.008
Left central	-0.48	0.13	-3.79	0.003
Midline central	-2.40	0.13	-17.82	< 0.001
Right central	-0.92	0.13	-7.21	< 0.001
Posterior	-1.50	0.10	-15.00	< 0.001
The N400 effect differences between with noise and no-noise at each ROI				
Left anterior	0.77	0.26	2.98	n.s.
Midline anterior	0.32	0.22	1.44	n.s.
Right anterior	0.49	0.26	1.90	n.s.
Left central	0.61	0.18	3.38	0.015
Midline central	0.50	0.19	2.64	n.s.
Right central	0.53	0.18	2.94	n.s.
Posterior	-0.19	0.14	-1.33	n.s.
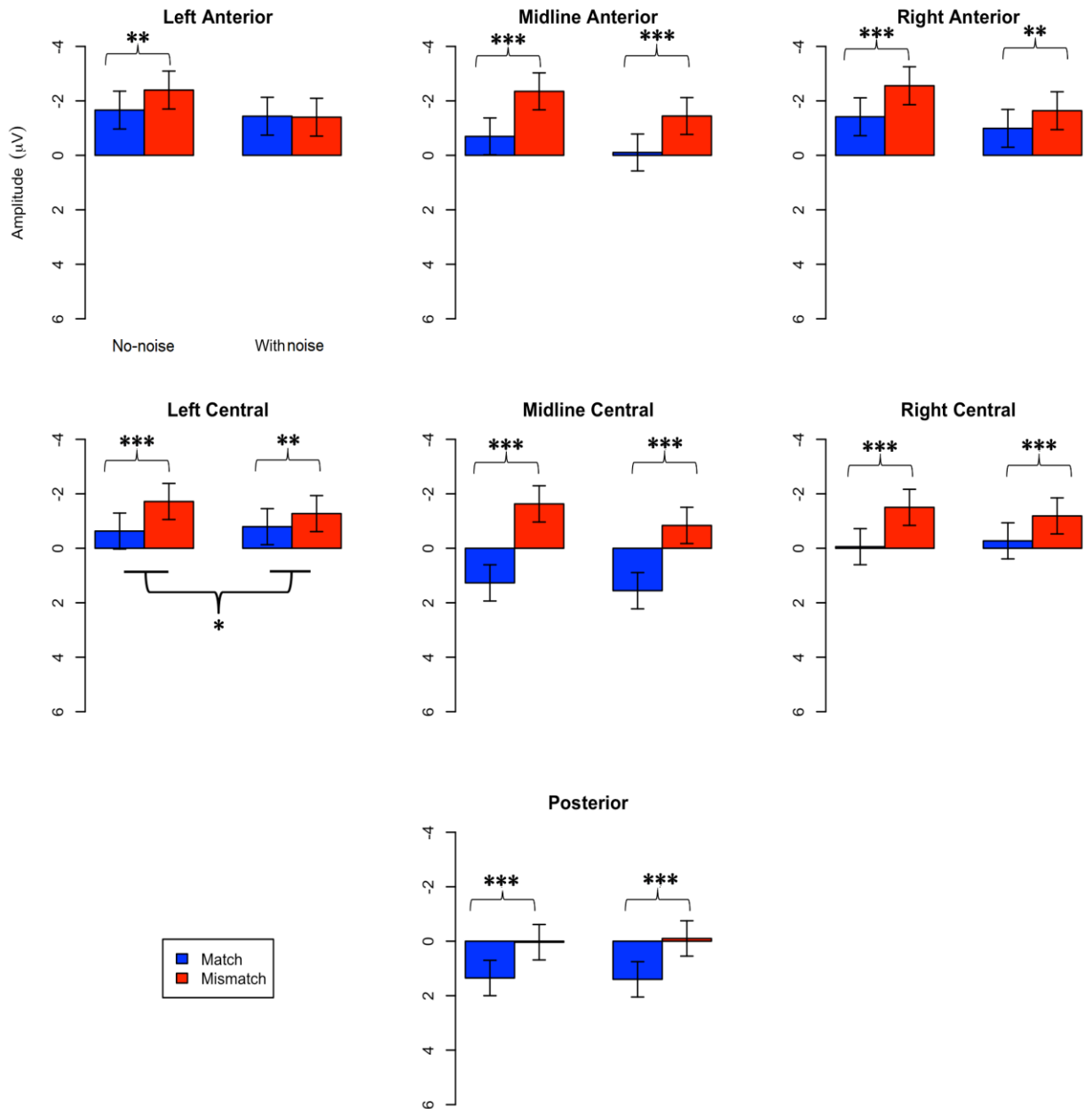
Figure 21. The amplitude (μV) for the SPIN task in response to match and mismatch sentences when there was no background noise and with background noise conditions, across the seven different ROIs. Negative amplitude is plotted in the upward direction. Error bars represent the 95% confidence interval. * indicates a significance of $p < 0.05$, ** indicates a significance of $p < 0.01$, and *** indicates a significance of $p < 0.001$.

3.2.3 Correlations between Statistical Learning and Speech Perception

GAMs were used to determine whether there was a relationship between the amplitudes of the SPIN and AGL tasks. For each model, the amplitude differences for each task (e.g., mismatch − match sentences for SPIN, ungrammatical − grammatical for AGL) were used. A GAM was fitted in which the amplitude differences of no-noise SPIN task were regressed as a function of AGL amplitude differences interacting with learning status on the AGL task, as well random intercepts for each subject and ROI. A second GAM was fitted in which the amplitude differences for the with background noise SPIN task were regressed as a function of the amplitude differences in the AGL task interacting with learning status, as well as random intercepts for each subject and ROI.

Figure 22 shows that in the no-noise condition, non-learners had a significant positive correlation between SPIN and AGL amplitudes, $\beta_{weight} = 0.48$, $t(78.09) = 2.0$, $p = 0.049$ — as the N400 in the SPIN task became larger, so did the negativity of the AGL grammaticality effect. Learners did not have a significant correlation between SPIN and AGL amplitudes, $\beta_{weight} = -0.08$, $t(78.09) = -0.73$, $p = 0.472$. The slopes of the correlations for learners and non-learners also significantly interacted, $\beta_{weight} = -0.56$, $t(78.09) = -2.13$, $p = 0.037$.

Figure 22. The correlation between SPIN and AGL amplitudes (μV) in the no-noise condition for the non-learners (red) and learners (blue). The points represent the amplitude differences for each subject at all ROIs. Lines represent the trend lines. Negative is plotted upward for the y-axis and on the right for the x-axis.

