8-13-2020 1:30 PM

The Role of Neural Entrainment in Statistical Learning

Jordan-Jerrica Mulgrew, *The University of Western Ontario*

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 Psychology

Recommended Citation

Mulgrew, Jordan-Jerrica, "The Role of Neural Entrainment in Statistical Learning" (2020). *Electronic Thesis and Dissertation Repository*. 7158.
https://ir.lib.uwo.ca/etd/7158

Abstract

Statistical learning (SL) refers to our ability to extract patterns from the environment. Research has long acknowledged the importance of SL in language; however, the neural mechanisms underlying SL remain largely unknown. One potential mechanism is neural entrainment, which refers to the tendency of endogenous neural oscillations to align with an ongoing rhythmic stimulus. In the context of SL, neural entrainment may align neural excitability to the ongoing structure of the speech stream, increasing sensitivity to underlying patterns and supporting the learning of word boundaries. This thesis tested the hypothesis that neural entrainment plays a causal role in SL by directly manipulating entrainment at specific frequencies cross-modally using a visual stimulus during learning. We found that boosting neural entrainment to match the frequency of the most informative moments of continuous speech (word onsets) resulted in better performance in an implicit SL task. These results support that neural entrainment plays a causal role in SL, as opposed to simply reflecting downstream effects of the learning process.

Keywords: neural entrainment, statistical learning, EEG, language acquisition, implicit learning

Summary for Lay Audience

Every day humans are bombarded by a wealth of sensory input that needs to be successfully navigated to make sense of the world. Through continuous exposure to this sensory input, humans can extract statistical relationships about stimuli in the world, a skill often referred to as statistical learning (SL). Although this skill is found across domains, it is thought to be particularly important for language acquisition. More specifically, SL is thought to play a key role in an important initial step to acquiring language – the learning of word boundaries. Within a language, syllables within words occur together more frequently than syllables that span a word boundary. Becoming sensitive to these statistical relationships between syllables may allow learners to discover word boundaries, especially when other cues such as pauses may be unavailable or unreliable. Although research supports the relationship between SL and word segmentation, the neural mechanisms contributing to SL remain largely unknown. One potential mechanism is neural entrainment, which refers to the tendency of brain activity (neural oscillations) to align with an ongoing rhythmic stimulus. This thesis investigated the potential role of neural entrainment in SL by using a visual stimulus to induce entrainment at frequencies congruent or incongruent with word onsets (boundaries) during learning. We found that boosting neural entrainment to match the frequency of word onsets resulted in better performance in an implicit task, which supports the potential causal role of entrainment in SL. Theoretically, this work helped to advance our understanding of the neural mechanisms contributing to SL. Practically, future work in this area could reveal novel ways to boost SL and improve language acquisition for adult second language learners and children with atypical language development. Moreover, since SL is found across domains, techniques used and discovered may inform other areas beyond language, such as visual processing.

Acknowledgements

I first want to thank my thesis supervisor, Dr. Laura Batterink, for her continued support, patience, and insight throughout this entire process. Dr. Batterink has been an immense source of knowledge and I cannot express how lucky and thankful I am to be one of her graduate students. I would also like to extend my thanks to all the individuals working in the Cognitive Neuroscience of Learning and Language lab, their assistance and company throughout the thesis process has been integral to the project. Further, I want to acknowledge and thank the individuals on my thesis committee who provided valuable insight to my thesis project, Dr. Marc Joanisse and Dr. Jessica Grahn. Special thanks to Nicolette Armstrong, the EEG maven, for all her help and support with the technical side of my project.

Finally, I want to thank my family and friends who have provided me with unfailing support and encouragement throughout my degree and the researching and writing of this thesis. This accomplishment would not have been possible without them.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

1 Introduction

Every day humans are bombarded by a wealth of sensory input that needs to be successfully navigated to make sense of the world. This sensory input is typically highly patterned, and humans have considerable expertise in pattern detection (Schapiro et al., 2014). Continuous exposure to this patterned input allows humans to, largely implicitly and automatically, extract statistical relationships about stimuli in the world. The ability to extract patterns from the environment is referred to as statistical learning (SL) and plays an important role across numerous areas of cognition (Sherman et al., 2020). For instance, SL has been shown to play a role in learning category-level relationships of complex real-world scenes (Brady & Oliva, 2008), higher-order spatial relations of complex shapes (Fiser & Aslin, 2001), and visuomotor sequences (Hunt & Aslin, 2001). It has also been shown to operate across auditory (Batterink, 2017; Saffran et al., 1999), visual (Fiser & Aslin, 2002), tactile (Conway & Christiansen, 2005), and multimodal domains (Mitchel & Weiss, 2011). Further, SL appears to occur across different ages, from neonates (Teinonen et al., 2009) and young children (Saffran et al., 1997), to younger and elderly adults (Campbell et al., 2012). Finally, evidence for SL has been found in animals other than humans, including cotton-top tamarins (Hauser et al., 2001) and rodents (Toro & Trobalón, 2005). Clearly, SL is ubiquitous and likely involved in many different cognitive processes, making it an important concept to study.

1.1 Statistical Learning in Language

While SL plays a role in many different aspects of cognition, it is thought to be especially important for language acquisition (Saffran, Newport, et al., 1996). Consider the following scenario: when we hear a conversation in a language we understand, we easily perceive speech as a sequence of discrete words. However, this is not the case when we listen to a conversation in an unknown language. Instead, we tend to perceive speech as one long continuous stream of phonemes, seemingly spoken at a rapid pace. Unlike in a familiar language, it is no longer clear where one word ends, and the next word begins. One reason why we are unable to parse an unknown language into individual words is that word boundaries are not reliably marked by acoustic cues, such as pauses or intonation. Further, although some languages may stress word-initial or word-final syllables, many languages do not; an examination of 444 languages found that over half did not have fixed stress patterns (Saffran, Newport, et al., 1996). If there are no reliable acoustic cues to mark word boundaries, then how are we able to segment continuous speech into component words? As language learners, how are we able to go from perceiving rapid-fire syllable units to whole, meaningful words?

Over the past two decades, researchers intrigued by this scenario have revealed that humans may be able to access deeper statistical patterns when “surface patterns” such as pauses and intonation are unreliable (Romberg & Saffran, 2010). Across a language sample, neighbouring syllables within words occur together more frequently than neighbouring syllables that span word boundaries. Becoming sensitive to these statistical patterns may allow us to discover word boundaries in continuous speech (Saffran, Newport, et al., 1996). These syllable patterns can be quantified in terms of “transitional probability”, the probability of occurrence of Y given that X has occurred (e.g., frequency of $X+Y$/frequency of X). A high transitional probability indicates that if X has occurred, Y is strongly predicted to occur, whereas a low transitional probability indicates a weaker relationship between the two. To illustrate, in the phrase “pretty baby”, the word internal syllable pairs (e.g., *ba* + *by*) occur together more frequently than the word external syllable pairs (e.g., *ty* + *by*) across the English language. To a fluent English speaker who has already learned these words, it is obvious that this phrase is made up of two distinct words. However, to someone who is learning English, differences in the transitional probabilities of neighbouring syllables could be an implicit cue to word boundaries (e.g., *ty* + *by* is not a word but a boundary).

To test whether learners could use transitional probabilities to identify words, Saffran, Newport, et al. (1996) exposed adult participants to a continuous artificial spoken language with a hidden statistical structure. The artificial language contained six trisyllabic words, played without any pauses or cues indicating where one word ended and the next began. As such, the only “cues” to word boundaries were the transitional probabilities between the syllables, with the transitional probabilities of syllables spanning a word boundary always lower than that of syllables within a word. After listening to the continuous speech stream for 21 minutes, participants completed a forced-choice test between words from the language and foil words (syllables from the language combined in a different order). Participants discriminated between words and foils at a level significantly better than chance, providing evidence of SL. Another study by Saffran and colleagues tested this ability in 8-month old infants, using a looking preference paradigm after exposure to assess learning (Saffran, Aslin, et al., 1996). The authors found that infants showed a novelty preference for foil items, indicating they had habituated to actual words, which provides evidence that they were able to learn the words in the stream. These results support the idea that SL allows learners to discover word boundaries simply through passive exposure to linguistic input by becoming sensitive to transitional probabilities between syllables.

Since these seminal studies by Saffran and colleagues, researchers have continued to corroborate the relationship between SL and language acquisition, showing it occurs across different ages and generalizes to natural language. In terms of word segmentation, studies have shown that neonates and children are able to use transitional probabilities to discover word boundaries similarly to adults, indicating SL is relatively stable across development (Choi et al., in press; Erickson & Thiessen, 2015; Saffran et al., 1997; Teinonen et al., 2009). To understand if findings from SL studies could generalize to a real-world setting, Pelucchi et al. (2009) conducted a natural language study with 8-month old English monolingual infants. Here, infants were exposed to Italian sentences where transitional probabilities between syllables were higher than those between word boundaries. A preferential looking task revealed the infants had a significant preference to look towards actual words from the language rather than the novel words, suggesting they became sensitive to the statistical relationships of the speech stream. Although preferential looking tasks using artificial language stimuli often find that infants look towards novel items more, studies of natural languages tend to find the opposite—that they look towards familiar items. These results suggest that SL findings can generalize to natural language and is found across development, making it a prime candidate for language acquisition, and thus an important ability to study.

For over 20 years, post-exposure behavioural measures have been used to assess learning outcomes in SL studies. In a typical SL paradigm, participants listen to a continuous speech stream of repeating trisyllabic words in which transitional probabilities serve as the only cue to word boundaries. After this exposure period, participants are given behavioural tasks that require them to discriminate between actual words from the language and foil words. For example, a popular task is the two-alternative forced choice task (2AFC test) in which participants hear an actual word from the language and a foil word and need to choose which word was from the language they heard. Another task includes rating words and foil words in terms of how familiar they seem (familiarity rating task). Even though these measures are quick and easy to administer, they are not without their limitations. First, these measures require an overt behavioural response, which can be difficult to reliably obtain in certain populations, such as infants or patients. Another issue is that post-exposure methods are sensitive to other cognitive processes such as long-term memory. Therefore, it is unclear whether results are solely reflecting learning or are measuring something else such as a participant's long-term memory, or a combination of the two. Furthermore, post-exposure measures are unable to tap into the time course of learning, which limits our understanding of the temporal dynamics involved in the learning process. As such, to move this body of work forward, clear and versatile measures of learning are of critical importance.

1.2 Neuroimaging Studies of Statistical Learning

Neuroimaging methods have been used to investigate SL and offer several advantages over traditional post-exposure behavioural measures. Neuroimaging techniques allow researchers to examine changes in brain activity in real-time without needing to rely on overt responses. Therefore, researchers can look at how SL unfolds over time, if there are any differences in how it unfolds between participants, and these methods can be used to study populations where behavioural measures may be difficult to use. Additionally, these methods offer the opportunity to examine which brain areas and neural mechanisms are involved in SL. Overall, neuroimaging techniques are a promising avenue to measure and better understand the neural mechanisms supporting SL.

In order to overcome the limitations of behavioural only measures, studies have attempted to examine the neural correlates of linguistic SL using functional magnetic resonance imaging (fMRI; Karuza et al., 2013; McNealy et al., 2006). For example, Karuza et al. (2013) measured brain activity using fMRI while participants listened to blocks of different auditory streams: a forward stream of trisyllabic words (typical SL continuous language stream) and a backward stream (same stream but played backwards). The use of a backward stream served as a control for auditory stimulation; piloting had revealed participants were unable to extract the statistical regularities when the stream is played backwards, but the stream still contains similar low-level acoustic properties as the forward stream. After each block, participants completed 2AFC tests to measure learning. The findings indicated the forward stream resulted in greater activation in the left superior temporal gyrus compared to the backward stream. Furthermore, when looking at participants’ learning across time, a significant cluster of activation appeared in the left inferior frontal gyrus in the forward stream condition, but not in the backward condition. These results are important as they were able to show that neural activation in auditory and language regions of the brain covaried with changes in performance over time (across blocks). This suggests that changes in brain activity may be able to capture learning as it is occurring, making neuroimaging techniques a powerful tool for examining SL.

In addition to fMRI, several other studies have used electroencephalography (EEG) and magnetoencephalography (MEG) frequency-tagging approaches to investigate SL (Batterink, 2017; Batterink & Paller, 2019; Batterink & Paller, 2017; Buiatti et al., 2009; Farthouat et al., 2017). These approaches offer better temporal resolution than fMRI and are well suited to capturing the neural response to a continuous sensory stream. Frequency-tagging approaches take advantage of the *steady-state response*, in which the brain’s electrophysical response entrains, or aligns, to the frequency and phase of a periodic stimulus (Buiatti et al., 2009; Thut et al., 2012). Due to the steady-state response, the frequency of a periodic stimulus can serve as a “tag” for stimulus-related brain responses. Frequency-tagging can be measured by computing either power or phase-locking of recorded neural activity as a function of frequency. These techniques represent a good tool to look at SL as they have a high signal to noise ratio and good temporal resolution (Farthouat et al., 2017).

The first study to use a frequency-tagging approach to examine SL was conducted by Buiatti and colleagues (2009). Participants were exposed to four different artificial language streams: a continuous structured stream, a continuous random stream, and then both streams with a 25ms pause in-between each syllable. Differing from the typical SL paradigm, words in the stream belonged to groups following a nonadjacent “AXC” structure, in which the first syllable and third syllable were always the same, but the second syllable was different for each word within the group. Participants were explicitly told the language contained words they needed to discover, and the streams were broken up into three blocks, separated by breaks in which participants wrote down any words they had discovered on a piece of paper. After each stream ended, participants completed a 2AFC test.

