N400 evidence for musical facilitation of word boundary identification in second language exposure

Dayna Andrea Moya Sepulveda

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018
ABSTRACT

N400 evidence for musical facilitation of word boundary identification in second language exposure

Dayna Andrea Moya Sepulveda

Lexical acquisition requires the ability to identify word boundaries in a continuous auditory speech stream. This complex task is even more challenging when learning a new language in adulthood. Previous studies have shown that word boundary identification can be enhanced by pairing musical tones with native language phonemes. The objective of this dissertation study was to investigate whether musical tones also have this effect in a novel pseudo-language that uses non-native speech sounds. The N400, a brain event-related potential that has been linked with familiarity responses and detection of statistical regularities during exposure to pseudowords, provides an index of brain activation associated with semantico-lexical processing. In this study, language-like stimuli incorporating a French phoneme (a high, front, rounded vowel that is not part of the English phonetic inventory) were presented to typically developing English monolingual adults. Participants were exposed to one of two training conditions for 7 minutes: monotone presentation of the concatenated language-like stimuli; or the same speech stream with a musical tone associated with each syllable. The training protocol was based on Schön, Boyer, Moreno et. al. (2008). Training was followed by a lexical decision task, requiring participants to distinguish “words” (heard during the training in a concatenated speech stream)
from “part words” (end of one word and the beginning of another, crossing word boundaries). High density EEG was recorded during the lexical decision, and analyzed offline to determine N400 event-related responses to the stimuli in each condition. Although behavioral measures did not reveal any significant differences between groups or conditions, we found a N4 significantly different response to “partword” in the tone-trained group, compared to the monotone. This difference only occurred in a frontal region with a right-hemisphere bias, and was not found to be significant over the left hemisphere. This difference suggests that participants in the tone group were supported in differentiating “words” from “partwords”, supporting the view that the inclusion of tonal information is beneficial in the early stages of L2 lexical learning.
Table of Contents

Listing of Tables and Figures........................................................................................................ii

ACKNOWLEDGMENTS................................................................................................................iv

DEDICATION................................................................................................................................vi

1. INTRODUCTION ..........................................................................................................................1

2. BACKGROUND ...........................................................................................................................4
   2.1. Statistical Learning ...................................................................................................................4
   2.1.1. Transitional Probability ........................................................................................................5
   2.1.2. Speech Segmentation .............................................................................................................7
   2.2. ERP Methods for investigations of lexical processing ............................................................11
      2.2.1. General methodology .........................................................................................................11
      2.2.2. Specific components: LAN ...............................................................................................13
      2.2.3. Specific components: P600 .............................................................................................14
      2.2.4. Specific components: N400 ..............................................................................................15
   2.3. Music and Language ..............................................................................................................21
      2.3.1. Pitch processing in music ...................................................................................................22
         2.3.1.1. Behavioral Studies ........................................................................................................22
         2.3.1.2. Neuroscientific Studies ...............................................................................................24
      2.3.2. Pitch processing in language .............................................................................................25
         2.3.2.1 Behavioral Studies ........................................................................................................27
         2.3.2.2 Neuroscientific Studies ...............................................................................................28
   2.4. Word boundary identification in L2: overview ......................................................................29
      2.4.1. Behavioral Studies ............................................................................................................30
      2.4.2. Neuroscientific Studies ....................................................................................................31

3. RESEARCH QUESTIONS AND HYPOTHESES ......................................................................33
   3.1 Research Questions and hypothesis .........................................................................................33
      3.1.1 Research Question 1: ........................................................................................................33
      3.1.2 Research Question 2: ........................................................................................................34

4. RESEARCH DESIGN AND METHODS ......................................................................................36
   4.1. Participants ............................................................................................................................36
   4.2. Experiment Design and Data Analysis ..................................................................................37
   4.3. Stimuli ..................................................................................................................................38
4.4. Training Methods ........................................................................................................41
4.5. Behavioral evaluations ............................................................................................42
4.5.1. Background Questionnaire ..................................................................................42
4.5.2. Working Memory (WAIS – IV) ..........................................................................42
4.5.3. Tone Test ............................................................................................................43
4.6. Neurophysiological measures ................................................................................44
4.6.1. Lexical Decision task .........................................................................................44
4.7. EEG Nets ................................................................................................................45
4.8. EEG Recording ........................................................................................................46
4.9. Experimental Procedures .......................................................................................46
4.10. Data Processing ....................................................................................................48
4.10.1. Behavioral data ...................................................................................................48
4.10.2. EEG data ..........................................................................................................48
5. RESULTS ....................................................................................................................51
5.1. Participants ..............................................................................................................51
5.2. Standardized Testing .............................................................................................52
5.3. Behavioral Results ..................................................................................................54
5.4. EEG Results ...........................................................................................................56
5.4.1. Usable Trials ........................................................................................................56
5.4.2. Group N400 Analyses .......................................................................................57
5.5. Results Summary ...................................................................................................60
6. DISCUSSION ...............................................................................................................62
6.1. Study Limitations and Delimitations ....................................................................66
6.2. Conclusions and Future Directions .....................................................................67
7 REFERENCES ............................................................................................................71
8 APPENDICES .............................................................................................................100
8.1. Teachers College, Columbia University IRB Approval ........................................100
8.2. Background Questionnaire ..................................................................................101
# Listing of Tables and Figures

<table>
<thead>
<tr>
<th>Figure #</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Illustration of the four experimental conditions associated with Koelsch et al.’s (2004)</td>
</tr>
<tr>
<td>Figure 2</td>
<td>N400 responses</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Sample stimuli from the François and Schön (2010) study</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Perception and Processing of music</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Music Processing</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Speech Processing</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Vowel Space English and French</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Pre-Processing EEG</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Segmentation and Averaging EEG</td>
</tr>
<tr>
<td>Figure 10</td>
<td>N400 Montage</td>
</tr>
<tr>
<td>Figure 11</td>
<td>ACC and RT</td>
</tr>
<tr>
<td>Figure 12</td>
<td>ACC and RT Tone Test</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Topo Plots</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Right Fronto-Central and Left Fronto-Central Montage</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Group Waveforms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table #</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Pseudo words and frequency</td>
</tr>
<tr>
<td>Table 2</td>
<td>Frequencies syllables in each condition</td>
</tr>
<tr>
<td>Table 3</td>
<td>Demographics</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to thank my family, friends, colleagues, and mentors for supporting me in the process to make this dissertation and the completion of my doctoral studies possible.

To my lab family, I am forever thankful of your support and help through this process. Lisa Levinson, thank you for the guidance, for being my official email editor and very close friend. I can’t imagine this process without your constant pep talks and encouragement.

To Lauren Shuffrey, thank you for your constant puppy videos, animal memes, stats support and friendship, this incredible combination helped me stay sane through this process.

To Laura Sanchez, I cannot thank you enough for being my first family and home in this process and for being my friend and academic peer helping solve every issue on the way.

To Paula Garcia, thank you for your friendship and academic advice throughout this process.