Figure 23 shows that in contrast to the no-noise condition, learners had a significant positive correlation between SPIN and AGL amplitudes in the with noise condition, $\beta_{weight} = 0.53$, $t(79.41) = 5.86$, $p < 0.001$ — as the N400 in the SPIN task became larger, so did the negativity of the AGL grammaticality effect. Similar to the no-noise condition, non-learners had a positive correlation between SPIN and AGL amplitudes in the with

70

noise condition, however this correlation trended towards significance, $\beta_{weight} = 0.34$,

$t(79.41) = 1.90$, $p = 0.062$. Unlike the no-noise condition, there was no significant

interaction for the with noise condition between the slopes of the correlations between

learners and non-learners, $\beta_{weight} = 0.18$; $t(79.41) = 0.91$, $p = 0.36$.

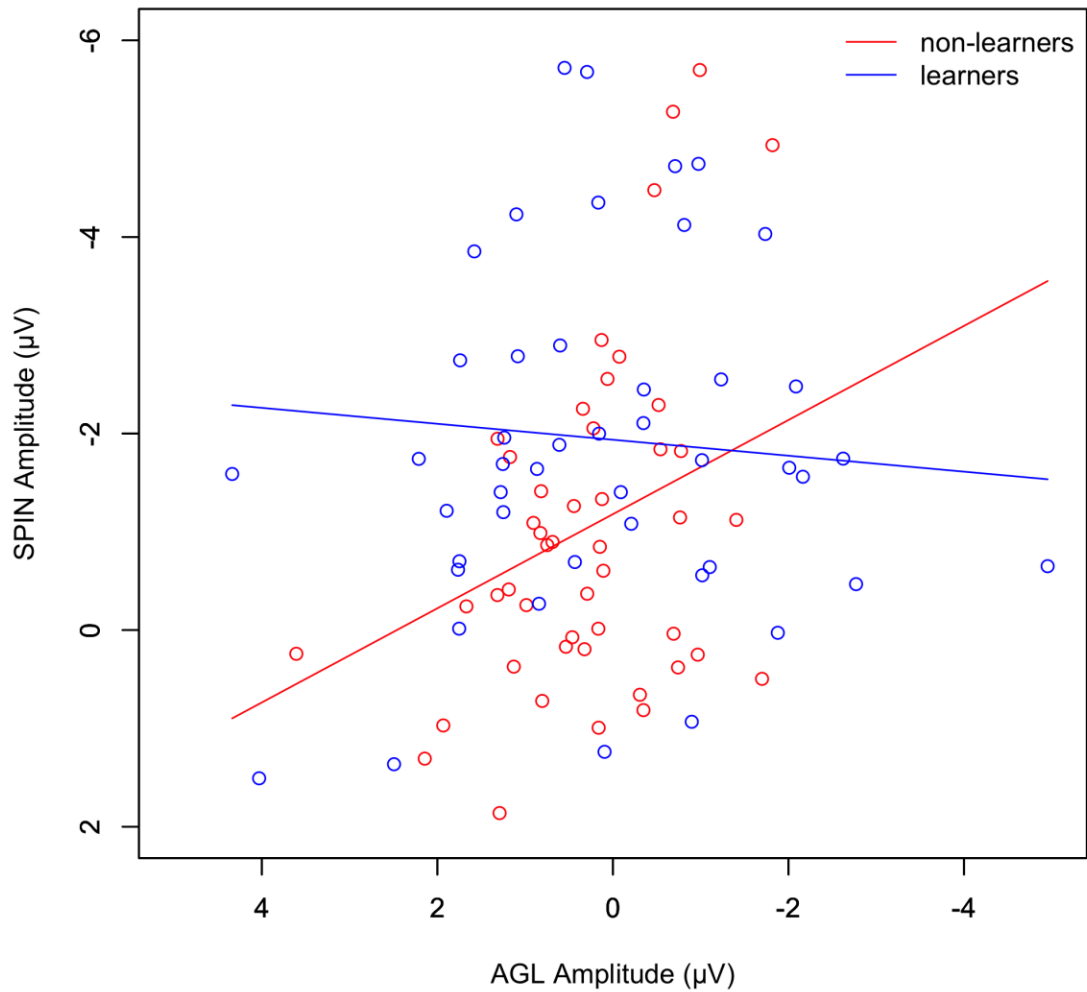**Amplitudes (μV) of SPIN (with noise) vs AGL for Mismatch - Match Sequences For All ROIs**
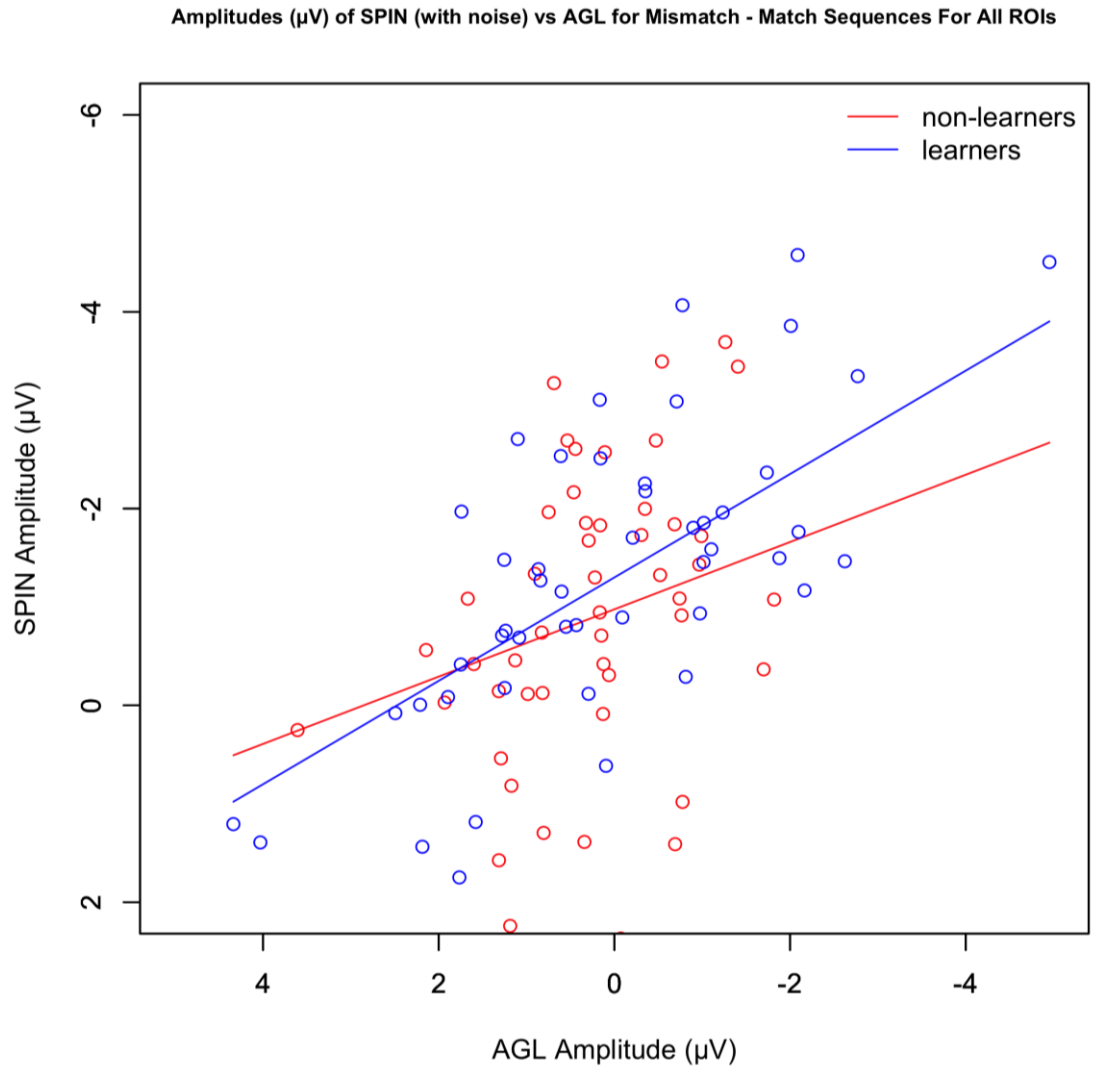
Figure 23. The correlation between SPIN and AGL amplitudes (μV) in the presence of background noise for non-learners (red) and learners (blue). The points represent the differences in amplitude for each subject for all seven ROIs. Lines represent the trend lines. Negative is plotted upward on the y-axis, and on the right for the x-axis.