At the syllable frequency, the authors found peaks in power for both random streams, whereas there appeared to be a suppression in power for both structured streams (Buiatti et al., 2009). At the word frequency, a clear peak was found only for the structured stream with 25ms pauses. This peak cannot be due solely to the 25ms pauses because there was no similar peak in the random stream, which also contained 25ms pauses. The authors suggest that the addition of pauses may alter perception to allow computations to focus on smaller segments, unbeknownst to the participants. Further, results indicated that power at the word frequency was significantly correlated with the number of words correctly reported after each block. However, power at the word frequency did not correlate with performance in the 2AFC test. Interestingly, the authors also found that the increase in power at the word frequency (trisyllabic) in the 25ms pause structured stream was accompanied by a decrease in power at the bisyllabic frequency. This result suggests becoming sensitive to the hidden trisyllabic word structure results in inhibition at frequencies corresponding to different word lengths. Overall, these results indicate that EEG can index SL, with changes in power spectra suggesting an increased sensitivity to the hidden structure of the language and a decreased sensitivity to alternate structure options (e.g., single syllables or bisyllabic chunks).

Although not language specific, a recent study used MEG to look at frequency-tagged responses in an auditory SL paradigm involving pure tones (Farthouat et al., 2017). Participants listened to two streams of pure tones: a structured stream in which tones were concatenated to create four tritones repeating in a pseudo-random order, and a random stream in which tones were presented in a random order. Therefore, the only cues to tritone boundaries were transitional probabilities, with transitional probabilities of tones spanning a boundary always lower than that of tones within a tritone (for the structured stream). The authors found that power at the tritone frequency increased linearly over time in the structured stream when compared to the random stream and became significantly different from the random stream after only three minutes of exposure. Further, while both streams elicited a frequency-tagged response at the tone frequency, power at this frequency was significantly lower in the structured stream. Although behavioural performance was at chance level, the neural measures showed evidence of learning, suggesting that neural frequency-tagging measures can be more sensitive to SL than behavioural measures.

Additional evidence supporting the utility of these approaches comes from a series of studies demonstrating that an EEG measure of neural entrainment can track SL during the learning process (Batterink, 2020; Batterink & Paller, 2019; Batterink & Paller, 2017). Batterink and Paller (2017) exposed participants to two different artificial language speech streams in a counterbalanced order while EEG was recorded. Following the typical SL paradigm, the “structured” stream contained syllables concatenated to create four trisyllabic words in which transitional probabilities within words was higher than between words. The “random” stream contained syllables presented in a pseudo-random order, such that there was no higher-order structure to the stream. After exposure, SL was measured using explicit tests such as a familiarity rating task and a 2AFC task, as well as an implicit response time (RT) task. Neural entrainment was quantified using inter-trial phase coherence (ITC), a measure of event-related phase-locking, which indicates the extent to which EEG phase is consistent across trials (in this case, across word presentations). A composite measure called the word learning index (WLI) was used to quantify participants’ sensitivity to the trisyllabic structure of the speech stream, relative to their sensitivity to the individual syllables (WLI = Word-ITC/Syllable-ITC). Across time, in the structured condition, neural entrainment to the syllable frequency decreased while entrainment to the word frequency increased. Logically, this resulted in an increased WLI as a function of exposure to the language. Importantly, there was no significant change in WLI over time during the random stream. Further, WLI predicted performance on the implicit learning task, with a higher WLI leading to better performance.

The results from Batterink and Paller (2017) suggest that neural entrainment is an effective tool for tracking SL as it is happening, as changes in entrainment reflect increases in sensitivity to the structure of a trisyllabic language over time, consistent with learning. Clearly, neural entrainment can be used to track SL as it is occurring, but what does the relationship between neural entrainment and SL mean mechanistically? On one hand, these prior correlational results could indicate that neural entrainment is simply reflecting the downstream consequences of the underlying mechanism of SL. On the other hand, it is possible that neural entrainment itself might be the mechanism underlying SL.

1.3 Neural Entrainment and Statistical Learning

To understand why neural entrainment may underlie SL, a deeper understanding of neural entrainment is needed. Broadly speaking, neural entrainment is the synchronization of an oscillatory system(s) to an external rhythm (Lakatos et al., 2019). More specifically, the rhythmicity of an external stimulus induces a series of phase-resets according to its rhythm; this series of phase-resets act to modulate the oscillatory system, causing it to stabilize and sync with the external rhythm over time. This differs from a simple phase-reset in which a single external stimulus interrupts an ongoing cycle and forces phase modulation in an oscillatory system at a specific point in time. Rather, because the external stimulus is rhythmic, it continuously induces phase-resetting, leading to an overall shift in the oscillatory system to match the rhythm. Synchronization between neural oscillations and external stimuli may allow for phases of high neural excitability to align with stimulus events, enabling more efficient processing (Bauer et al., 2020; Thut et al., 2012). Considering that language input is structured, by aligning neural signals with the speech stream, neural entrainment may allow for enhanced processing at the most informative moments of the speech signal. In the context of a typical SL paradigm, neural entrainment may function to align excitability to the ongoing statistical patterns of the speech stream, allowing individuals to pick up on the hidden words.

1.4 Causal Role of Neural Entrainment in Other Language Areas

Supporting the possibility that neural entrainment may contribute to SL, recent evidence has shown that neural entrainment plays a causal role in other aspects of language processing, such as speech comprehension (Kösem et al., 2017; Peelle & Davis, 2012; Riecke et al., 2017;

Wilsch et al., 2018). These studies use non-invasive brain stimulation, specifically transcranial alternating current stimulation (tACS), to directly manipulate entrainment and study causal effects. Riecke et al. (2017) found that entrainment to speech-envelope information resulted in better performance on a speech intelligibility measure. This study involved the use of a novel brain-stimulation technique termed “envTCS”, which is the application of an electric current carrying speech-envelope information, over auditory cortex. Participants listened to sentences in which critical amplitude speech-envelope information was largely attenuated. Instead, some participants received the speech-envelope information through envTCS while listening to the attenuated sentence. The results indicated participants who received envTCS were able to verbally repeat more words of the sentence. These results suggest that neural tracking of the speech-envelope is important in being able to discern speech, as when tracking is intact, people demonstrate a greater ability to reproduce attended speech. More evidence for the causal role of neural entrainment to the speech envelope in speech intelligibility comes from Wilsch et al. (2018), who adopted a similar approach as Riecke et al. (2017). In this study, each participant listened to sentences masked in noise while receiving speech-envelope information through tACS to auditory regions. Again, the results indicated that entrainment to the speech-envelope resulted in better sentence comprehension.

Along a similar vein, Kösem et al. (2017) found that neural entrainment to speech-envelope dynamics directly influenced perception of ambiguous words. The authors presented sentences of varying speech rates, where the beginning of the sentence (carrier window) was manipulated to be fast or slow, and the end of the sentence (target window) was presented at normal speed. Participants’ task was to report the last word of each sentence. Critically, the last word in the target window could be interpreted as two different words depending on the perception of an ambiguous vowel (e.g., short versus long /a/). The results showed that perception of the last word was influenced by the contextual speech rate of the carrier window; participants were biased to perceive the word with a long /a/ after a fast speech rate and a short /a/ after a slow speech rate. Given this growing body of work showing a causal role of entrainment in speech intelligibility, a reasonable hypothesis is that entrainment may also play a causal role in SL, especially when considering studies support a correlation between SL and neural entrainment (Batterink, 2017; Batterink & Paller, 2017).

1.5 Cross-modal Entrainment Through Visual Stimulation

As evidenced by the prior studies, neural entrainment can be induced using non-invasive brain stimulation methods. However, research also supports the idea that neural entrainment can be induced through cross-modal sensory stimulation, which has certain pragmatic and logistical advantages over direct brain stimulation methods (Bauer et al., 2020; Park et al., 2016; Romei et al., 2012). In the context of an auditory SL paradigm, a promising method of cross-modal entrainment is through visual stimulation. Cross-modal interactions between these domains is supported by animal research which finds visual information is represented in early auditory areas and directly impacts auditory activity (Atilgan et al., 2018; Kayser et al., 2008). Additional evidence for cross-modal effects between visual and auditory domains comes from behavioural studies of humans (Park et al., 2016; Romei et al., 2012). For example, an EEG study found that viewing a video of lip movements enhanced entrainment in auditory areas (Park et al., 2016). Further, a study that used TMS to probe visual cortex excitability found that phosphene perception was phase-aligned to the presentation of a short sound (Romei et al., 2012).

Not only can entrainment be induced cross-modally, but recent work also suggests that visual stimuli can successfully induce entrainment in a way that could be used to address causality. More specifically, Notbohm et al. (2016) investigated whether rhythmic light stimulation resulted in a pattern of entrainment that could be predicted by a theoretical concept of entrainment, *the Arnold Tongue*, or simply by the superposition of steady-state responses. In order to use visual stimuli to causally investigate effects of entrainment, the visual stimuli need to truly manipulate an internal oscillator, rather than simply producing intermittent phase-locking. The Arnold Tongue predicts the degree of entrainment of an internal oscillator and an external driving stimulus depending on the amplitude of the external stimulation (stimulus intensity) and frequency of the external stimulation. Entrainment is more likely to occur the closer the frequency of the external stimulation gets to the intrinsic frequency. However, increasing the stimulus intensity can widen the window of entrainment around the intrinsic frequency, meaning more distant stimulation frequencies can entrain the internal oscillator.

To test whether visual steady-state responses follow the pattern predicted by the Arnold Tongue or intermittent phase-locking, participants were presented with 35 different rhythmic light stimulation conditions that were the combination of five levels of light intensity and seven different stimulation frequencies distributed in 1 Hz intervals around the individual alpha frequency (Notbohm et al., 2016). Critically, the rhythmic stimulation was preceded by a jittered frequency that had the same intensity and average number of flashes. If steady-state responses of visual stimuli are the result of superposition, then phase-locking to jittered sequences and rhythmic stimulation would not differ, nor would the responses resemble the pattern of entrainment predicted by the Arnold Tongue. The results indicated that phase-locking followed the pattern predicted by the Arnold Tongue. Further, within the Arnold Tongue, phase-locking from rhythmic stimulation was significantly stronger than that of the jittered sequence; outside the Arnold Tongue there was no significant difference in phase-locking. Overall, these results strongly support that steady-state responses invoked by visual stimuli are the result of entrainment of an internal oscillator rather than superposition of neural responses, suggesting visual stimulation can be used to probe causal effects of entrainment.

1.6 Current Study

The goal of the current study is to test the hypothesis that neural entrainment plays a causal role in SL by directly manipulating entrainment during SL. As such, we used a repetitive visual sensory stimulus to manipulate neural entrainment during the exposure period, and then tested whether this manipulation influenced post-learning measures of SL. If entrainment plays a causal role in SL, we expect that SL should be enhanced when neural entrainment is boosted at the word frequency.

In the current study, adult native English speakers listened to a continuous speech stream of repeating trisyllabic nonsense words while viewing a repetitive visual stimulus designed to elicit neural entrainment at different frequencies, assigned between participants. Specifically, participants viewed a looping video of a water drop falling off a leaf that either matched the hidden word frequency (“match” condition, 1.1 Hz: 3-syllable units) or was faster than the hidden word frequency (“mismatch” condition, 1.6 Hz: 2-syllable units). Furthermore, a third control group was included, in which participants viewed a static image of the leaf (“static” condition). The use of a naturalistic video was meant to conceal the true nature of the visual manipulation and reduce the potential of the stimulus being used strategically to segment the stream. During this exposure period, EEG was recorded to index neural entrainment at the word and syllable frequencies. After exposure to the speech stream, participants completed two behavioural tests of learning: an explicit task that examined participants’ ability to discriminate between words from the language and foil items, and an implicit task in which participants made speeded responses to target syllables occurring across different triplet positions (first, second and third).

The main hypothesis of the study is that when visual stimulation aligns with the hidden word frequency, neural entrainment at the word frequency should increase, resulting in better performance on post-exposure measures of SL. In contrast, visual stimulation at an incongruent frequency should interfere with neural entrainment at the word frequency, impairing SL performance. We expect participants in the static group should show intermediate performance on post-exposure tests, above that of the mismatch group but lower than performance in the match group, as they are not benefitting from an entrainment “boost” at the word frequency nor experiencing interference due to increased entrainment at a competing frequency. Further, we expected to observe a shift in entrainment from the syllable frequency to the word frequency over the exposure period, reflecting learning. We anticipated that this shift should be strongest for participants in the match condition, followed by the static condition, with those in the mismatch condition showing the weakest shift across time. Finally, we expected to see a positive correlation between word entrainment and behavioural performance.

Chapter 2

2 Methods

2.1 Participants

Participants in the current study included native English-speaking adults between the ages of 18 to 35, recruited from the University of Western Ontario’s research participant pool and through responses to posters placed around campus (N = 60; 39 females, mean age = 19.15). All participants provided written consent to take part in the study and received either course credit or monetary compensation at a rate of $14/hr. Participants completed a screening questionnaire, which included questions about native language, neurological history, vision, hearing, and medication use, either via e-mail or through the SONA system before participating. All participants in the current study had normal hearing and normal or corrected-to-normal vision, as well as no prior history of neurological disorders. Participants were randomly assigned to one of the three experimental conditions (match, mismatch, static), determined by the order run. There were no significant differences in age or sex of the participants across the three conditions (age: $F(2,53) = 0.50$, $p = 0.61$; sex: $X^2(4, N = 60) = 0.45$, $p = 0.98$,).

2.2 Stimuli

Auditory stimuli.

