8-19-2024 9:00 AM

# Harnessing Implicit Learning to Support the Discovery of Second Language Phoneme Patterns in Adult Learners

Émilie R. Hoeppner,

Supervisor: Batterink, Laura J., *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Neuroscience
© Émilie R. Hoeppner 2024

## Recommended Citation

Hoeppner, Émilie R., "Harnessing Implicit Learning to Support the Discovery of Second Language Phoneme Patterns in Adult Learners" (2024). *Electronic Thesis and Dissertation Repository*. 10351.
https://ir.lib.uwo.ca/etd/10351

# Abstract

Statistical learning studies have focused primarily on artificial languages, despite having an overall goal of providing insight into how we learn natural languages. Here, we investigate the impact of statistical learning on processing phonemic regularities within a second language in adult learners. Participants passively listened to Italian (L2 exposure group) or English (control group) podcasts for three weeks and completed a word rating task with ERP recorded before and after this listening period. Behaviourally, the L2 exposure group showed a nonsignificant trend towards increased sensitivity to phonotactic probability over the three weeks. At the ERP level, only the L2 group showed a significant change in nonword processing from session 1 to session 2, with significantly greater ERP negativity from 300-800ms to nonwords compared to words at session 2. These results provide preliminary indications that statistical learning may be leveraged to learn phonemic regularities in natural language, extending artificial language research.

**Keywords**: Statistical Learning, Second Language Acquisition, Phonotactic Probability, Word Frequency, EEG, ERP

# Summary for Lay Audience

This thesis delves into how we can pick up a new language without actively trying. Traditionally, research in second language (L2) learning has emphasized explicit methods, where learners consciously study vocabulary and grammar rules. However, recent studies suggest that our brains possess the ability to passively absorb types of linguistic patterns through implicit learning akin to how we naturally acquire our first language.

In this study, I explore the impact of implicit learning on the processing of L2 sound patterns in adult learners. Participants were assigned to either the L2 exposure group, who passively listened to Italian podcasts daily over three weeks, or the control group, who listened to English podcasts. Before and after the three-week listening period, participants engaged in a word rating task while their brainwave activity was recorded. Specifically, they rated how confident they were that the word they heard was a real Italian word. I hypothesized that those exposed to Italian would show improvements in differentiating between Italian words and nonwords after the three weeks.

Following the listening period, the L2 exposure group demonstrated an increased sensitivity to L2 sound patterns compared to the control group. Test items that contained more probable sound sequences were rated higher than test items with lower probability sequences; however, this trend was not significant. In contrast, participants in the control group performed significantly worse on the task after the listening period.

Brainwave recordings revealed significant differences in brain activity between words and nonwords in the L2 exposure group before and after exposure, where the nonwords and words elicited different responses at post-test, but not at pre-test. This result indicates that even without intentional learning, prior experience with a language’s pattern influences neural processing of linguistic stimuli, demonstrating that the brain can pick up on sound regularities in an unfamiliar language just through listening.

These findings shed light on the potential utility of passive exposure to facilitate language learning in adult learners. Understanding how our brains rapidly adapt to new linguistic environments helps demonstrate the importance of immersive language experiences in second language acquisition, informing new second language training approaches.

# Co-Authorship Statement

The larger research project described in this thesis was conducted in partnership with Amiya Aggarwal. The experimental design of the exposure phase and podcast selection was done together, and she computed word frequency independently.

Data analysis and writing found in this thesis were done independently by me, with feedback from Dr. Laura Batterink.

# Acknowledgements

I wish to sincerely thank my supervisor Dr. Laura Batterink. Your support and guidance have been instrumental in my development not only as a grad student but also as a researcher. Your assistance in teaching me how to conduct thorough research and emphasizing attention to detail has been invaluable. Additionally, thank you for sharing laughs over my bad jokes and memes; your ability to find humour in our interactions has made the journey all the more enjoyable.

Thank you to my committee for your invaluable insight and feedback: Dr. Ken McRae, Dr. Deanna Friesen and Dr. Paul Gribble.

A special thanks to the News in Slow Italian Podcast team for providing us with free access to their podcasts and corresponding transcripts.

Thank you to all of my fellow CNLL Lab members. To Aaron and Priya, thank you for your advice and insight. Thank you to my research assistants, Chelsea, Shveta and Maream; without you, data collection would not have been as efficient, nor would have the year been as fun. I would also like to extend a special thanks to my lab partner, Amiya. I truly could not think of a better person to go through this journey with. I will always cherish our brainstorming sessions and laughter-filled travels. Your sense of humour and constant snacks carried me through the stressful periods.

Thank you to my friends, both near and far. Kiera and Michael, for your abilities to find humour in any situation and your willingness to facetime at any hour of the day or night. Mahdiyar, Bruno, and Mehrdad for always lending an ear and book, movie and show recommendations, that I promise I will eventually read and watch. To the rest of my London family: the 1 pm lunch club, the trivia losers, and my DnD groups, as well as my many other friends, thank you for welcoming and embracing me.

A very special thanks to the musical artists that have been the soundtrack to the last 2 years of studying and writing: John Powell, Marcus Wagner, and La Bottine Sourriante.

To my family, Jim, Joanne, and Sarah, *merci*. Your support as I left home to follow my calling in neuroscience has meant the world to me. The care packages, recipes, cat pictures, and family gossip make being away from home just a little bit easier and that much harder.

Lastly, I would like to thank my true partner, Mackenzie. Thank you for listening to me talk about statistical learning for many, many hours and for being my biggest cheerleader, dance moves and all.

## Table of Contents

Contents

## List of Tables

## List of Figures

# List of Appendices

# 1. Introduction

Many adults face the challenge of learning a second language (L2) later in life. For example, immigrants who move to a country where the local language is unfamiliar must quickly learn the new language to integrate into society. However, acquiring a new language as an adult can be a challenging and time-consuming endeavour. Unlike children, who have the time and specific environment (school) in which to immerse themselves in language learning, adults often have other commitments that limit their time and attention. Traditional language learning methods, such as classroom instruction, are time-intensive and may not always be feasible. This creates a pressing need for accessible and cost-effective solutions that can seamlessly integrate into adults' daily lives. Implicit learning, specifically statistical learning, offers a relatively underexplored avenue that could be exploited to supplement more traditional explicit language learning approaches.

## 1.1 Statistical Learning and its Role in Language

A key challenge faced by language learners is the identification of words in continuous speech, also known as speech or word segmentation. When listening to a continuous speech stream, one does not inherently know when a word starts and ends. In natural, fluent speech, there is no single, reliable cue to indicate word boundaries, such as pauses or changes in intonation (Field, 2003). This absence of reliable cues to word onsets often leads L2 learners to struggle to identify word boundaries and to properly segment, and therefore understand, speech. This inability to identify words within a stream creates the perception that the speech rate is faster than the native language, despite speech rates largely remaining constant across languages (Pellegrino & Coupé, 2011).

Statistical learning provides a potential solution to the speech segmentation problem. Statistical learning is the process by which patterns in the environment are detected and learned without conscious effort or feedback (Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). The first study investigating statistical learning in speech established the standard paradigm for laboratory statistical learning studies: the “triplet learning paradigm” (Saffran, Aslin et al., 1996). Participants encounter a sequence of stimuli organized into repeating triplets that co-occur consistently over time, with *transitional probabilities* cuing word boundaries. Transitional probability refers to the likelihood of a specific event or item occurring given the occurrence of a previous event or item within a sequence or context. In the context of speech segmentation, it often is used to describe the probability of one syllable occurring given the occurrence of a preceding syllable. Relatively lower transitional probabilities between syllables may signal a potential word boundary. In the context of auditory language studies, these triplet "words" consist of syllables from an artificial language, such as "*ba fu ro*" and "*ko ga tu*" (Saffran, Aslin, et al., 1996). In this example*, ba* and *fu* have a high transitional probability, while *ro* and *ko* would have a low transitional probability, cuing a potential word boundary. After exposure to a stream of artificial language triplets, infants exhibited a looking-time preference for part-words or nonwords, compared to words from the stream. This finding suggests they had gained sensitivity to the transitional probabilities of the stream as transitional probability was the only cue to word boundaries (Saffran, Aslan et al., 1996).

Children and adults have demonstrated the capacity to use statistical learning similarly to infants (Choi et al., 2020; Moreau et al., 2022; Saffran et al., 1997; Saffran, Newport, et al., 1996). Replicating the infant experiment conducted by Saffran, Aslan, and Newport (1996) with adults yielded compelling evidence of statistical learning in adult learners (Saffran, Aslin, et al, 1996). Following exposure to a continuous speech stream, adults demonstrated the ability to distinguish between words and non-words in a 2 Alternative Force-Choice recognition task that required them to discriminate between words and foils (recombined syllables presented in a novel order). This above-chance performance indicated their ability to extract linguistic patterns similarly to infants, providing early evidence that adults retain the same mechanism from infancy for pattern discovery.

Importantly, participants do not need to pay explicit attention to the stimuli to learn the speech stream's regularities. Saffran and colleagues (1997) found that participants learned the words of an unsegmented artificial language speech stream despite engaging in a distractor task at the same time. Participants completed an online graphic design task while the speech stream played in the background. Following this task, participants performed significantly above chance level at identifying the words from the stream, indicating that statistical learning can occur in the absence of focused attention.

Statistical learning is not restricted to syllable and word learning; studies have observed similarities in how patterns are learned across visual and auditory domains. Similar to how infants can detect statistical regularities in tone sequences from a continuous stream (Saffran et al., 1999), they can also pick up on regularities in visually presented scenes of pairs of symbols, paying more attention to more frequently occurring pairs (Fiser & Aslin, 2002). Conway & Christiansen (2005) expanded on this finding, demonstrating that statistical learning can occur across tactile, visual, and auditory stimuli. While statistical learning mechanisms may operate similarly across different cognitive domains at a high level, there are still modality-specific nuances in how these mechanisms are implemented at a lower level (Frost et al., 2015). This thesis will focus on statistical learning in the domain of speech.

As outlined in the above-mentioned papers, much research has focused on participants' ability to learn miniature, highly artificial languages in a single laboratory session (Henin et al., 2021; Moreau et al., 2022; Saffran, Aslin, et al., 1996; Saffran, Newport, et al., 1996). However, while the triplet experimental paradigm has been instrumental in elucidating the mechanisms of statistical learning, the use of artificial languages—lacking semantic meaning, grammar, and sentence structure—has limitations in studying the processing and acquisition of natural language patterns. Therefore, further research into how statistical learning can scale up to support natural language learning is crucial.

## 1.2 Phonemes and Phonotactic Patterns in Second Language Learning

A key aspect of language is learning the sounds (or phonemes) that are specific to each language, along with their temporal regularities, which govern how sounds can be combined. Phonemes are the fundamental building blocks of language and represent the smallest unit of speech. Represented by unique characters in the International Phonetic Alphabet (IPA), phonemes encompass the diverse sounds utilized in language production. Phonemes are combined to make morphemes, which are the smallest *meaningful* unit of speech. These are then combined to make words (Morais & Kolinsky, 1994). For instance, the word 'balloon' comprises six phonemes (/bəlʉ́wn/), making up only one morpheme ('ball-'), as the '-oon' segment lacks independent meaning.

The focus of this study is on changes in sensitivity to phoneme sequences, as opposed to the distributional probability of the individual phonemes. How phonemes are combined exhibits language-specific patterns, contributing to the distinct phonemic structures of individual languages. Phonotactic probability quantifies the frequency of occurrence of a specific phoneme or phoneme combinations within a language (Storkel et al., 2006). Similar to how transitional probabilities refer to the probability of a syllable occurring based on the preceding one, phonotactic probability is calculated based on the likelihood of a phoneme occurring given the context of previously encountered phonemes. Higher phonotactic probability indicates a greater likelihood of the phoneme occurring, rendering the word more representative of the language's typical patterns. Phonotactic regularities have been shown to be useful for speech segmentation in native listeners (McQueen, 1998; Yip, 2020) and L2 speakers alike (Katayama, 2022). Thus, just as learners in artificial language experiments have been shown to use transitional probabilities to discover words in the continuous speech stream, learners of a natural language may use phonotactic probabilities to discover word boundaries in natural conversation. Similar to learners in artificial language experiments, L2 learners may also be able to simply listen to the L2 to learn the language’s phonotactic regularities, which may in turn support early word form learning. This idea is supported by several studies showing that passive exposure alone can support aspects of phonemic learning (learning phonemes and phoneme patterns) in adult L2 learners, as follows in the section below.

### 1.3 Background Exposure Facilitates Learning of the L2

Ambient exposure to a second language has been shown to be sufficient to detect its phonemic regularities. A study conducted with English monolinguals in New Zealand exemplifies the impact of passive lifetime exposure to a language. In this study, participants who had never explicitly engaged in learning the Māori language but were regularly exposed to it in everyday contexts developed a Māori proto-lexicon—a repertoire of known words with no accompanying semantic knowledge (Oh et al., 2020). In one task, participants were presented with written Māori words and nonwords ranging on phonotactic probability and frequency of occurrence in Māori and were asked to rate their confidence that the word was a real Māori word. Participants reliably distinguished between real Māori words and non-words, rating words higher than nonwords across phonotactic probabilities. Furthermore, participants demonstrated a basic sensitivity to Māori phonology, as reflected in higher ratings of high phonotactic probability words than low phonotactic probability words, across word frequency. The control group made up of participants from the United States, with no exposure to Māori, showed no sensitivity to Māori phonotactic probability, establishing that it was background exposure driving the effect rather than some intrinsic sensitivity to the Māori statistics in the New Zealand group.

A second experiment with a different group of participants evaluated participants' ratings of Māori nonwords to investigate participants' knowledge of what phonemes and phoneme combinations made up the Māori language (Oh et al., 2020). The authors found that participants' ratings could be predicted by the words' phonotactic probabilities, and their ratings did not differ significantly from native Māori speakers. This further established a sensitivity of Māori phonotactics independently of semantic knowledge of the language, due to the absence of real words presented in this experiment. Panther and colleagues (2023) extended this research by establishing a direct link between proto-lexicon size and phonotactic knowledge.

Further investigations into ambient exposure reveal its remarkable capacity to facilitate incidental language learning even over relatively short timeframes. Kittleson and colleagues (2010) demonstrated that just three weeks of ambient exposure may lead to significant improvements in language processing. Non-Norwegian speaking international university students took part in a word identification task at two time points, separated by three weeks of ambient exposure to Norwegian while living in Norway. They were exposed to short sentences containing a target word in Norwegian and then were asked to identify whether individual test items were true words or nonwords in Norwegian. Following the exposure period, participants exhibited enhanced abilities to correctly reject nonwords. Nonwords were created by either combining a syllable from a target word with a syllable from an adjacent bisyllabic word within the familiarization sentence or were two monosyllabic words from the familiarization sentences combined. Although there was no control group included, this improvement suggests that three weeks of exposure to the Norwegian language environment resulted in an improvement in participants' abilities to correctly reject nonwords, providing evidence that participants showed an increase in sensitivity to characteristic Norwegian sound sequences.

More recently, Alexander and colleagues (2022) provided further evidence of the efficacy of short-term ambient exposure in language learning. In this study, participants completed a pre- and post-testing session separated by a listening period of 14 days, during which they listened to Italian podcasts for one hour a day. A control group listened to English podcasts for the same duration. During testing sessions, they took part in an exposure and word detection task, in which they had to identify predetermined target words within an Italian speech stream. This task also contained multiple repetitions of so-called "trained" words for the subsequent word familiarity task. In this word familiarity task, participants rated their familiarity of the auditorily presented Italian words, which consisted of the "trained" words, "nontrained" words that were not found in the word detection task and nonwords, which did not exist in Italian. While the Italian and control groups provided similar scores for both the nonword foils and the trained words, only the L2 exposure group showed a gain in sensitivity to the nontrained words after the three-week listening period, suggesting a gain in sensitivity to sound features of Italian words. These findings again suggest that adult learners can extract relevant speech patterns from a natural language simply by consistently listening to a new language.

Taken together, these studies highlight the potential role of ambient, incidental exposure to an L2 in facilitating the learning of certain sound-based aspects of language, offering promising avenues for adult learners seeking to enhance their language proficiency.

### 1.4 Neuroimaging Research in Statistical Learning and Second Languages

Complementing behavioural research, neuroimaging research has proven invaluable in understanding the neural mechanisms of language learning. Given the dynamic and fast-paced nature of language processing, EEG is a common method used to investigate second language learning (Luk et al., 2020). EEG's high temporal resolution is useful in investigating temporal aspects of language processing. EEG enables the collection of continuous electrical activity recorded at the scalp, from which time-locked event-related potentials (ERPs) can be extracted to reflect brain responses to specific stimuli (McWeeny & Norton, 2020). These ERPs are then analyzed for components, known waveforms associated with different perceptual and cognitive processes.

Of relevance to the current study, violations of phonological expectations have been observed to elicit an ERP component known as the Phonological Mapping Negativity (PMN), previously known as the Phonological Mismatch Negativity (Lewendon et al., 2020). The PMN is characterized by a negative deflection reaching its peak around 300ms following stimulus onset, observed across frontocentral and centroparietal electrodes (Connolly & Phillips, 1994; Lee et al., 2012). It is elicited by violations of pre-lexical phonological expectations and processing (Desroches et al., 2009; Lewendon et al., 2020), even within second languages (Desroches et al., 2022). Classic research by Connolly and Phillips (1994) revealed that the PMN is elicited when the closing word of a sentence is phonologically unexpected but semantically congruent, such as in the sentence "*The pig wallowed in the **pen***," where mud would be expected. This result shows that when a sentence-final word starts with a phoneme that does not match the listener's expectations, a more negative PMN waveform is elicited, even if the word fits semantically within the sentence. Importantly, the PMN has been found to be dissociated from semantic expectations. Connolly and colleagues (2001) demonstrated that manipulations of lexicality did not affect the PMN, distinguishing it from the N400 component associated with semantic processing. In contrast to the PMN, the N400 component is elicited by semantically unexpected but phonologically congruent sentences, such as "*The gambler had a streak of bad* **luggage,**" where luck would be expected (Connolly & Phillips, 1994). In a study by Lee and colleagues (2012), participants were instructed to indicate if the second syllable within an auditorily presented nonsense syllable pair was expected or unexpected. After participants were trained on correct pairings, the PMN was evoked to unexpected second syllables. As no semantic meaning was associated with the nonword syllable pairs, the PMN was elicited in the absence of an N400 response. This finding demonstrates that the PMN can be elicited outside of sentence contexts and is functionally distinct from the N400.