CHAPTER 4 DISCUSSION

This study aimed to determine whether there was a relationship between the neurocognitive systems supporting statistical learning and speech perception in noise. This relationship was investigated by testing whether people's performance on an AGL task correlated with their ability to accurately predict incoming information in a SPIN task. The results of the present study confirmed many of the hypotheses, and posed several new questions regarding this relationship.

Based on the literature, I hypothesized that the AGL task would elicit a P600 effect in response to the differences in amplitudes between ungrammatical and grammatical sequences in the central parietal regions of the scalp (Christiansen et al., 2012). I also predicted that people who were more sensitive to the transitional probabilities of the stimuli — measured by the learning scores — would have an enhanced P600 effect in comparison to those who were less sensitive.

Secondly, I predicted that both SPIN conditions would elicit an N400 effect, regardless of whether there was background noise or not. It was expected that the addition of background noise to the SPIN task would impair behavioural performance, because it would be more difficult to hear the target word. Additionally, it was hypothesized that the presence of background noise with a low SNR would reduce the observed N400 effect, in comparison to when there was no background noise.

Lastly, I predicted that there would be a significant correlation between the P600 in the AGL task and the N400 in the SPIN task. Specifically, I predicted that as the amplitude of the P600 increased (became more positive) in the AGL task, the amplitude of the N400 in the SPIN task would also increase (become more negative). It was also

predicted that in the presence of background noise, this correlation would be reversed in people who were better at the AGL task.

4.1 Artificial Grammar Learning

Behavioural Performance

Behaviorally, I expected to see variation in learning scores across participants for the AGL task. This hypothesis was confirmed by the behavioural results as the learning scores ranged from the negatives to the positives, and this range was similar to those observed in other AGL tasks (Conway et al., 2010, 2011). Due to the range of learning scores — and in particular the fact that learning scores less than zero indicate that the person did not extract the statistical probabilities of the sequences — I was concerned that combining data from all participants might dilute any possible P600 effect at the group level. To circumvent this issue, participants were split into two groups, non-learners or learners, based on the median of the learning scores. This follows the practice of other statistical learning studies e.g., Sanders, Newport, and Neville (2002) . The *t*-tests on the learning scores also supported this split, as learners had significantly higher learning scores than non-learners. Additionally, learners had higher accuracy (65%) than non-learners (50%), but this difference did not survive multiple comparison correction.

ERP Responses to Artificial Grammar Learning

Based on the literature, I hypothesized that the AGL task would elicit a P600 in the midline central ROI, and that the P600 would be modulated by learning scores (Christiansen et al., 2012). People who were more sensitive to the statistical frequencies of the stimuli were expected to have an enhanced P600 effect in comparison to people who were less sensitive. The grand average waveforms of non-learners did not differ in

response to grammatical and ungrammatical stimuli. Statistical analysis demonstrated that for non-learners, there were no significant differences between the amplitudes for grammatical and ungrammatical stimuli across any ROI. In contrast the difference waveforms at the midline central ROI for learners demonstrated a significant negative deflection at approximately 100 ms that was sustained for the duration of the epoch. Additionally, this negativity for learners was significantly larger than the ungrammatical – grammatical difference wave for bad learners.

While the expected P600 at the central parietal region was not observed, the sustained negativity for learners but not non-learners suggests that it is an effect of the AGL grammatical violations. Another study investigating ERPs in response to statistical learning paradigms found a similar sustained negativity (Lang & Kotchoubey, 2000). This study reported an N2 effect 200 – 360 ms post-stimulus onset, and another large negative deflection beginning at 390 ms, which was sustained until approximately 700 ms and was also largest at midline anterior electrode. Lang and Kotchoubey (2000) hypothesized that this "slow negative wave" (SNW) may be a member of an N400 family, though separate from the N400 itself, as it was elicited in response to incongruous stimuli. The effect in the present study was later than the typical 300 – 500 ms time window in which the N400 tends to occur. However, other studies have shown that the N400 effect can be delayed by a multitude of factors, including accuracy and confidence in responses (Cansino & Tellez-Alanis, 2000). Furthermore, Kutas and Federmeier (2011) note that the N400 effect can occur between a longer time period of 200 – 600 ms.

It is possible that the sustained negativity observed in learners may be the SNW or another N400-like effect. It seems more likely that this sustained negativity is

representative of an N400-like effect, because it was maximal over the midline central and posterior ROIs, rather than the anterior ROIs where the SNW was observed by Lang and Kotchoubey (2000). Learners had higher learning scores on the AGL task compared to non-learners and it is possible that learners assimilated enough of the transitional probabilities of the stimuli to develop expectations as to what the sequences should look like. This may be true, especially in light of the research done by Lang and Kotchoubey (2000), who found that even passive statistical learning paradigms can elicit ERP effects. Thus, learners may have acquired enough of the statistical frequencies of the stimuli to be "surprised" when presented with an unexpected stimulus (an ungrammaticality). The evidence suggests that learners may have developed expectations about what a grammatical sequence should look like and when the target stimulus violated this expectation, an N400-like response was elicited.