Twelve syllables generated using an artificial speech synthesizer with a sampling rate of 44100 Hz were concatenated to create 4 “nonsense” words (*pautoki*, *mailone*, *nurafi*, *gabalu*; taken from Batterink & Paller, 2017). Critically, these words were designed to have an internal transitional probability of 1.0 and an external transitional probability of .33. That is, syllables within a given word always occurred together, but the first or last syllable of that word could be followed by the first or last syllable of any of the other three words. To illustrate, for the word *mailone*, *mai* was always followed by *lo* and *lo* was always followed by *ne* (internal probability = 1.0), whereas the final syllable *ne* was equally likely to be followed by the beginning syllable of any of the other trisyllabic words (e.g., *nu*, *pau*, or *ga*; external probability = .33).

To create the continuous speech stream, these nonsense words were presented in a predefined pseudorandom order, with the constraint that the same word did not repeat consecutively, at a rate of 300ms per syllable (3.3 Hz). The total speech stream contained 1,200 syllables (400 words), with each word represented an equal number of times throughout the stream (100 presentations for each word).

Visual stimulus.

The visual stimuli were created by capturing frames (still images) from a slow-motion looping video, which depicted a droplet of water falling off a leaf and hitting the surface of a body of water below. The video was obtained from an online stock image website called Deposit Photos (https://depositphotos.com/). Image sets were created by calculating the total number of images needed for a full video cycle (i.e., the time between the appearance of a new droplet on the leaf until it hits the water) at the desired frequency, given the monitor refresh rate (60 Hz). For the 1.1 Hz video, a full cycle consisted of 54 images; for the 1.6 Hz video, a full cycle consisted of 72 images. In the static condition, a single image of the leaf was presented. For all three conditions, images were presented at a rate of 60 images per second.

2.3 Procedure

First, participants signed a consent form and completed a demographics form that provided detailed information about our screening criteria. After this, participants were fitted with a 64-electrode elastic cap from the Biosemi system (Biosemi, Amsterdam, the Netherlands). Four additional exogenous electrodes were placed, one behind the left and right mastoids, as well as one at the outer canthi and bottom eyelid of the left eye. Participants were seated in a sound attenuated booth at a comfortable viewing distance (approximately 70 cm) from the monitor. The speech stream was presented at a comfortable listening level using a pair of speakers connected to the computer, positioned in front of the participant (to the right and left of the monitor).

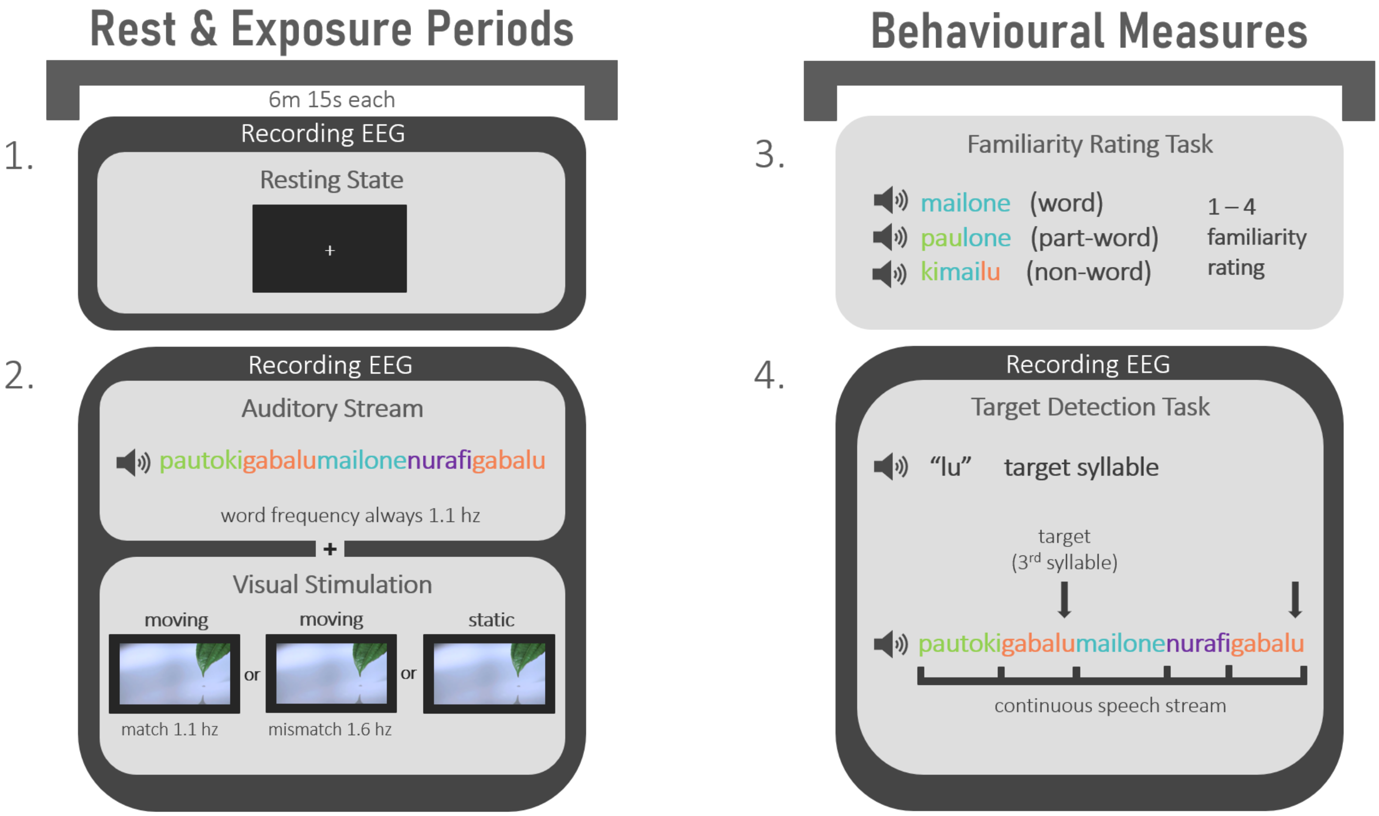


Figure 1. Schematic showing the study procedure. Here, numbers refer to the order in which tasks were completed.

Resting state.

Participants were instructed to relax and maintain focus on a fixation cross in the center of the screen for 6 minutes and 15 seconds while their EEG was recorded. No auditory stimuli were presented during this time.

Exposure task.

Participants completed one of the three exposure conditions. Across all three conditions, participants were exposed to the continuous speech stream for a total of 6 minutes while viewing the visual stimuli. The visual stimuli began first, with the auditory stimuli beginning 15 seconds later, for a total task time of 6 minutes and 15 seconds. In the match and mismatch conditions, participants watched a looping video of a droplet of water falling off a leaf and hitting the surface of a body of water below. Participants were told that the purpose of the experiment was to test whether viewing a video from nature helps people to relax while listening to a nonsense language. Further, participants were instructed to visually fixate on the location in which the droplet of water hits the surface below. To avoid drawing attention to the true purpose of the visual manipulation, this instruction was phrased to be for the purpose of tracking eye movements, as they cause large distortions in data. As mentioned previously, the match and mismatch conditions differ in the temporal relationship between the video and the syllables in the speech stream. In the match condition (n = 20), the onset of each hidden trisyllabic nonsense word aligned temporally with the water droplet hitting the surface of the water below (video frequency 1.1 Hz, word frequency 1.1 Hz). For the mismatch condition (n = 20), the water droplet did not reliably hit the surface of the water at the onset of each word in the stream; it was equally likely to co-occur with syllables across the three triplet positions (video frequency 1.6 Hz, word frequency 1.1 Hz, n = 20). In the static condition (n = 20), participants viewed a single static image and were instructed to maintain fixation on the tip of the leaf.

Rating task.

Following the exposure task, participants completed a familiarity rating task to assess explicit knowledge of the nonsense words. For each trial, participants heard a trisyllabic item and had to indicate via a button press how familiar that item sounded based on the language they had just heard (with 1 being very unfamiliar and 4 being very familiar). The item could be a word from the language (e.g., *mailone*), a part-word consisting of a syllable pair from the language and an additional unpaired syllable (e.g., *maito* + *ki*), or a non-word consisting of syllables from the language that never occurred together (e.g., *ki* + *mai* + *lu*). As in the exposure period, the syllable presentation rate was kept at 300ms, with the rating prompt appearing on the screen 770ms after the final syllable of a given word. The next trial began 1500ms after response selection. In total, the task consisted of 12 trials, including 4 words, 4 part-words, and 4 non-words.

Target detection task.

Following the rating task, participants completed a speeded target detection task to assess implicit knowledge of the nonsense words. Each trial consisted of a short speech stream containing the four trisyllabic words, each presented four times in pseudorandom order, and presented at a rate of 300ms per syllable, just as in the original speech stream. For each trial, participants were instructed to press a button every time they heard a specific target syllable in the speech stream. Participants were told to respond as quickly as possible while still maintaining accuracy. At the beginning of each trial, participants pressed a button to hear the target syllable and then started the speech stream via a second button press. Throughout the trial, the phonetic spelling of the syllable remained on the screen to remind participants which target syllable they were listening for. In total, each syllable in the language served as a target 3 times (12 syllables, 36 streams in total). Furthermore, each stream contained 4 target syllables, resulting in 48 trials for each syllable-position condition (word-initial syllable, word-medial syllable, and word-final syllable). This task is considered to tap into participants' implicit knowledge of the words as participants can show RT time prediction effects (i.e., faster RTs to predictable syllables) even in the absence of conscious or explicit knowledge of the words.

Post-task interview.

Following the behavioural tasks, participants completed an oral post-task interview to assess their explicit knowledge and awareness of the artificial language and the visual manipulation (see Appendix G). The questionnaire began with open-ended questions to avoid prompting participants to presence of a visual manipulation, with later questions providing more information about the manipulation and asking more specifically about participants’ awareness of the link between visual and auditory stimuli. Further, a question asking whether they noticed any words or patterns in the audio stream was included to see if participants could explicitly recall any of the hidden nonsense words. Participants’ responses were recorded on paper by the experimenter.

2.4 Analyses

Behavioural data analysis.

Rating task. Average familiarity ratings for words, part-words, and non-words were calculated for each participant. For this task, we anticipated that participants should rate words as most familiar, followed by part-words, with non-words rated as least familiar. Additionally, we hypothesized that the match group would show the greatest discrimination ability in this task, followed by the static group, and lastly the mismatch group. Rating scores were assessed using a mixed repeated measures ANOVA with word type (word, part-word, non-word) as a within-subjects factor and condition (match, mismatch, static) as a between-subjects factor. As a composite measure for further correlational analyses, a “rating score” was calculated for each participant by taking the average score to words and subtracting the average score for part-words and non-words. For this measure, a score of 3 would indicate perfect sensitivity, with all values above 0 providing evidence of learning (Batterink & Paller, 2017).

Target detection task. Adopting the same criterion as previous studies in our lab (Batterink & Paller, 2017), responses that occurred before 0ms or after 1200ms were considered false alarms and were not included in further analyses. RTs were analysed using a linear mixed effects model. Given the length of this task, it is possible that performance may either improve (due to online learning) or decline (due to fatigue) as a function of trial number. Furthermore, previous work by Batterink (2017) found that RTs to target syllables increased for targets occurring later in a given stream. To account for these sources of variance, RTs were modeled using a linear mixed effects model, including fixed effects of triplet position (word-initial, word-medial, word-final), condition (match, mismatch, static), trial (1 – 36), stream position (4 – 48, i.e., in which position the target occurred within a single trial), and the interaction between triplet position and condition. Subject intercept was included as a random effect to account for pre-existing differences between participants.

For this task, we expected a linear trend in RTs, with participants responding slowest to word-initial syllables, followed by word-medial, and fastest to respond to word-final syllables. This effect was tested by examining the main effect of syllable position. Furthermore, if entrainment plays a role in SL, we would expect participants in the match group to show the largest RT prediction effect in this task, followed by the static group, with the mismatch group showing a similar or smaller effect than the static group. These hypothesized differences in RT across groups were tested by examining the interaction between condition and triplet position.

As a composite measure for later correlational analyses, mean RTs for each syllable position (word-initial, word-medial, word-final) were calculated for each participant, and a relative RT score was created by taking the mean RT for word-initial syllables, subtracting the mean RT for word-final syllables, and dividing this by the mean RT for word-initial syllables ($(RT_1 – RT_3)/RT_1$). This relative RT score accounts for baseline differences in RT (e.g., some participants have faster RTs than others, regardless of condition). If participants have implicitly learned the nonsense words, then we would expect faster RTs for word-final syllables as they are more predictable. Therefore, greater implicit learning of the nonsense words would be indicated by a larger relative RT score.

EEG recording and analysis.

EEG was recorded during the exposure period. EEG was recorded at a sampling rate of 512 Hz using a 64-channel Active-Two Biosemi system (Biosemi, Amsterdam, the Netherlands), set up according to the 10/20 system. Additional electrodes were placed around the participant's left eye (one placed at the outer canthi and one on the bottom eyelid) as well as on the left and right mastoids. During recording, signals were recorded relative to the Common Mode Sensor (CMS) active electrode and then re-referenced offline to the average of the left and right mastoid electrodes.

All EEG analyses were conducted using EEGLAB and ERPLAB. First, bad channels were identified visually and interpolated, resulting in an average of 1.3 channels per participant being interpolated (*SD* = 1.2). Next, data were band-pass filtered using a Butterworth filter with a 0.1 to 30 Hz cut-off. The continuous data were visually inspected and artifacts (e.g., long periods of muscle activity, abnormal eye movements) were manually rejected. After artifact rejection, data were time locked to the onset of each word and epoched into overlapping windows of 0 to 10.8s, relative to word onset, corresponding to 12-word durations per epoch.