An additional component that has been related to phonological processing is the N100, which has also been specifically tied to statistical learning and early language acquisition (Sanders et al., 2002; Tong et al., 2020). The N100, also known as the N1, is characterized by a negative peak occurring between 80ms and 120ms following stimulus onset at frontal and central electrodes (Heidlmayr et al., 2021). Primarily associated with bottom-up auditory processing, the N100 is believed to reflect early perceptual encoding with minimal top-down influences (Luck, 2023). In the field of statistical learning, Sanders and colleagues (2002) demonstrated that after explicitly learning a nonsense language, participants who exhibited higher accuracy in a behavioural word discrimination task exhibited better neural word segmentation, as indexed by increased N100 amplitudes to word onsets embedded in a continuous speech stream of the nonsense words. Additionally, listen-and-repeat paradigms, in which participants hear a word and then must voice it themselves, have been shown to reduce N100 latency for words that were voiced, suggesting improved phoneme perception leading to a shortened latency (Saloranta et al., 2020). The N100 can also index phoneme learning in L2 learners. Heidlmayr and colleagues (2021) observed heightened N100 responses to phonologically incongruent words within sentences in French-English bilinguals who acquired English later in life. For example, substituting "ship" with "sheep" in the sentence "The anchor of the *ship* was let down" elicited N100 responses, indicating increased neural processing due to deviations from phonological expectations. This suggests that listeners had internalized the phonological rules of English and that changes in the N100 reflect this learning.

Past studies have also linked the P200 component in the processing of phonotactic probability and phonotactic frequency. The P200, also referred to as the P2, is a positive going waveform, peaking approximately 200ms following stimulus onset within anterior electrodes (Luck, 2023). Although primarily associated with acoustic properties of sound and language (Remijn et al., 2014), Hunter (2013) found evidence that it also reflects phonotactic probability processing. Words with high probability initial phonemes produced shorter P200 latencies, suggesting a facilitating effect of phonotactic probability. Regarding amplitude, when passively listening to a speech stream following training within an L2, L2 learners showed a larger P200 to legal pseudowords—words that followed phonotactic patterns of the L2—than illegal pseudowords, words that did not follow the L2 phonotactic patterns (Rossi et al., 2013). These results suggest that training through passive exposure to an L2 facilitates phonotactic probability learning, as indexed by increased directed attention to the familiar sequences.

One final ERP component worth delving into is the N400, a negative-going waveform occurring approximately 200 to 600ms post-stimulus that serves as a robust indicator of semantic and lexical processing (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980, 1984;

Luck, 2023). The N400 is larger in amplitude when words are encountered that violate semantic expectations, compared to words that are semantically expected, such as in the sentence "Don't touch the wet *dog*", where *paint* would be more semantically expected (Kutas & Hillyard, 1984). Words that were unexpected, but semantically similar to the target word elicited smaller N400s, bolstering its role in semantic processing. For example, "He liked lemon and sugar in his *coffee*", where *tea* would be a better fit, would result in an N400 larger than if the sentence finished with *coffee*, but still smaller than to a semantically unrelated word, demonstrating the high sensitivity of the N400. However, the N400 has also been observed in several other aspects of language processing. Outside of semantics, MEG evidence suggests that the N400 may be sensitive to phonological expectations. Dobel and colleagues (2009) conducted a magnetoencephalography (MEG) study with German participants, in which participants engaged in a semantic picture priming procedure with words and nonwords, with the nonwords distinguished by a novel phoneme from a different language. Participants completed this priming procedure both before and after a 5-day training period, during which they were exposed to both words and nonwords. The authors found increased N400 amplitudes in response to non-native novel phonemes post-training. This change in amplitude following training suggests enhanced processing of the new phoneme, indicating successful learning and integration.

The N400 has also been shown to be sensitive to word learning even without semantic contexts. Explicit language training has demonstrated increased N400 responses to nonsense words when segmenting speech streams following training, indicating word learning (Sanders et al., 2002). This demonstrated that the N400 is also sensitive to word segmentation, outside of any influence of semantics. Even a time frame as short as only 14 hours of classroom instruction can induce neural changes to pseudoword (nonword) processing (McLaughlin et al., 2004). Following 14 hours of learning, when presented with word pairs, L2 pseudowords in the second position elected larger negative-going N400s than both the related and unrelated word pairs. This indicated that the participants had learnt some aspects of the languages as the pseudowords were more surprising than they had been before classroom instruction. Finally, considering only pseudowords, Rossi and colleagues (2013) found that trained pseudowords showed a smaller N400 post-training than untrained pseudowords, both when they followed the rules of the L2 and when they did not. This suggested a whole-word familiarization effect rather than a learning effect of phonotactic sequences. Therefore, the N400 may not be sensitive to more complex regularities, such as phonemic sequences, but may instead reflect whole-word learning.

In summary, EEG research has provided valuable insights into the neural mechanisms underlying statistical learning and language processing. This research has revealed distinct components such as the N100, P200, PMN, and N400, which reflect various stages of phonological and semantic processing. These neuroimaging insights lay the groundwork for our current study, which aims to explore how adult second language learners acquire a new language through passive exposure in a naturalistic setting,

### 1.5 The Current Study

This study seeks to investigate whether statistical learning scales up to support natural language acquisition, extending findings from artificial language research. To address this question, English monolinguals were passively exposed to Italian for one hour daily over three weeks. A control condition involved participants listening to English podcasts instead over the same period. Podcasts were chosen as the exposure medium to emulate the traditional auditory statistical learning methodology of continuous exposure streams. Additionally, they offer a clear and continuous stream of input and do not have instrumentation. Before and after this exposure period, participants completed an auditory word rating task, in which they were presented with Italian words and nonwords, varying in phonotactic probability and frequency, and had to rate their confidence that the word they heard was a real Italian word. EEG responses were recorded throughout this task, offering a comprehensive examination of the neural mechanisms underlying second language acquisition as a result of passive exposure to L2 speech. I hypothesized that the L2 exposure group (participants exposed to Italian) would show increased sensitivity to phonotactic probability in the word-rating task, post-exposure, as indicated by higher ratings for high probability words than low probability words. Additionally, I hypothesized that following the exposure period, the L2 exposure group would also show greater negativity to nonwords when compared to words as revealed by the PMN and N400, but greater positivity within the P200 to words compared to nonwords. No such difference was expected to be observed in the control group.

# 2. Methods

## 2.1 Participants

A total of 57 participants were initially recruited from the Western student population and surrounding London, ON community. Of these 57 participants, 4 participants dropped out of the study during the listening period and informed us of their unwillingness to continue before session 2, 7 were removed due to non-compliance with the experimental protocol, and two participants' data was removed due to technical issues. Non-compliance was defined as failure to complete the daily surveys associated with each podcast more than 2 days in a row, or having a gap larger than 4 days between receiving the daily podcast and listening completion. This resulted in a final sample of 44 participants who completed both sessions 1 and 2 (mean age = 22.9, SD = 2.7; 28 women, 16 men). Within this sample, 42 were righthanded and 2 were lefthanded. All participants were between 17 and 35 years old, were fluent in English, and reported no previous exposure to Italian. Additionally, participants reported no fluency in any other second language, nor any significant previous classroom or other experience with Romance languages, including French, Spanish, Italian, Romanian, and Portuguese. For French, lack of significant experience was defined as not having taken any French language classes beyond Grade 9, as mandated by Canadian education requirements. None of the participants reported any history of learning, hearing, or neurological disorders, and all reported normal or corrected-to-normal vision and hearing. A table with all demographic information can be found in Table 7 in Appendix D. Participants were not informed of the research hypothesis until after completing the study. Participants were compensated $40 for each testing session, and $7.50 for each day of podcast listening during the exposure period. This resulted in a total of $240 for the full completion of the study. Ethics approval was obtained from the Western University Research Ethics Board.

## 2.2 Stimuli

### *2.2.1 Daily Exposure Podcasts*

Daily exposure podcasts for the L2 Exposure group were composed of excerpts from the podcast "News in Slow Italian," a single-speaker podcast providing detailed coverage of recent news events in Italian. Excerpts were carefully selected to exclude any English words or easily identifiable names, such as those of politicians or media figures. The podcasts featured an equal representation of both female and male speakers, with a total of five different voices. Podcasts were sped up to more closely resemble typical, clear, Italian speech as the original files held Italian speech that was slower than typical. In Audacity, the excerpts underwent a 10 percent increase tempo change, to refrain from changing the pitch. A native Italian speaker listened to excerpts and indicated that they did resemble native Italian speech. For the control group, daily exposure podcasts were composed of English news podcasts, including "Newsworthy," and "Times the Brief," which were chosen due to being single-speaker podcasts that reported current events, which matched our Italian Exposure podcasts. Due to a lack of single-speaker female news podcasts, “The Lazy Genius” podcast was included to ensure female representation. This podcast talked about tips and tricks to improve productivity and productiveness in life, matching the factual approach of the news podcasts and staying away from a narrative structure. Since this podcast remained factual and was not a narrative story, it was deemed complimentary enough to the news podcast. In both languages, each individual exposure podcast was created by concatenating excerpts from different episodes to achieve a duration of approximately one hour (mean = 60.49 minutes, range = 53.57 - 66.46 minutes). All podcasts in both languages were edited to contain three to five embedded "hidden" English words, which were played after the sound of a chime and had to be reported in a Qualtrics survey following each day’s listening session.

### *2.2.2 Stimuli in Word Rating Task*

**2.2.2.1 Word Selection.** Words selected for the Word Rating task were extracted from transcripts of the 21 Italian daily exposure podcasts. Using the spaCy package (Honnibal & Montani, 2017) within Python, all bisyllabic nouns and their frequencies (i.e., the overall appearance count across the 21 podcasts) were identified from the Italian exposure podcast transcripts. These words were filtered to exclude any English words, cognates, or proper nouns. For Piloting, the 150 highest frequent words were assigned to the High-Frequency Condition, while the 150 lowest frequency words were assigned to the Low-Frequency Condition. Following piloting (section 2.2.4), 120 high frequency words and 120 low frequency words were used as test stimuli.

These words were converted into their phonetic representations using the BAS Web Service Grapheme to Phoneme (G2P) tool (Reichel, 2012), which transforms graphemes into their corresponding phonetic components based on the International Phonetic Alphabet (IPA). Subsequently, the phonotactic score of each word, indicating how typical the word is in Italian based on its phonemes, was calculated, as follows (section 2.2.2.2).

**2.2.2.2 Model Training and Calculating Phonotactic Score.** To calculate the phonotactic probability of each word, the exposure transcripts were converted into their phonetic components in IPA and concatenated into a single dataset. Punctuation, including hyphens, was removed to treat the transcript as a continuous, unsegmented stream, except for paragraph breaks denoting speaker or podcast changes. Phonotactic models were trained on the Italian exposure podcasts using the SRI Language Modeling Toolkit (SRILM), a specialized toolkit for statistical language models (Stolcke, 2002). A trigram model with Whitten-Bell smoothing and no tokenization was trained on the exposure podcast dataset, containing the sequence of phonemes of all the exposure podcasts. This was used to then generate log probabilities of the phonetic sequence of each test item using the trained model. The resulting log probabilities of the phonetic sequence of each word were then normalized by dividing them by the number of phonemes in the sequence plus one, accounting for the end-of-word symbol (Oh et al., 2020). This normalization process yielded the phonotactic score of each word.  A higher (less negative) score signifies a higher likelihood of the phonetic sequence occurring in Italian. Further details regarding the commands employed are available in Appendix A.  Finally, our two frequency conditions were segregated into low and high phonotactic score conditions based on the median split by phonotactic score. This resulted in a 2x2 design for items in the word category, corresponding to high and low frequency by high and low phonotactic scores in which each cell held a total of 60 items (Table 1) for a total of 240 words.

**2.2.2.3 Nonword Creation.** Non-words were generated following the non-tokenization method outlined by Oh et al. (2020). Nonwords were found neither within the exposure podcast nor in the actual Italian language. This approach involved selecting words from our real word condition and changing up to 3 phonemes within the sequence. All nonwords had a corresponding word that was chosen at random. These alterations were made to produce words with lower phonotactic scores than all real words, resulting in phonetic sequences with an extremely low probability of occurring in the Italian language, particularly in bisyllabic words.

A total of 160 non-words were created and then divided based on their average phonotactic scores to establish high and low phonotactic score conditions within the nonword condition. In sum, this process yielded five distinct cells: High-Frequency High Phonotactics (HFHP), High-Frequency Low Phonotactics (HFLP), Low-Frequency High Phonotactics (LFHP), Low-Frequency Low Phonotactics (LFLP), and Nonword Very Low Phonotactics (NVP) (Table 1).

**Table 1**

*Frequency and Range of Phonotactic Score of all Word Categories*

| Phonotactic Probability | High Frequency<br>Count = 7-252 | | Low Frequency<br>Count = 1 | | Nonwords<br>Count = 0 | |
|---|---|---|---|---|---|---|
| | n | PS | n | PS | n | PS |
| High | 60 | -0.57 to -0.93 | 60 | -0.62 to -0.99 | - | |
| Low | 60 | -0.93 to -1.32 | 60 | -1.23 to -1.77 | - | |
| Very Low/ Nonwords | - | | - | | 120 | -1.85 to -3.79 |

*Note:* N represents the number of test items within the word category.  PS represents the range of phonotactic scores of the test items.

### *2.2.3 Text-to-Speech Procedure*

Each word's and nonword's phonetic transcripts were converted into orthographic transcripts using the BAS Web Services Pho2Syl service (Reichel, 2012). This service translates phonological transcripts into syllables, which are then concatenated to create the orthographic representation of each word. Subsequently, these orthographic representations were adjusted to ensure correct pronunciation by the text-to-speech program.  Audio files for each word were then generated using the Google Cloud Console's text-to-speech functionality. The audio files were saved in WAV format at a sampling rate of 48000 Hz (see Appendix A for details). Finally, a phoneme-to-speech website, the *IPA Reader* with the setting “Carla [Italian]” (Linero, 2018) which uses Amazon’s Polly text-to-speech service, was utilized to verify the accuracy of the Cloud Console's generated words. The Google Cloud Console words were auditorily checked against the IPA reader for all test items and the orthography of nonwords was changed to fit the phonetic representation. This step was necessary because Google Cloud Console software requires orthographic input, and thus, the phonetic transcripts needed to be converted into their corresponding orthographic representations. However, the generated orthographic input did not always suitably represent the nonword, or the software could not suitably pronounce the nonword, and therefore the orthography was changed to reflect the phonetic pronunciation.

To ensure that the created non-words did not inadvertently resemble real Italian words, all nonword audio files were reviewed by a native Italian speaker. Within this step, the native Italian speaker confirmed that nonwords resembled natural Italian. Any nonword that was accidentally a real Italian word, or sounded similar to real Italian words was flagged and subsequently replaced with another nonword. This step ensured the integrity of the nonword stimuli.

### *2.2.4 Piloting*

A total of 493 test items (including words and nonwords) underwent initial pilot testing through an online word rating task administered on Prolific. Participant recruitment adhered to the same criteria as outlined previously, resulting in a sample of 22 participants (Mage = 28.23, SD = 4.6). Words that received an average rating outside of 2 standard deviations of the mean were excluded from further analysis and were not included in the final pool of stimuli. Additionally, non-words with very low ratings were excluded, as this indicated that the non-word was too easily discernible from real words. The objective was to ensure that the scores of the nonword words resembled those in the low-frequency category, as judged by English speakers without significant exposure to Italian.  The 60 words with the middle-most ratings, excluding the highest and lowest-rated words were chosen for each category. This was done to remove too-obvious real words and remove real words that seemingly resembled nonwords too closely; phonotactic score was not of importance in this step. This meant that a total of 133 words were removed from the stimuli pool to leave a total of 360 words and nonwords.

Next, the chosen words were counterbalanced into two versions (A or B) based on their phonotactic scores. Care was taken to match the average ratings of the six categories between sessions, ensuring consistency across experimental conditions. The results of the piloting phase guided the selection of the final 240 real words and 120 non-words words, with each version of the final word rating task containing 120 words and 60 non-words. For comprehensive details, including the phonotactic score, frequency, and category of all words and non-words, please refer to Appendix B.

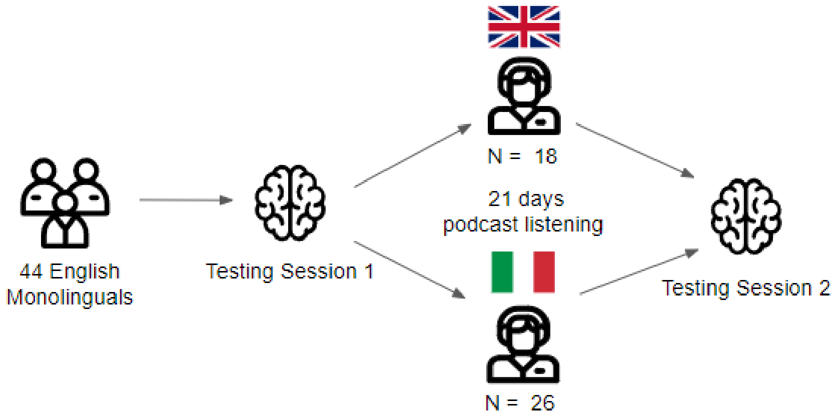
## 2.3 Procedure

### *2.3.1 Overview*

Participants visited the laboratory for an initial testing session (session 1) that included 3 tasks: a word rating task, a word mapping task, and a continuous listening task (data from these two latter tasks will not be analyzed in the current thesis). They then went home and listened to one podcast a day for 21 days in either Italian (L2 exposure group) or English (control group) (Figure 1). Within 10 days of completing the final podcast (mean = 3.4, range = 1-10 days), they returned to the laboratory to do a second round of testing, with the same tasks as session 1.