And to all my friends in our lab “village”, it takes one to actually achieve such an important step (Dr. Phillip (Trey) Avery, Dr. Felicidad Garcia, Dr. Paula Garcia, Dr. Heather Green, Dr. Reem Khamis-Dakwar, Airey Lau, Lisa Levinson, Lauren Shuffrey, Dr. Chaille Maddox, Dr. Melissa Randazzo, Dr. Guannan Shen, Grace Kwak Wang, and our PI Dr. Karen Froud). Thank you to each and every one of you, your creativity, love, support and amazing heart have made this possible. I am excited to continue working with you and changing this world for the best.
To the extended lab members Alexis Becerra and Sushma Narayan, for helping me with recruitment, with the runs and data management, thank you for your time and willingness to spend hours in the lab running participants.

To the academic staff of the Biobehavioral Sciences Department and everyone that made this possible. Maria Lamadrid, thank you for the endless support, in so many way I could not list them all, you are a great part of making this possible. To Erynn Lowery, thank you for going above and beyond to accommodate and expedite any lengthy process with a smile. To Yvonne Wallace, thank you for the healthy snacks and endless printing ASAP favors. To Dr. Lisa Edmonds, thank you for the departments support in this process, specially this last semester. I definitely could not have made it without it.

To all my friends back home, that supported me with endless conversations and sleepless nights. Thank you for your patience and making me feel it was o.k. to just follow your dreams every step of the way.

To Dr. Peter Gordon, thank you for being my first academic advisor and for your support and guidance through these years.

Special thanks to my dissertation committee members: Dr. Karen Froud, Dr. Peter Gordon, Dr. Levy, Dr. Remez and Dr. Phillip Avery.
DEDICATION

To my family, thank you for all the patience and care through this process.

To my parents Sonia Sepulveda and Jose Moya for the continuous push to “get it done” and “make it work” no matter what and for understanding at the end, that my stubbornness/determination, could have wonderful results. Thank you for your unconditional love and support through this lengthy process.

To my brothers Deivy and Danny Moya, thank you for the late messages and posts that for a brief moment gave me my sanity back. Thank you for you love and for “having my back” when I needed it the most.

To my wonderful husband Eduardo, thank you for the cooking, cleaning, working, and taking care of Ben, through sleepless nights powered by caffeine and so much more that allowed me to get this done. I love you and appreciate every little bit of sacrifice you have made to accompany in this wonderful adventure.

To my little Benjamin, thank you for those smiles, hugs and “talks” that kept me going without any sleep for days. I love you to infinity and beyond.

To my wonderful advisor, Dr. Karen Froud, thank you for your patience, your guidance, professionalism and for being so talented in translating my Spanglish. Your academic training has made me a very rigorous researcher and your life vision a better person. Thank you for making this possible and I hope we can keep making this world a better place, one research at a time.
1. INTRODUCTION

In a multilingual world, a second language is often needed for cultural and communicative participation. This is especially true for the increasing numbers of people living in diaspora populations or immigrating to countries where their language is not the dominant language. In the U.S. alone, there were around 41.3 million new immigrants in 2013 (U.S. Census Bureau's American Community Survey, 2013). This number accounts for around 20% of the world immigrant population, and suggests that around an eighth of the population of the US speaks a first language (L1) other than English (Ryan, 2011).

This situation makes it increasingly important to understand the underlying mechanisms of second language (L2) learning, to ensure rapid and meaningful acquisition of the speech sounds, syntax and vocabulary found in different languages. Research shows that one of the first steps in learning a new language is being able to identify when one word ends and another one begins, permitting the separation of individual words from a continuous speech stream (e.g., Jusczyk & Aslin, 1995). Literature suggests that this ability may be related to the unconscious detection of statistical regularities within sensory input (statistical learning) (e.g., Saffran et al., 1996). Statistical learning is thought to underpin the detection of transitional probabilities within a language – i.e., the probability of two phonemes occurring together, as predicted by the probability of the first phoneme’s occurrence (Saffran et al., 1996).

A question arises, however, when one considers the linguistic input and knowledge of second language learners. What cues are useful when sounds in the speech stream are not similar to those in a speaker's native language? For those attempting to learn a second language as
adults, word identification in a speech stream is extremely difficult (Ellis, 1997). Only after long exposures to speech streams are the statistical regularities salient enough to help extract these cues (Schön, Moreno, Besson et al. 2008), making it challenging to achieve proficiency after short periods of training.

Research has shown that musical information can support learning of different aspects of language. For example, musical expertise can facilitate word identification in tonal languages in adult English monolinguals (Alexander, Wong & Bradlow, 2005), even those who are amateur, rather than professional, musicians (at least 6 years of formal musical training). There is also evidence showing that monolingual speakers detect word boundaries in an artificial language faster if musical tones are added to the syllables (Francois et al. 2012; Schön et al. 2008). There is still a need for research to investigate whether these specific properties may facilitate foundational aspects of language learning, such as word segmentation in L2 learners, when the phonemic properties of the L2 input are different from those of the native language.

In order to address this gap in the literature, this dissertation study examined how musical information can facilitate word boundary identification in second language learning when properties of the speech stream differ systematically from those of the L1. The study investigated the impact of musical tones on learners' word detection abilities for a novel second language (using French vowels) in more naturalistic cross-linguistic learning situations. By using novel word-like stimuli that are either monotonous or mapped onto musical tones, and using electroencephalography (EEG) to measure brain responses, it is possible to determine whether participants are able to automatically detect word boundaries under monotonous and musical listening conditions. The use of EEG means that no overt behavioral response is necessary, so
that we can capture the extent to which participants may be processing the speech stream without conscious awareness even when associated behavioral measures of processing do not reach significance. This could help us understand how learning mechanisms (statistical learning) in combination with acoustic cues (musical tones) can facilitate word boundary detection, a crucial precursor skill for language acquisition. In turn, findings could inform pedagogical and clinical approaches for those learning an L2 or needing support to acquire their L1.

This dissertation is organized as follows. Chapter 2 provides a background for the research: how we extract words from a speech continuum, and how we carry out this same task when we encounter a second language. Chapter 2 also reviews relevant research on the relationship between music and language, and on how musical tone information might facilitate word recognition. Chapter 3 outlines the research questions, hypotheses and expected outcomes of the dissertation study. Chapter 4 discusses the event-related potential (ERP) technique that is implemented in this study along with the dissertation study design and methods. This last chapter will also describe the equipment and electroencephalography (EEG) recording parameters to be used, along with a detailed study protocol, data processing procedures and the approaches used for data analysis. Finally, Chapter 5 presents the results and chapter 6 the discussion of the study findings, the study limitations and future directions.
2. BACKGROUND

In this chapter, background knowledge and supporting evidence will be presented to establish the theoretical framework for several key aspects of this study, including statistical learning, second language acquisition and music as a second language learning facilitator. Each of these topics represents a vast body of knowledge; however, for this study each will be considered specifically in relation to word boundary identification, as this relates to the initial stages of second language acquisition. To study the benefits of using musical tones as a facilitator of word boundary identification, electroencephalography (EEG) and the related Event-Related Potential (ERP) technique were selected because of their superior temporal resolution, that is well-suited to investigating extremely rapid cognitive processes such as those involved in word boundary identification, without relying on the analysis of overt, behavioral responses from participants. Specifically, the N400 ERP component will be discussed, to provide a basis for the research questions that are stated in Chapter 3.