Neural entrainment to the statistical structure of the language was measured by computing the inter-trial coherence (ITC) at the syllable and word frequencies. ITC is a measure of event-related phase-locking that ranges from 0 to 1. In this measure, 0 indicates non-phase locked activity and 1 indicates strictly phase-locked activity. ITC from 0.2 to 5.0 Hz (44 linear-spaced frequencies) was calculated using a continuous Morlet wavelet transformation, computed with the *newtimef* function in EEGLAB. A scaling factor of 0.5 was used, such that 1 cycle was used in the wavelet transformation at the lowest frequency (.2 Hz), and 11.25 cycles were used at the highest frequency (5.0 Hz).

Topographic plots of entrainment at our frequencies of interest revealed strong entrainment at the word frequency in occipital areas for the match group. Although we assume our visual manipulation should result in cross-modal entrainment effects beyond occipital regions, strong entrainment in visual areas is nonetheless to be expected, given that the stimulation was visual. Since we were interested in neural entrainment differences between groups over auditory regions, we wished to exclude visual entrainment effects and more directly compare neural entrainment effects over auditory regions specifically. As such, we selected a subset of electrodes that showed maximal entrainment at our frequencies of interest in the static group, in which there was no dynamic visual stimulus. This subset of electrodes was used for all groups in all subsequent EEG analyses except for the exploratory correlation analysis.

If participants became sensitive to the statistical structure of the language, then we would expect to see higher entrainment at the word frequency relative to the syllable frequency. As mentioned in the Introduction, this trade-off in neural entrainment has been quantified using a measure called the Word Learning Index (WLI; Batterink & Paller, 2017). WLI is computed by dividing ITC at the word frequency by ITC at the syllable frequency (WLI = Word-ITC/Syllable-ITC); a higher value in this measure would indicate stronger entrainment at the word frequency relative to the syllable frequency. A positive increase in the WLI over time represents the shift from perceiving individual syllable units to perceiving whole words, an indication of SL. The WLI was computed across the electrode subset using our respective word and syllable frequencies (1.1 Hz and 3.3 Hz). To assess whether the visual manipulation resulted in significant differences in entrainment, three separate ANOVAs were used to compare neural entrainment at our frequencies of interest (syllable and word frequency) as well as WLI, with group as a between-subjects factor. We hypothesized the match group would experience higher entrainment to the word frequency relative to the syllable frequency, resulting in a larger WLI than the static and mismatch groups. Further, the mismatch group was expected to entrain less strongly at the word frequency and more strongly at the syllable frequency, leading to a smaller WLI. This is because the visual stimulus should interfere with the ability to lock onto the word frequency, which might bias perception towards the syllable frequency.

Additionally, exploratory correlational analyses were conducted to examine the relation between the WLI across all 64 scalp electrodes and behavioural performance, across all participants, as well as the match group alone. Given that our entrainment manipulation is cross-modal, we were interested in examining the relationship between entrainment and behaviour across brain regions. If our manipulation results in strong entrainment-behaviour correlations in visual areas rather than frontal or auditory regions, it might suggest a stronger role of bottom-up entrainment. In contrast, stronger correlations between behaviour and entrainment in auditory or frontal regions might suggest there are more complicated interactions between primary sensory areas and domain-general high-level regions.

Finally, linear mixed effects models were used to examine neural entrainment at the frequencies of interest over time, and whether this time course differs as a function of condition. As such, ITC at our entrainment frequencies were calculated for each epoch in overlapping bundles of 100 epochs, starting at the second epoch. A bundle size of 100 represents a reasonable trade-off between temporal resolution over time and signal-to-noise ratio in an individual bundle and has been used in prior work in our lab. To first examine changes within each group separately, nine models were created, one for each group at each frequency of interest (word, syllable, and the WLI). Here, syllable presentation number (running sum of how many syllables had been presented) was entered as a fixed effect and subject intercept was entered as a random effect. Next, to examine differences between groups, three models were created, one for each entrainment frequency of interest (word, syllable, and the WLI composite measure). Across these models, fixed effects included condition and syllable presentation, with subject intercept entered as a random effect to account for pre-existing differences between subjects.

Based on findings from Batterink and Paller (2017), we would expect that in our match and static groups, syllable entrainment should decrease over time and word entrainment should increase, leading to a larger WLI over time. As participants begin to implicitly pick up on the hidden trisyllabic structure of the language, they should increasingly perceive the words, and as a result, perceive the syllables to a lesser extent. This relative shift in entrainment is thought to represent an increased sensitivity to the structure of the language over time— in other words, statistical learning. In comparison, in our mismatch group, the visual stimulation should impair participants’ ability to entrain at the word frequency, resulting in weaker changes in entrainment to the word frequency over time, as compared to the static and match groups. To the extent that the mismatch group shows weaker entrainment at the word frequency, we expect stronger entrainment at the syllable frequency over time compared to the other groups. This stronger entrainment at the syllable frequency relative to the word frequency would lead to a smaller WLI over time.

Chapter 3

3 Results

Behavioural results.

Rating task. Across all groups, words were rated as the most familiar (*M* = 3.36 , *SD* = 0.44), followed by part-words (*M* = 2.70, *SD* = 0.49), and non-words as the least familiar (*M* = 2.29, *SD* = 0.49; Effect of word type (*F*(2,114) = 94.61, *p* < .001); Linear contrast (*F*(1,57) = 152.12, *p* < .001); see Figure 2), providing evidence of explicit knowledge accrued during SL. However, in contrast to our hypothesis, ratings did not significantly differ across the three groups (Condition: *F*(2,57) = 0.99, *p* = 0.38; Word Type x Condition: *F*(4,114) = 0.95, *p* = 0.44).

Target detection task. Overall, accuracy for targets was high at 91.32%, with approximately 10.54 (*SD* = 8.82) false alarms per participant, with no significant differences between groups (Accuracy: *F*(2,59) = 0.91, *p* = 0.40; False Alarms: *F*(2,59) = 2.48 , *p* = 0.093). Across all groups, RTs were slowest for word-initial syllables (*M* = 427.58ms, *SD* = 50.10ms), intermediate for word-medial syllables (*M* = 392.91ms, *SD* = 54.81ms), and fastest for word-final syllables (*M* = 341.08ms, *SD* = 64.57ms; see figure 3). Numerically, these results follow the hypothesized linear trend of RTs decreasing for later positioned syllables. This result indicates that overall, participants appeared to respond increasingly more quickly to more predictable syllables, providing evidence of SL.

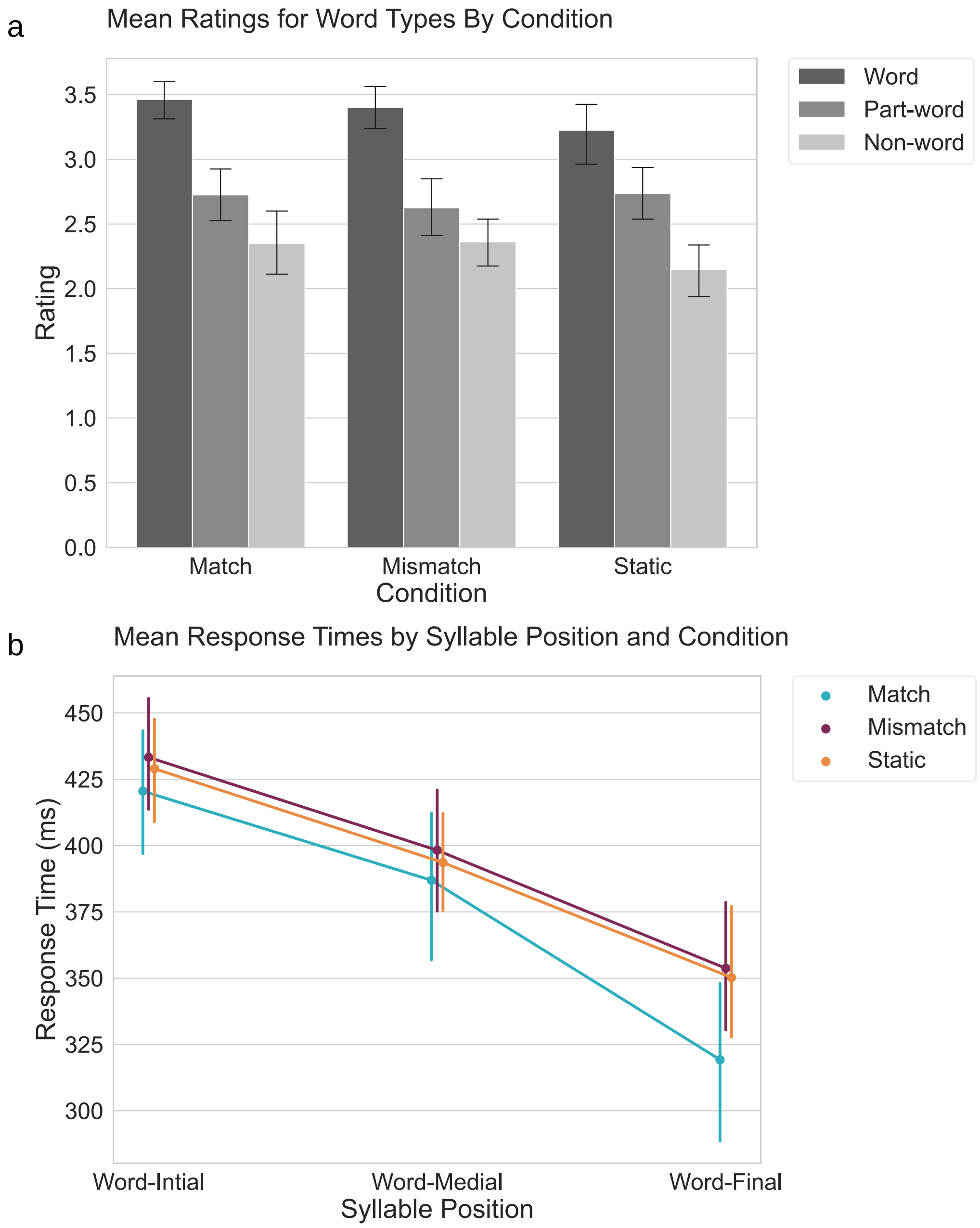


Figure 2. Behavioural results for familiarity rating task and target detection task. Error bars represent the bootstrapped 95% CI. (a) Average familiarity ratings for different word types by condition. (b) Average RTs for each syllable position in the target detection task across conditions.

The linear mixed effects model examined this effect statistically, and further tested whether there were RT differences across the three groups. In the initial model, stream position (4 – 48, i.e., in which position the target occurred within a single trial) was not found to be significant ($F(1,7783) = 0.171$, $p = 0.68$). We thus conducted a follow-up model that included all the same factors as the original model, except for the nonsignificant factor of stream position. The second, simpler model revealed a significant effect of triplet position (Triplet Position Effect: $F(1,7783) = 496.20$, $p < .001$) and an interaction between condition and triplet position (Condition x Triplet Position: $F(2,7783) = 3.42$, $p = 0.033$). In line with our hypothesis, the match group showed a significantly stronger triplet position effect compared to the static group ($t(17784) = -2.51$, $p = 0.012$; parameter estimate syllable position: $M = -12$ ms, $SE = 4.78$). In contrast, the RT triplet position effect did not significantly differ between the mismatch and static groups ($t(17784) = -0.61$, $p = 0.54$).

EEG results.

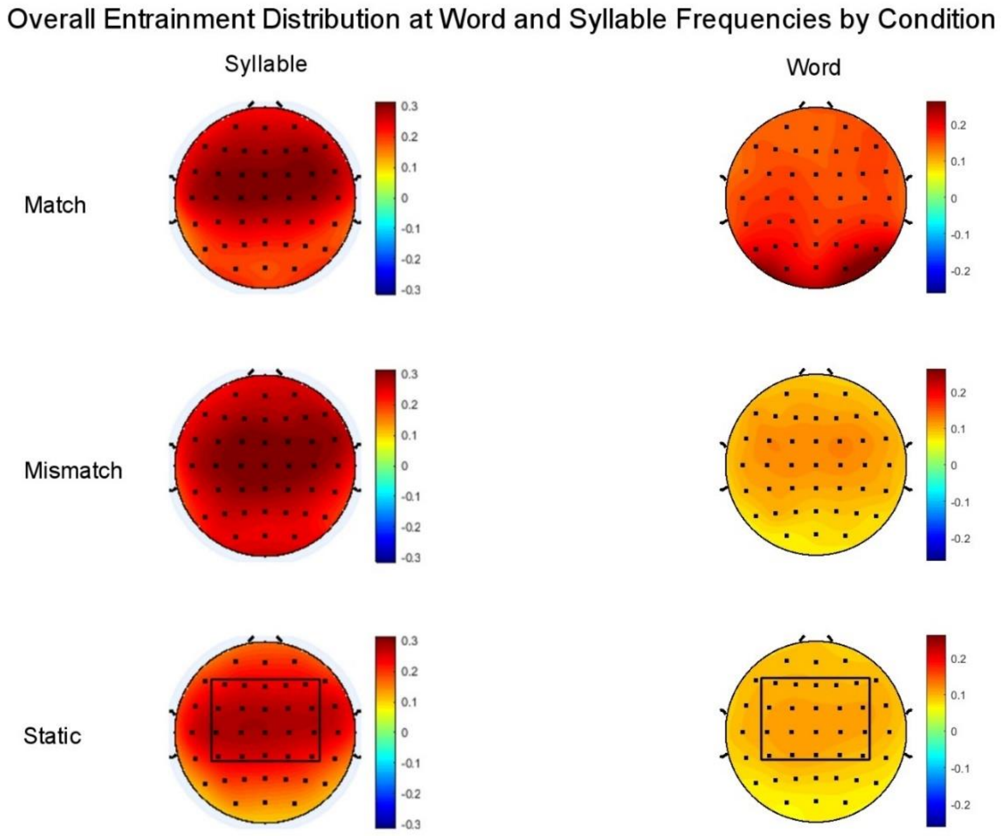


Figure 3. Overall entrainment distribution at word (1.1 Hz) and syllable (3.3 Hz) frequencies for all conditions. The black box indicates the subset of electrodes used for all EEG analyses (except for the exploratory correlational analysis).