**Figure 1**

*A Diagram of the Study Design, Including Participant Numbers, for Both Groups*



### *2.3.2 Testing Sessions*

Upon arrival to the lab for the first testing session, participants provided informed consent and completed a demographic questionnaire via Qualtrics, which included inquiries about their language usage. Once eligibility was confirmed, participants were then assigned to either the control group or the L2 exposure group, based on a pre-set order. During the second session, participants first completed a survey regarding their activities while listening to the podcasts, along with the Need for Cognition questionnaire. Otherwise, the two testing sessions were very similar in structure. The EEG cap was placed on the participant's head during the completion of the demographic questionnaire (session 1) or the Need for Cognition questionnaire (session 2).

### *2.3.3 Word Rating Task*

As mentioned, the word rating task was the first task that participants completed within each testing session. Throughout the task, participants were seated comfortably in a dimly lit, sound-attenuated room. Participants were instructed to minimize eye movements and other physical movements while fixating on a central crosshair displayed on a computer monitor and to keep their fingers on the 1-5 keys at the top of a computer keyboard.

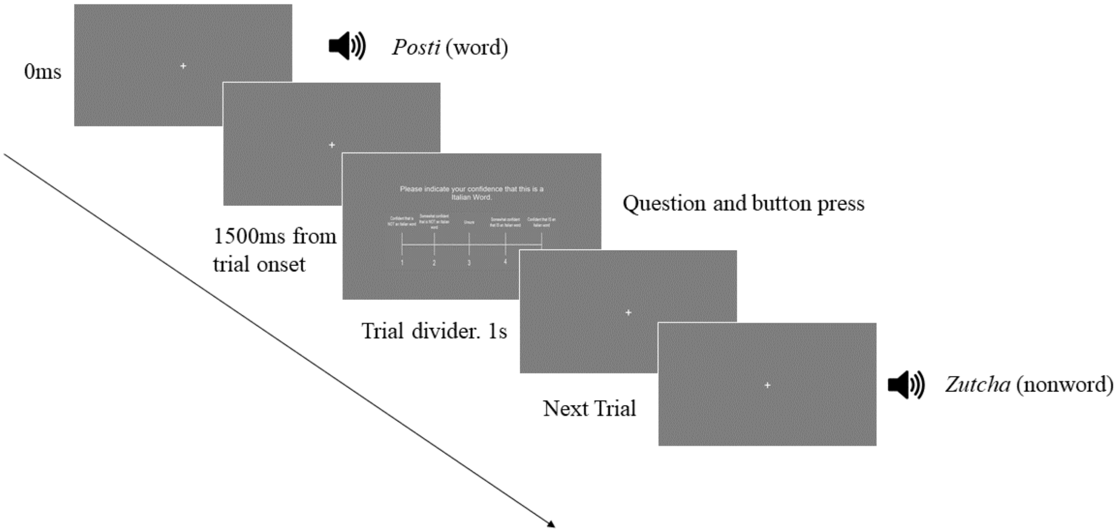
On each trial, participants were presented with a spoken Italian word or nonword and asked to indicate their level of confidence that each item was an Italian word. The written word was not displayed on the screen; participants only heard the stimulus auditorily. The confidence rating scale ranged from 1 to 5, as follows: 1 represented "Very confident that this is NOT an Italian word," 2 indicated "Somewhat confident that this is NOT an Italian word," 3 was "Unsure," 4 meant "Somewhat confident that this is an Italian word," and 5 represented "Very confident that this is an Italian word." Each trial began with a fixation cross, followed immediately by the presentation of the word or nonword (Figure 2). Word and nonwords ranged in length from 229 ms to 748 ms, (mean = 485.26 ms). Participants then saw the rating scale and were asked to input their responses using the top row of the keyboard. Participants were instructed to position their fingers before the start of the block and reduce looking down at the keyboard. They had unlimited time to submit their answers. Following each response, there was a 500ms interval before the start of the next trial.

The task consisted of 9 blocks, each containing 20 trials. Between blocks, participants could choose to take a 30-second break or continue directly after a mandatory 5-second break. Blocks and trials within each block were presented in random order for each participant, and task version assignment (A vs B) to each testing session (session 1 versus session 2) was counterbalanced across participants.

Prior to the main task, participants completed 10 practice trials using English words. Practice trials provided participants with immediate feedback on the correct choice to ensure their comprehension of the task. The task was run using PsychoPy software version 2023.1.0, presented on a monitor within a sound-attenuated, electrically attenuated booth.

**Figure 2**

*A Diagram of the Trial and Task Design*



*Note:* Stimuli were of variable length, therefore the second fixation cross before the onset of the question and button press was of variable length.

### *2.3.4 Exposure Period*

During the exposure period, participants received the exposure podcasts via email as downloadable MP3 files each day at 8 am. They were instructed to listen to the daily podcast without interruption, meaning they could not break it up into chunks throughout the day.

Additionally, participants were advised to listen to the podcast while engaging in undemanding, non-verbal activities, such as driving, cooking, and cleaning, and were instructed to refrain from activities requiring high cognitive effort and/or activities that were verbal in nature, such as homework or playing video games.

Participants were required to complete a survey accompanying each podcast, in which they needed to identify the hidden words. Each podcast had 3 to 5 English words, which were matched between the English and Italian podcasts (i.e., the same words for the day 1 English podcast would be presented for the day 1 Italian podcast). A chime would play to indicate the hidden word, and the podcast would resume directly after the hidden word was played. Participants had to identify all hidden words in a multiple selection format with 6 choices, which included distractor items. On average, participants who completed both testing sessions indicated that they listened to 20.6 out of the 21 podcasts. The signal detection theory measure, d', was calculated to assess participants' ability to identify hidden words embedded within podcasts. The mean d' score across participants was 4.86 (SD = 1.72), indicating a high level of accuracy in detecting these hidden words. Failure to complete the podcasts at the prescribed rate of one podcast and survey per day, missing more than two consecutive days, or consistently failing to correctly identify more than one hidden word resulted in participants being dropped from the study. Participants dropped from the study were compensated for the time already spent listening to the podcasts, but were not invited to complete session 2. Detailed information about the reported activities can be found in Table 2.

**Table 2**

*Distribution of Self-Reported Activities Done While Listening to the Exposure Podcasts*

| Activity | n | % of Participants Reported |
|---|---|---|
| Cleaning | 32 | 76.2% |
| Cooking/Eating | 28 | 66.7% |
| Commuting | 32 | 76.2% |
| Consuming media | 4 | 9.5% |
| Self-Care Routine | 4 | 9.5% |
| Exercising | 3 | 7.1% |
| Art | 2 | 4.8% |

*Note:* Values do not sum to 100% because participants were free to endorse more than one item.

## 2.4 Analysis

### *2.4.1 Behavioural Data*

Linear mixed effects models (LMEs) were used to analyze the behavioural data. The dependent variable consisted of participants' 1-5 confidence ratings for each test item. Fixed effects at the trial level included Phonotactic Score (range, -0.57 to 3.97), Session (Session 1, Session 2), Group (English and L2 Exposure), Frequency (High, Low, nonword) and Word Type (Word and Nonword). Session, Group, Frequency and Word Type were sum coded.

First, we tested whether L2 exposure resulted in a gain of sensitivity to L2 phonotactics, with a stronger effect of phonotactic score on word ratings indicating greater sensitivity to Italian phonotactics. An LME model was computed using Phonotactic Score, Session and Group as fixed factors, with Participant and Stimulus included as random intercepts. Follow-up analyses were done using emtrends (Lenth, 2024). Emtrends was used to evaluate significant interactions using numerical predictors (Phonotactic Score) as opposed to Emmeans, which is better suited for categorical predictors (such as Word Type and Frequency).

```
modelPhonoScore <-lmer(Rating_Resp ~ PhonoScore * Group * Session +
(1|participant) + (1|Word), data = completed)
```

```
emtrends(modelPhonoScore,pairwise ~ Session | Group,var= "PhonoScore" )
```

Next, we investigated whether L2 exposure resulted in learners becoming sensitive to word frequency, specific to the 21 podcasts. An LME model was conducted with Frequency, Session and Group as fixed factors, and Participant and Stimulus as random effects. Post hoc comparisons were conducted using pairwise tests with Tukey adjustments on significant interactions using the categorical factors using the emmeans package (Lenth, 2024).

```
modelFreq <-lmer(Rating_Resp ~ Frequency * Group * Session + (1|participant)
+ (1|Word), data = completed)
```

```
posthoc_Freq <- emmeans(modelFreq, pairwise ~ Frequency * Group * Session,
adjust = "tukey")
```

Lastly, we were interested in whether exposure would inform participants of what makes a true Italian word, collapsed across frequency and phonotactics. A Word Type × Session × Group LME model with Participant and Stimulus as random effects was performed to capture any changes in learners' sensitivity to words versus nonwords following exposure. Post hoc comparisons were conducted using pairwise tests with Tukey adjustments using the emmeans package (Lenth, 2024). These analytical approaches allowed a detailed exploration of how exposure to Italian language podcasts influenced participants' perceptions of word authenticity throughout the study.

```
modelWordType <-lmer(Rating_Resp ~ WordType * Group * Session +
(1|participant) + (1|Word), data = completed)
```

```
posthoc_WordType <- emmeans(modelWordType, pairwise ~ WordType * Group
*Session, adjust = "tukey")
```

Data was preprocessed using JupyterLab in Python, using the NumPy (Harris et al., 2020), Pandas (McKinney & et al, 2010) and SciPy (Virtanen et al., 2020) packages.  Linear mixed effect modelling was conducted in R. First, we used the lmer() command from lme4 (Bates et al., 2015) package to fit our LMEs. To obtain hypothesis tests and p-values for the fixed effects in our LMEs, we utilized the lmerTest package (Kuznetsova et al., 2017). For model selection and model averaging, we employed the MuMIn package (Bartoń, 2023).

## 2.5 EEG Data

### *2.5.1 EEG Data Recording and Preprocessing*

Throughout the word rating task, EEG data was recorded at a sampling rate of 512 Hz using a 64-channel BioSemi ActiveTwo system, recorded relative to the Common Mode Sensor (CMS) active electrode. Electrode placement followed the international 10-20 system. Additional external electrodes were placed on the right and left mastoids (for later offline referencing), above and below the left eye, as well as on the outer canthi of both eyes for EOG monitoring. Event triggers, audio onsets, and keyboard responses were captured using the Cedrus StimTracker Duo.

EEG preprocessing and analyses were conducted using EEGLAB and ERPLAB toolboxes in MATLAB. First, data were re-referenced to the average of both mastoid channels. A broadband filter (0.1 Hz - 30 Hz) was then applied to the data. Next, epochs were extracted from -1000 to 2000 ms relative to stimulus onset. A preliminary manual epoch rejection stage removed artifacts, with an average of 7.09 epochs rejected and 2.7 channels flagged as bad. Independent Component Analysis (ICA), including the eye channels but excluding channels flagged as bad, was then applied. Components reflecting eye movements, muscle activity, noise in single channels ("channel-pop" artifacts) and line noise were removed manually with the help of ICLabel (Pion-Tonachini et al., 2019) and removed from the data. Bad channels were interpolated, and data then underwent a second round of epoch rejection to remove any residual artifacts not entirely removed from the ICA procedure. Artifact-free epochs were averaged to generate ERPs, with baseline correction using the -200 to 0 ms prestimulus interval.

Of our sample of 48 participants, two participants were excluded from EEG analyses, one for excessive eye blinking and the other for signal loss midway through the experiment, resulting in a final sample of 46 participants for ERP analyses (L2 exposure = 24, control = 18).

### *2.5.2 ERP Analysis*

Each stimulus was binned based on phonotactic category following the category breakdown in Table 1. Visual inspection of the waveforms revealed a difference in negative amplitude between conditions within the frontal electrodes at the 300ms to 800ms latency. Therefore, we identified a frontal group of electrodes (Fp1, Fp2, Fpz, AF3, AF4, AF7, AF8, AFz, Fz, F1, F2, F3, F4, F5, F6, F7, and F8) as a region of interest (ROI) for further statistical analysis. Further, we identified the 300ms to 800ms latency as our window for analysis, given the observed latency of the effect (Figure 7). Mean amplitudes were measured from the 300ms to 800ms window, baseline corrected from -200 ms to 0, and extracted for each bin for each participant, averaged across the channels of interest. Repeated measure ANOVAs were conducted to examine the effect of word type, phonotactic category, and frequency on the ERPs.

**2.5.2.1 Words versus Nonwords.** We expected to see group differences in ERP effects between session 1 and session 2, such that the L2 exposure group would show a greater increase in neural sensitivity to Italian words versus nonwords from session 1 to session 2 than the control group. We were also interested in solely the L2 exposure group's changes in processing over time, irrespective of the profile shown by the control group. As such, we conducted two analyses. First, we conducted a repeated measures ANOVA testing using Word Type, Group, and Session as factors, using pair-wise comparisons as follow-up analyses. In addition, we ran a Word Type x Session repeated measures ANOVA within the L2 exposure group. Pair-wise comparisons were conducted as a follow-up on significant interactions. We also report the same analysis within the control group. All syntax used for the ERP analyses can be found in Appendix A.1.

**2.5.2.2 Phonotactic Category.** We further were interested in the impact of the Phonotactic Category (High, Low, Nonword) on the ERPs. A categorical approach was used when analyzing the ERPs for simplicity and power rather than the continuous approach detailed in the behavioural section. Additionally, the very low and extremely low categories were combined for increased power. We ran a Phonotactic Category x Group x Session repeated measure ANOVA, using pairwise comparison for follow-up analyses on significant interactions. Similar to when we investigated word type, we were also interested in solely the L2 exposure groups' changes in processing phonotactics over time. As such, we ran a Phonotactic Category x Session repeated measures ANOVA within the L2 exposure group. Pair-wise comparisons were conducted as a follow-up on significant interactions. The same analysis was done within the control group as well.

**2.5.2.3 Frequency Category.** Lastly, we were interested in the effect of frequency on the ERP's. We ran a Frequency (High, Low, Nonword) x Group x Session repeated measures ANOVA to test the difference in response to word frequency following exposure between groups. As we did for Word Type and Phonotactic Category, we ran a Frequency x Session repeated measures ANOVA within the L2 exposure group using pairwise comparisons as follow-up tests on significant interactions.

# 3. Results

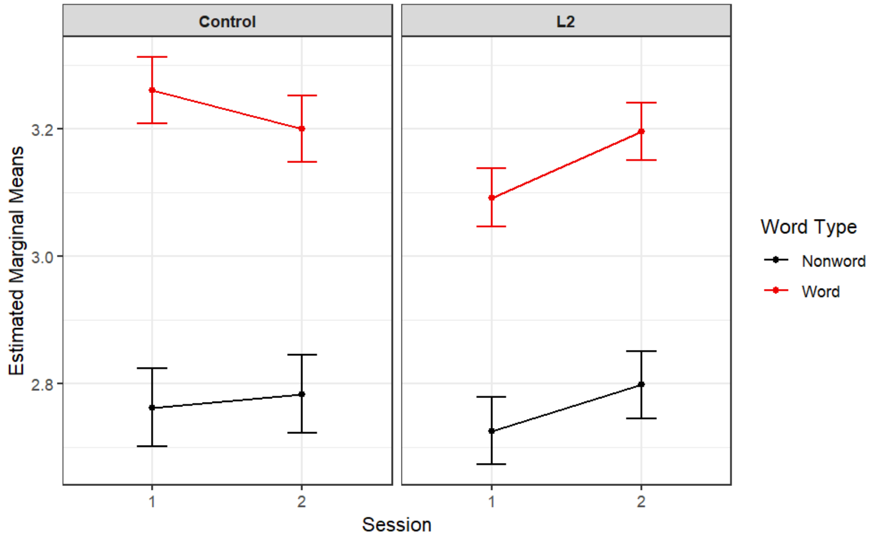
## 3.1 Behavioural Results

### *3.1.1 Word Type (Word versus Nonword)*

Across both groups and both sessions, words were rated as significantly higher (more word-like) than nonwords (Word Type: $F(1, 362)= 105.930$, $p < 0.001$). This finding indicates that participants showed an overall "baseline" level of sensitivity to words versus nonwords in the L2. In contrast to our hypotheses, from session 1 to session 2, the change in rating differences between words and nonwords did not differ between the two groups (Word Type x Session x Group: $F(1, 15448) = 2.633$, $p = 0.105$) (Figure 3). We did find a significant Word Type by Group interaction, $F(1, 15448) = 4.6031$, $p = 0.031$, indicating that the overall rating difference between words and non-words was smaller in the L2 exposure group compared to the control group. Finally, we found that the L2 exposure group showed a significant overall increase in overall ratings (for both words and nonwords) from session 1 to session 2, relative to the control group (Group x Session: $F(1, 15448) = 9.47$, $p = 0.0021$). Within the L2 group, the increase from session 1 to session 2 was significant (emmeans, Session1 L2 – Session2 L2: estimate = -0.088, SE= 0.022, $z$ ratio = -3.94, $p < 0.001$). Full model results can be seen in Table 4, Appendix C.

## Figure 3

*Word Ratings as a Function of Group, Session and Word Type*



*Note:* Values represent estimated marginal means. Error bars represent standard error.