While there was an absence of a P600 effect in learners at the midline central ROI, scalp topography maps and statistical modelling indicated a significant positivity that began 650 ms post-stimulus onset in the anterior ROIs. Statistical modelling demonstrated that the difference in amplitudes between ungrammatical and grammatical sequences was significantly positive in the left and right anterior ROIs, though not at the midline anterior ROI. While the P600 elicited by syntactic violations is typically found in the midline central and posterior regions of the scalp (Coulson et al., 1998; Friederici et al., 2002; Hagoort & Brown, 2000; Kaan & Swaab, 2003a; Osterhout & Holcomb, 1992); other studies have reported more anterior positivities (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Friederici et al., 2002; Kaan & Swaab, 2003a, 2003b; Osterhout & Holcomb, 1992). Kaan and Swaab (2003b) reported an anterior positivity for

syntactically more complex sentences compared to simpler constructions, in contrast to a typical posterior P600 for syntactic violations. Federmeier et al. (2007) found that sentences containing semantic violations elicited a frontal P600-like effect in response to sentences that were strongly constrained and contained a mismatch, following the N400. The authors hypothesized that the frontal P600 may instead reflect costs associated with processing sentences that violated large expectations, because this effect was only found for highly constrained sentences containing a mismatch. Other studies, however, have found a more posterior P600 following highly-constrained semantic violations (Coulson & Van Petten, 2002; Juottonen, Revonsuo, & Lang, 1996; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Moreno & Kutas, 2005; Newman, Tremblay, Nichols, Neville, & Ullman, 2012; Ojima, Nakata, & Kakigi, 2005; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009). Following Kaan and Swaab's (2003b) finding of a greater anterior P600 for increased grammatical complexity, it is possible in the present study that learners — although not consciously aware of the statistical probabilities of the stimuli — at some level processed the grammar but because this was not well-learned (as evidenced by their generally poor accuracy) the system processing the grammar was taxed, thus leading to the anterior positivity.

The combination of the frontal P600-like effect along with the N400-like effect in the midline central and posterior ROIs suggests that learners were somewhat implicitly aware of the underlying grammar, especially as they more accurately reproduced grammatical sequences (as indexed by their learning scores). Learners may have engaged a system that was involved in the processing of the transitional probabilities of the stimuli, and this

system was highly activated in response to ungrammatical stimuli because the system was sensitized to the grammar.

Implicit versus Explicit Learning

Based on the literature, I expected that the current AGL task — which was a modified version of the AGL paradigm used by Christiansen et al. (2012) — would elicit a central parietal P600 effect. Christiansen et al. (2012) investigated whether language and statistical learning relied on a similar system to process syntactic violations. The authors found that there was a P600 effect in the central parietal region of the scalp that was common between both the statistical learning paradigm (an AGL task) and the language task. These results suggest that statistical learning relies on a similar system as language to process syntactic violations. However, unlike the AGL paradigm used in the current study, Christiansen et al. (2012) informed participants before they began the task that the sequences they would see followed specific patterns. Participants were shown different grammatical segments of the stimuli and eventually were exposed to the full sequences. Additionally, participants repeatedly observed the same sequences. This was to ensure that participants would have high levels of accuracy and thus develop implicit expectations about the sequences. After training, participants were tested on how well they acquired the grammar and were asked to judge whether a sequence was grammatical or not. Participants were highly accurate in their judgments (>90%) (Christiansen et al., 2012). The AGL paradigm used in the current study did not follow these methods: participants received less training than those in the Christiansen et al. (2012) experiment, were naïve about the grammar, and were asked to reproduce the sequences.

These differences between the AGL tasks are arguably why the hypothesized P600 effects were not elicited in this study. It is possible that the P600 effect may only be elicited when participants are proficient enough to know whether a violation occurred. Most research on the P600 involves people contrasting sentences that are syntactically correct to sentences that either have a syntactic violation or are ambiguous (e.g., garden path sentences) (Federmeier et al., 2007; Kaan & Swaab, 2003b). In these types of experiments, people are generally aware of the violation and can likely identify where the error occurs. For example in the following sentence, *She shook the sands off of her beach towel*, most people can identify the location of the syntactic violation, though they may not know the precise reason why it is wrong (e.g., sand is a mass noun). Similarly, Christiansen et al. (2012) trained people to high accuracy on the AGL task so that they could develop expectations about the sequences, as well as informed participants that the sequences did follow grammatical rules. Statistical learning was measured only by the accuracy of their responses, but participants likely could have articulated where the violation in an ungrammatical sequence occurred, if not the reason why the sequence was incorrect. Thus, the P600 effect may only be found in experiments in which people 1) are aware of the violation and 2) can identify the violation. This theory could explain why the expected P600 was not found for this study: participants were not informed about the underlying grammar and thus while they may have developed an expectation about what a target stimulus should look like, they may not have been aware enough to know what the issue was.

In support of this theory, van de Meerendonk et al. (2009) theorized that when a strong expectation of a stimulus is violated, such as in the Christiansen et al. (2012) AGL

task, the P600 reflects a general reanalysis rather than a syntactic reanalysis. van de Meerendonk et al. (2009) speculate that the P600 effect is elicited by semantic violations when stimuli do not match people's strong expectations. The participants in the Christiansen et al. (2012) AGL task likely had strong expectations about what the frequencies of the stimuli were because they were trained to have high levels of accuracy, and a P600 effect was observed when their expectations were violated. In contrast, participants for the current AGL task may not have developed strong enough expectations to elicit this effect. This may explain why I instead saw an N400-like effect in the central parietal and posterior electrodes.

Other EEG experiments on statistical learning have found N400s in response to grammatical violations (Abla, Katahira, & Okanoya, 2008; Abla & Okanoya, 2009). Abla and Okanoya (2009) examined ERP responses as people watched streams of shapes in which certain shape triplets had a higher transitional probability than others. Afterwards, participants were shown two pairs of shape triplets, one of which had appeared during the continuous stream, and were asked to indicate which was familiar. Participants were split into high and low learners. High learners had an N400 effect in response to the first shape in triplets with lower transitional probabilities; this effect decreased over time as high learners saw more triplets that had lower occurrences. In contrast, low learners had no N400 effect at any point. These results suggest that people sensitive to the transitional probabilities of stimuli experience an N400 effect when they observe stimuli with lower transitional probabilities (Abla & Okanoya, 2009). The results of the Abla and Okanoya (2009) experiment are similar to the current AGL paradigm for both accuracy and that an N400-like effect was observed in learners. Similar results were also observed when