Overall level of entrainment. As previously mentioned, to more directly capture neural entrainment effects over auditory regions, a subset of 20 electrodes (see Figure 3) was selected that showed maximal entrainment in the static group for all subsequent EEG analyses, except for the exploratory correlational analysis. All groups showed clear peaks in entrainment at the syllabic and word frequencies in this electrode subset (see Figure 4a).

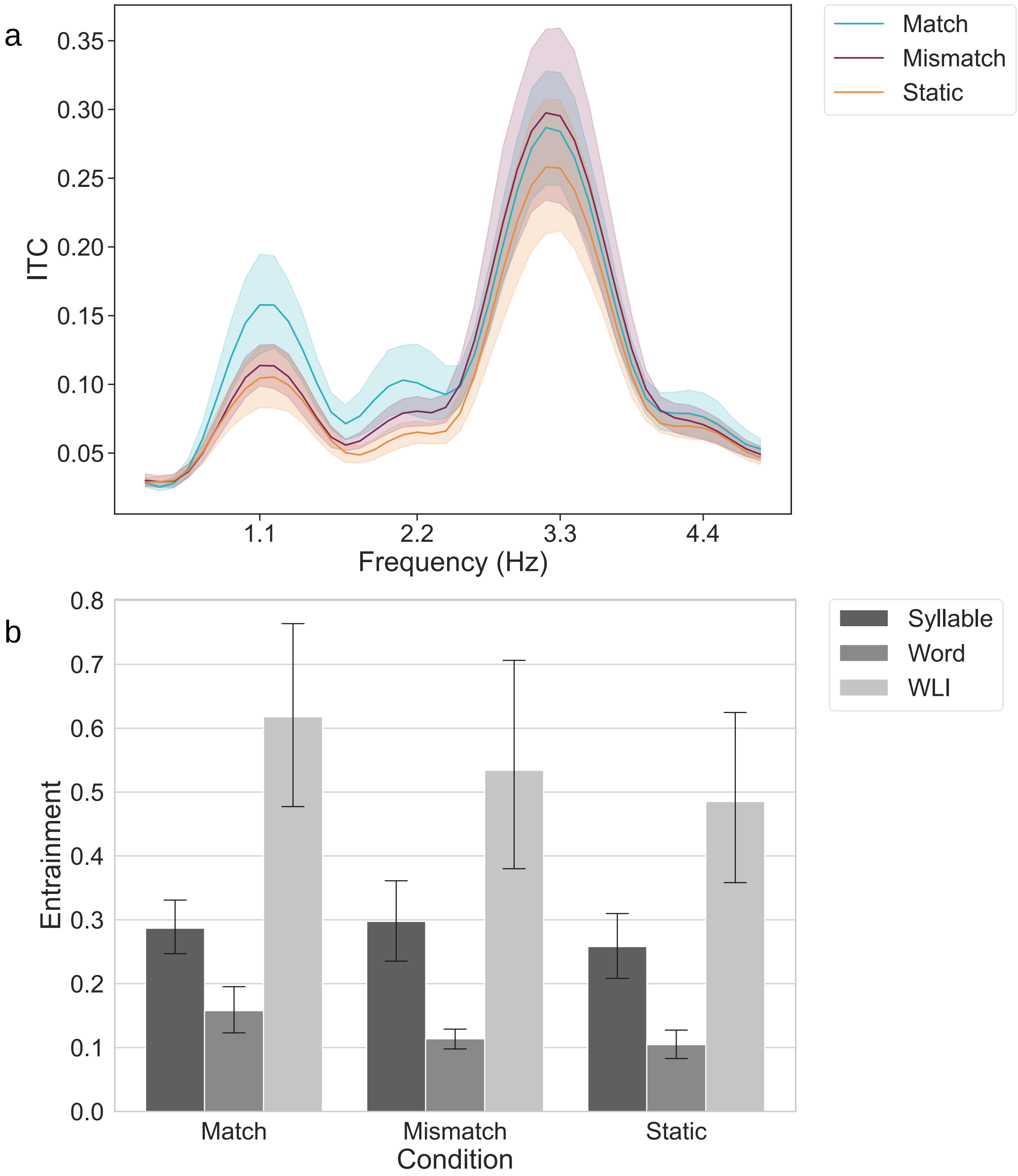


Figure 4. Overall entrainment across frequencies and conditions. Shaded areas and error bars represent bootstrapped 95% CI. (a) Inter-trial coherence (ITC) across conditions. (b) Overall level of entrainment across frequencies of interest by condition (syllable frequency = 3.3 Hz, word frequency = 1.1 Hz)

As predicted, groups differed significantly in their overall level of entrainment at the word frequency ($F(2,57) = 4.40$, $p = 0.017$). Planned contrasts revealed the match group showed stronger word frequency entrainment ($M = .16$, $SD = 0.08$) than both the static group ($M = .10$, *SD* = .05, *t*(57,59) = 2.77, *p* = 0.008) and the mismatch group (*M* = .11, *SD* = .04, *t*(57, 59) = 2.30, *p* = 0.025). In contrast, the three groups did not significantly differ in entrainment at the syllable frequency or WLI (*F*(2,57) = .55, *p* = .58; *F*(2,57) = .72, *p* = .49, respectively).

Entrainment over time across conditions. As hypothesized, the static group experienced an increase in entrainment at the word frequency and a decrease in syllable frequency over time, leading to an increase in WLI as a function of exposure. The mismatch group experienced the same pattern of entrainment as the static group, but the match group did not. Unexpectedly, while the match group experienced an increase in entrainment at the word frequency over time, they also experienced an increase in entrainment at the syllable frequency and a decrease in WLI. The results of each model can be found bellow in Table 1.

Table 1
Summary of changes in entrainment at frequencies of interest for each condition

Variable	Syllable Frequency	Word Frequency	WLI
Static Condition	Decrease t(5179) = -15.87 , p < .01	Increase t(5179) = 5.48, p < .01	Increase t(5179) = 12.42, p < .01
Match Condition	Increase t(5521) = 13.41, p <.01	Increase t(5521) = 4.63, p <. 01	Decrease t(5521) = -10.22, p < .01
Mismatch Condition	Decrease t(5325) = -12.79, p < .01	Increase t(5325) = 19.08, p < .01	Increase t(5325) = 18.14, p < .01

Entrainment over time between conditions. Next, we compared neural entrainment over time across the three groups, using the static group as the baseline. Overall, and in contrast to our hypothesis, the match group did not show stronger shifts in entrainment over time relative to the static group. While the static group experienced a decrease in syllable entrainment (*t*(16025) = -16.27, *p* < .001), the match group experienced a significant increase in entrainment at the syllable frequency over time relative to the static group (*t*(16025) = 21.52, *p* < .001). Further, the match group did not show a stronger increase in word entrainment over time relative to the static group (*t*(16025) = 0.170, *p* = 0.865). Thus, relative to the static group, the match group showed a greater decrease in the WLI over time (*t*(16026) = -15.11 , *p* < .001).

In line with our hypothesis, the mismatch group experienced a significant increase in entrainment at the syllable frequency over time relative to the static condition (*t*(16025) = 3.59, *p* < .001). However, the mismatch group also experienced a significant increase in entrainment at the word frequency compared to the static condition (*t*(16025) = 8.92 , *p* < .001). These higher levels of entrainment at the syllable and word frequencies in the mismatch condition resulted in a significant increase in WLI over time, larger than what was seen in the static condition (*t*(16026) = 5.63 , *p* <.001).

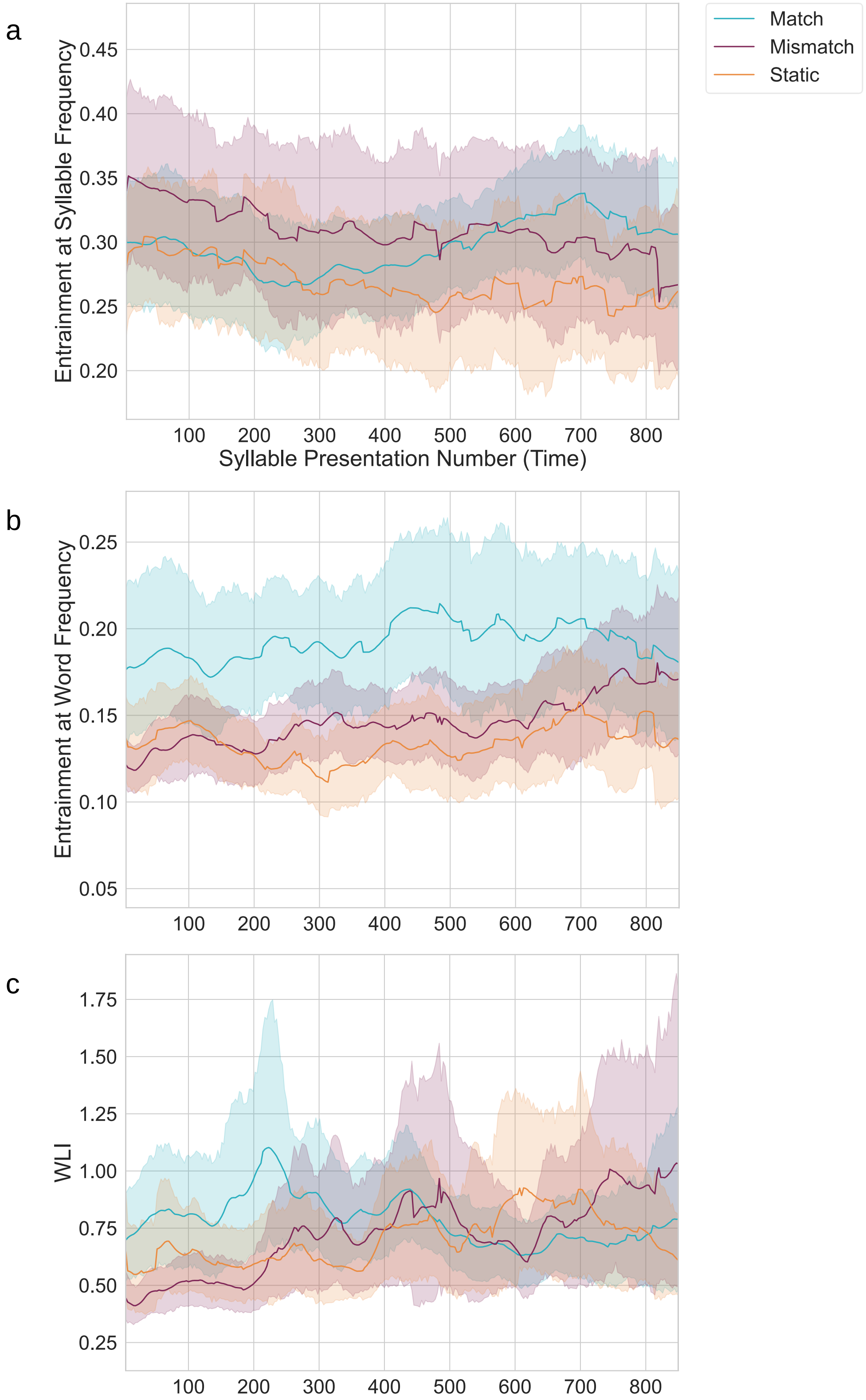


Figure 5. Entrainment over time across frequencies of interest and conditions. Shaded areas represent bootstrapped 95% CI. (a) Entrainment to syllable frequency over time across conditions. (b) Entrainment to word frequency over time across conditions. (c) WLI over time across conditions.

EEG and behavioural correlations.

Pearson's correlations were computed to examine the relationship between entrainment measures and learning outcomes in our fronto-central electrode subset (see Table 2). Collapsed across groups, WLI across the entire exposure period was significantly positively associated with rating score, indicating that participants with a higher WLI during the exposure period also showed a better ability to discriminate between actual words from the language and foil words in the familiarity rating task ($r(60) = .30$, $p = 0.020$). In contrast to our predictions, there were no significant associations between any entrainment measures and the relative RT score. However, the rating score and relative RT score were significantly positively associated ($r(60) = .32$, $p = 0.012$). The better participants were able to discriminate actual words from the language from foils, the faster they were to respond to more predictable syllables (faster for word-final in comparison to word-initial). Correlation matrices separated by group can be found in Appendices A, B, and C.

Table 2

Means standard deviations, and Pearson correlations of entrainment and learning measures

Variable	Rating Score	Relative RT Score	Syllable Entrainment	Word Entrainment	WLI	*M*	*SD*
Rating Score		.32 *	-.23	.10	.30 *	.87	.52
Relative RT Score			-.07	.04	.11	.20	.12
Syllable Entrainment				.20	-.57 **	.28	.12
Word Entrainment					.53 **	.12	.06
WLI						.54	.35

*Note: * p < .05; ** p < .01*

Exploratory correlational analysis. Pearson's correlations were computed to examine the relationship between WLI at each electrode (64 in total) and behavioural performance. Across all groups, WLI at several frontal electrodes was found to significantly positively correlate with both the relative RT score and rating score (see Figure 6). This indicates that stronger entrainment at the word frequency relative to the syllable frequency in frontal regions is associated with better implicit and explicit knowledge of the nonsense words. These findings are echoed in the match group with strong correlations found across frontal electrodes.