### *3.1.2 Phonotactic Score*

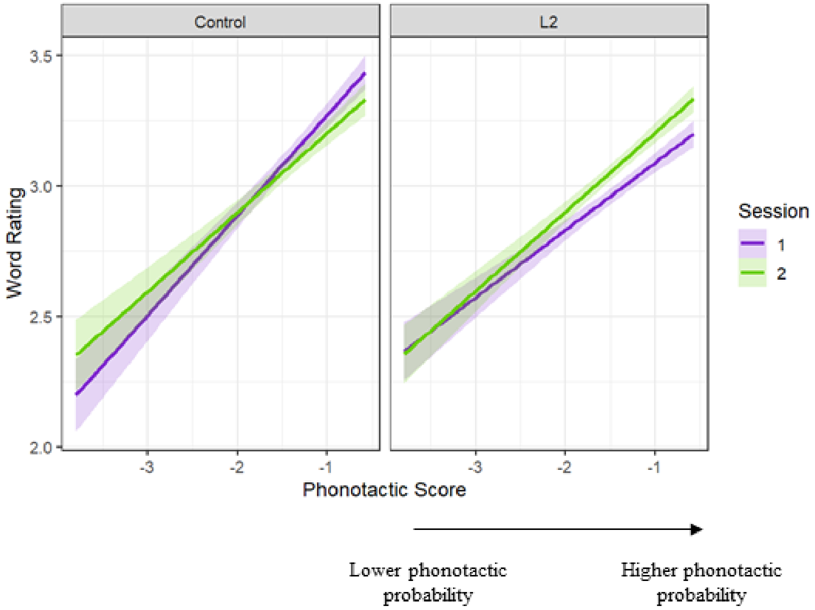
Across the two groups and two sessions, participants demonstrated a baseline sensitivity to phonotactic probability, where words with high phonotactic scores were rated higher (more word-like) than words with low phonotactic scores (Main Effect, Phonotactic Score: $F(1, 362) = 120.58$, $p < 0.001$).

Of direct interest to our hypothesis, the two groups differed in their sensitivity to phonotactic scores from the first to the second session (Phonotactic Score x Group x Session: $F(1, 15449) = 6.941$, $p = 0.008$) (Figure 4). The interaction indicates that the L2 group showed a greater increase in sensitivity to L2 phonotactics from session 1 to session 2 compared to the control group. Follow-up tests with emtrends indicated that the control group showed greater sensitivity to phonotactic probabilities (i.e., a stronger correlation between phonotactic probability and rating) at session 1 compared to session 2 (emtrends, Control; Session 1 – Session 2: estimate = 0.084, SE = 0.038, $z$ ratio = 2.20, $p = 0.028$). For the L2 group, although there was a numerical increase in sensitivity in session 2 when compared to session 1, the difference was not significant (estimate = -0.047, SE = 0.032, $z$

ratio = -1.47, $p = 0.14$). Finally, we found that the L2 exposure group showed a greater increase in overall ratings from session 1 to session 2, relative to the control group (Group x Session: $F(1, 15449) = 6.94$, $p < 0.001$). Full model results can be seen in Table 5, Appendix C.

**Figure 4**

*Word Rating Across Phonotactic Score Between Session and Group*



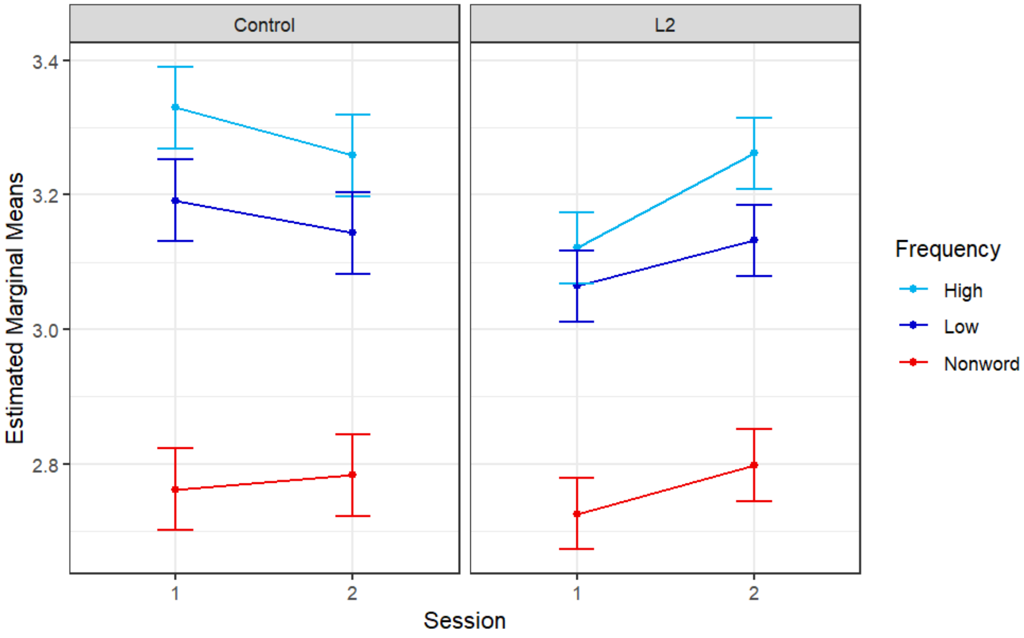
*Note:* Values represent the average rating for each word across participants. The shaded area represents standard error.

### *3.1.3 Word Frequency*

Across groups and sessions, participants showed overall sensitivity to word frequency (Frequency: $F(2, 361) = 56.31$, $p < 0.001$). High frequency words were rated the highest, followed by low frequency words, with nonwords rated the lowest (High – Nonword: estimate = 0.475, SE = 0.047, $z$ ratio = 10.13, $p < 0.001$; Low – Nonword: estimate = 0.37, SE = 0.047, $z$ ratio = 7.79, $p < 0.001$; High – Low Frequency: estimate = 0.011, SE = 0.047, $z$ ratio = 2.34, $p = 0.05$). In contrast to results observed for phonotactic score, the two groups did not differ in their word ratings from session 1 to session 2 in regards to their sensitivity to frequency (Frequency x Group x Session: *F*(2, 15468) = 1.98, *p* = 0.14) (Figure 5). Full model results can be seen in Table 6, Appendix C.

**Figure 5**

*Word Rating as a Function of Frequency, Session and Group*



*Note:* Values represent estimated marginal means. Error bars represent standard error.

## 3.2 ERP Results

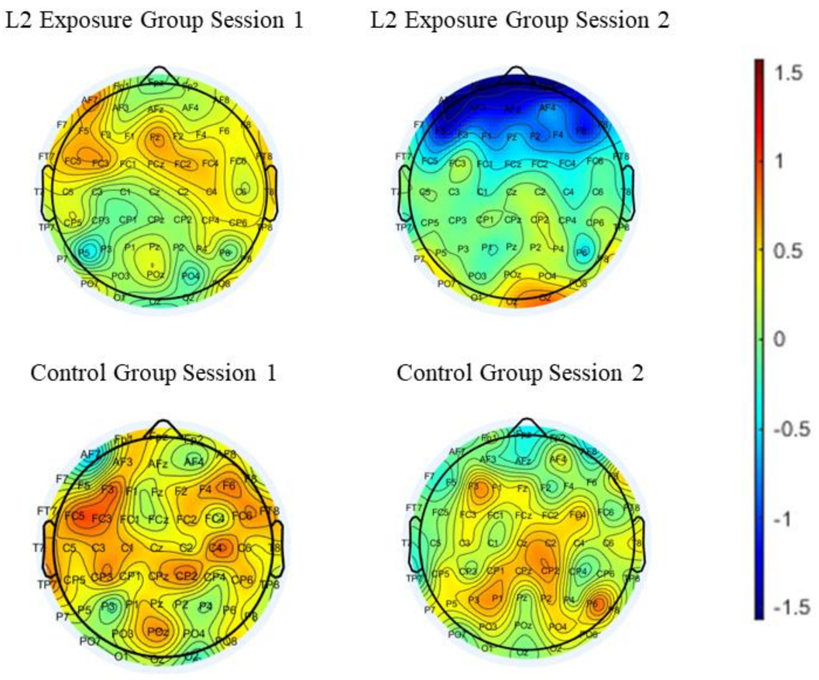
### *3.2.1 Word Type (Word vs Nonword)*

From 300ms to 800ms within the frontal electrodes, the two groups did not significantly differ in their ERP response to nonwords versus words from the first to the second session (Word Type X Group x Session: *F*(1, 40) = 1.61, *p* = 0.21) (Figure 8). We did find that across both groups, there was a marginally significant difference in amplitude between words and nonwords from session 1 to session 2 (Word Type x Session: *F*(1, 40) = 3.76, *p* = 0.06), with nonwords eliciting marginally significant greater negative amplitude than words at session 2 (pairwise t-test, Session 2 Word – Nonword: *t*(1, 41) = 1.76, *p* = 0.086) (Figure 7).

Despite not finding a significant group interaction, we were interested in testing the specific hypothesis that the L2 exposure group would develop increases in sensitivity to nonwords versus words over the three-week listening period. As such, we tested the interaction of Word Type x Session within the L2 exposure group alone. We found a significant interaction (Word Type x Session: $F(1, 23) = 4.381$, $p = 0.048$), suggesting an increase in neural sensitivity to nonwords. Follow-up pairwise t-tests indicated that there was a marginally significant ERP difference between word types at session 2 (Session 2, Word – Nonword: $t(1, 23) = 2.01$, $p = 0.056$), in which nonwords elicited greater negativities than words (Figure 6). At session 1, we did not find any significant difference between words and nonwords (Session 1, Word – Nonword: $t(1, 23) = -0.85$, $p = 0.41$). In contrast, the same analysis within the control group did not reveal a significant Word Type x Session interaction (Word Type x Session: $F(1, 17) = 0.39$, $p = 0.54$).

**Figure 6**

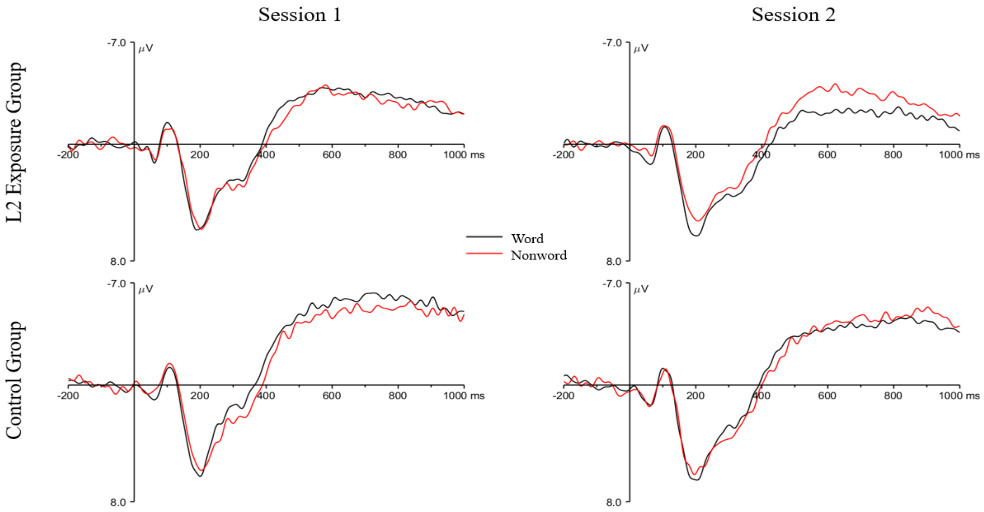
*Scalp Topography plots of the nonword-word difference wave from the 300ms to 800ms latency*



*Note:* Scale is in microvolts

## Figure 7

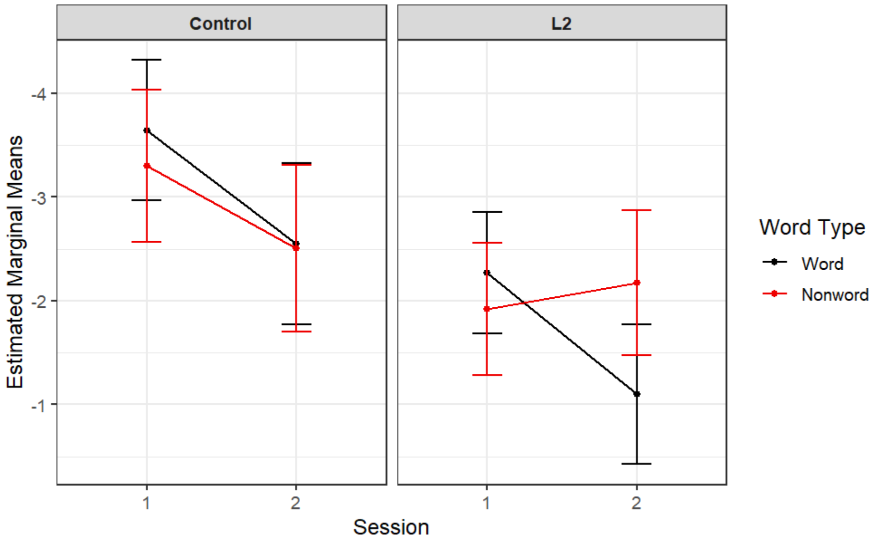
*ERPs as a function of Group, Session and Word Type*



*Note:* Figure represents the F6 Electrode. Y axis is in microvolts and plotted negative up. The x-axis is in milliseconds.

## Figure 8

*ERP Amplitudes as a function of Word Type, Session and Group*



*Note:* Plotted values are estimated marginal means. Error bars represent standard error. The Y-axis is plotted negative up.

### *3.2.2 Phonotactics*

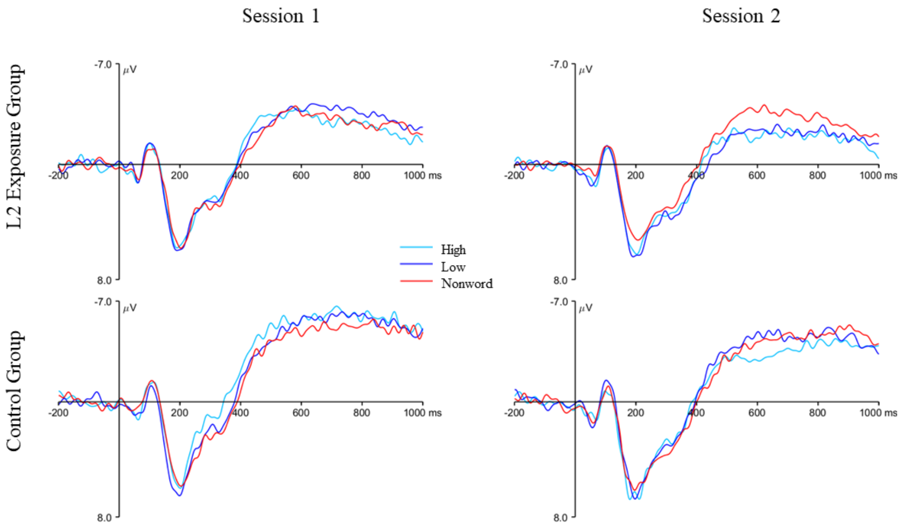
We looked at amplitudes from 300ms to 800ms within the frontal electrodes and found that the two groups did not differ in their change in sensitivity to phonotactic categories (High, Low, Nonword) from session 1 to session 2 (Phonotactic Category x Session x Group: $F(2, 80) = 1.58$, $p = 0.21$) (Figure 10). We did observe a main effect of Session (Session: $F(1, 40)= 4.96$, $p = 0.032$), such that across groups and phonotactic categories, ERP amplitudes were overall greater at session 1 than session 2 (Pairwise t-test, Session 1 – Session 2: estimate = -0.84, SE = 0.38, $t(1, 40) = -2.23$, $p= 0.032$).

When analyzing the L2 exposure group response from session 1 to session 2, there was a marginally significant interaction between Phonotactic Category and Session (Phonotactic Category x Session: $F(2, 46) = 2.62$, $p = 0.083$), indicating that L2 participants showed a somewhat different profile of ERP responses to words of different phonotactic scores from session 1 to session 2 (Figure 9). Follow-up tests indicated that the three phonotactic score categories did not differ at session 1 (all $p$s > 0.28). In contrast, at session 2, nonwords elicited a marginally significantly more negative ERP than words with both low and high phonotactic scores (Session 2, High – Nonword: $t(1, 23) = 1.82$, $p = 0.081$; Session 2, Low – Nonword: $t(1, 23) = 1.91$, $p = 0.069$). Only the low phonotactic score category differed significantly from session 1 to session 2, with decreased negativities at session 2 compared to session 1 (Low, Session 1 – Session 2: $t(1,23) = -2.35$, $p = 0.027$).

The control group did not show any significant changes in amplitude to words of different phonotactic categories from session 1 to session 2 (Phonotactic Category x Session: $F(2, 34) = 0.79$, $p = 0.46$).