2.1. Statistical Learning

Statistical learning is a relatively novel concept in the study of learning. It is based on the observation that we, as humans, rely unconsciously on the statistical patterns that we encounter in everyday life, in order to learn new information. Statistical learning is therefore based on repetition, how many times we encounter a particular stimulus; for example, if a child sees a vertical line enough times in one context meaning a specific thing, s/he will learn to associate that context with that kind of line (Hu, Xiao, Fu, Xie, Tan & Maybank, 2006). It is hypothesized that the same principle holds for language learning: if you encounter enough repetitions of a
specific linguistic stimulus in a similar context, you start to determine what chunks of information are meaningful in that context and which are not (Saffran, Aslin, & Newport, 1996; Thompson & Newport, 2007; He & Tong, 2017). To be more specific, if one combination of phonemes is heard frequently enough, a listener will assume that it is statistically more probable that those combinations mean something important, compared to phonemes that are rarely or never heard in that specific context. But how many repetitions are necessary? How can we actually calculate how much exposure we need to extract the information? This is where the notion of transitional probability (TP) can help us understand and calculate the necessary degree of exposure for learning (Saffran, Newport & Aslin, 1996).

2.1.1. Transitional Probability

The notion of TP and its application to statistical learning in language dates back to Harris (1955), who first presented the idea that humans can calculate the numbers of phonemic combinations possible in the context of a first phoneme. In his paper, Harris focused on morphemic segments in sentences, but introduced the idea of calculating the numbers of needed repetitions for effective learning based on the perception of an initial phoneme: “for counting at each phonemic position n of a test utterance, all the phonemes that occur in the (n + I) the place (in any utterances) after the particular string of phonemes from the beginning of the test utterance up to n” (Harris, 1955, p212).

Later, Hayes and Clark (1970) elaborated on Harris’ proposal with the suggestion that word segmentation might be initially approached by learners as clusters of information. Saffran, Newport and Aslin (1996) were the first to associate the specific term Transitional Probability
(TP) with the probabilistic calculation of phonemic combinations due to the occurrence of the first phoneme, though some authors still use the older term “conditional probability” instead. Saffran, Newport and Aslin (1996) defined TP as “the probability of one event (e.g., of one syllable) given the occurrence of another event. This statistic refers to more than the frequency with which one element follows another, as it adjusts for the base rate of the first event or element”. The TP of Y given X is represented by the following equation (Saffran, Newport & Aslin, 1996, p. 610):

\[
TP = P(Y|X) = \frac{\text{frequency}(XY)}{\text{frequency}(X)}.
\]

Evidence for the use of TP in statistical learning has been observed in both children and adults, who track these statistical variables in syllables, thereby unconsciously detecting when words end and begin (Aslin et al., 1998; Hauser et al., 2001; Saffran, Aslin, et al., 1996a; Saffran, Newport, et al., 1996b; Saffran et al., 1999; Toro and Trobalón, 2005; Pelucchi, Hay, & Saffran, 2009). Because this mechanism is available at a young age (Thiessen & Saffran, 2003), it is thought to be very important during language acquisition (Aslin et al., 1998; Saffran, Aslin, et al., 1996; Swingley, 2005).

In order to evaluate the use of transitional probabilities for word boundary detection, researchers have used artificial languages to simulate language acquisition (e.g., Saffran et al. 1996a, 1996b; Aslin et al. 1998; Kuhl 2004; Gervain et al. 2008; Teinonen et al. 2009). These studies, a few of which are described in greater detail below, ask participants to listen to wordlike stimuli from an artificial language without acoustic cues (that would normally aid the
process of word boundary identification), thereby forcing participants to rely on transitional probabilities only. Even without available acoustic cues, and with minimal exposure to novel word forms, participants demonstrate the ability to segment words correctly by processing statistical information only (Saffran, Newport, et al., 1996b; Gordon et al., 2010; Francois & Schöen, 2010; Schöen & Francois, 2011).

2.1.2. Speech Segmentation

Speech segmentation is the initial challenge encountered in comprehension of verbal discourse in either a first or a second language, during the early stages of acquisition or learning (e.g. Cole and Jakimik, 1980). Trying to understand a continuous verbal stream, when listening to someone speak in an unfamiliar language, is known to constitute a learning and cognitive challenge. Two lines of research on this phenomenon have dominated the field: developmental work has focused on babies and toddlers learning their first language (e.g., Khul, Williams, Lacerda, Stevens & Lindblom, 1992; Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993; Jusczyk & Aslin, 1995), while other researchers have targeted second language learners to determine whether they rely on the same abilities as children when learning a language (e.g., Stein, Dierks, Brandeis, Wirth, Strik & Koenig, 2006; Wang, Khul, Chen & Dong, 2009; Cunillera et al., 2009). When listening to a familiar language, the task of speech stream segmentation is easier, because there are several cues available to support this task. For example, prosody, context, grammar, and lexical knowledge will help in identifying words that otherwise would (perceptually) merge together (Echols, Crowhurst & Childers, 1997; Levy, 2009; Lany & Saffran, 2011).
One explanation for this complex ability emerges from Bloomfield’s notion of “minimum free form”, the idea that children avoid word segmentation in spoken language and just identify minimal meaningful aspects of utterances to gain initial insights into the semantics of their L1 (Bloomfield, 1933). This approach breaks down, however, when we consider all the articles and function words that are present in verbal discourse, even in motherese, that have no concrete meaning; and the fact that closed class words generally are acquired after open class words (Saffran, Newport & Aslin, 1996). Bloomfield’s approach also places an undue emphasis on the speaker, seeming to imply that words initially should be presented in isolation, in order to facilitate language acquisition. On the contrary, studies have shown that parents do not speak in single words even to young babies and toddlers (Thiessen, Hill & Saffran, 2005). What parents do emphasize in child-directed speech, though, is the stress of the word, usually at the end of a word. This characteristic of motherese, which is applied even though it sometimes results in atypical utterances that would be considered ungrammatical if adult-directed, enhances the importance of the post-utterance pause, giving cues to word segmentation (Ferguson, 1964; Fernald & Khul, 1987; Thiessen, Hill & Saffran, 2005).