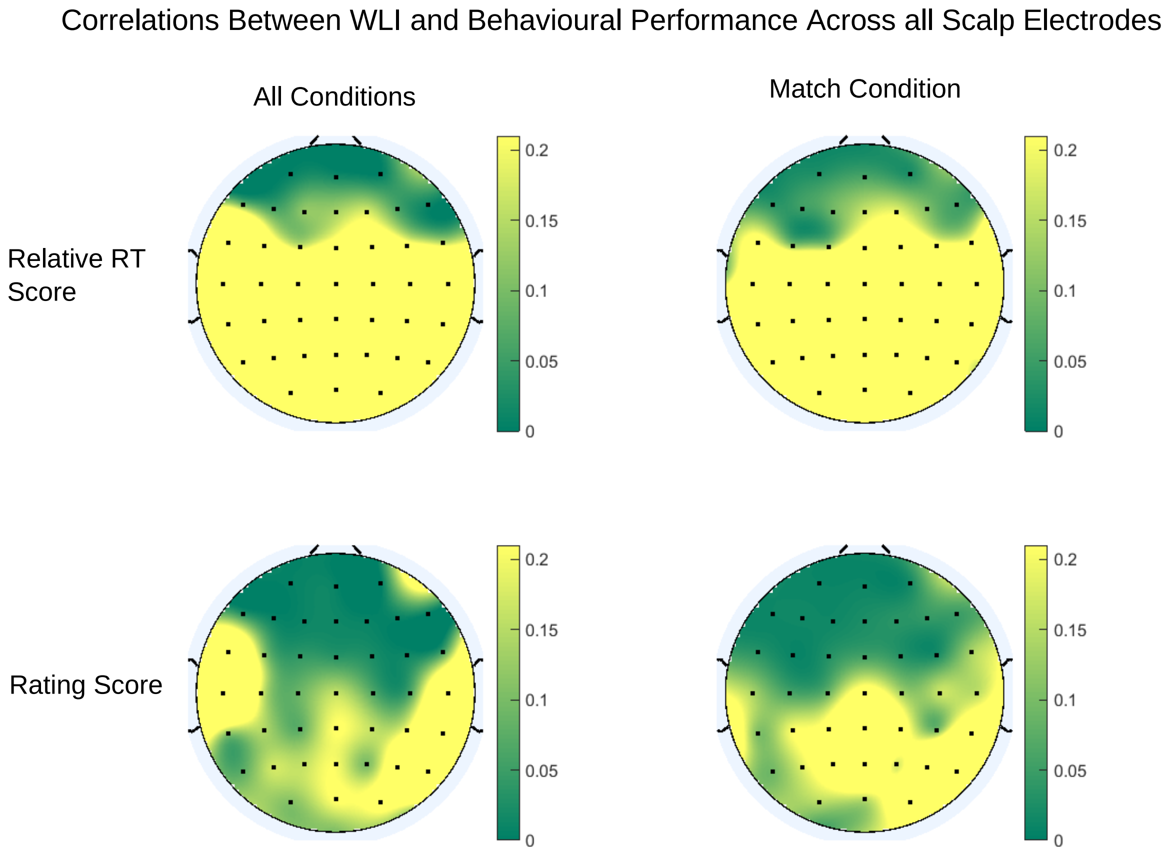


Figure 6. Correlations between WLI and behavioural performance across all scalp electrodes. The scale indicates the p-value of the correlations, with the darkest green indicating the most significant values. All correlation values were positive.

Chapter 4

4 Discussion

The aim of the present study was to test the hypothesis that neural entrainment plays a causal role in SL by directly manipulating entrainment using a visual stimulus during learning. Here, EEG was recorded while participants listened to a trisyllabic artificial language and simultaneously viewed a repetitive visual stimulus designed to elicit entrainment at specific frequencies. After exposure, SL was measured using two behavioural tasks: an explicit task indexing participants' ability to discriminate between actual words from the language and foil words, and an implicit RT task measuring participants' sensitivity to the trisyllabic structure. Overall, our results indicate the visual manipulation successfully increased entrainment at the word frequency; participants in the match group showed a significant boost in overall level of entrainment at the word frequency compared to the other groups. We also found evidence that boosting entrainment at the word frequency facilitates SL. Consistent with our hypothesis, participants in the match group were faster at detecting more predictable (later positioned) syllables, showing increased sensitivity to the hidden words of the language. Contrary to our hypothesis, groups did not differ in performance on the familiarity rating task. Correlational analyses of entrainment at our frequencies of interest in our a priori defined fronto-central region of interest revealed participants who entrain more strongly at the word frequency relative to the syllable frequency (higher WLI) were better able to explicitly discriminate words from foil words. Additional exploratory correlational analysis across all scalp electrodes revealed significant positive relationships between WLI in anterior areas and performance on both behavioural tasks.

4.1 Visual Stimulation Boosts Neural Entrainment at Word Frequency

Although we expected the visual stimulation to induce entrainment in visual areas, we also expected that it would have cross-modal effects, as unimodal sensory information has been found to directly influence neural activity across modalities (Atilgan et al., 2018; Kayser et al., 2008; Luo et al., 2010). For example, Atilgan et al. (2018) found evidence that visual stimulation produces reliable changes in local-field potential phase in the auditory cortex of ferrets, indicating visual information is present in and influences early auditory areas. As seen in Figure 3, the visual stimulation resulted in strong entrainment over occipital regions but also had a robust, broad impact on entrainment across the scalp. This suggests our visual manipulation successfully induced cross-modal effects and influenced entrainment across multiple brain areas beyond occipital regions. The overall level of word entrainment was significantly higher for the match group than the other groups, even when considering only electrodes most sensitive to auditory entrainment effects (as included in our fronto-central region of interest), indicating the visual stimulation was effective at boosting entrainment at the word level.

4.2 Visual Stimulation at the Word Frequency Increases RT Prediction Effect

Boosting neural entrainment at the word frequency resulted in an increased RT prediction effect in the implicit target detection task. Relative to the other two groups, participants in the match group showed faster RTs for more predictable, later positioned syllables. This behavioural result suggests participants in the match group were more sensitive to the hidden structure of the speech stream. Therefore, increasing entrainment at the word frequency during exposure had a significant effect on SL, which was detected after the exposure period using an implicit measure. This result provides support for a causal role of neural entrainment in SL.

As mentioned previously, synchronization between neural oscillations and external stimuli allows for phases of high neural excitability to coincide with stimulus events, enabling more efficient processing (Bauer et al., 2020; Peelle & Davis, 2012; Thut et al., 2012). For the match group, we propose that boosting entrainment at the word frequency aligned peaks of neural excitability with the structure of the speech stream. In turn, this could result in enhanced processing at the most informative moment of the signal (for example, word onsets). This enhanced processing likely facilitates the perceptual decoding of the speech stream; what was once perceived as a continuous stream of individual syllables begins to be perceived as larger temporal chunks, supporting word learning.

How exactly boosting entrainment at the word frequency could result in better perceptual decoding of a speech stream can be explained through speech processing models proposed by Giraud and Poeppel (2012) and Peelle and Davis (2012). These models posit that higher-frequency neural activity is modulated by low-frequency oscillations, resulting in nested oscillations that are sensitive to stimulation at multiple timescales (e.g., phonemes, syllables, words, and phrases). This coupling is thought to support the integration of smaller linguistic segments into larger meaningful chunks, which likely impact neural computations across different frequencies and brain areas. These models of speech processing are supported by a recent SL study from Batterink (2020), which found words with higher levels of phase-locking during exposure were more successfully recognized in a subsequent 2AFC test as compared to words with lower levels of phase-locking. This result suggests that successfully recognized words were likely extracted as meaningful word units and tracked by word-rate oscillations, whereas poorly recognized words were processed as a string of unrelated syllables. Although we did not look at entrainment to individual words, we can see that overall, our visual stimulation resulted in a significantly higher level of entrainment in fronto-central electrodes at the word frequency in our match group. This difference in word entrainment is especially prominent early in the task (see Figure 5b), suggesting that our visual manipulation may have hastened the coupling between low-frequency oscillations and higher-frequency neural activity in the match group. The fact that our match group experienced significantly higher levels of entrainment at the word frequency and experienced an increased RT prediction effect suggests the visual stimulation enhanced the coupling process, leading to better extraction of the meaningful units of the speech stream (words).

The high level of entrainment at the word frequency suggests this coupling process was strengthened in the match group, but what might the increase in entrainment at the word frequency over time suggest for this coupling process mechanistically? If phase-locking was driven purely by bottom-up processes, then one would not necessarily expect to see increases or decreases in entrainment at frequencies corresponding to specific linguistic units over time, as the stimulus remains constant (Batterink, 2020). Rather, this increase in entrainment at the word frequency over time could reflect temporal information coming down from higher-order areas to influence phase-locking to what has been learned to be a relevant and meaningful unit. As such, entrainment to meaningful linguistic units could influence higher-level representations of these units, which then influence future processing of the same or similar units, cyclically. Although EEG does not provide the necessary spatial precision to untangle how different brain regions interact to contribute to SL over time, evidence for the involvement and influence of higher-order brain areas is supported by our exploratory correlational analyses, which revealed that WLI in frontal regions was associated with better SL performance.

To this point, we have provided evidence that stronger entrainment at the word frequency in the match group likely indicates better decoding of the speech stream and that the increase in word entrainment over time could reflect top-down influence from higher-order brain areas. Together, these two ideas work cohesively to explain why our visual stimulation resulted in an increased RT prediction effect. As proposed by Batterink (2020), it is plausible that linguistic representations in higher-order brain areas could exert top-down influence on phase-locking in primary sensory areas, facilitating predictive processing. Therefore, because the match group entrained more strongly to the word frequency during exposure, higher-order areas may have stored more complete or accurate representations of the underlying patterns. During the target detection task, neural phase-locking consistent with the word frequency might be more quickly reinstated in the match group due to influence from higher-order areas, leading to a stronger RT prediction effect than the other groups. The idea that interactions between higher-order regions and auditory regions play a role in auditory SL is supported by neuroimaging work that finds activity in the left inferior frontal cortex and superior temporal gyrus increase in SL paradigms and correlate with individual rates of SL (Abla & Okanoya, 2008; Karuza et al., 2013; McNealy et al., 2006). As such, we propose that information from sensory areas (e.g., superior temporal gyrus) may be sent to domain-general higher-order areas (e.g., left inferior frontal cortex) where it is bound into a percept that can then influence activity in primary areas, cyclically. Future work should examine entrainment during the target detection task to look for differences that may exist between groups that could explain this behavioural benefit.

Contrary to our hypothesis, the visual stimulation did not result in better performance in the familiarity rating task, our explicit measure of learning. A possible explanation for the dissociation between the two tasks is that boosting entrainment at the word frequency may have enhanced implicit knowledge of the nonsense words, but not explicit knowledge. As mentioned earlier, the manipulation may have resulted in better linguistic representations in higher-order areas, influencing phase-locking at the word frequency to quickly reinstate during the target detection streams, which then enhanced perception and decoding of the stream. Since the familiarity rating task does not directly capture online sensitivity to the structure of the language, but rather relies on explicit memory representations that are extracted based on this sensitivity, this benefit may not have been particularly useful. However, logically, stronger entrainment to the word frequency and structure at some point would be expected to result in stronger explicit knowledge. The finding that our visual manipulation did not result in better performance in the familiarity rating task might mean it was not strong enough to drive the transfer of knowledge from implicit to explicit.

4.3 Correlations Between Frontal Regions and SL

The exploratory correlational analysis of all scalp electrodes revealed significant positive correlations between WLI in frontal regions and performance on both SL tasks. That is, participants who entrained more strongly at the word frequency in frontal regions showed better SL performance in both implicit and explicit measures. The fact that these significant correlations were found mostly over frontal regions and not occipital regions suggest that it is the extent to which entrainment occurred in higher level regions that influences or drives SL. As such, entrainment seen in occipital channels (see Figure 3) likely reflects more low-level responses to the visual stimulus that are not related to SL specifically.

4.4 Neural Entrainment Over Time

In support of our hypotheses, the static group followed the expected shift in entrainment over time, decreasing at the syllable frequency and increasing at the word frequency as a function of exposure. This pattern of entrainment was also found in work by Batterink and Paller (2017) and was found to significantly predict performance in an implicit SL measure. The results of the current study and Batterink and Paller (2017) are in line with an idea put forth by Buiatti et al. (2009) that an increase in processing at the word frequency is accompanied by a decrease in processing at alternate structure frequencies (e.g., single syllables). Together, these results strongly suggest that this pattern of entrainment reflects changes in sensitivity to structure over time that would be predicted by SL; as participants pick up on the hidden words in the stream, they experience an increase in entrainment at the word structure and a relative decrease to the syllable structure.