## Figure 9

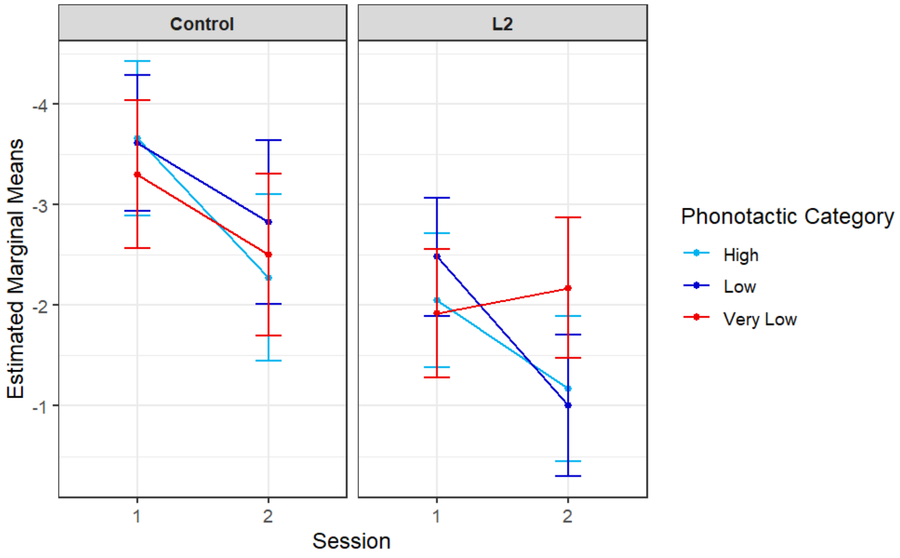
*ERPs of F6 Electrode Divided by Group, Session and Phonotactic Category*



*Note:* Figure represents the F6 Electrode. Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

## Figure 10

*ERP Amplitudes as a Function of Phonotactic Category, Session and Group*



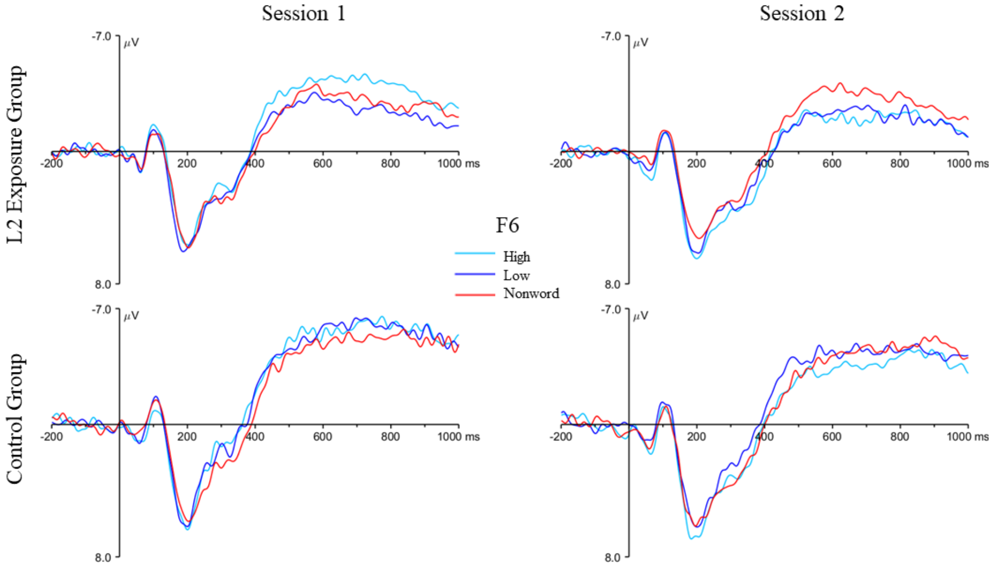
*Note:* Error bars represent standard error and the y-axis is plotted negative up.

### *3.2.3 Frequency*

We did not find that the two groups showed differences from session 1 to session 2 in their sensitivity to Frequency (Frequency x Group x Session: $F(2, 80) = 1.67$, $p = 0.20$) (Figure 12)**.** When focusing on the L2 exposure group, they showed a significant change in neural processing as a function of Frequency at session 2 when compared to session 1 (Frequency x Session: $F(2, 46) = 3.96$, $p = 0.026$). Within the L2 exposure group, follow-up tests revealed that at session 2, only the high frequency words differed marginally significantly from the nonwords, with nonwords eliciting a more negative response (Session 2, High – Nonword: $t(1,32) = 2.33$, $p = 0.086$) (Figure 11). When investigating changes over time within the L2 exposure group, only the high frequency words had significant differences from session 1 to session 2, with decreased negativity in session 2 when compared to session 1 (High Frequency, Session 2 – Session 1: $t(1, 41) = -3.73$, $p < 0.001$). The control group did not differ across sessions in their sensitivity to word frequency (Control, Frequency x Session: $F(2, 34)= 0.19$, $p = 0.82$).

**Figure 11**

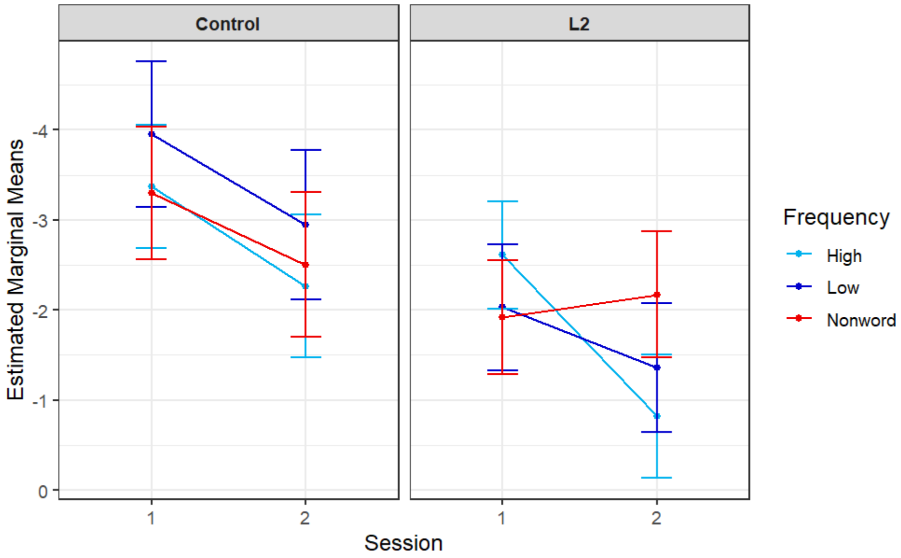
*ERPs of F6 Electrode Divided by Group, Session and Frequency*



*Note:* Figure represents the F6 Electrode. Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

**Figure 12**

*ERP Amplitude as a Function of Frequency, Session and Group*



*Note:* Plotted values are Estimated Marginal Means. Y axis is plotted negative up. Error bars represent standard error.

# 4. Discussion

This project sought to investigate the effect of passive exposure on phoneme regularity learning within a natural language, extending findings from artificial language paradigms. We found that participants who listened to Italian for 1 hour a day for 3 weeks showed a gain in sensitivity to words versus nonwords at the neural level, as indicated by an increased negativity within the 300 to 800ms latency to nonwords over frontal electrodes. In contrast, at the behavioural level, L2 learners did not show a significant change over time on our explicit word rating measure, with stable performance across the two sessions. By comparison, the control group performed significantly *worse* on the word rating task at session 2 when compared to session 1, exhibiting a decline in performance. They also did not show significant neural changes from session 1 to session 2.

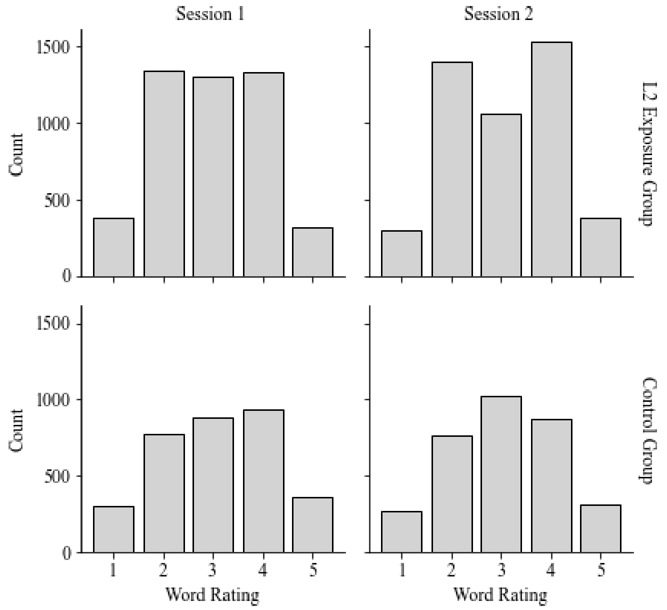
## 4.1 No Significant Changes in Behavioural Sensitivity in the L2 Exposure Group

Although we did not find a significant change in behavioural sensitivity to Italian phonotactics within the L2 exposure group, our results indicate that the 3-week listening experience affected the learner's sensitivity over time relative to control, indicated by a significant interaction between group, session and phonotactic score. The current trend indicates that the association between phonotactic score and ratings increased from session 1 to session 2 within the L2 exposure group, indicating that passive exposure can at least prevent a decline in sensitivity when compared to no exposure. This decline in sensitivity was shown within the control group's significantly lower sensitivity to phonotactic probability in session 2 when compared to session 1. These findings resemble Alexander's and colleagues (2023) results, in which the English control group also demonstrated worse performance at session 2 when compared to session 1 (Figure 4). It is possible that the control group guessed that they were not the group of interest as they listened to English podcasts but completed Italian tasks, and thus lost interest or motivation in the task. It is also possible that since the tasks were identical in structure (but not in content) across sessions, all participants grew bored and therefore did not put as much effort into responding to the task. Therefore, both the behavioural and neural responses might be modulated by engagement to the task, attenuating sensitivity to the phonotactic structure of the language.

In contrast to our hypotheses, within the L2 exposure group, we failed to find any significant differences in behavioural sensitivity to L2 phonotactics over time. These findings contrast Alexander and colleagues' significant results in which the L2 exposure group rated the untrained real Italian words significantly higher at session 2 than at session 1. These discrepancies between studies could be due to the nature of the phrasing of the question we used within the task. We adapted the design of the study from Oh et al (2020) which used a confidence scale ranging from unsure to very confident in their confidence that it was an Italian word. The use of "very confident" on the extreme ends of our scale (ratings 1 and 5) may have led participants to use these responses less, as their unfamiliarity with Italian may have led them to avoid endorsing any items with confidence (Figure 13). The participants' reluctance to use these extremes could be why we did not see a significant change in behavioural sensitivity as there were no big changes in rating responses.   Additionally, in our study, the word rating task was completed first in a series of three tasks during the testing session, without any immediate prior familiarization with Italian. Alexander and colleagues had training sentences before their rating task, possibly providing an opportunity for learning before the familiarity task.

**Figure 13**

*Histograms of Word Rating Responses by Session and Group*



Our results also did not resemble Oh and colleagues' (2020) results, who found that participants who were exposed to Māori rated words with high phonotactic probabilities as more word-like. These authors also found significant differences in performance between their Māori exposure group and control group responses on this word rating task. However, these differences were the result of a lifetime of exposure. Furthermore, their sample had several modes of input due to living in proximity to the L2 in a country in which Māori is an official language, meaning that signage and government postings are also presented in Māori alongside English, supplementing auditory input. In contrast, our participants' sole exposure to Italian between the two testing sessions was background speech. Therefore, the difference in input and the volume of L2 exposure between Oh and colleagues' participants and ours is considerable, which may explain why our results did not mirror theirs.

## 4.3 L2 Exposure Group Demonstrates Changes in Neural Processing Following Exposure

### *4.3.1 Significant Differences of Neural Processing Within the L2 Over Time*

Despite no significant behavioural results, we demonstrated that short-term exposure is sufficient to induce neural changes to single words and nonwords presented in isolation in a natural language. At session 2, participants who listened to Italian showed a greater ERP negativity to Italian nonwords compared to words from the 300 to 800ms latency, within the frontal electrodes (Figure 7). This effect was particularly robust when comparing high frequency words to nonwords, indicating that highly repeated words elicited a reduced negativity relative to less frequent words, with nonwords eliciting the most negative responses. This discrepancy between behavioural and neural results aligns with previous research within the domain of L2 learning, in which neural effects have been observed before explicit behavioural effects (McLaughlin et al., 2004; Tokowicz & MacWhinney, 2005). This dissociation between behavioural and neural changes could be due to the nature of early language learning. The brain may be picking up on implicit patterns which can lead to changes at the neural level, but the learning effect is not robust enough to capture differences at the behavioural level.

### 4.3.2 Functional Significance of the Observed ERP Word Versus Nonword Effect

Our observed waveform is not easily identifiable as it does not have clear links within language or statistical learning literature. Although our identified waveform extends past the typical N400 window of 200ms to 600ms, this observed response could potentially reflect a later going N400 wave (Kutas & Hillyard, 1984; Luck, 2023). The distribution of the observed effect over the frontal electrodes differs from the classic distribution of a typical N400, which is maximal over centroparietal electrodes (Kutas & Hillyard, 1984). Nonetheless, Bermúdez-Margaretto and colleagues (2022) found that late bilinguals demonstrated N400 effects within frontocentral electrodes when processing phonological input. L1 written words primed by L2 phonologically dissimilar words (*Russian* Дичь (game) [dʲˈit͡ɕ] - *English* steel) elicited a greater negative amplitude at approximately 400ms compared to when the preceding prime was phonologically similar (*Russian*. Дичь (game) [dʲˈit͡ɕ] – *English.* ditch). Their results indicate that phonological processing within late learners of an L2 may happen later and recruit more anterior regions of the brain, an effect that the authors attribute to less automaticity of lower proficiency bilinguals. Additionally, Gallagher and colleagues (2022) found that when presented with auditory sentences, low-proficiency bilinguals showed morpho-syntactic violation effects within the frontal regions, contrasting the effects in posterior regions demonstrated by high-proficiency bilinguals and native speakers. Therefore, in the current study, given that our learners could be considered extremely low proficiency, they may have recruited the frontal regions to process unfamiliar L2 (non)words.

The more negative ERP response to nonwords compared to words may partially reflect a training or frequency effect independent of phonotactic probability, in which the more frequent words elicit a decreased N400 following exposure. Following exposure, our L2 group demonstrated a difference in negativity between nonwords and high frequency words, but not nonwords and low frequency words (Figure 11). This theory of a familiarization effect is particularly supported by the low frequency words seeing little change in negativity over time, while the high frequency words elicited significantly decreased negativities from session 1 to session 2. This change in neural response to high frequency words over time is similar to Rossi, Hartmüller, Vignotto, & Obrig's (2013) findings in which L2 pseudowords trained over three days through passive exposure showed a decrease in the N400 post-training when compared to untrained words. Furthermore, Dufour and colleagues (2013) found that in a single word presentation paradigm within monolinguals, high frequency words elicit lower N400s within the frontocentral and centroparietal electrodes when compared to low frequency words, indicating easier processing. Taken together, these findings suggest a familiarization effect in our L2 exposure group to L2 words that were frequently presented. However, this theory would not fully explain why the unfamiliar nonwords are eliciting greater negativities at session 2 when compared to session 1. If the ERP effect were driven entirely by a reduced negativity to high frequency words, the ERP to nonwords should remain stable. Therefore, something more than a frequency effect is modulating the neural response.

### *4.3.3 ERPs may reflect increased L2 phonemic sensitivity in L2 Learners*

The N400 findings discussed in the previous studies were primarily found within semantic contexts, as in these studies, monolinguals and bilinguals had sufficient proficiency to understand semantically similar and dissimilar aspects. Our learners did not hold any semantic knowledge of Italian, nor was there any context since it was a single word presentation. Although the N400 studies provide valuable insight, it alone may not be a sufficient explanation for our results. The possibility of a late-going Phonological Mapping Negativity (PMN) wave may provide an additional explanation. The PMN's sensitivity to phonological expectations and processing, distinct from semantics is more suited to a sample that holds no semantic knowledge within the language (Connolly et al., 2001; Connolly & Phillips, 1994). While the PMN typically peaks around 300ms, which is earlier than our identified latency, our results are similar to its pattern of activation. Within session 2 in our L2 exposure group, nonwords elicited marginally significantly larger negative deflections than high phonotactic probability words within the frontocentral region of the scalp (Figure 9), which is typical of a PMN. Although our results were only marginally significant, these results suggest that our participants learned some basics of Italian phonotactics, as they now had some phonological expectations to identify the less-surprising high phonotactic words from the more-surprising very low phonotactic probability nonwords.

The lack of significant difference in phonological ERP effects from session 1 to session 2 could be due to our participants' lack of L2 proficiency. Previous research has found that PMN amplitude to phonological violations in phoneme discrimination tasks is inversely correlated with L2 proficiency in late bilinguals (Heidlmayr et al., 2021), such that violations or surprising items elicit smaller ERP amplitudes in less proficient bilinguals compared to more proficient bilinguals. Our learners were extremely low in proficiency, having only started becoming familiar with Italian during this study. Therefore, the PMN responses within our L2 exposure group could have been very attenuated compared to higher proficiency learners.

Interestingly, the low phonotactic, less word-like, words experienced a significant decrease in negativity from session 1 to session 2 within the L2 exposure group, leading to a very similar amplitude as the high phonotactic, more word-like, words (Figure 9). It is possible that the low phonotactic words were not sufficiently separable from the high phonotactic category to elicit robust neural differences between the phonotactic categories of Italian words, as the lowest high phonotactic word and the highest low phonotactic probability word differed by only 0.0023 points. Furthermore, the very low phonotactic probability words were made to violate phonotactic regularities, whereas the low phonotactic words still followed them, just to a lesser degree.

### 4.4 Limitations

One limitation of the study is that, by design, phonotactics were somewhat confounded with word frequency.  As the model used to generate phonotactic probability was trained on the corpus of exposure podcasts, high-frequency words were more likely to have higher phonotactic probabilities. As such, the more frequently a word was presented, the higher the likelihood of its combination of phonemes. Despite dissociating frequency and phonotactics to the best of our abilities within this study's design, it is difficult to completely separate frequency from phonotactic probability. However, this is similar to natural language, in which highly frequent words are more likely to be more phototactically typical.

Furthermore, the choice to not have any overlap in phonotactic probability between the words and nonwords, nor have any "untrained" high and low phonotactic words (real Italian words to which they were not exposed) furthered the difficulty of disentangling frequency and phonotactic effects. The low frequency, high phonotactic category was thought the best to disentangle the frequency and phonotactic effects, as they only appear once. However, they still do appear and therefore could have been encoded. In future studies, replacing the low frequency words with words that have a high frequency in Italian but do not appear within the podcasts, or adding high phonotactic probability nonwords would help tease apart the relationship between word familiarization and phonotactic learning.

Previous research has found that exposure to as short as 2 minutes (Saffran, Aslin et al, 1996) can lead to statistical learning of an artificial language. In second language acquisition research, Gullberg (2010) found that as little as 7 minutes of listening to the L2 led to increased identification of illegal L2 syllable combinations, indicating phonotactic learning. Within our study, it is possible that within the word rating task in the first session, there was initial learning of the phonotactic probabilities of Italian before the podcast exposure. A third of the stimuli were nonwords, therefore participants may have learnt and applied the "wrong" phonotactic patterns when making decisions. The control group would have been particularly influenced by these incorrect regularities as they did not have further exposure to Italian outside of the testing sessions. This could explain why they performed worse in the behavioural measures at session 2.