Another possibility is that word segmentation could be facilitated through prosodic cues. Echols (1993) examined the idea that some prosodic elements, such as stress patterns, could make speech segmentation easier. She concluded that stress patterns are more salient to children than to adults for word identification, but that stress remains a primary cue for segmentation even into adulthood. Other research has shown that especially mothers use a higher pitch in infant-directed speech (Fernald & Mazzie, 1991) and stress some syllables that correspond usually with the first words of infants (Echols and Newport, 1992). A large amount of research supports the
importance of prosodic cues for language acquisition (Aslin, Saffran, & Newport, 1998; Endress, & Hauser, 2010; Buiatti, Peña, & Dehaene-Lambertz, 2009; Tremblay, Coughlin, & Bahler, 2012). Unfortunately, prosodic cues alone cannot support word boundary identification for language acquisition, since we are exposed to individual variations in pronunciation, speech rate, pitch, stress placement and other features; not to mention the variability between each language (Norris, McQueen, Cutler, & Butterfield, 1997; Aslin, Saffran, & Newport, 1998; Strange, Weber, Levy, Shafiro, Hisagi & Nishi, 2007; Thiessen & Erickson, 2013). For example, Hyman (1977) pointed out that only 211 of 444 languages use stress as a fixed component either at the beginning or end of the word. Even if stress cues were reliably associated with word boundaries, relying on stress patterns alone would require innate recognition of the stress patterns in an L1. Even worse, the learner (child or adult, for an L1 or an L2) would have to know where in the word the stressed syllable is located (initially, medially or finally) – an improbable learning challenge for those acquiring a language, at whatever stage in the lifespan (Chobert & Besson, 2013).

A third possibility for word boundary identification comes from the idea that a semantic component is initially associated with a phonemic combination, due to exposure to both the word and to specific concrete elements of the environment (Osgood & Sebeok, 1954). For example, a child knows what “cat” means because they encounter this acoustic percept in the presence of their family pet. A challenge to this account comes from the observation that semantic associations in language acquisition come much later in development (Wojcik & Saffran, 2013), but children at even 8 months of age are able to respond to word-like units in continuous speech.
without any additional nonlinguistic information being present in the environment (Jusczyk & Aslin, 1995; Saffran, Newport & Aslin, 1996a).

All the theoretical approaches presented above show the importance of all these components of language – meaning, prosody, stress, environmental cues – that are used in the process of learning a new language. Unfortunately, none of these approaches provide a definitive account of the initial stages of word identification in a continuous speech stream; nor do they shed light on cross-linguistic commonalities that could facilitate this learning task.

A fourth possibility, and the one upon which this study is based, is that there is an innate ability of humans to calculate the statistical regularities associated with phonemes in a speech stream, and to use this information to determine when a word begins and another one ends. This specific statistical regularity that can be detected is what we refer to as transitional probability (TP). Saffran, Newport and Aslin (1996) provided the first behavioral evidence that transitional probabilities are important for word segmentation in language acquisition. Twenty-four undergraduate monolingual English speakers were exposed to a continuous speech stream in a novel language, for 21 minutes. The stimuli were constructed from combinations of four consonants and three vowels, resulting in 11 syllables that were then combined to give six trisyllabic words. The transitional probabilities between phonemes across words (between 0.1 and 0.2) and within words (between 0.31 and 1.0) were controlled. After exposure, participants were asked to determine whether specific words belonged or did not belong to the newly learned language. The authors proposed that correct discrimination between words and non-words in the novel language would indicate that participants were able to access statistical properties of the language for word segmentation, since no other possible cues (such as prosody) were provided.
Indeed, findings indicated that participants were able to distinguish words from non-words, with better recognition for the words, hence demonstrating identification of the word boundaries using statistical probabilities alone.

This study shows that even adults can extract statistical information from a new language as a first approach to determining word segmentation, when no other linguistic information is available. It is important to note that the authors described this paradigm as a second language learning task, but their stimuli in fact made use of the same phonemes as the participants’ native language, English.

2.2. ERP Methods for investigations of lexical processing

Because this dissertation study uses Event-Related Potential (ERP) methodology, in what follows I will provide an overview of these methods and the specific ERP components that provide indices of aspects of linguistic and musical processing. Against this background, I will review the similarities and differences between musical and linguistic processing in the brain. Next, I will review existing behavioral and neuroscientific research that relates to lexical acquisition – specifically, word boundary identification – in L2, and the processing of pitch in both musical and linguistic contexts. Finally, I will summarize the gap in the literature that this study was designed to address.

2.2.1. General methodology

Speech is a rapid process that unfolds on the scale of milliseconds. Because of this, it is fundamental to examine what is happening in the brain (where speech is both prepared and
decoded) on a very precise temporal scale. Electroencephalography (EEG) is a neuroscientific approach that responds to this demand for timing precision. This method allows analysis of cortical brain responses to word recognition in adults and children. Event-Related Potentials (ERPs) can be derived from the EEG recording, and can provide evidence about the neural responses that are correlated with processing non-native words in adulthood. Both EEG and the ERP method are described in more detail below.

Electroencephalography (EEG) is a method for indexing brain functions by recording electrical activity generated by large populations of neurons as summed voltages at the scalp. It is a non-invasive procedure that can be applied repeatedly in individuals, and carries minimal risk. The high temporal resolution of EEG recordings makes it a very suitable method for examining rapid responses to auditory speech stimuli. Even though the spatial resolution of this method is not as precise as that provided by other neuroimaging methods such as functional magnetic resonance imaging (fMRI), it is possible to identify signal generators of specific components recorded through EEG, especially through the use of high-density systems (Luck, 2005).

Event-related Potentials (ERPs) are derived from the continuously recorded EEG, by averaging together the time-locked or synchronized electrical activity with responses to multiple instances of a cognitive event, such as a sound being heard, or a word being recognized. This averaging process enhances the signal-to-noise ratio of recordings and removes non-phase-locked activations (such as eye blinks and heart beats) so that (in principle) only activation related to the event of interest is represented in the averaged data. In this way, the ERP method provides a means for examining the brain’s time-locked responses to specific cognitive events
(Handy, 2005). ERP components are characterized by simultaneous multi-dimensional measures of polarity (negative or positive voltage deflection), amplitude, latency, and scalp distribution. Several ERP components have been identified as reflecting activation associated with particular cognitive processes. A few such components that are known to be associated with the processing of language and music are described below.

2.2.2. Specific components: LAN

One ERP component that has been used to research language and music is the left anterior negativity (LAN) response, a negative voltage deflection over left anterior (language processing) regions observed around 100 to 500 milliseconds after stimulus onset (Kutas & Besson, 2002). This component has been linked to working memory tasks and syntactic processing. Within this extended time window, two related ERPs have been identified: the LAN itself and an earlier component referred to as the early left anterior negativity (ELAN). The ELAN appears to be primarily associated with the rapid detection of word-category errors in sentence processing (Neville, Nicol, Barss, Foster & Garret, 1991; Hahne & Jescheniak, 2001). The LAN on the other hand has been related to detection of morphosyntactic errors (Osterhout & Mobley, 1995; Coulson, King & Kutas, 1998).