people listened to streams of nonlinguistic auditory stimuli, such as tones (Abla et al., 2008). Another study used triplets of non-verbal sounds, such as glass breaking, but an N100 ERP in response to the initial sound of the triplet was found rather than an N400 (Sanders, Ameral, & Sayles, 2009). Mueller, Oberecker, and Friederici (2009) presented native and non-native Italian speakers with syntactically correct spoken Italian sentences that had non-adjacent dependencies between words. After training, participants listened to syntactically correct and incorrect sentences, and had to judge whether the sentence was grammatically correct or not. Native speakers had an N400 followed by a P600 effect in response to syntactically incorrect sentences; non-native speakers showed a similar N400-like effect followed by an anterior P600-like effect (Mueller et al., 2009). These results are in accordance with those of the present study, in that learners showed both an N400-like effect and an anterior P600-like effect similar to those observed in non-native speakers. It appears that statistical learning paradigms tend to elicit ERP responses other than the P600 (Abla et al., 2008; Abla & Okanoya, 2009; Lang & Kotchoubey, 2000; Mueller et al., 2009; Sanders et al., 2009). Interestingly, a similar pattern has been found in second language learning, which can be considered in part a much more complex statistical learning task. Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, and Molinaro (2006) investigated how ERP responses to semantically anomalous and syntactically incorrect sentences changed over time in adults who were learning a second language. The authors found that participants initially had an N400 effect in response to syntactically incorrect sentences, rather than a P600. However, the N400 effect was replaced by a P600 after four months of language learning, after which time the amplitude of the P600 increased with proficiency (Osterhout et al., 2006).

In light of the results of these studies (Abla et al., 2008; Abla & Okanoya, 2009; Lang & Kotchoubey, 2000; Mueller et al., 2009; Sanders et al., 2009) as well as those from Osterhout et al. (2006), it appears that the N400 effect is initially elicited in response to stimuli with lower transitional probabilities. As people receive further exposure to stimuli and become more accurate in identifying grammatical stimuli, it appears that the N400 effect is gradually replaced by a P600 (Osterhout et al., 2006). This could explain why Christiansen et al. (2012) observed a P600 effect in response to ungrammatical sequences in their statistical learning paradigm: people received extensive training and were highly accurate (>90%) at identifying which sequences were grammatical. It may be that more extensive training, higher accuracy, as well as explicit awareness of the artificial grammar for the current study would have elicited a more typical central parietal P600 rather than an N400-like effect (Abla et al., 2008; Abla & Okanoya, 2009; Mueller et al., 2009; Osterhout et al., 2006; Sanders et al., 2009). Future studies should investigate whether there is a difference in the type of ERP response elicited in participants when they are naïve to the grammar versus when they are informed that the sequences follow a pattern.

4.2 Speech Perception in Noise

Behavioural Performance

It was hypothesized that behavioural performance on the SPIN task would be impaired through the addition of background noise. The behavioural results for the SPIN task supported this hypothesis, as participants were both more accurate and faster at repeating the target word of the sentence in the absence of background noise. Participants were faster at repeating the target word when the target word was a match rather than a

mismatch, regardless of condition. This result was expected as previous studies have demonstrated that people are slower at repeating target words that are unrelated to the preceding context (e.g., mismatches) (Holcomb, 1993; Neely, 1991). In the presence of background noise, people accurately identified match target words significantly more than mismatch target words. The behavioural results for the SPIN task supported the hypothesis that the addition of background noise would make the task more difficult.

Speech Perception in Noise and the N400

ERPs to the final words in sentences showed the predicted N400 mismatch effect, in both the quiet and noise conditions. The amplitude of the N400 mismatch effect was somewhat larger in quiet than background noise, however this difference did not reach statistical significance except in the left central ROI. The significance of this finding is unclear, because this was not where the N400 amplitude was maximal. An additional observation, which was not anticipated and so no statistical analyses were conducted, was that the onset and offset of the N400 in background noise was delayed by approximately 100 ms. This finding warrants further investigation but is outside of the scope of the current study.

The lack of widespread differences in N400 amplitude between quiet and noise conditions was surprising, as I had predicted that the N400 mismatch should be reduced in the presence of background noise. Other studies have shown that the addition of background noise reduces the N400 effect (Aydelott et al., 2006; Strauß et al., 2013). This is likely because the addition of background noise makes the task more difficult, and thus people may not perceive the mismatch targets as a mismatch. A low SNR (-1 dB) was used for the SPIN task in the current study because I wanted to ensure that accuracy

was above chance (this was the case, as demonstrated by the behavioural results), as well as to ensure that the N400 effect was still elicited (which it was). However, it appears that the low SNR of the background noise did not reduce the N400 effect as much as expected. The SNR chosen for this task may not have been low enough to induce a significant reduction in the N400 mismatch effect in comparison to when there was no background noise. This is also reflected in the behavioural data, in which the differences in accuracy between both SPIN conditions did not survive multiple comparison correction.

4.3 Correlations between Statistical Learning and Speech Perception

I hypothesized that there would be a significant correlation between the size of the AGL P600 and SPIN N400 effects. It was predicted that in the absence of background noise, as the N400 effect increased (became more negative), so should the P600 effect (more positive). The reasoning behind the hypothesis for the AGL-SPIN correlation was that higher learning scores in the AGL task reflected greater ability to use prior, relevant information to predict upcoming information – and given this ability, these same people would show a larger N400 effect to semantic violations since strong predictions about the sentence endings were made but then violated. These people may also be unconsciously expecting target words with higher Cloze probabilities, resulting in a larger N400 effect when the target word did not meet their expectations

It was hypothesized that adding background noise to the SPIN task would make the SPIN task more difficult and reduce the N400 mismatch effect. The N400 mismatch effect was expected to reduce because by making it more difficult to hear, people would rely more on the context of the sentence and top-down processing to guess what the final

word of the sentence was. Thus, if participants did not hear the final word and had to guess, they would be more likely to guess a match word, and this would reduce the N400 mismatch effect. It was expected that people with higher learning scores would have an even more reduced N400 mismatch effect, because they would be better at using the context of the sentence to predict what the final word should have been. In contrast, people with low learning scores should experience a greater N400 effect than people with high learning scores, since people with low learning scores should be less adept at using the context of the sentence to predict what the final word should have been. Instead, they may have predicted an equally wrong word, rather than the match word thus eliciting an N400 mismatch effect.

The results of the EEG data and statistical modelling provided partial support for these hypotheses. Since the predicted P600 effect was not obtained for AGL violations, clearly the predicted correlation could not be found. However, the effect that was observed for AGL — a greater negativity in the midline central ROI — did correlate with the N400 mismatch effect from the SPIN task. Interestingly, the statistical modelling for the behavioural data demonstrated that there was no significant correlation between AGL learning scores and accuracy on the SPIN task. This suggests that the effects observed in the ERP data did not translate into behaviour. This demonstrates the importance of using EEG for these types of paradigms, as EEG is capable of bringing to light how processing can be affected by experimental manipulations even in the absence of behavioural changes. Non-learners had a significant correlation between SPIN and AGL ERP effects in the quiet condition, suggesting that non-learners may have had less expectations about what the final word of the sentence should have been, this correlation was non-significant

(though still marginal) in the presence of noise. Thus, the correlation for people who were less sensitive to the transitional probabilities of the grammatical stimuli were not strongly affected by the addition of background noise. This supports the hypothesis that non-learners would still misperceive the target word when background noise was added, eliciting an N400 mismatch effect. However, learners showed the opposite trend: when there was no background noise, learners showed no significant correlation but in the presence of background noise the correlation became significant — as the N400 mismatch effect increased, so did the AGL grammaticality effect. Thus, the hypothesis that people who were more sensitive to the transitional probabilities of the AGL stimuli would be more affected by the addition of background noise was supported.