## 4.5 Future Directions

One direction for future research would be investigating the impact of limited explicit knowledge on learning through passive exposure. This study tested participants who had no base knowledge of Italian, and therefore were learning all regularities "from scratch". Those who may have limited knowledge of Italian, such as early classroom learners of Italians, might benefit more from the exposure than those with no experience (such as our sample), as it would supplement their existing knowledge. Previous research has found that explicit knowledge of an L2 significantly contributed to the acquisition of implicit knowledge within the L2 (Kim & Godfroid, 2023; Suzuki & Dekeyser, 2017). As such, future studies may investigate how classroom learners differentiate from novel participants within the design of this study.

An additional research question that we could examine would be whether passive exposure can be used to learn more etymologically distinct languages. As this study used

Italian, which shares many phonemes and even words (cognates) with English, doing a similar study but with a language that does not share as many phonemes would further inform our understanding of phoneme integration and pattern learning. This could remove the baseline effect we found within our study in which participants showed sensitivity to Italian phonotactics even in the first session. I would expect to find similar results, with changes seen with those exposed to the L2. I also expect that these changes may be bigger due to less of a baseline effect. Additionally, using a language with more phonemes or more complex regularities, such as a tonal language, could test the limits of statistical learning within natural language processing.

## 5. Conclusion

To conclude, we have found trends indicating that passive auditory exposure to a second language may facilitate learning of phonotactic regularities in untrained adult learners. Participants exposed to Italian demonstrated marginally increased sensitivity to phonotactic probability, indicated by the neural results. Nonwords, which violated phonemic L2 regularities, elicited larger amplitudes in the 300 to 800ms latency in the frontal electrodes compared to words, indicating the beginning of phonological expectations formation within the L2 learners.  Additionally, we found that word frequency within the exposure podcasts led to differences in word processing within the L2 exposure group, with high frequency words eliciting lower negativities at session 2 than nonwords. These results add to the field of statistical learning by suggesting that statistical learning can scale up to natural language acquisition.

If additional research confirms our initial findings, these insights may be used to inform and supplement language learning in adults, facilitating receptive processing through regularity detection and learning. Future studies could investigate how regular, explicit language learning could be supplemented with passive listening during periods of low-effort cognitive activities, increasing the time spent exposed to the language. Particularly in cases in which immersion is not possible, podcasts could serve as a substitute for natural exposure. For instance, learners could listen to target language podcasts, audiobooks, or radio broadcasts while commuting, exercising, or performing household tasks. This approach not only maximizes language exposure but also integrates learning seamlessly into daily routines, potentially accelerating the acquisition process and enhancing overall proficiency.

# References

Alexander, E., Van Hedger, S. C., & Batterink, L. J. (2023). Learning words without trying: Daily second language podcasts support word form learning in adults. *Psychonomic Bulletin & Review*, *30*(2), 751–762. https://doi.org/https://doi.org/10.3758/s13423-022-02190-1

Bartoń, K. (2023). *MuMIn: Multi-Model Inference* (R package version 1.47.5). https://cran.r-project.org/package=MuMIn.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bermúdez-Margaretto, B., Gallo, F., Novitskiy, N., Myachykov, A., Petrova, A., & Shtyrov, Y. (2022). Ultra-rapid and automatic interplay between L1 and L2 semantics in late bilinguals: EEG evidence. *Cortex*, *151*, 147–161. https://doi.org/10.1016/J.CORTEX.2022.03.004

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal Infants Discover Statistical Word Patterns at Similar Rates as Adults: Evidence From Neural Entrainment. *Psychological Science*, *31*(9), 1161–1173. https://doi.org/10.1177/0956797620933237/FORMAT/EPUB

Connolly, J. F., & Phillips, N. A. (1994). Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266. https://doi.org/10.1162/jocn.1994.6.3.256

Connolly, J. F., Service, E., D'Arcy, R. C. N., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *Neuroreport*, *12*(2), 237–243. https://doi.org/10.1097/00001756-200102120-00012

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. https://doi.org/10.1037/0278-7393.31.1.24

Desroches, A. S., Friesen, D. C., Teles, M., Korade, C. A., & Forest, E. W. (2022). The dynamics of spoken word recognition in bilinguals. *Bilingualism: Language and Cognition*, *25*(4), 705–710. https://doi.org/10.1017/S1366728922000049

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the Time Course of Spoken Word Recognition: Electrophysiological Evidence for the Influences of Phonological Similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893–1906. https://doi.org/10.1162/jocn.2008.21142

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the N400m. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3697–3709. https://doi.org/10.1098/rstb.2009.0158

Dufour, S., Brunellière, A., & Frauenfelder, U. H. (2013). Tracking the Time Course of Word-Frequency Effects in Auditory Word Recognition With Event-Related Potentials. *Cognitive Science*, *37*(3), 489–507. https://doi.org/10.1111/cogs.12015

Field, J. (2003). Promoting perception: lexical segmentation in L2 listening. *ELT Journal*, *57*(4), 325–334. https://doi.org/10.1093/elt/57.4.325

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826. https://doi.org/10.1073/pnas.232472899

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. https://doi.org/10.1016/j.tics.2014.12.010

Gallagher, D., Yano, M., & Ohta, S. (2022). The Neurophysiological Modality Effect in Native and Second Language Processing: An ERP Study. *BioRxiv*.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning*, *60*, 5-24.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Heidlmayr, K., Ferragne, E., & Isel, F. (2021). Neuroplasticity in the phonological system: The PMN and the N400 as markers for the perception of non-native phonemic contrasts by late second language learners. *Neuropsychologia*, *156*, 107831. https://doi.org/10.1016/j.neuropsychologia.2021.107831

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*(8).

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.*

Hunter, C. R. (2013). Early effects of neighborhood density and phonotactic probability of spoken words on event-related potentials. *Brain and Language*, *127*(3), 463–474. https://doi.org/10.1016/j.bandl.2013.09.006

Katayama, T. (2022). Effects of L1 Phonotactic Constraints on Word Segmentation with Clear and Ambiguous English Syllable Boundaries. *Journal of the Phonetic Society of Japan*, *26*, 63–71.

Kim, K. M., & Godfroid, A. (2023). Bilingualism: Language and Cognition The interface of explicit and implicit second-language knowledge: A longitudinal study. *Bilingualism: Language and Cognition*, *26*, 709–723. https://doi.org/10.1017/S1366728922000773

Kittleson, M. M., Aguilar, J. M., Tokerud, G. L., Plante, E., & Asbøjrnsen, A. E. (2010). Implicit language learning: Adults' ability to segment words in Norwegian. *Bilingualism*, *13*(4), 513–523. https://doi.org/10.1017/S1366728910000039

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. https://doi.org/10.1146/ANNUREV.PSYCH.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205. https://doi.org/10.1126/SCIENCE.7350657

Kutas, M., & Hillyard, S. A. (1984). Brain Potentials During Reading Reflect Word Expectancy and Semantic Association. *Antzelevitch, C. Science*, *305*(4), 25–43.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.

Lee, J. Y., Harkrider, A. W., & Hedrick, M. S. (2012). Electrophysiological and behavioral measures of phonological processing of auditory nonsense V–CV–VCV stimuli. *Neuropsychologia*, *50*(5), 666–673. https://doi.org/10.1016/j.neuropsychologia.2012.01.003

Lenth, R. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means.* (version 1.10.0). https://cran.r-project.org/package=emmeans.

Lewendon, J., Mortimore, L., & Egan, C. (2020). The Phonological Mapping (Mismatch) Negativity: History, Inconsistency, and Future Direction. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01967

Linero, K. (2018, September 13). *Pronouncing things with Amazon's Polly*. Cuttlesoft.

Luck, S. J. (2023). Event-related potentials. In *APA handbook of research methods in psychology: Foundations, planning, measures, and psychometrics (Vol. 1) (2nd ed.).* (pp. 605–630). American Psychological Association. https://doi.org/10.1037/0000318-028

Luk, G., Pliatsikas, C., & Rossi, E. (2020). Brain changes associated with language development and learning: A primer on methodology and applications. *System*, *89*, 102209. https://doi.org/10.1016/j.system.2020.102209

McKinney, W., & et al. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference* , *445*, 51–56.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704. https://doi.org/10.1038/nn1264

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*(1), 21–46. https://doi.org/10.1006/JMLA.1998.2568

McWeeny, S., & Norton, E. S. (2020). Understanding event-related potentials (ERPs) in clinical and basic language and communication disorders research: a tutorial. *International Journal of Language & Communication Disorders*, *55*(4), 445–457. https://doi.org/10.1111/1460-6984.12535

Morais, J., & Kolinsky, R. (1994). Perception and awareness in phonological processing: the case of the phoneme. *Cognition*, *50*(1–3), 287–297. https://doi.org/10.1016/0010-0277(94)90032-9

Moreau, C. N., Joanisse, M. F., Mulgrew, J., & Batterink, L. J. (2022). No statistical learning advantage in children over adults: Evidence from behaviour and neural entrainment. *Developmental Cognitive Neuroscience*, *57*, 101154. https://doi.org/10.1016/j.dcn.2022.101154

Oh, Y. M., Todd, S., Beckner, C., Hay, J., King, J., & Needle, J. (2020). *Non-Māori-speaking New Zealanders have a Māori proto-lexicon Detailed Materials and Methods*.

Oh, Y., Todd, S., Beckner, C., Hay, J., King, J., & Needle, J. (2020). Non-Māori-speaking New Zealanders have a Māori proto-lexicon. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-78810-4

Panther, F., Mattingley, W., Todd, S., Hay, J., & King, J. (2023). Proto-Lexicon Size and Phonotactic Knowledge are Linked in Non-Māori Speaking New Zealand Adults. *Laboratory Phonology*, *14*(1). https://doi.org/https://doi.org/10.16995/labphon.7943

Pellegrino, F., & Coupé, C. (2011). A CROSS-LANGUAGE PERSPECTIVE ON SPEECH INFORMATION RATE. *Language*, *87*(3), 539–558.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Reichel, U. D. (2012). PermA and Balloon: Tools for string alignment and text processing. *Proc. Interspeech*, 346.

Remijn, G. B., Hasuo, E., Fujihira, H., & Morimoto, S. (2014). An introduction to the measurement of auditory event-related potentials (ERPs). *Acoustical Science and Technology*, *35*(5), 229–242. https://doi.org/10.1250/ast.35.229

Rossi, S., Hartmüller, T., Vignotto, M., & Obrig, H. (2013). Electrophysiological evidence for modulation of lexical processing after repetitive exposure to foreign phonotactic rules. *Brain and Language*, *127*(3), 404–414. https://doi.org/10.1016/j.bandl.2013.02.009

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. In *JOURNAL OF MEMORY AND LANGUAGE* (Vol. 35).

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and Learning) out of the Corner of Your Ear. *Psychological Science*, *8*(2), 101–105. https://doi.org/10.1111/J.1467-9280.1997.TB00690.X

Saloranta, A., Alku, P., & Peltola, M. S. (2020). Listen-and-repeat training improves perception of second language vowel duration: Evidence from mismatch negativity (MMN) and N1 responses and behavioral discrimination. *International Journal of Psychophysiology*, *147*, 72–82. https://doi.org/10.1016/j.ijpsycho.2019.11.005

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. https://doi.org/10.1038/nn873

Stolcke, A. (2002). SRILM-an extensible language modeling toolkit. *Interspeech*.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1175–1192. https://doi.org/10.1044/1092-4388(2006/085)

Suzuki, Y., & Dekeyser, R. (2017). The Interface of Explicit and Implicit Knowledge in a Second Language: Insights From Individual Differences in Cognitive Aptitudes. *Language Learning*, *67*, 747–790. https://doi.org/10.1111/lang.12241

Tokowicz, N., & MacWhinney, B. (2005). IMPLICIT AND EXPLICIT MEASURES OF SENSITIVITY TO VIOLATIONS IN SECOND LANGUAGE GRAMMAR: An Event-Related Potential Investigation. *Studies in Second Language Acquisition*, *27*(02). https://doi.org/10.1017/S0272263105050102

Tong, X., Wang, Y., & Tong, S. X. (2020). The Neural Signature of Statistical Learning of Orthography. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00026

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Yip, M. C. W. (2020). Spoken word recognition of L2 using probabilistic phonotactics in L1: evidence from Cantonese-English bilinguals. *Language Sciences*, *80*, 101287. https://doi.org/10.1016/J.LANGSCI.2020.101287

# Appendices

## Appendix A: Commands Used for Stimuli Making

### *A.1 Commands used to make the ngram model and compute log probabilities*

Commands were done within C++, using Ubuntu.

Counting ngrams
```
ngram-count -text <corpus.txt> -order 3 -write <CorpusCount.txt>
```

Making the model
```
ngram-count -text <corpus.txt> -order 3 -wbdiscount -interpolate -lm <model1.lm>
```

Computing Log probabilities
```
ngram -lm <model1.lm> -ppl test.txt -debug 1 > <lm1output.txt>
```

### *A.2 Commands for Google Cloud Text -to-speech software*

The Text to speech software used the following settings:

Language: Italian (Italy) - it-IT
Voice: it-IT-Neural2-A
Audio encoding - Linear16
Speed: 1
Pitch: 0
Volume gain: 0
Sample rate (Hertz): 48000

## Appendix B: All Test Items

**Table 3**

*All test items used for the word rating task, including corresponding Phonotactic Score, Word Category, Word Type, and Frequency.*

| Word | PhonoScore | WordCat | WordType | Freq_Count |
|---|---|---|---|---|
| mente | -0.574 | HFHP | Word | 32 |
| conto | -0.600 | HFHP | Word | 13 |
| conti | -0.640 | HFHP | Word | 9 |
| fatto | -0.659 | HFHP | Word | 144 |
| parte | -0.664 | HFHP | Word | 150 |
| parti | -0.668 | HFHP | Word | 14 |
| stati | -0.693 | HFHP | Word | 11 |
| mano | -0.702 | HFHP | Word | 26 |
| costo | -0.706 | HFHP | Word | 13 |
| svolta | -0.714 | HFHP | Word | 15 |
| canto | -0.720 | HFHP | Word | 8 |
| costi | -0.738 | HFHP | Word | 13 |
| testa | -0.744 | HFHP | Word | 12 |
| corte | -0.744 | HFHP | Word | 15 |
| passo | -0.748 | HFHP | Word | 35 |
| giugno | -0.750 | HFHP | Word | 51 |
| storia | -0.750 | HFHP | Word | 97 |
| posti | -0.756 | HFHP | Word | 10 |
| scontro | -0.757 | HFHP | Word | 10 |
| gente | -0.758 | HFHP | Word | 44 |
| mostra | -0.759 | HFHP | Word | 21 |
| segno | -0.764 | HFHP | Word | 17 |
| testo | -0.765 | HFHP | Word | 11 |
| porte | -0.782 | HFHP | Word | 15 |
| volto | -0.785 | HFHP | Word | 12 |
| lista | -0.799 | HFHP | Word | 10 |
| carta | -0.799 | HFHP | Word | 8 |
| atto | -0.817 | HFHP | Word | 39 |
| scontri | -0.823 | HFHP | Word | 7 |
| pace | -0.863 | HFHP | Word | 13 |
| casi | -0.863 | HFHP | Word | 16 |
| arte | -0.864 | HFHP | Word | 11 |
| arti | -0.869 | HFHP | Word | 6 |
| passi | -0.871 | HFHP | Word | 15 |
| vero | -0.872 | HFHP | Word | 22 |

| | | | | |
|---|---|---|---|---|
| voti | -0.875 | HFHP | Word | 33 |
| fascia | -0.877 | HFHP | Word | 8 |
| base | -0.879 | HFHP | Word | 74 |
| voglia | -0.884 | HFHP | Word | 12 |
| folla | -0.887 | HFHP | Word | 17 |
| sera | -0.891 | HFHP | Word | 32 |
| pena | -0.899 | HFHP | Word | 18 |
| stelle | -0.900 | HFHP | Word | 8 |
| sorta | -0.903 | HFHP | Word | 7 |
| scelta | -0.904 | HFHP | Word | 73 |
| causa | -0.909 | HFHP | Word | 36 |
| studi | -0.914 | HFHP | Word | 8 |
| gioco | -0.917 | HFHP | Word | 29 |
| senso | -0.920 | HFHP | Word | 41 |
| messa | -0.920 | HFHP | Word | 7 |
| stile | -0.922 | HFHP | Word | 23 |
| crisi | -0.923 | HFHP | Word | 87 |
| rischi | -0.924 | HFHP | Word | 8 |
| piedi | -0.925 | HFHP | Word | 8 |
| stadi | -0.925 | HFHP | Word | 8 |
| colpi | -0.926 | HFHP | Word | 11 |
| nota | -0.926 | HFHP | Word | 10 |
| cose | -0.926 | HFHP | Word | 42 |
| ruolo | -0.927 | HFHP | Word | 49 |
| vite | -0.929 | HFHP | Word | 7 |
| nome | -0.929 | HFLP | Word | 77 |
| fasce | -0.930 | HFLP | Word | 10 |
| luogo | -0.930 | HFLP | Word | 38 |
| gioia | -0.932 | HFLP | Word | 7 |
| coppie | -0.937 | HFLP | Word | 16 |
| scuole | -0.938 | HFLP | Word | 26 |
| donna | -0.944 | HFLP | Word | 21 |
| foto | -0.946 | HFLP | Word | 10 |
| leggi | -0.948 | HFLP | Word | 9 |
| strage | -0.952 | HFLP | Word | 15 |
| padre | -0.959 | HFLP | Word | 17 |
| calcio | -0.963 | HFLP | Word | 31 |
| sfera | -0.963 | HFLP | Word | 16 |
| musei | -0.964 | HFLP | Word | 12 |
| cerca | -0.973 | HFLP | Word | 11 |
| sguardo | -0.973 | HFLP | Word | 14 |
| gusto | -0.976 | HFLP | Word | 8 |
| rabbia | -0.981 | HFLP | Word | 7 |