A related component that has been used more in investigations of music processing is referred to as the early right anterior negativity, or ERAN. This negativity has been associated with violations of melodic context (Brattico, Tervaniemi, Naatanen & Petretz, 2006; Miranda and Ullman, 2007), being elicited for example by harmonically unexpected chords (Fritz et al., 2013). The ERAN has also been seen to interact with LAN, linking music-syntactic processing
of chord functions and the processing of morpho-syntactic information in language (Koelsch, 2011; Maidhof & Koelsch, 2011).

### 2.2.3. Specific components: P600

Another component used to register linguistic processes and music processing is the P600 that usually follows the ELAN or LAN. Compared to (E)LAN, the P600 appears to reflect a higher level of control, as it may be influenced by the percentage of incorrect items (such as morphosyntactic violations) presented within an experiment (Hahne and Friederici, 1999). The P600 is also called the syntactic positive shift (SPS), because it represents a generalized response to syntactic errors (Hagoort, Brown & Groothusen, 1993). The P600 ERP is usually seen in response to a syntactic violation in a sentential context, and it constitutes a large positive-going voltage deflection around 500 to 800 milliseconds after stimulus onset (Osterhout & Holcomb, 1992).

Recently the P600/SPS component has also been associated with the processing of semantic anomalies in sentences (Kim & Osterhout, 2005). Furthermore, this component has been elicited by the presentation of syntactically erroneous musical information, such as harmonic violations (Patel, Gibson, Ratner, Besson & Holcomb, 1998). Similarly to language, music is thought of as having a syntactic congruency, structural rules that are followed and maintained within musical phrases and pieces (e.g., Piston, 1978). When these musical syntactic congruencies are violated, a P600 response can be observed (Patel et al., 1998).

The use of the P600 as a measure of music processing in the brain started with Besson and Faïta (1995), who compared ERP responses to deviant notes in musical melodies presented
at the end of a short musical phrase. This experiment elicited significant P600 responses from listeners to such harmonic violations. Later, Patel et al. (1998) tried to establish whether the P600 could be considered an index of uniquely linguistic processes, or whether it indexes more general cognitive operations that involve processing basic relationships in rule-governed sequences. For this they examined music and language processing in two experiments comparing simple, complex and ungrammatical syntactic structures in language and in-key, nearby-key and distant-key chords for music. When comparing results between tasks and conditions the syntactically incongruous words in language and harmonically incongruous chords in music were statistically indistinguishable in terms of amplitude and scalp distribution for the P600.

The P600 offers a useful tool to research the processing of syntactic information within a phrase, whether that phrase is musical or linguistic. For the purposes of this study, however, this component is associated with a level of complexity that new L2 learners cannot achieve as a first step in second language acquisition. In order to observe brain responses that are more closely associated with lexical knowledge, we will focus on an earlier ERP component, referred to as the N400.

2.2.4. Specific components: N400

The N400 is an ERP component described as a negative deflection that peaks around 400 milliseconds after stimulus presentation, thought to reflect the processing of semantic anomalies in linguistic stimuli (e.g., Young & Rugg, 1992). This component has also been elicited in tasks requiring musical processing. For example, Koelsch et al. (2004) used musical excerpts and sentences to prime semantically related and unrelated words. Participants (nonmusicians)
listened to known musical excerpts that were associated with specific meanings, and they also listened to sentences. The musical excerpts were followed by words that were associated with the intended meaning of the music, or anomalous words; and the sentences likewise were completed with either semantically expected or unexpected words (see figure 1 below). EEG was recorded during this listening task and processed offline to derive ERP responses to the musical and linguistic stimuli. In both conditions the N400 component was enhanced in response to the unexpected word, either following a musical phrase or at the end of a sentence (see Figure 2). Koelsch et al. concluded that their findings indicated that both linguistic and musical primes set up semantic expectancies that can be realized linguistically, and that N400 acts as an index of expectancy violation in both domains.

Figure 1. Illustration of the four experimental conditions associated with Koelsch et al.’s (2004) Experiment that demonstrated priming of words through either musical or linguistic primes. To the right are shown corresponding grand-averaged ERPs comparing the response to target words, following presentation of semantically related and unrelated prime sentences and musical phrases. Reprinted from “Music, language and meaning: brain signatures of semantic processing” by Koelsch et al., 2004, Nature Neuroscience, 7, p. 303. Copyright 2004 by Nature Publishing Group. Reprinted with permission.
The N400 has also been used as a marker of brain activation associated with cognitive processes bridging music and language. Gordon, Schön, Magne, Astésano and Besson (2010) presented two conditions (linguistic and musical) to non-musician speakers of French. Participants were trained to associate specific words with specific sequences of notes, and then EEG was recorded while they listened to 120 words in four different conditions: words that were the same with melodies that were the same; words that were the same although the melodies were varied; words that differed although their associated melodies were kept the same; and words that differed and that were played with different melodies (see figure 2 below). During the experiment, participants were asked to monitor, and indicate via button press, whether the linguistic information or the melodic information varied from expectation. Both the musical and the linguistic conditions elicited a larger N400 component when the melody and word were not congruent, indicating that the N400 ERP can act as an index of neural processing when a musical or a linguistic expectancy is violated. The N400 component that was observed in the musical and linguistic conditions was similar in terms of scalp distribution, latency and duration (Gordon et al., 2010).
François and Schön (2010) used a similar method to associate musical tones with language-like stimuli. They used French phonemes to develop novel wordlike stimuli controlling for transition probability (TP) for both the phonemic and tonal combinations. Participants, all native speakers of French, were first exposed to a training period, where they heard the wordlike stimuli for 15 minutes. After the training they were asked to respond a lexical decision task while EEG was recorded. Participants had to determine whether they heard a sequence of syllables or tones to which they had previously been exposed (a “word”), or a novel syllable or tone sequence that bridged a boundary (a “part-word”). The authors found that N400 amplitudes (between 400 and 900ms) were enhanced for part-words compared to words, whether these were enhanced for part-words compared to words.
sequences of syllables or of tones (see figure 3 below). This N400 enhancement reflects the statistical learning that participants achieved during the learning phase of the experiment.

In a subsequent study, François and Schön (2011) reduced the learning phase time to 5.5 minutes, and examined statistical learning performance in both musicians and non-musicians. They found that an N400 response was enhanced, as before, in response to part-words for the musicians compared to the non-musicians in both linguistic and musical tests. The main difference between the two groups was that non-musicians showed a later N400 peak in response to the unfamiliar “words” compared to the musicians, whose ERP responses were shifted earlier in latency. François and Schön concluded that both groups processed familiar and unfamiliar items very differently, with musicians having an advantage for the detection of word boundaries overall.

Schön, Boyer, Moreno, Besson, Peretz and Kolinsky (2008) compared linguistic processing that was associated with exposure to a monotone speech stream or to a speech stream that incorporated musical tones (pitch information). Behavioral (reaction time and accuracy) data
revealed a significant advantage for participants in the tone-training condition. Though these findings are suggestive, note that this study has not been replicated with the addition of neuroimaging methods; nor has it been carried out in such a way that participants were exposed to phonemes in the wordlike stimuli that were not a part of their L1 (in this case, French). Hence, it is difficult to claim that this study really provided an L2-like learning experience.