The results suggest that learners rely on sentence context to develop expectations about what the final word of the sentence for the SPIN task should be. In the absence of noise, learners had large N400 mismatch effects suggesting that they formed expectations about what the final word of the sentence should have been. However, the relationship between SPIN and AGL ERP amplitude for learners was non-significant, suggesting that the N400 mismatch effect did not change regardless of the AGL amplitude. In the presence of background noise, learners had a significant positive relationship between N400 and AGL ERP amplitudes. This relationship for learners in the presence of background noise may indicate the use of top-down processing to predict the final word of the SPIN task. In contrast, the significant positive correlation in non-learners in the absence of noise may reflect greater reliance on top-down processing even in quiet; the reduced, non-significant correlation in noise may reflect an interruption of the processing system. However, since non-learners as a group did not show a significant ERP

difference between grammatical and ungrammatical sequences in the AGL task, the correlation in this group is difficult to interpret.

4.4 Limitations of the Study

One important consideration in interpreting the results of the present study is that for the EEG tasks, all trials were analyzed regardless of whether participants were correct or not for each trial. This type of analysis is generally acceptable, as other EEG tasks investigating the ERP responses to AGL paradigms have analyzed all trials (Christiansen et al., 2012; Lang & Kotchoubey, 2000; Sanders et al., 2002). It is unlikely that the results of the SPIN task in the absence of background noise would be affected by this method of analysis, due to high accuracy. However, it is possible that the N400 effect when background noise was added to the SPIN task could be affected because accuracy was significantly lower and for some mismatch trials, participants supplied the original correct word. For example, when participants heard the following mismatch sentence with background noise, *To prevent football injury, all players must wear shoulder pads / bats*, most participants repeated the final word as the actual match word (*pads*), rather than the mismatch word (*bats*). While this is a clear example of top-down processing, it means that on that trial, the expected enhancement of the N400 would not be seen because the person "heard" a matching word rather than a mismatch. This may explain the somewhat reduced amplitude of the N400 mismatch effect in the noise condition.

Additionally, this analysis method may have affected the EEG results for the AGL task. By analyzing all trials, rather than only correct trials, sensitivity to either the P600 or N400-like ERPs may have been lost since incorrect responses may reflect a lack of sensitivity to the underlying grammar. While it could be argued that the P600 should

have been elicited regardless of accuracy, as Lang and Kotchoubey (2000) demonstrated that passive statistical learning paradigms still elicit ERP responses, it seems more likely that accuracy on the AGL task affected sensitivity to the P600 ERP. Overall accuracy for the AGL task was less than 70% (65% for learners, 50% for non-learners). Thus, the lack of a midline central P600 effect may be due to low accuracy (compared to the accuracy observed by Christiansen et al., 2012) on the AGL task. This may also explain why I observed an N400-like effect at the midline central ROI for learners; learners may have been accurate enough to develop some expectations about the grammar and when thus an N400-like effect was observed when these expectations were violated. Future studies should investigate whether high accuracy is required to elicit a P600 effect on an AGL task.

The current analysis method did not attempt to parse out these effects out and future analysis should examine the ERP responses for correct trials rather than all trials for both the SPIN and AGL tasks. This method of analysis would likely also impact the results of the correlations, and it is possible that the non-significant negative relationship between SPIN and AGL amplitudes for learners in the absence of background noise may become significantly negative. A significant positive correlation would support the hypothesis that people who are better at statistical learning are also better at using their prior knowledge about sentence context to predict incoming information.

4.5 Future Studies

The present study aimed to determine whether there was a relationship between statistical learning and speech perception in noise abilities. It was hypothesized that the AGL task would elicit a P600 in the central parietal region of the scalp, and that the P600

would be modulated by learning scores. People who were more sensitive to the transitional probabilities of the sequences were predicted to have an enhanced P600 effect in comparison to people who were less sensitive. These hypotheses were not supported. Once people were divided into learners and non-learners, there was some evidence of a P600-like effect in the anterior ROIs as well as an N400-like effect in the midline central ROI for learners. Future studies should investigate whether the absence of P600 effects elicited by the current AGL paradigm is due to either low accuracy on the AGL task (lower than Christiansen et al., 2012) or if it is due to the naivety of participants regarding the grammaticality of the sequences.

To determine whether the lack of a P600 effect in the current AGL task is due to lower accuracy than that observed by Christiansen et al. (2012), participants should receive more extensive training that what they received in the current paradigm and should perhaps use the same training method as used by Christiansen et al. (2012). Unlike the Christiansen et al. (2012) AGL paradigm though, participants should not be informed at any point that the sequences follow any type of grammar. Christiansen et al. (2012) started training with small segments of grammatical sequences (two stimuli), and incrementally lengthened the sequences until people were observing full sequences. Additionally, people observed the same sequences more than once during the training phase. This training method should be used to investigate whether the P600 effect is only elicited in an AGL task when people have extensive training, as second language studies have demonstrated that the N400 is gradually replaced by a P600 (Osterhout et al., 2006). After training, people can be tested on how well they learned the grammar using the same method as the current AGL paradigm. If overall accuracy increases and a P600 effect is

observed, this would suggest that training people until they have high levels of accuracy is necessary in order to observe a P600 effect.

An additional study should examine the differences in ERP responses between two AGL tasks: one in which participants are informed before beginning the AGL task that the sequences follow specific patterns and another in which participants are not informed about the underlying grammaticality of the sequences. This would help determine whether there are differences in the type of statistical learning abilities that are being tested by these different AGL paradigms. If a P600 is elicited in response to an AGL task where participants are aware of the grammar, this suggests that having the ability to create concrete rules about the sequences is important for eliciting P600s in response to syntactic violations. If an N400-like response similar to those observed in the current AGL paradigm and by others (Abla et al., 2008; Abla & Okanoya, 2009; Lang & Kotchoubey, 2000; Mueller et al., 2009; Osterhout et al., 2006) is found when participants are naïve to the grammar, this would suggest that an N400-like effect is found when participants are unconsciously aware of grammaticality but still generate expectations based on the statistical frequencies of the stimuli. If an N400 effect was found, this would suggest that participants need to be able to ascribe a reason to why the sequences violate the grammar in order to elicit a P600 effect.