|  |  |  |  |  |
|---|---|---|---|---|
| scelte | -0.982 | HFLP | Word | 19 |
| lancio | -0.983 | HFLP | Word | 7 |
| cuore | -0.983 | HFLP | Word | 30 |
| corsa | -0.985 | HFLP | Word | 12 |
| bordo | -0.986 | HFLP | Word | 9 |
| stampa | -0.986 | HFLP | Word | 59 |
| flusso | -0.987 | HFLP | Word | 14 |
| occhi | -1.035 | HFLP | Word | 38 |
| casse | -1.036 | HFLP | Word | 9 |
| reti | -1.036 | HFLP | Word | 13 |
| voce | -1.044 | HFLP | Word | 30 |
| tema | -1.054 | HFLP | Word | 59 |
| legno | -1.058 | HFLP | Word | 9 |
| nomi | -1.071 | HFLP | Word | 17 |
| uomo | -1.073 | HFLP | Word | 39 |
| dubbio | -1.081 | HFLP | Word | 22 |
| beni | -1.083 | HFLP | Word | 9 |
| blocco | -1.092 | HFLP | Word | 18 |
| bocca | -1.097 | HFLP | Word | 8 |
| peso | -1.101 | HFLP | Word | 11 |
| lotta | -1.107 | HFLP | Word | 32 |
| scena | -1.112 | HFLP | Word | 38 |
| moglie | -1.115 | HFLP | Word | 11 |
| sede | -1.119 | HFLP | Word | 30 |
| corpo | -1.124 | HFLP | Word | 12 |
| nave | -1.130 | HFLP | Word | 22 |
| fede | -1.135 | HFLP | Word | 7 |
| zona | -1.148 | HFLP | Word | 14 |
| sangue | -1.159 | HFLP | Word | 7 |
| tasso | -1.165 | HFLP | Word | 9 |
| flussi | -1.178 | HFLP | Word | 12 |
| gamma | -1.207 | HFLP | Word | 9 |
| giro | -1.225 | HFLP | Word | 25 |
| pesca | -1.230 | HFLP | Word | 8 |
| rotta | -1.232 | HFLP | Word | 13 |
| tipo | -1.252 | HFLP | Word | 38 |
| schermo | -1.254 | HFLP | Word | 8 |
| leva | -1.289 | HFLP | Word | 10 |
| ciclo | -1.295 | HFLP | Word | 13 |
| linee | -1.299 | HFLP | Word | 9 |
| droga | -1.313 | HFLP | Word | 8 |
| cibo | -1.322 | HFLP | Word | 23 |
| hanno | -0.617 | LFHP | Word | 1 |

| | | | | |
|---|---|---|---|---|
| mille | -0.657 | LFHP | Word | 1 |
| fare | -0.698 | LFHP | Word | 1 |
| conche | -0.707 | LFHP | Word | 1 |
| quinte | -0.732 | LFHP | Word | 1 |
| visti | -0.751 | LFHP | Word | 1 |
| turche | -0.780 | LFHP | Word | 1 |
| solo | -0.780 | LFHP | Word | 1 |
| monte | -0.787 | LFHP | Word | 1 |
| stiva | -0.791 | LFHP | Word | 1 |
| forni | -0.797 | LFHP | Word | 1 |
| finta | -0.823 | LFHP | Word | 1 |
| tante | -0.828 | LFHP | Word | 1 |
| scritto | -0.831 | LFHP | Word | 1 |
| pura | -0.835 | LFHP | Word | 1 |
| monti | -0.837 | LFHP | Word | 1 |
| scarto | -0.838 | LFHP | Word | 1 |
| duro | -0.838 | LFHP | Word | 1 |
| dura | -0.840 | LFHP | Word | 1 |
| cinta | -0.854 | LFHP | Word | 1 |
| sbando | -0.856 | LFHP | Word | 1 |
| cure | -0.862 | LFHP | Word | 1 |
| spinto | -0.863 | LFHP | Word | 1 |
| torta | -0.866 | LFHP | Word | 1 |
| tratte | -0.869 | LFHP | Word | 1 |
| casco | -0.872 | LFHP | Word | 1 |
| perso | -0.876 | LFHP | Word | 1 |
| sardo | -0.877 | LFHP | Word | 1 |
| salti | -0.885 | LFHP | Word | 1 |
| parchi | -0.893 | LFHP | Word | 1 |
| resti | -0.897 | LFHP | Word | 1 |
| banda | -0.900 | LFHP | Word | 1 |
| ratto | -0.906 | LFHP | Word | 1 |
| rata | -0.911 | LFHP | Word | 1 |
| pasti | -0.912 | LFHP | Word | 1 |
| palle | -0.917 | LFHP | Word | 1 |
| uova | -0.929 | LFHP | Word | 1 |
| pala | -0.934 | LFHP | Word | 1 |
| turni | -0.938 | LFHP | Word | 1 |
| piste | -0.939 | LFHP | Word | 1 |
| densi | -0.942 | LFHP | Word | 1 |
| fitta | -0.944 | LFHP | Word | 1 |
| pranzo | -0.949 | LFHP | Word | 1 |
| stasi | -0.951 | LFHP | Word | 1 |

| | | | | |
|---|---|---|---|---|
| vecchio | -0.959 | LFHP | Word | 1 |
| inno | -0.960 | LFHP | Word | 1 |
| lite | -0.964 | LFHP | Word | 1 |
| degna | -0.965 | LFHP | Word | 1 |
| segni | -0.967 | LFHP | Word | 1 |
| ricco | -0.968 | LFHP | Word | 1 |
| ditta | -0.968 | LFHP | Word | 1 |
| sfere | -0.973 | LFHP | Word | 1 |
| saldi | -0.974 | LFHP | Word | 1 |
| vene | -0.975 | LFHP | Word | 1 |
| stracci | -0.980 | LFHP | Word | 1 |
| marna | -0.980 | LFHP | Word | 1 |
| cancro | -0.981 | LFHP | Word | 1 |
| fasi | -0.986 | LFHP | Word | 1 |
| bara | -0.989 | LFHP | Word | 1 |
| musi | -0.991 | LFHP | Word | 1 |
| niente | -1.237 | LFLP | Word | 1 |
| ciance | -1.238 | LFLP | Word | 1 |
| chiodi | -1.241 | LFLP | Word | 1 |
| cortei | -1.242 | LFLP | Word | 1 |
| pelo | -1.244 | LFLP | Word | 1 |
| socio | -1.246 | LFLP | Word | 1 |
| onde | -1.247 | LFLP | Word | 1 |
| giacca | -1.249 | LFLP | Word | 1 |
| pesche | -1.251 | LFLP | Word | 1 |
| zaini | -1.262 | LFLP | Word | 1 |
| scudo | -1.270 | LFLP | Word | 1 |
| gamba | -1.273 | LFLP | Word | 1 |
| targhe | -1.273 | LFLP | Word | 1 |
| truffa | -1.275 | LFLP | Word | 1 |
| svago | -1.279 | LFLP | Word | 1 |
| agost | -1.281 | LFLP | Word | 1 |
| gonfie | -1.283 | LFLP | Word | 1 |
| urlo | -1.289 | LFLP | Word | 1 |
| buio | -1.308 | LFLP | Word | 1 |
| dubbie | -1.308 | LFLP | Word | 1 |
| irto | -1.310 | LFLP | Word | 1 |
| sacche | -1.313 | LFLP | Word | 1 |
| enne | -1.326 | LFLP | Word | 1 |
| svelta | -1.337 | LFLP | Word | 1 |
| stoffa | -1.352 | LFLP | Word | 1 |
| coppe | -1.352 | LFLP | Word | 1 |
| giacche | -1.357 | LFLP | Word | 1 |

|  |  |  |  |  |
|---|---|---|---|---|
| tonno | -1.359 | LFLP | Word | 1 |
| norcia | -1.361 | LFLP | Word | 1 |
| albi | -1.382 | LFLP | Word | 1 |
| tocco | -1.389 | LFLP | Word | 1 |
| lupo | -1.394 | LFLP | Word | 1 |
| buie | -1.400 | LFLP | Word | 1 |
| sveglio | -1.410 | LFLP | Word | 1 |
| tubo | -1.415 | LFLP | Word | 1 |
| stoffe | -1.415 | LFLP | Word | 1 |
| freddo | -1.423 | LFLP | Word | 1 |
| soia | -1.424 | LFLP | Word | 1 |
| gomma | -1.431 | LFLP | Word | 1 |
| mesa | -1.437 | LFLP | Word | 1 |
| bivio | -1.438 | LFLP | Word | 1 |
| piombo | -1.453 | LFLP | Word | 1 |
| buffa | -1.453 | LFLP | Word | 1 |
| topi | -1.456 | LFLP | Word | 1 |
| schizzo | -1.460 | LFLP | Word | 1 |
| valzer | -1.461 | LFLP | Word | 1 |
| pieghe | -1.467 | LFLP | Word | 1 |
| torre | -1.468 | LFLP | Word | 1 |
| bacio | -1.471 | LFLP | Word | 1 |
| droghe | -1.475 | LFLP | Word | 1 |
| cadmio | -1.492 | LFLP | Word | 1 |
| manna | -1.494 | LFLP | Word | 1 |
| colse | -1.561 | LFLP | Word | 1 |
| succo | -1.566 | LFLP | Word | 1 |
| voga | -1.584 | LFLP | Word | 1 |
| buccia | -1.642 | LFLP | Word | 1 |
| ghiotta | -1.649 | LFLP | Word | 1 |
| scotso | -1.671 | LFLP | Word | 1 |
| cursus | -1.716 | LFLP | Word | 1 |
| poppa | -1.769 | LFLP | Word | 1 |
| doche | -1.847 | NHP | Nonword | 0 |
| jukko | -1.853 | NHP | Nonword | 0 |
| valtse | -1.855 | NHP | Nonword | 0 |
| zuzo | -1.856 | NHP | Nonword | 0 |
| mummi | -1.864 | NHP | Nonword | 0 |
| zegge | -1.866 | NHP | Nonword | 0 |
| chommo | -1.876 | NHP | Nonword | 0 |
| rjonnia | -1.883 | NHP | Nonword | 0 |
| zitte | -1.891 | NHP | Nonword | 0 |
| vutsha | -1.892 | NHP | Nonword | 0 |

| | | | | |
|---|---|---|---|---|
| guatso | -1.897 | NHP | Nonword | 0 |
| ceinni | -1.900 | NHP | Nonword | 0 |
| lassi | -1.903 | NHP | Nonword | 0 |
| brogu | -1.903 | NHP | Nonword | 0 |
| mesche | -1.904 | NHP | Nonword | 0 |
| cearse | -1.905 | NHP | Nonword | 0 |
| scacie | -1.908 | NHP | Nonword | 0 |
| nubu | -1.910 | NHP | Nonword | 0 |
| vesi | -1.922 | NHP | Nonword | 0 |
| rotu | -1.922 | NHP | Nonword | 0 |
| rore | -1.927 | NHP | Nonword | 0 |
| zvuvo | -1.928 | NHP | Nonword | 0 |
| mufe | -1.937 | NHP | Nonword | 0 |
| medgio | -1.939 | NHP | Nonword | 0 |
| walo | -1.939 | NHP | Nonword | 0 |
| pamgue | -1.943 | NHP | Nonword | 0 |
| chimmo | -1.946 | NHP | Nonword | 0 |
| bivao | -1.955 | NHP | Nonword | 0 |
| piedze | -1.970 | NHP | Nonword | 0 |
| vawe | -1.975 | NHP | Nonword | 0 |
| meigni | -1.980 | NHP | Nonword | 0 |
| uzpe | -1.995 | NHP | Nonword | 0 |
| scoupi | -2.008 | NHP | Nonword | 0 |
| iutje | -2.010 | NHP | Nonword | 0 |
| aebei | -2.010 | NHP | Nonword | 0 |
| tcholi | -2.016 | NHP | Nonword | 0 |
| mudda | -2.035 | NHP | Nonword | 0 |
| cevi | -2.041 | NHP | Nonword | 0 |
| chevi | -2.041 | NHP | Nonword | 0 |
| potco | -2.050 | NHP | Nonword | 0 |
| sluoi | -2.052 | NHP | Nonword | 0 |
| foppo | -2.060 | NHP | Nonword | 0 |
| nolne | -2.065 | NHP | Nonword | 0 |
| sise | -2.073 | NHP | Nonword | 0 |
| sbiagge | -2.103 | NHP | Nonword | 0 |
| carsu | -2.110 | NHP | Nonword | 0 |
| tebba | -2.112 | NHP | Nonword | 0 |
| cebbe | -2.115 | NHP | Nonword | 0 |
| zutcha | -2.126 | NHP | Nonword | 0 |
| nache | -2.131 | NHP | Nonword | 0 |
| nake | -2.131 | NHP | Nonword | 0 |
| pisve | -2.133 | NHP | Nonword | 0 |
| morle | -2.148 | NHP | Nonword | 0 |

| | | | | |
|---|---|---|---|---|
| gnedi | -2.179 | NHP | Nonword | 0 |
| dlaglia | -2.185 | NHP | Nonword | 0 |
| sciulla | -2.207 | NHP | Nonword | 0 |
| votza | -2.211 | NHP | Nonword | 0 |
| vubi | -2.211 | NHP | Nonword | 0 |
| cego | -2.211 | NHP | Nonword | 0 |
| bimfi | -2.219 | NHP | Nonword | 0 |
| slossa | -2.221 | NLP | Nonword | 0 |
| leca | -2.223 | NLP | Nonword | 0 |
| shugge | -2.232 | NLP | Nonword | 0 |
| gnizo | -2.239 | NLP | Nonword | 0 |
| scionfi | -2.260 | NLP | Nonword | 0 |
| piomts | -2.279 | NLP | Nonword | 0 |
| pjonts | -2.279 | NLP | Nonword | 0 |
| gnora | -2.284 | NLP | Nonword | 0 |
| dzaci | -2.289 | NLP | Nonword | 0 |
| tecu | -2.295 | NLP | Nonword | 0 |
| garuf | -2.296 | NLP | Nonword | 0 |
| gljasse | -2.301 | NLP | Nonword | 0 |
| mjetche | -2.310 | NLP | Nonword | 0 |
| vutche | -2.318 | NLP | Nonword | 0 |
| dzoglio | -2.344 | NLP | Nonword | 0 |
| psesju | -2.347 | NLP | Nonword | 0 |
| gnunno | -2.353 | NLP | Nonword | 0 |
| muo | -2.370 | NLP | Nonword | 0 |
| shaptcha | -2.382 | NLP | Nonword | 0 |
| wuga | -2.389 | NLP | Nonword | 0 |
| gnoia | -2.395 | NLP | Nonword | 0 |
| sascio | -2.398 | NLP | Nonword | 0 |
| slupo | -2.409 | NLP | Nonword | 0 |
| peggion | -2.426 | NLP | Nonword | 0 |
| zrelgio | -2.451 | NLP | Nonword | 0 |
| sciube | -2.479 | NLP | Nonword | 0 |
| tchaglio | -2.495 | NLP | Nonword | 0 |
| lona | -2.513 | NLP | Nonword | 0 |
| sgida | -2.519 | NLP | Nonword | 0 |
| cuppa | -2.551 | NLP | Nonword | 0 |
| gioca | -2.554 | NLP | Nonword | 0 |
| tchesa | -2.555 | NLP | Nonword | 0 |
| tsesa | -2.555 | NLP | Nonword | 0 |
| tsufa | -2.555 | NLP | Nonword | 0 |
| zunba | -2.557 | NLP | Nonword | 0 |
| kilku | -2.564 | NLP | Nonword | 0 |

|  |  |  |  |  |
|---|---|---|---|---|
| frumi | -2.592 | NLP | Nonword | 0 |
| dzedzo | -2.599 | NLP | Nonword | 0 |
| dzeggio | -2.599 | NLP | Nonword | 0 |
| diba | -2.638 | NLP | Nonword | 0 |
| luzza | -2.652 | NLP | Nonword | 0 |
| durbu | -2.658 | NLP | Nonword | 0 |
| gliumo | -2.706 | NLP | Nonword | 0 |
| tchepa | -2.740 | NLP | Nonword | 0 |
| vocciu | -2.745 | NLP | Nonword | 0 |
| djeba | -2.788 | NLP | Nonword | 0 |
| dzigne | -2.792 | NLP | Nonword | 0 |
| gneie | -2.792 | NLP | Nonword | 0 |
| tsolli | -2.792 | NLP | Nonword | 0 |
| svofe | -2.887 | NLP | Nonword | 0 |
| goscio | -2.942 | NLP | Nonword | 0 |
| dudze | -2.958 | NLP | Nonword | 0 |
| gluio | -2.978 | NLP | Nonword | 0 |
| gnassi | -2.981 | NLP | Nonword | 0 |
| gliepe | -3.008 | NLP | Nonword | 0 |
| tchuzu | -3.086 | NLP | Nonword | 0 |
| dzotta | -3.114 | NLP | Nonword | 0 |
| gliecchi | -3.335 | NLP | Nonword | 0 |
| dzasci | -3.794 | NLP | Nonword | 0 |
| dzeshi | -3.794 | NLP | Nonword | 0 |

*Note*: Table is sorted by Phonotactic Score within Word Category.