The studies reviewed in this section provide evidence that the N400 ERP component is a viable target to act as an outcome measure in studies of word recognition at very early stages of language acquisition (specifically, word boundary identification). More research is needed to understand whether exposing monolinguals to a completely new set of word-like stimuli that contain non-L1 phonemes would provide a more realistic model of L2 lexical acquisition processes. Such an approach could provide valuable insights into the early processes of second language acquisition. Given the research to date, it is possible that the addition of musical tones to novel word-like stimuli supports word boundary identification only in an artificial learning task that maintains the features of L1 phonology (a native phoneme inventory). In order to understand whether this strategy generalizes to a more realistic L2 lexical acquisition task, it is necessary to evaluate the impact of musical tone addition to wordlike stimuli that make use of non-native phonemes. Furthermore, the addition of brain imaging to these behavioral paradigms would provide additional insights into the underlying processes that support word boundary identification in a novel, L2-like, speech stream. Such insights could provide a foundation for development of more effective pedagogical strategies to support second language learners in the early stages of L2 lexical acquisition.
2.3. Music and Language

Music and language have long been known to be associated with similar neural pathways and systems (e.g., Patel, 2007). These observations came originally from evolutionists and linguists, dating back to the 1700s. Rousseau, for example, suggested that music was the precursor of language, and that both musical and linguistic phenomena emerged as a response to the need to organize human society (Rousseau, 1781, reprinted 1986). Darwin similarly proposed that language emerged from music, but he was supportive of the idea that language had emerged – evolved – from a need for sexual communication, much like that engaged in by animals (Darwin, 1871). The long-standing debate about the origins of, and relationships between, music and language has distilled into three main points of view: music came first; language came first; or they co-occurred. All three perspectives, however, agree that there is an intertwined relationship between music and language.

Several brain studies have examined parallels between music and language, in terms of similar ERP components or overlapping neural pathways (e.g. Zatorre, 2013; Fritz, Poeppel, Trainor & Schlaug, 2013; Peretz, Vuvan, Lagrois & Armony, 2015; Patel & Morgan 2016). Understanding the brain mechanisms associated with music processing has been a long quest (e.g., Critchley & Henson 1977), though in the last couple of years it has gained renewed attention. According to Peretz and Zatorre (2005) there are two main reasons for this resurgence. Firstly, the investigation of music in the brain provides a window into brain organization for acquired complex skills – for example, in musicians. Secondly, investigating the underlying mechanisms of music processing could elucidate the extent to which music-related networks are distinct from the ones involved in language. In the next section, some neuroscientific studies of
music perception will be reviewed, in order to elucidate some of the neural relationships between music and language.

### 2.3.1. Pitch processing in music

There are several studies that have shed light on the neural mechanisms of music perception. Early findings came from lesion deficit studies, case studies of brain damage that impacted musical processing. Milner (1962), for example, found that right temporal lobe damage would result in a greater deficit in pitch processing, compared to similar damage that affected the left hemisphere of the brain. This led to the understanding that the right temporal neocortex is important for the interpretation and understanding of pitch. Later studies showed that pitch deficits, such as an inability to perceive pitch of a complex tone with missing fundamentals\(^1\),\(^2\) (Zatorre, 1988) and increased perception of pitch direction change (Johnsrude, Penhun & Zatorre, 2000) are associated with damage to the right anterolateral part of Heschl’s gyrus (primary auditory cortex).

#### 2.3.1.1. Behavioral Studies

Lesion deficit research has provided insights into the complexity of music perception and processing in the brain (see figure 4). The study of amusia, acquired or congenital, has helped us

---

\(^1\) "The perceived pitch of a periodic complex tone need not correspond to a harmonic partial physically present in the signal" (p.566) Zatorre, 1988.

\(^2\) "An auditory illusion of pitch perception that arises if the set of overtones or harmonics associated with a certain fundamental frequency are presented to a listener without the fundamental frequency: the listener perceives the sound as having the pitch of the fundamental although the fundamental frequency is not present in the sound". Colman dictionary of Psychology, 2015.
to understand how pitch related difficulties are alike or differ. People with congenital amusia are unable to perceive dissonance (Ayotte et al., 2002; Peretz, 2001a, b, c), for instance, but all amusic individuals who present with difficulty in recognizing or producing music are systematically impaired on particular dimensions such as pitch or temporal organization (Peretz, 2006).

![Figure 4. Perception and Processing of music](image)

Figure 4. Perception and Processing of music

Contrary to expectations, research has shown that individuals who have perfect pitch do not necessarily outperform others in making judgments of musical interval (Burns & Campbell, 1994; Miyazaki, 1992, 1993). This line of research has explicated one kind of link between language and music – the naming of musical notes (Schulkin & Raglan, 2014). Research into musical expertise has also provided insights into the influence that musical training can have on
speech and language perception and learning. For example, Marie, Delogu, Lampis, Belardinelli, Olivetti and Besson (2011) looked into the relative abilities of musicians and non-musicians who were trying to learn a tonal language (Chinese). The authors found that musicians showed lower error rates on a tone identification task compared to nonmusicians, suggesting an advantage in musical skill that generalized to the processing of specific linguistic information.

All this information has shed light on necessary mechanisms needed to understand how pitch is processed in music and how this can relate to language. Neuroimaging research has also been used to investigate the brain regions and activations that are associated with these complex cognitive processes.

2.3.1.2. Neuroscientific Studies

Currently the general consensus is that pitch perception (Zatorre & Gandour, 2008; Hyde, Peretz & Zatorre, 2008) and production (Ozdemir, Norton & Schlaug, 2006) are processed in right temporal neocortex and the right anterolateral part of Heschl’s gyrus (primary auditory cortex). More complex information – such as the identification of pitch height (Warren, Uppenkamp, Patterson & Griffiths, 2003) or contour (Tervaniemi, 2003) – requires recruitment of association areas. These associations are mediated through two projection pathways, the anteroventral and the posterodorsal pathways (Zatorre & Salimpoor, 2015). As shown in figure 5 below, the ventral stream projects from the primary auditory cortex towards the superior and inferior temporal gyri and sulci to the inferior frontal cortex. The dorsal stream, on the other hand, also projects from the primary auditory cortex but to parietal, premotor and dorsolateral frontal cortices. Research has shown that music processing has a slight right hemisphere bias,
though both hemispheres are involved (Peretz, Gosselin, Belin, Zatorre, Plailly & Tillmann, 2009).