Finally, it would be interesting to combine the methods for each of the future studies to determine whether extensive training or prior knowledge about grammar is important in order to elicit a P600 effect. This could be investigated through a between-subject paradigm. Two groups would receive extensive training on the AGL paradigm following the methods used by Christiansen et al. (2012). A third group would perform the same

AGL task as that used in the current paradigm. Those who receive minimal training should be informed beforehand that the sequences follow a grammar. One of the groups who receive extensive training should also be informed about the grammaticality of the sequences prior to beginning the task; the other group should be naïve. This type of experiment would allow researchers to elucidate whether extensive training, prior knowledge, or a combination of both is required in order to elicit a P600 effect.

4.6 Conclusions

The present study aimed to determine whether there was a relationship between an individual's sensitivity to the statistical frequencies of stimuli in their environment and the ability to predict incoming information. I found evidence that suggests that such a relationship exists in brain responses during these two tasks, especially under degraded listening conditions. However, this relationship did not appear to manifest at the level of behavioural performance. When there was no background noise, people sensitive to the transitional probabilities of the AGL task had a non-significant negative correlation between the N400 and AGL ERP amplitudes. When background noise was added, learners had a significant positive correlation between the amplitudes. These results suggest that learners were more likely relying on sentence context and top-down processing to predict what the final word of the SPIN task should be. In the absence of noise, a large N400 mismatch effect was elicited because the mismatch word violated their expectations, but in the presence of background noise, learners most likely perceived a mismatch word as the accurate match word. In contrast, the positive correlations for non-learners only marginally differed between SPIN conditions. These

results suggest that non-learners were less adept at using the context of the sentence to predict what the final word of the SPIN task should be.

REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*(6), 952–964. http://doi.org/10.1162/jocn.2008.20058

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research*, *64*(2), 185–190. http://doi.org/10.1016/j.neures.2009.02.013

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural timing is linked to speech perception in noise. *The Journal of Neuroscience*, *30*(14), 4922–4926. http://doi.org/10.1523/JNEUROSCI.0107-10.2010

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. http://doi.org/10.1111/1467-9280.00063

Aydelott, J., Dick, F., & Mills, D. L. (2006). Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology*, *43*(5), 454–464. http://doi.org/10.1111/j.1469-8986.2006.00448.x

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2001). Aetiology and clinical presentations of auditory processing disorders--a review. *Archives of Disease in Childhood*, *85*(5), 361–365. http://doi.org/10.1136/adc.85.5.361

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *47*(3-4), 296–312. http://doi.org/10.1111/j.1469-7610.2006.01611.x

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior Research Methods*, *42*(3), 665–670. http://doi.org/10.3758/BRM.42.3.665

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound* (2nd ed.). Cambridge, MA: MIT Press.

Bressler, S. L., & Ding, M. (2006). Event-Related Potentials. In M. Akay (Ed.), *Wiley Encyclopedia of Biomedical Engineering* (pp. 1–8). Hoboken: Wiley. http://doi.org/10.1016/j.psychres.2011.02.020

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. http://doi.org/10.1016/j.brainres.2012.01.055

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44. http://doi.org/10.1162/jocn.1993.5.1.34

Cansino, S., & Tellez-Alanis, B. (2000). ERPs elicited by a cognitive incongruity paradigm: A semantic memory study. *Neuroreport*, *11*(5), 977–981.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*, 572–582.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*(2), B69–77. http://doi.org/10.1016/s0010-0277(02)00233-0

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *27*(2), 231–256. http://doi.org/10.1080/01690965.2011.606666

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371. http://doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39. http://doi.org/10.1037/0278-7393.31.1.24

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and across modalities: Abstract versus Stimulus-Specific Representations. *Psychological Science*, *17*(10), 905–912.

Conway, C. M., Ellefson, M. R., & Christiansen, M. H. (2003). When less is less and when less is more: Starting small with staged input. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 270–275). Mahwah, NJ.

Conway, C. M., Karpicke, J., & Pisoni, D. B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of Deaf Studies and Deaf Education*, *12*(3), 317–334. http://doi.org/10.1093/deafed/enm019

Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*, *14*(1), 69–82. http://doi.org/10.1111/j.1467-7687.2010.00960.x

Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, *120*(5), 2421–2424. http://doi.org/10.1121/1.2229005

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*(1), 21–58. http://doi.org/10.1080/016909698386582

Coulson, S., & Van Petten, C. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968. http://doi.org/10.3758/BF03195780

Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2012). The N400 and Late Positive Complex (LPC) effects reflect controlled rather than automatic mechanisms of sentence processing. *Brain Sciences*, *2*(4), 267–297. http://doi.org/10.3390/brainsci2030267

Damasio, A. R. (1991). Signs of aphasia. In M. Sarno (Ed.), *Acquired aphasia* (2nd ed., pp. 27–43). San Diego, CA: Academic Press.

Damasio, A. R. (1992). Aphasia. *The New England Journal of Medicine*, *326*(8), 531–539. http://doi.org/10.1056/NEJM199202203260806

Damasio, H. (1991). Neuroanatomical correlates of the aphasias. In M. Sarno (Ed.), *Acquired aphasia* (2nd ed., pp. 45 – 71). San Diego, CA: Academic Press.

De Weirdt, W. (1988). Speech perception and frequency discrimination in good and poor readers. *Applied Psycholinguistics*, *9*(2), 163–183. http://doi.org/10.1017/S0142716400006792

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. http://doi.org/10.1016/j.jneumeth.2003.10.009

Dodge, Y. (2006). REML. In *The Oxford Dictionary of Statistical Terms* (p. 512). Oxford, UK: Oxford University Press.

Elliott, L. L. (1995). Verbal auditory closure and the speech perception in noise (SPIN) Test. *Journal of Speech, Language, and Hearing Research*, *38*(6), 1363–1376.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*(1), 75–84. http://doi.org/10.1016/j.brainres.2006.06.101

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. http://doi.org/10.1037/0278-7393.28.3.458

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, *31*(1), 45–63. http://doi.org/10.1023/A:1014376204525

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. http://doi.org/10.1016/j.tics.2014.12.010

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*(7), 1243–1252. http://doi.org/10.1177/0956797612472207

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436. http://doi.org/10.1111/1467-9280.00476

Goodglass, H. (1993). *Understanding aphasia*. San Diego, CA: Academic Press.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The assessment of aphasia and related disorders* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531–1549.

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews*, *6*(3), 121–143. http://doi.org/10.1016/j.plrev.2009.06.001

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience*, *8*(5), 393–402. http://doi.org/10.1038/nrn2113

Holcomb, P. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61. http://doi.org/10.1111/j.1469-8986.1993.tb03204.x

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4), 411–430. http://doi.org/10.1016/S0893-6080(00)00026-5

Jarvis, B. G. (2012). DirectRT (Version 2014.1.114). New York, NY: Empirisoft Corporation.