## Appendix C: Results from Behavioural Models

**Table 4**

*Word Type x Group x Session LME Model Output*

| | SumSq | Mean Sq | NumDF | DenDF | Fvalue | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| WordType | 110.432 | 110.432 | 1 | 362.4 | 105.93 | < .000 | *** |
| Group | 0.72 | 0.72 | 1.00 | 42.80 | 0.69 | 0.41 | |
| Session | 4.00 | 4.00 | 1.00 | 15448.40 | 3.84 | 0.05 | . |
| WordType:Group | 4.80 | 4.80 | 1.00 | 15432.00 | 4.60 | 0.03 | * |
| WordType:Session | 0.53 | 0.53 | 1.00 | 15448.70 | 0.50 | 0.48 | |
| Group:Session | 9.87 | 9.87 | 1.00 | 15448.40 | 9.47 | 0.00 | ** |
| WordType:Group: Session | 2.75 | 2.75 | 1.00 | 15448.70 | 2.63 | 0.10 | |

**Table 5**

*Phonotactic Score x Group x Session LME Model Output*

| | SumSq | Mean Sq | NumDF | DenDF | Fvalue | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| PhonoScore | 125.65 | 125.65 | 1.00 | 362.50 | 120.58 | < .000 | *** |
| Group | 5.23 | 5.23 | 1.00 | 80.60 | 5.01 | 0.03 | * |
| Session | 0.01 | 0.01 | 1.00 | 15449.50 | 0.00 | 0.94 | |
| PhonoScore:Group | 6.77 | 6.77 | 1.00 | 15432.00 | 6.50 | 0.01 | * |
| PhonoScore:Session | 0.59 | 0.59 | 1.00 | 15449.60 | 0.57 | 0.45 | |
| Group:Session | 16.61 | 16.61 | 1.00 | 15449.50 | 15.94 | 0.00 | *** |
| PhonoScore:Group: Session | 7.23 | 7.23 | 1.00 | 15449.60 | 6.94 | 0.01 | ** |

**Table 6**

*Frequency x Group x Session LME Model Output*

| | SumSq | Mean Sq | NumDF | DenDF | Fvalue | Pr(>F) | |
|---|---|---|---|---|---|---|---|
| Frequency | 117.42 | 58.71 | 2.00 | 361.40 | 56.32 | < .000 | *** |
| Group | 1.15 | 1.15 | 1.00 | 42.00 | 1.11 | 0.30 | |
| Session | 3.55 | 3.55 | 1.00 | 15445.20 | 3.40 | 0.07 | . |
| Frequency:Group | 5.52 | 2.76 | 2.00 | 15429.00 | 2.65 | 0.07 | . |
| Frequency:Session | 0.89 | 0.45 | 2.00 | 15469.90 | 0.43 | 0.65 | |
| Group:Session | 15.22 | 15.22 | 1.00 | 15445.20 | 14.60 | 0.00 | *** |
| Frequency:Group: Session | 4.13 | 2.06 | 2.00 | 15469.90 | 1.98 | 0.14 | |

## Appendix D: Additional Demographic Information

**Table 7**

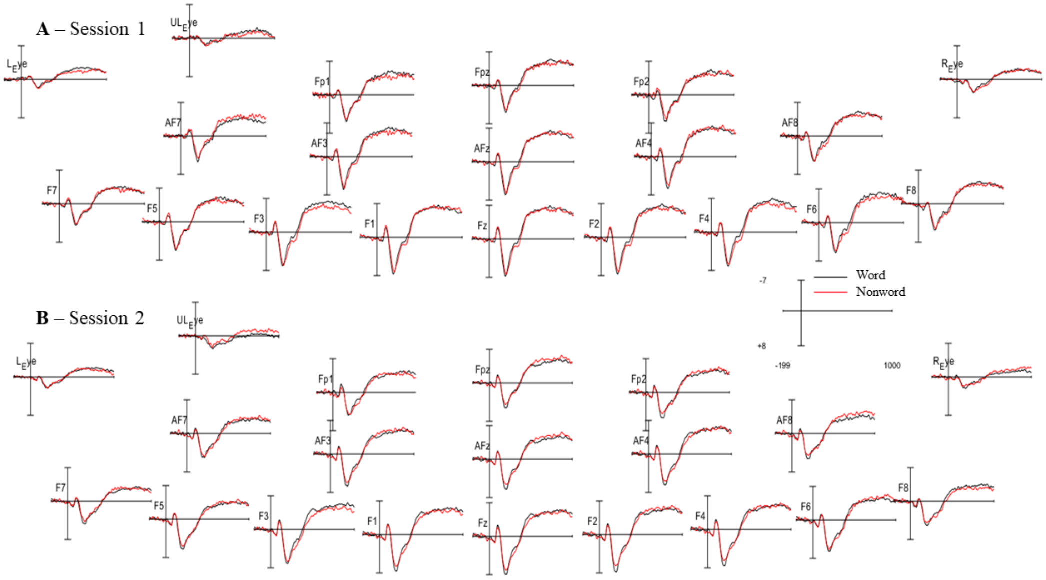
*Demographic Information of Participants Included in the Study*

| Characteristic | n |
|---|---|
| Gender | |
| Male | 16 |
| Female | 28 |
| Handedness | |
| Right | 42 |
| Left | 2 |
| Romance Language Exposure | |
| None | 37 |
| Required French | 6 |
| Other | 1 |
| M age | 22.86 |
| SD | 2.7 |

## Appendix E: Additional ERPs

### Figure 14

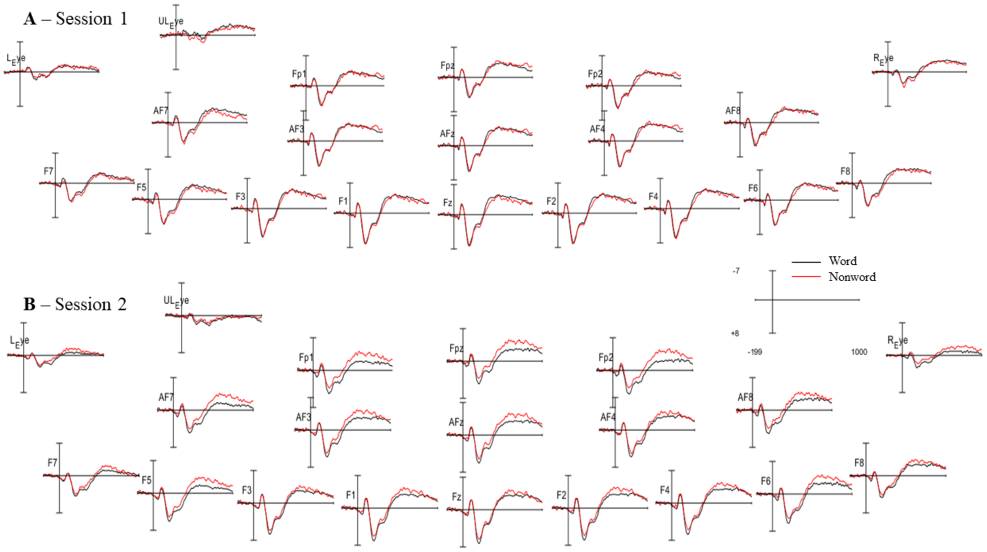
*ERPs of the Frontal Electrodes of the Control Group, Divided by Session and Word Type.*



Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

### Figure 15

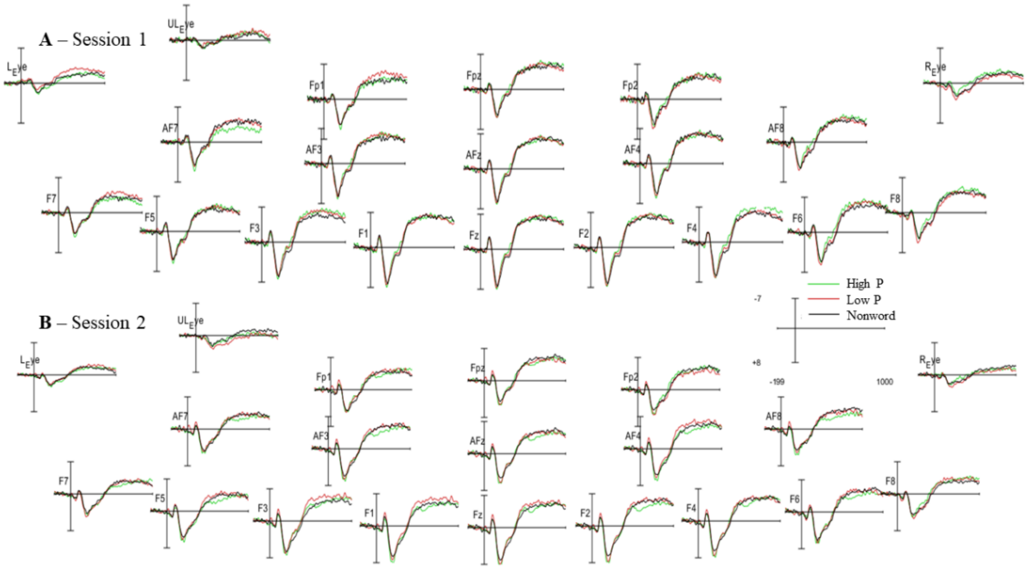
*ERPs of the frontal electrodes of the L2 Exposure group separated by Session and Word Type.*



*Note:* Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

**Figure 16**

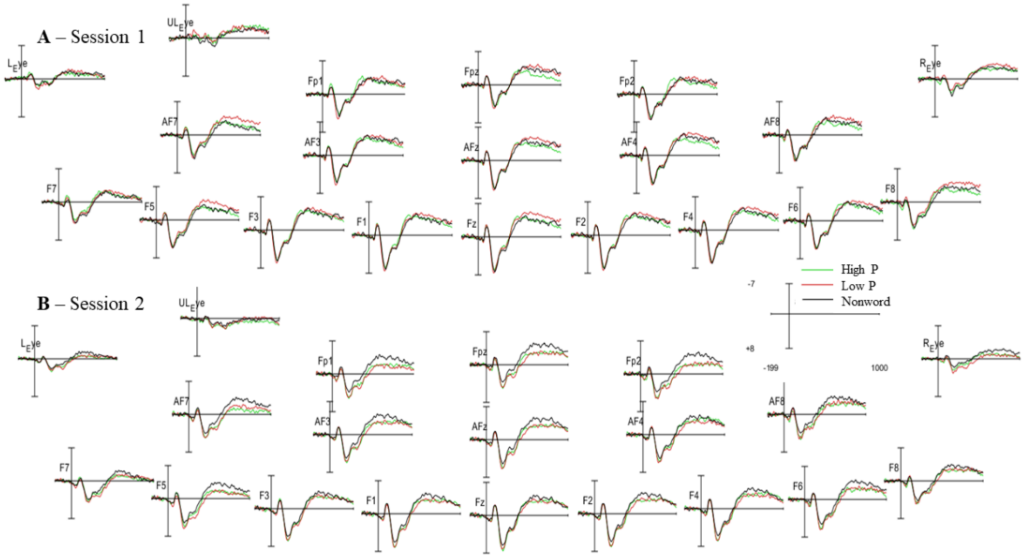
*ERPs of the Frontal Electrodes of the Control Group, Divided by Session and Phonotactic Category.*



*Note:* Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

**Figure 17**

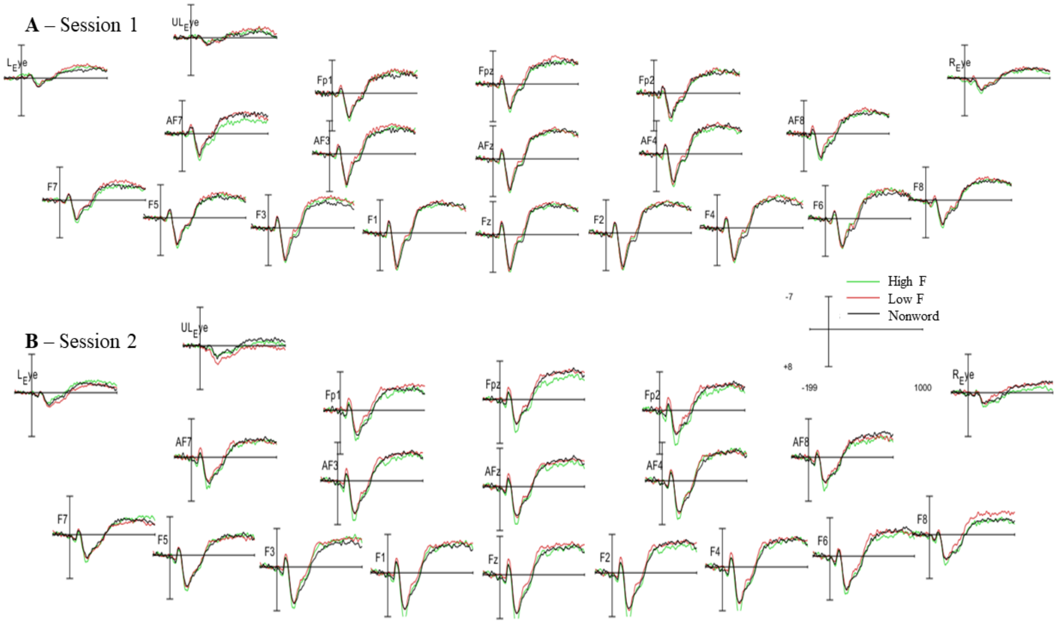
*ERPs of the Frontal Electrodes of the L2 Exposure Group, Divided by Session and Phonotactic Category.*



*Note:* Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

**Figure 18**

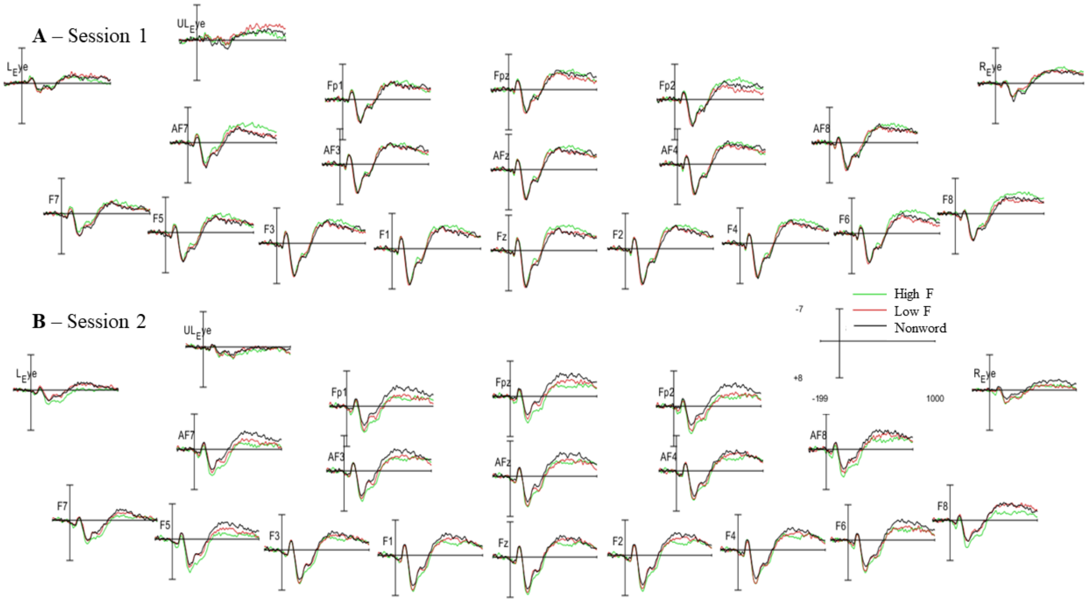
*ERPs of the Frontal Electrodes of the Control Group, Divided by Session and Frequency.*



*Note:* Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

**Figure 19**

*ERPs of the Frontal Electrodes of the L2 Exposure Group, Divided by Session and Frequency.*



*Note:* Y axis is in microvolts and plotted negative up. X axis is in milliseconds.

# Curriculum Vitae

| | |
|---|---|
| **Name:** | Émilie Rae Hoeppner |
| **Post-secondary Education and Degrees:** | University of Winnipeg<br>Winnipeg, Manitoba, Canada<br>BSc Hons. Neuroscience<br>Certificate in Applied Psychology<br>2017-2022<br><br>The University of Western Ontario<br>London, Ontario, Canada<br>Masters in Neuroscience<br>2022- Present |
| **Honours and Awards:** | Ontario Graduate Scholarship<br>2024-2025<br><br>Gold Medal (Honours Neuroscience)<br>The University of Winnipeg<br>2022 |
| **Related Work Experience** | Teaching Assistant<br>The University of Western Ontario<br>2022-2024<br><br>Grad Club Trivia Host<br>The University of Western Ontario<br>2023-2024 |

## Conference Publications:

Hoeppner, É. R., Aggarwal, A. S., Batterink, L J.., (April 2024) Harnessing Statistical Learning for Second Language Acquisition [Poster Presentation] Cognitive Neuroscience Annual Meeting 2024, Toronto, ON, CA

Aggarwal, A. S., Hoeppner, É. R., Batterink, L. J., (April 2024) Can statistical learning support speech segmentation of a natural language in adult learners? Cognitive Neuroscience Annual Meeting 2024 [Poster Presentation] Toronto, ON, CA

Hoeppner, É. R., Aggarwal, A. S., Batterink, L. J., (February 2024) Harnessing Statistical Learning for Second Language Acquisition [Poster Presentation] Lake Ontario Visionary Establishment Conference, Niagara Falls, ON, CA

Friesen, D.F., Kalirao, T., Hoeppner, É. R., & Desroches, A.S. (November, 2023). Language Expectations Influence Cross-language Activation in Bilinguals. [Poster Presentation]. Psychonomic Society, San Francisco, CA

Hoeppner, É. R., Aggarwal, A. S., Batterink, L. J., (March 2023) Harnessing Statistical Learning for Second Language Acquisition [Poster Presentation] Cognitive Neuroscience Annual Meeting 2023, San Francisco, CA, USA

Aggarwal, A. S., Hoeppner, É. R., Batterink, L. J., (March 2023) Can statistical learning support speech segmentation of a natural language in adult learners? Cognitive Neuroscience Annual Meeting 2023 [Poster Presentation] San Francisco, CA, USA

Hoeppner, É. R., (February 2023) Harnessing Statistical Learning for Second Language Acquisition [Conference Session] Neuroscience Research Day, The University of Western Ontario, London, ON, Canada

Hoeppner, É. R., (June 2022), The Women's Health Clinic Provincial Eating Disorder Recovery Program Improves Psychological Symptoms of Eating Disorders, [Poster Presentation] 2022 Canadian Evaluation Society Conference, Winnipeg, MB

Hoeppner, É. R., Desroches, A. S., (April 2022) Spoken Word Recognition in Bilinguals: The Effects of Explicit Between-Language Competition [Conference Session] Prairie Undergraduate Psychology Research Conference, Winnipeg, MB, Canada

Hoeppner, É., R., Hinds, A., (2021) Psychological Symptoms in Eating Disorder Recovery, Unpublished manuscript, The University of Winnipeg, Women's Health Clinic