**Figure 5. Music Processing**
An illustration of the two bidirectional association pathways (dorsal and ventral) involved in music processing. These pathways originate from the primary auditory cortex, and project either to dorsal areas like parietal and premotor cortices, or to ventral areas such as the inferior temporal cortex and gyrus. Reprinted from “From perception to pleasure: music and its neural substrates” by Zatorre and Salimpoor, 2013, *Proceedings of the National Academy of Sciences of the United States of America*, 110, p. 10431. Copyright 2013 by PloS ONE. *Reprinted with permission pending.*

### 2.3.2. Pitch processing in language

Similar to the proposal that musical information is processed via dual streams in the brain (Zatorre and Salimpoor, 2015), Hickok and Poeppel (2004; 2007) proposed that speech processing requires at least two association streams in order for acoustic information to be mapped onto linguistic meaning. This dual-stream model (see figure 6 below) suggests that speech processing is bilaterally organized in the brain, and that there are multiple routes to lexical access that act as parallel channels to permit processing of multiple and often redundant spectral and temporal cues in the speech signal. Hickok and Poeppel support their proposal with evidence from patients with unilateral hemispheric damage, split-brain patients (patients who have no corpus callosum as a result of surgical sectioning), and individuals who have
experienced Wada procedures (a procedure during which one hemisphere of the brain is anaesthetized to examine lateralization of various functions). The findings suggest that there is at least one pathway in each hemisphere to access the mental lexicon (the individual’s mental repertoire of words). This model of speech processing has been evaluated experimentally in multiple studies, a few of which are described below.

**Figure 6. Speech Processing**
2.3.2.1 Behavioral Studies

In speech, there is a wide variety of cues that can give us information about the auditory speech stream. Variation over time can provide information that is important for utterance interpretation – for example, prosodic cues that indicate whether an utterance is a question or a statement, or that carry emotional content (Moore, 2008). Fairbanks and Pronovost (1939), for example, examined the characteristics of pitch that are related to sadness. They found that usually lower pitch and a smaller pitch excursion are associated with utterances that are interpreted by listeners as being sad.

Pitch perception in speech has been long studied in terms of fundamental frequencies and vowel or consonant perception, giving us more information into how speech is perceived. For example, one line of research has examined infant directed speech (IDS), sometimes referred to as motherese, a speech pattern that purposely modifies aspects of pitch (among other linguistic properties) and that is known to support the acquisition of speech and language in infancy and childhood. Specifically, it has been found that this type of speech has generally a higher and wider pitch range, compared to adult-directed speech patterns (Fernald & Simon, 1984; Snow & Ferguson, 1977; Stern, Spieker & MacKain, 1982; Stern, Spieker, Barnett & MacKain, 1983).

We can see that pitch changes in speech relate to different kinds of information that can be extracted from the continuous speech signal. Research using neuroimaging methods has helped to identify the brain activations and connections that underpin this complex mapping.
Hickok and Poeppel’s (2004; 2007) model of speech processing proposes that early cortical stages at the Superior Temporal Gyrus (STG) are bilateral (figure 6). Then, two streams are identified as processing different types of data: The ventral stream, left-lateralized with a bilateral component, is related to speech comprehension, mapping sound into meaning. The dorsal stream, left-dominant, is related to mapping sound into articulatory-based representations. These claims have been evaluated using brain imaging, especially in bilingual populations.

Second language acquisition research has shown volumetric differences between bilinguals, who have larger Heschl's gyri (HG, primary auditory cortex), and monolinguals (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2006; Wong, Warrier, Penhune, Roy, Sadeh, Parrish, & Zatorre, 2008). Results also suggested that differences in the size of HG were related to experience with a second language and not to an innate feature (Ressel, Pallier, Ventura-Campus, Diaz, Avila and Sebastian-Galles, 2012). As well as anatomical changes, L2 acquisition is also associated with activation differences. Hesling, Dilharreguy, Bordessoules, and Allard (2012) used fMRI to examine differences between the ventral and dorsal pathways in two groups of adult late bilinguals with different L2 proficiency levels (high and moderate) while performing a comprehension task in both L1 (French) and L2 (English). Hesling et al. examined activation in the ventral and dorsal pathways to determine differences in processing of L1 and L2 stimuli between highly proficient and moderately proficient listeners. The authors found that L1 and L2 connected prosodic speech stimuli were associated with activations in the same dorsal and ventral networks in highly proficient L2 listeners however, moderately proficient L2...
listeners only exhibited common L1 and L2 activations in the superior temporal sulcus and the medial temporal gyrus. The moderately proficient listeners showed no ventral pathway activation while processing connected speech in either their L2 or their L1. Hesling et al. concluded that L2 processing at different levels of proficiency is supported by differences in the integrated activity of distributed networks that include the left posterior superior temporal sulcus, the left Sylvian parietal temporal region, and the left pars triangularis.

Studies like these have demonstrated that learning a second language results in structural and functional brain changes. However, it is still unknown whether learning a second language in adulthood results in the same or similar changes in language-related brain structures (e.g., Ressel et. al., 2012). Studies using neuroimaging techniques such as MRI, fMRI, e- fMRI, TMS, and PET are contributing to a more detailed understanding of similarities and differences among monolingual and bilingual children and adults, while controlling for factors such as age of acquisition and experience with the language. The next section discusses the overlapping literature looking at neural correlated of music and language using the Event-Related Potential (ERP) technique to study non-native speech perception. In this section I will focus on one specific component of the ERP, the N400, and present research evidence on the suitability of this component for elucidation of the processes that underpin word boundary identification.

2.4. Word boundary identification in L2: overview

Second language learning is a complex process that has gained much attention in the last century (Block, 2003; Loewen and Gass, 2009). Globalization of languages and the increasingly multicultural make-up of populations across the world have enhanced the importance of
understanding and facilitating this process, especially in adults. Usually second language acquisition research, or L2 research, focuses on complex and advanced language functions, such as phonemic accuracy or lexical acquisition (Fricke, Kroll & Dussias, 2016; Granlund, Hazan, Baker & Rachel, 2012; Hoshino & Guillaume, 2011; Cuetos, Bonin, Alameda & Caramazza, 2010). This dissertation study focused instead on the initial states of second language acquisition; specifically, word boundary identification, which is a necessary step for lexical learning.

2.4.1. Behavioral Studies

Just as babies encounter the significant processing task of learning to segment words in their first language, L2 learners face the same challenge when encountering a new language (Jusczyk & Aslin, 1995). Because natural speech does not typically incorporate pauses between words, and it is thought that there are no common rules to identify word boundaries (Aslin et al., 1998; Saffran, Aslin, et al., 1996 and Saffran; Newport, et al., 1996), the challenge of segmenting words in a second language is considerable. The difficulty of this task is exemplified by the wide variability of evidence regarding the onset of word segmentation abilities during the course of L2 acquisition; research has shown that it varies depending on the language (Newman, Bernstein Ratner, Jusczyk & Dow, 2006; Singh, Reznick & Xuehua, 2012) and depending on individual proficiency and ability (Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008). This variability is attributed to differences in phonological saliency and consistency in language-specific phonological patterns (Kooijman, Junge, Johnson, Hagoort & Cutler, 2013).
Another challenge associated with L2 word boundary identification is the extent to which an individual L2 learner is able to identify similarities between the L2 and their L1, and apply this information to the L2 input when necessary (Al-Jasser, 2008; Tremblay, Coughlin, Bahler & Gaillard, 2012; Tremblay & Spinelli, 2014). For example, Weber and Cutler (2006) looked at highly proficient German users of English as a second language, and native speakers of American English. Participants were required to listen to nonsense sequences and respond whenever they detected an embedded English word. Results showed that both German and English speakers had good mastery of English boundary information, though German speakers were unable to suppress their native boundary constraint probabilities, resulting in higher error rates.