Jay, T. B. (2003). *The Psychology of Language*. (L. Jewell, Ed.) (1st ed.). Upper Saddle River, New Jersey: Prentice Hall.

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, *4*(2), 99–107. http://doi.org/10.1016/0926-6410(96)00022-5

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. http://doi.org/10.1080/016909600386084

Kaan, E., & Swaab, T. Y. (2003a). Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations. *Cognitive Brain Research*, *17*(3), 621–635. http://doi.org/10.1016/S0926-6410(03)00175-7

Kaan, E., & Swaab, T. Y. (2003b). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98–110. http://doi.org/10.1162/089892903321107855

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351. http://doi.org/10.1121/1.381436

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*(3), 321–340. http://doi.org/10.1016/j.cognition.2010.05.011

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, *100*(3), 223–237. http://doi.org/10.1016/j.bandl.2005.12.006

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. http://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Lang, S., & Kotchoubey, B. (2000). Learning effects on event-related brain potentials. *Neuroreport*, *11*(15), 3327–3331.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Brian, C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, *103*(49), 18866–18869.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36. http://doi.org/10.1097/00003446-199802000-00001

Luria, A. R. (1966). *Higher cortical functions in man* (1st ed.). Oxford, UK: Basic Books Inc.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*. http://doi.org/10.1016/0010-0277(80)90015-3

McBride-Chang, C. (1995). Phonological processing, speech perception, and reading disability: An integrative review. *Educational Psychologist*, *30*(3), 109–121.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. http://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369. http://doi.org/10.1016/j.tics.2006.06.007

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, *485*(7397), 233–236. http://doi.org/10.1038/nature11020

Meulman, N., Wieling, M., Sprenger, S. A., Stowe, L. A., & Schmid, M. S. (2014). Age effects in L2 grammar processing as revealed by ERPs and how (not) to study them. In *6th Annual Meeting of the Society for the Neurobiology of Language* (p. 29). Amsterdam.

Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, *41*(5), 329–335. http://doi.org/10.1037/h0062491

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*(31), 1–9. http://doi.org/10.3389/fpsyg.2010.00031

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, *22*(2), 205–220. http://doi.org/10.1016/j.cogbrainres.2004.08.010

Mueller, J. L., Oberecker, R., & Friederici, A. D. (2009). Syntactic learning by mere exposure - An ERP study in adult learners. *BMC Neuroscience*, *10*(1), 89. http://doi.org/10.1186/1471-2202-10-89

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (1st ed., pp. 264–336). Hillsdale, NJ: Erlbaum.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, *24*(2006), 1205–1223. http://doi.org/10.1162/jocn_a_00143

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., … Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, *29*(49), 15564–15574. http://doi.org/10.1523/JNEUROSCI.3065-09.2009

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*(8), 1212–1228. http://doi.org/10.1162/0898929055002436

Onnis, L., Christiansen, M. H., Chater, N., & Gómez, R. L. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886 – 891). Mahwah, NJ.

Onnis, L., Farmer, T. A., Baroni, M., Christiansen, M. H., & Spivey, M. (2008). Generalizable distributional regularities aid fluent language processing: The case of semantic valence tendencies. *Italian Journal of Linguistics*, *20*, 125–152.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and Event-Related Potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199–230. http://doi.org/10.1111/j.1467-9922.2006.00361.x

Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, *103*(1), 577–587. http://doi.org/10.1121/1.421128

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608. http://doi.org/10.1121/1.412282

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://doi.org/http://www.R-project.org/

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*, *125*(2), 123–138.

Rintelmann, W. F. (1985). Monaural speech tests in the detection of central auditory disorders. In M. L. Pinheiro & F. E. Musiek (Eds.), *Assessment of central auditory dysfunction: Foundations and clinical correlates* (pp. 173–200). Baltimore, MD: Williams & Wilkins.

Robertson, E. M. (2007). The serial reaction time task: Implicit motor skill learning? *Journal of Neuroscience*, *27*(38), 10073–10075. http://doi.org/10.1523/JNEUROSCI.2747-07.2007

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. http://doi.org/10.1126/science.274.5294.1926

Sanders, L. D., Ameral, V., & Sayles, K. (2009). Event-related potentials index segmentation of nonsense sounds. *Neuropsychologia*, *47*(4), 1183–1186. http://doi.org/10.1016/j.neuropsychologia.2008.11.005

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience*, *5*(7), 700–703. http://doi.org/10.1038/nn873

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *298*(1089), 199–209. http://doi.org/10.1098/rstb.1982.0082

Sörqvist, P., & Rönnberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: What is the role for working memory capacity? *Journal of Speech, Language, and Hearing Research*, *55*(1), 210–218. http://doi.org/10.1044/1092-4388(2011/10-0353)

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: Revisiting the N400. *Journal of Cognitive Neuroscience*, *25*(8), 1383–1395. http://doi.org/10.1162/jocn_a_00389

Stuart, A. (2008). Reception thresholds for sentences in quiet, continuous moise, and interrupted noise in school-age children. *Journal of the American Academy of Audiology*, *19*(2), 135–146. http://doi.org/10.3766/jaaa.19.2.4

Stuart, A., Givens, G. D., Walker, L. J., & Elangovan, S. (2006). Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise. *The Journal of the Acoustical Society of America*, *119*(4), 1946–1949. http://doi.org/10.1121/1.2178700

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *335*(1273), 71–78. http://doi.org/10.1098/rstb.1992.0009

Tallal, P., Stark, R., Kallman, C., & Mellits, D. (1981). A reexamination of some nonverbal perceptual abilities of language-impaired and normal children as a function of age and sensory modality. *Journal of Speech, Language, and Hearing Research*, *24*(3), 351–357.

Tremblay, A., & Ransijn, J. (2013). LMERConvenienceFunctions: A suite of functions to back-fit fixed effects and forward-fit random effects, as well as other miscellaneous functions.

Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, *6*(2), 302–324. http://doi.org/10.1075/ml.6.2.04tre

Tunney, R. J., & Altmann, G. T. M. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1322–1333.

Van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, *3*(5), 1211–1224. http://doi.org/10.1111/j.1749-818X.2009.00163.x

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*. http://doi.org/10.1126/science.167.3917.392

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology*, *41*(1), 48–61. http://doi.org/10.1037/h0084150

Wood, S. N., Shaw, S., & Goude, Y. (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *64*(1), 139–155. http://doi.org/10.1111/rssc.12068

Yost, W. (2003). Audition. In I. B. Weiner, A. F. Healy, & R. W. Proctor (Eds.), *Handbook of psychology: Experimental psychology* (2nd ed., p. 130). Hoboken, NJ: John Wiley & Sons Ltd.

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732–745. http://doi.org/10.1111/j.1467-7687.2009.00817.x