Although the match and mismatch groups did not show the hypothesized patterns of entrainment over time, the results are still consistent with SL. The mismatch group showed a similar trajectory to the static group, while those in the match group also experienced an increase in entrainment at the word frequency over time. However, it appears that our visual manipulation did not result in a continued cumulative improvement (or deterioration) in entrainment over time. To illustrate, when looking at word entrainment over time (see Figure 5), although early on groups differed in their level of entrainment, group trajectories did not continue to diverge over time. One possible explanation for this finding is that the match group was maximally entrained at the word frequency earlier on during the stream and experienced a ceiling effect. Given that increases in entrainment are typically marginal and cannot increase exponentially, an individual may reach a maximal level of entrainment at a given frequency, and then show a plateau or even decrease in entrainment. For the static and mismatch groups, starting at a lower level of entrainment would allow for a greater increase in entrainment before hitting a potential “ceiling”. This idea is supported by a recent study by Batterink and Paller (2019) in which entrainment at the word frequency was found to reach a plateau after six minutes of exposure. Further, Choi et al. (in press) found evidence that entrainment in both infants and adults occurs rapidly during the first two minutes of exposure to a statistical speech stream and then appears to slow, following a logarithmic curve. The finding that the match group experienced an increase in entrainment at the syllable frequency was unexpected but may reflect habituation or a loss of attention to the visual stimuli over time, which then allowed for more processing dominance by the raw auditory input.

4.5 Failure to Replicate Previous EEG and Behavioural Correlations

Our results did find a positive relationship between WLI and the familiarity rating task in fronto-central electrodes, such that participants who experienced stronger entrainment to the word frequency relative to the syllable frequency showed a better ability to discriminate between actual words from the language and foil words. This supports the general finding that stronger entrainment towards the word frequency is related to better performance on post-exposure tests. Unexpectedly, our results failed to replicate prior results in the literature of a positive relationship between WLI and the RT prediction effect (Batterink & Paller, 2019; Batterink & Paller, 2017). In the current study, WLI in fronto-central electrodes was not significantly correlated with the performance in the target detection task. However, our exploratory analysis revealed strong correlations between WLI and both of our behavioural measures in more anterior electrodes. As described earlier, it could be that these effects are more prominent in frontal areas, which is supported by our exploratory analysis and a body of work finding SL-related activity in frontal regions (Abla & Okanoya, 2008; Karuza et al., 2013; McNealy et al., 2006). Another possibility why we did not see this correlation in fronto-central electrodes is that our data might be nosier than previous studies due to a reduced number of participants. In the current study there are only

20 participants in each condition, whereas the previous studies mentioned had a group size over 20.

4.6 Limitations and Future Directions

One limitation of the present study is that entrainment was manipulated cross-modally using a visual stimulus. The hope was that the visual stimulation would manipulate neural entrainment outside of participants' awareness, allowing us to look directly at the relationship between neural entrainment and SL. However, there is a possibility the visual stimulus could have served as an explicit cue to word boundaries for participants. If this were the case, entrainment would not have been manipulated outside of awareness and top-down influences on entrainment would not allow us to address causality. Without using direct brain stimulation methods, the only option that does not introduce another stimulus would be to modify the speech stream directly. This is problematic because manipulating the speech stream (e.g., adding pauses or changing intonation) introduces additional cues to word boundaries outside of transitional probabilities, which is the main cue being investigated. Therefore, we decided using a visual stimulus to manipulate entrainment would be the best option, outside of brain stimulation methods, to investigate the relationship between entrainment and SL.

Regardless, the current study cannot rule out the possibility that participants may have been more sensitive or aware of the manipulation and could have used the visual stimuli to strategically segment the words. Behavioural results obtained from the survey after testing revealed some participants noticed a link between the two stimuli, with most descriptions along the lines of "the drops and syllables had the same rhythm". Anecdotally, this statement was also frequently endorsed by participants in the mismatch condition, in which the two stimuli were not temporally coherent. Although work does support that visual stimuli can be used to probe causal effects of entrainment (Notbohm et al., 2016), this does not negate conscious top-down strategies that can impact entrainment. For example, Okawa et al. (2017) had participants imagine three types of rhythms while viewing a rhythmic visual stimulus that induced entrainment at a specific frequency (the "beat" frequency). The authors found evidence of entrainment to the harmonics (imagined rhythms) of the beat frequency provided by the visual stimulus.

However, if our visual stimulus served as an explicit cue, then we would expect participants in the match group to have performed significantly better in the familiarity task. This is because, if participants are aware the visual stimulus coincides with the beginning of every word, they would be expected to use the visual stimulus as an explicit cue for decoding words. In fact, this was not the case as participants in the match group did not perform better on the familiarity rating task than the other two groups. Given these results, evidence from the current study strongly suggests that entrainment plays a causal role in SL but cannot exclude the possibility that the visual stimulus acted as an implicit or explicit cue to facilitate learning. Future studies should manipulate entrainment using non-invasive brain stimulation methods to avoid the confounds associated with cross-modal entrainment.

4.7 Conclusions

The current study provides a solid step forward in the literature of the relationship between neural entrainment and SL. Our findings indicate that boosting neural entrainment at the word frequency results in better segmentation of a trisyllabic speech stream, as evidenced by better performance in an implicit measure of SL Further, although exploratory, the results also suggest that it is the extent to which entrainment occurs in higher-order brain areas that drives SL. Overall, this study provides evidence for the potential causal role of neural entrainment in SL.

On a theoretical level, this work helps to advance our understanding of neural mechanisms that could contribute to SL, namely neural entrainment. This work also lends support to oscillatory models of speech processing and a body of work that suggests frontal regions play an important part in SL. Future work in this area could reveal novel ways to boost SL as well as methods to improve language acquisition for adult second language learners or children with atypical language development. Lastly, since SL is domain-general, these findings can help to inform other areas beyond language, such as visual processing.

References

Abla, D., & Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia*, *46*(11), 2787–2795. https://doi.org/10.1016/j.neuropsychologia.2008.05.012

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron*, *97*(3), 640-655.e4. https://doi.org/10.1016/j.neuron.2017.12.034

Batterink, L. (2020). Syllables in Sync Form a Link: Neural Phase-locking Reflects Word Knowledge during Language Learning. *Journal of Cognitive Neuroscience*, 1–14. https://doi.org/10.1162/jocn_a_01581

Batterink, L. J. (2017). Rapid Statistical Learning Supporting Word Extraction From Continuous Speech. *Psychological Science*, *28*(7), 921–928. https://doi.org/10.1177/0956797617698226

Batterink, L. J., & Paller, K. A. (2019). Statistical learning of speech regularities can occur outside the focus of attention. *Cortex*, *115*, 56–71. https://doi.org/10.1016/j.cortex.2019.01.013

Batterink, L. J., & Paller, Ken. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Bauer, A. K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of Neural Oscillations and Cross-modal Influences. *Trends in Cognitive Sciences*, *24*(6), 481–495. https://doi.org/10.1016/j.tics.2020.03.003

Brady, T. F., & Oliva, A. (2008). Statistical Learning Using Real-World Scenes: Extracting Categorical Regularities Without Conscious Intent. *Psychological Science*, *19*(7), 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Buiatti, M., Pena, M., & Dehaenelambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, *44*(2), 509–519. https://doi.org/10.1016/j.neuroimage.2008.09.015

Campbell, K. L., Healey, M. K., Lee, M. M. S., Zimerman, S., & Hasher, L. (2012). Age Differences in Visual Statistical Learning. *Psychology and Aging*, *27*(3), 650–656. https://doi.org/10.1037/a0026780

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (in press). *Preverbal infants discover statistical word patterns at similar rates as adults: Evidence from neural entrainment*. Psychological Science.

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

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Farthouat, J., Franco, A., Mary, A., Delpouve, J., Wens, V., Op de Beeck, M., De Tiège, X., & Peigneux, P. (2017). Auditory Magnetoencephalographic Frequency-Tagged Responses Mirror the Ongoing Segmentation Processes Underlying Statistical Learning. *Brain Topography*, *30*(2), 220–232. https://doi.org/10.1007/s10548-016-0518-y

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467. http://dx.doi.org.proxy1.lib.uwo.ca/10.1037/0278-7393.28.3.458

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64. https://doi.org/10.1016/S0010-0277(00)00132-3

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658–680. http://dx.doi.org.proxy1.lib.uwo.ca/10.1037/0096-3445.130.4.658

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, *127*(1), 46–54. https://doi.org/10.1016/j.bandl.2012.11.007

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual Modulation of Neurons in Auditory Cortex. *Cerebral Cortex*, *18*(7), 1560–1574. https://doi.org/10.1093/cercor/bhm187

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2017). *Neural entrainment determines the words we hear*. https://doi.org/10.1101/175000

Lakatos, P., Gross, J., & Thut, G. (2019). A New Unifying Account of the Roles of Neuronal Entrainment. *Current Biology : CB*, *29*, R890–R905. https://doi.org/10.1016/j.cub.2019.07.075

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory Cortex Tracks Both Auditory and Visual Stimulus Dynamics Using Low-Frequency Neuronal Phase Modulation. *PLOS Biology*, *8*(8), e1000445. https://doi.org/10.1371/journal.pbio.1000445

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the Language Code: Neural Mechanisms Underlying Speech Parsing. *The Journal of Neuroscience*, *26*(29), 7629–7639. https://doi.org/10.1523/JNEUROSCI.5501-05.2006

Mitchel, A. D., & Weiss, D. J. (2011). Learning Across Senses: Cross-Modal Effects in Multisensory Statistical Learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(5), 1081–1091. https://doi.org/10.1037/a0023700

Notbohm, A., Kurths, J., & Herrmann, C. S. (2016). Modification of Brain Oscillations via Rhythmic Light Stimulation Provides Evidence for Entrainment but Not for Superposition of Event-Related Responses. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00010

Okawa, H., Suefusa, K., & Tanaka, T. (2017). Neural Entrainment to Auditory Imagery of Rhythms. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00493

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *ELife*, *5*. https://doi.org/10.7554/eLife.14521

Peelle, J. E., & Davis, M. H. (2012). Neural Oscillations Carry Speech Rhythm through to Comprehension. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00320

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3). https://doi.org/10.1111/j.1467-8624.2009.01290.x

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2017). Neural Entrainment to Speech Modulates Speech Intelligibility. *Current Biology*, *28*(2), 161–169.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews. Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Romei, V., Gross, J., & Thut, G. (2012). Sounds Reset Rhythms of Visual Cortex and Corresponding Human Visual Perception. *Current Biology*, *22*(9–2), 807–813. https://doi.org/10.1016/j.cub.2012.03.025

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928. JSTOR.

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. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

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. JSTOR.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The Necessity of the Medial Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747. https://doi.org/10.1162/jocn_a_00578

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, *32*, 15–20. https://doi.org/10.1016/j.cobeha.2020.01.015

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21. https://doi.org/10.1186/1471-2202-10-21

Thut, G., Miniussi, C., & Gross, J. (2012). The Functional Importance of Rhythmic Activity in the Brain. *Current Biology*, *22*(16), R658–R663. https://doi.org/10.1016/j.cub.2012.06.061

Toro, J. M., & Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–875. https://doi.org/10.3758/BF03193539

Wilsch, A., Neuling, T., Obleser, J., & Herrmann, C. S. (2018). Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *NeuroImage*, *172*, 766–774. https://doi.org/10.1016/j.neuroimage.2018.01.038

Appendices

Appendix A: EEG and Behavioural Correlations for Match Condition

Means standard deviations, and Pearson correlation of entrainment and learning measures in the match condition

Variable	Rating Score	Relative RT Score	Syllable Entrainment	Word Entrainment	WLI	*M*	*SD*
Rating Score		.45*	-.40	.23	.48*	.93	.66
Relative RT Score			-.21*	-.10	.16	.24	.15
Syllable Entrainment				.07	-.46*	.29	.10
Word Entrainment					.78**	.16	.08
WLI						.62	.35

*Note: * p < .05; ** p < .01*

Appendix B: EEG and Behavioural Correlations for Mismatch Condition

Means standard deviations, and Pearson correlations of entrainment and learning measures in the mismatch condition

Variable	Rating Score	Relative RT Score	Syllable Entrainment	Word Entrainment	WLI	*M*	*SD*
Rating Score		.50*	-.28	-.34	.24	.91	.46
Relative RT Score			.10	-.02	.08	.18	.10
Syllable Entrainment				.32	-.73**	.30	.15
Word Entrainment					.19	.11	.04
WLI						.49	.31

*Note: * p < .05; ** p < .01*

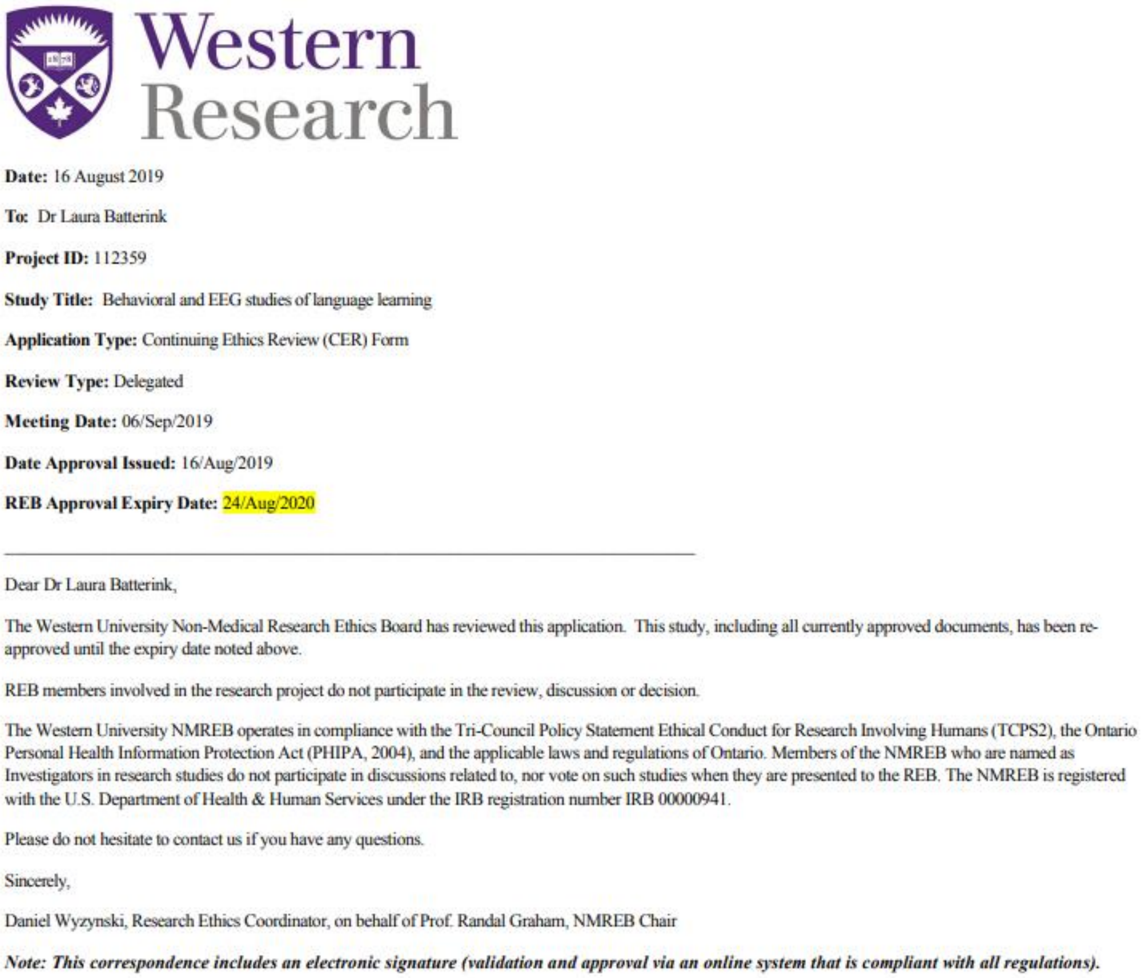
Appendix C: EEG and Behavioural Correlations for Static Condition

Means standard deviations, and Pearson correlations of entrainment and learning measures in the static condition

Variable	Rating Score	Relative RT Score	Syllable Entrainment	Word Entrainment	WLI	*M*	*SD*
Rating Score		-.18	-.11	-.003	.07	.78	.45
Relative RT Score			-.16	-.10	-.04	.18	.10
Syllable Entrainment				.34	-.53*	.26	.12
Word Entrainment					.46*	.10	.05
WLI						.49	.31

*Note: * p < .05; ** p < .01*

Appendix D: Ethics Approval



Western Research

Date: 16 August 2019

To: Dr Laura Batterink

Project ID: 112359

Study Title: Behavioral and EEG studies of language learning

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

Meeting Date: 06/Sep/2019

Date Approval Issued: 16/Aug/2019

REB Approval Expiry Date: 24/Aug/2020

Dear Dr Laura Batterink,

The Western University Non-Medical Research Ethics Board has reviewed this application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University NMREB operates in compliance with the Tri-Council Policy Statement Ethical Conduct for Research Involving Humans (TCPS2), the Ontario Personal Health Information Protection Act (PHIPA, 2004), and the applicable laws and regulations of Ontario. Members of the NMREB who are named as Investigators in research studies do not participate in discussions related to, nor vote on such studies when they are presented to the REB. The NMREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000941.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Daniel Wyzynski, Research Ethics Coordinator, on behalf of Prof. Randal Graham, NMREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix E: Consent Form

Project Title: Behavioral and EEG Studies of Language Learning

Principal Investigator:
Dr. Laura Batterink
Department of Psychology, The University of Western Ontario, London, ON
Telephone: [redacted]; Email: [redacted].

1. Invitation to Participate

You are being invited to participate in a research study about how adults learn new languages.

The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research. It is important for you to understand why the study is being conducted and what it will involve. Please take the time to read this carefully, and feel free to ask questions if anything is unclear or if there are words or phrases you do not understand.

2. Why is this study being done?

The purpose of the study is to investigate how people pick up on different aspects of language, such as vocabulary and grammar. Compared to children, adults often struggle to acquire a new language. By understanding the neural mechanisms that adults rely on when they are exposed to a new language, we may better understand why adults often have more difficulty acquiring new languages. We may also get useful information about the best learning practices and types of training to help adults acquire a new language.

3. How long will you be in this study?

It is expected that this study will take 1.5 hours to complete.

4. What are the study procedures?

The experiments conducted as part of this study will test how humans process and learn about different types of linguistic stimuli, such as syllables, words, phrases and sentences. If you agree to participate, you will be asked to listen to language-related auditory stimuli and/or read words and sentences on a screen. You may be asked to perform different tasks associated with the stimuli, such as responding to targets by pressing a button, or making different judgments or ratings about your impressions of the stimuli. You may be asked to respond using your voice, and your voice may be recorded using an audio recorder. If you do not wish to be recorded, you can still participate in other parts of the study.

Your brain activity may be recorded using a technique called electroencephalography (EEG), where electrodes placed on the scalp measure electrical signals that brain cells use to communicate. An elastic cap will be placed on your head. The cap will be strapped down to fit snugly and comfortably. The sensors, which look like white pieces of plastic about 1 inch in diameter attached to the cap, will be filled with a small amount of conductive gel. To monitor blinking and eye movements, the experimenter will place similar sensors on the skin surface near your eyes. These sensors will be secured in place using tape. When the sensors are removed, the gel will be wiped off using tissue. Some gel may remain in your hair, but it can easily be removed by rinsing with water. You will be given the opportunity to wash your hair at the end of the study.

The task(s) will be conducted in the Brain and Mind Institute in the Western Interdisciplinary Research Building (WIRB) on the University of Western Ontario campus.

5. What are the risks and harms of participating in this study?

There are no known or anticipated risks or discomforts associated with participating in this study. However, you may experience a minor inconvenience as some gel may remain in your hair at the end of the study. The gel can easily be removed by washing your hair. You will be given the opportunity to wash your hair at the end of the study.

6. What are the benefits?

You do not directly stand to benefit from this study. Although you may not directly benefit from your participation, the information gathered may provide benefits to society as a whole which include enhancing our scientific understanding of language, learning, and the brain, and leading to advancements in second language training and treatment of language-related disorders (for example, specific language impairment and autism).

7. Can participants choose to leave the study?

You may refuse to participate, refuse to answer any questions or withdraw from the study at any time. If you decide to withdraw from the study, you have the right to request withdrawal of information collected about you. If you wish to have your information removed please let the researcher know. Withdrawing or refusing to answer questions will not result in loss of promised compensation.

8. How will participants' information be kept confidential?

Any personal or identifying information obtained from this study will be kept confidential and will be accessible only to the investigators of this study. Identifiable information that will be collected during the study includes your full name, telephone number, email address, partial date of birth (month and year) and, in some cases, audio voice recordings. In the event of publication, any data resulting from your participation will be identified only by case number, without any reference to your name or personal information. Only the research team will have access to information that identifies you to carry out this research study.

If files are shared with other researchers or the results are made public, any personal information that could identify you will be removed. Only anonymized data will be shared outside the research team (e.g., in an open access repository for publication purposes, or for other researchers to verify the findings or re-analyze).

Any documents identifying you by name will be kept separately from your data, and will be destroyed after 7 years. De-identified and anonymous study records will be maintained for a minimum of 7 years. A list linking your study number with your name will be kept by the researcher in a secure place, separate from your study file.

Representatives of the University of Western Ontario Non-Medical Research Ethics Board may require access to your study-related records to monitor the conduct of the research.

9. Are participants compensated to be in this study?

You will receive course credit (1 credit per hour) or monetary compensation ($14 per hour) for your participation in this study. If you do not complete the entire study you will still be compensated a pro-rated amount (based on the same rates specified above: 1 credit/h or $14/h). When calculating prorated compensation, your total participation time will be rounded up to the nearest half hour. For example, if you withdraw after 1 hour and 15 minutes, your participation time will be rounded to 1.5 h and you will receive 1.5 credits or $21. Therefore, even if you withdraw prior to completing study, you will still be compensated for the amount of time you spent participating.

10. What are the rights of participants?

Your participation in this study is voluntary. You may decide not to be in this study. Even if you consent to participate you have the right to not answer individual questions or to withdraw from the study at any time. If you are a student at Western and you choose not to participate or to leave the study at any time, it will have no effect on your academic standing.

We will give you new information that is learned during the study that might affect your decision to stay in the study.

You do not waive any legal right by signing this consent form

11. Whom do participants contact for questions?

If you have questions about this research study please contact Laura Batterink, Principal Investigator, Telephone: [redacted] Email: [redacted]

If you have any questions about your rights as a research participant or the conduct of this study, you may contact [redacted]

This letter is yours to keep for future reference.

Consent Form

Project Title: Behavioral and EEG studies of language learning
Study Investigator's Name: Dr. Laura Batterink

I agree to be audio-recorded in this research.

☐ **YES** ☐ **NO**

I agree to be contacted for future research studies.

☐ **YES** ☐ **NO**

I have read the Letter of Information, have had the nature of the study explained to me and I agree to participate. All questions have been answered to my satisfaction.

Participant's Name (please print): ____

Participant's Signature: ____

Date: ____

My signature means that I have explained the study to the participant named above. I have answered all questions.

Person Obtaining Informed Consent (please print): ________________________________

Signature: ________________________________

Date: ________________________________

Appendix F: Participant Information Form

ADULT PARTICIPANT INFORMATION SHEET

Subject Code: ________________ Birth date: ________________________ Sex:______

Do you consider yourself: ☐ right-handed ☐ left-handed ☐ ambidextrous

LANGUAGE BACKGROUND

Is English the first language that you learned? ☐ yes ☐ no

If *No*, what language did you first learn?____

If *No*, at what age did you first begin learning English?_____________________________

In what context?____

If *No,* in which language (English or your native language) are you more comfortable?___

Are you fluent in any language other than English? ☐ yes (list language),______________ ☐ no

Are you regularly exposed to any language other than English? ☐ yes,________________ ☐ no

If *Yes*, in what context? ____

Are there are other languages not asked about above that you know? ☐ yes,____________ ☐ no

If *Yes*, please list and describe how you learned them. ___________________________________

____

____

What is your field of study/major?____

NEUROLOGICAL HISTORY

Have you ever had brain surgery? ☐ yes ☐ no

Have you ever had, or do you currently have, any neurological disorders (e.g., seizures, schizophrenia)?

☐ yes ☐ no If *Yes*, please explain: ____

Are there any known neurological problems in your family? ☐ yes ☐ no

If *Yes*, please explain: ____

Are you currently taking any medication(s) that may affect brain functioning (including but not limited to anti-depressants, anti-psychotics, anti-seizure)? ☐ yes,___________ ☐ no

Have you ever had, or do you currently have, any speech, hearing, learning, or psychiatric disorders?

☐ yes ☐ no If *Yes*, please explain: ___________________________________

VISION AND HEARING

Do you have normal or corrected-to-normal vision? ☐ yes ☐ no

Do you have normal hearing? ☐ yes ☐ no

CURRENT STATE

How many hours of sleep did you get last night? __________

How many hours of sleep do you typically get per night? __________

Do you feel like you got enough sleep last night to function normally both physically and mentally?

☐ yes ☐ no If no, please explain: ____

Is there any *other* circumstance (not asked about above) that makes you feel like you are not at your mental best right now? ☐ yes ☐ no If yes, please comment: ____________________

Please rate your level of current fatigue on a 1-10 scale, where 1 is “so tired I can barely function today” and 10 is “I feel super rested, I’ve never felt better.” (Circle 1-10)

(very tired) 1 2 3 4 5 6 7 8 9 10 (feel great)

Appendix G: Post-task Interview

Post-Task Interview

C1 – all questions
C2 – all questions except 4
C3 – question 3 and 5

1. Did notice any link or connection between the video you observed and the sounds you were hearing?
   YES / NO
   DESCRIBE:

2. Did you notice any association between the timing of the video and the syllables in the stream.
   YES / NO
   DESCRIBE:

3. Did you notice any words or patterns in the audio stream.
   YES / NO
   DESCRIBE:

4. Did you notice that the water droplet hit the fixation point at the beginning of each word?
   YES / NO

5. Were any of the syllables in the target detection task difficult to hear? If so, which ones?

Curriculum Vitae

Name:	Jerrica Mulgrew
Post-secondary Education and Degrees:	Trent University Peterborough, Ontario, Canada 2014-2018 B.Sc. Psychology
	The University of Western Ontario London, Ontario, Canada 2018-2020 M.Sc. Psychology Candidate
Honours and Awards:	Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award 2017
	Natural Sciences and Engineering Research Council of Canada (NSERC) Canada Graduate Scholarship Masters 2019
Related Work Experience:	Research Assistant and Lab Coordinator Cognitive Ethology Lab Trent University 2016-2018
	Graduate Teaching Assistant The University of Western Ontario 2018-2020

Publications:

Kramer, S. S. R., Mulgrew, J., Anderson, N. C., Vasilyev, D., Kingston, A., Reynolds, M. G., Ward, R. (2020). Physically attractive faces attract us physically. Cognition, 198.

Kramer, S. S. R., Mulgrew, J. (2018). Displaying red and black on a first date: A field study using the “First Dates” television series. Evolutionary Psychology, 16(2).

Kramer, S. S. R., Mulgrew, J., & Reynolds, M. G. (2018). Unfamiliar face matching with photographs of infants and children. PeerJ, 6(e5010).