2.4.2. Neuroscientific Studies

At the lexical level, one challenge for word learning in an L2 would be the processing of cognates – that is, words that have a common etymological root. Acheson, Ganushchak, Christoffels, and Hagoort (2012) developed an EEG experiment to evaluate the brain responses associated with a picture naming task. Stimuli were bilingual (Dutch-German) cognates, and they targeted the error-related negativity (ERN) component, an ERP signature that is associated with conflict resolution and can be elicited during language production tasks. Acheson et al. showed that ERN amplitudes were enhanced in response to cognates, compared to non-cognates. These results suggest that a cognate naming task might involve lexical and phonological access in both languages, generating a response conflict. Hoshino and Thierry (2011) also showed this processing conflict in an EEG study, observing that amplitudes of the targeted ERPs varied when elements of L1 and L2 were lexically related, whether the relationship was phonological,
semantic or phono-translational (that is, targets that were phonologically related to the translated picture name). Prosodic differences between L1 and L2 can also lead to interference effects that apparently prevent non-native listeners from attending to these patterns (Tremblay, Broersma, Coughlin & Choi, 2012).

Research into the early stages of L2 is subject to extreme variability, based on L1 interference, individual proficiency, learning methods and many other factors. Hence, it is very difficult to investigate the early stages of L2 acquisition in a controlled manner. This dissertation study aims to support investigation of the earliest stages of L2 acquisition, by attempting to identify cues that could facilitate aspects of early language acquisition independent of prosodic or phonological patterns associated with the L1 or L2. In the next chapter, I specify the research questions, hypotheses and predictions associated with the dissertation study, given the background provided up to this point.
3. RESEARCH QUESTIONS AND HYPOTHESES

This chapter presents the research questions, related hypotheses, and expected outcomes of the dissertation study.

3.1 Research Questions and hypothesis

The purpose of this study is to investigate the mechanisms underlying one aspect of the complex relationship between music and language, using neurophysiological methods to gain new insights into the processing of word boundaries in a lexical learning task. This research also contributes to our understanding of the ways in which statistical regularities can be manipulated to impact language learning, as reported by Gordon, Schön, Magne, Astésano, & Besson (2010).

In addition, this study seeks to replicate behavioral findings reported by Schön et al. (2008), supporting the idea that music facilitates second language acquisition by providing reliable cues to segment speech. By gaining a deeper understanding of the relationships between musical (tone) processing and lexical acquisition in a language learning situation, findings from the dissertation study could facilitate the development of educational strategies to support the teaching and learning of L2, independent of the sociocultural background or phonemic inventory of the learners.

3.1.1 Research Question 1:

Does training with musical tones associated with phonemes of a second language alter the brain indices of lexical access in learners, following limited exposure to a novel L2 speech stream?
• Hypothesis 1: Representation of musical tone is associated with lexical representation in the brain, in such a way that even limited exposure to a novel speech stream will, when paired with musical tones, result in the identification of statistical regularities in the speech stream.

Prediction: Based on this hypothesis, we predict that English monolinguals trained with musical tones associated with the phonemes of a pseudo-language (that uses non-L1 phonemes) will show a significantly greater N400 amplitude in response to pseudowords (part-words) than whole words from the pseudo-language in a lexical decision task, compared to a monotone-trained group.

3.1.2 Research Question 2:

Does training with musical tones associated with phonemes of a second language help learners to consciously identify lexical items in a novel language, following limited exposure to an L2 speech stream?

• Hypothesis 2: The facilitation of word boundary identification by the addition of musical tones as seen in the brain data will support statistical learning in a way that extends to the behavioral domain, allowing for the more accurate and faster identification of words vs pseudowords (part-words) in a novel L2-like speech stream.

Prediction: Based on this hypothesis, it can be predicted that word boundary identification will be more effective in a condition where English monolinguals are trained with
musical tones associated with the (non-L1) phonemes of a pseudo-language, compared to those trained in a monotone pseudo-language condition. This is predicted to result in the tone-trained group showing significantly shorter reaction times and increased accuracy in a task requiring them to identify words and part-words of the pseudo-language, compared to the monotone-trained group.

In chapter 4, I describe the dissertation research study that was designed to address these questions and hypotheses.
4 RESEARCH DESIGN AND METHODS

This chapter describes the research design and methods, including participants, behavioral and neurophysiological measures, experiment design, stimuli, and procedure. The data reported in Chapter 5 were collected using the following procedures.

4.1. Participants

Twenty-two participants were recruited for this study, 19 – 39 years of age, all typically developing (TD) monolingual English speakers. The inclusion criteria for all participants were as follows:

- Must pass a hearing screening the day of the experiment (500 Hz, 1000 Hz, 2000 Hz at 20 dB)
- Must be a monolingual English speaker (not able to hold a conversation in any L2 by self-report)
- Do not have formal musical training (more than 3 years), or practice any instrument regularly (three times or more a week with any kind of training).

Volunteers were recruited from the tri-state area through advertising at Teachers College, Columbia University and nearby institutions, word of mouth, social media postings (e.g., Facebook, online Teachers College, Columbia University message boards) and outreach to local Universities. Informed consent was obtained from all participants, and they were assured that participation was voluntary and that consent could be withdrawn at any time without penalty. All
procedures were carried out with approval from the Teachers College Institutional Review Board.

**4.2. Experiment Design and Data Analysis**

The study followed a 2 X 2 mixed experimental design. Two factors were built into the study: Group, a between-subjects factor determined by the type of training session (tone trained vs. monotone trained), and Condition, a within-subjects factor that reflected stimuli constituting either Words or Part-Words (words: word-like stimuli from the novel “language” with high transitional probability; part-words: part of the end of a word-like stimulus presented together with the beginning of another one – such stimuli were heard in the training but with lower transitional probabilities).

Dependent variables in this study were as follows. Behavioral measures looked at accuracy of responding (percentage of correct identification of a word or a part-word, via button press) and response times in milliseconds, during a lexical decision task (described in detail below). Neurophysiological measures examined brain event-related responses over fronto-central and midline electrodes, since amplitude differences in the target ERP (N400) have been reported over these areas (e.g., Kutas & Federmeier, 2011). Specific measures of interest were amplitude (peak negative amplitude across electrodes of interest within a 250–600 ms time window) measured in microvolts (µV), and latency (time in milliseconds post-stimulus onset to the N400 peak). Amplitude was measured in two ways, first for the peak response (looking for greatest negative deflection in the relevant time window), and secondly for mean peak amplitude in a time window of 60 milliseconds centered around the identified negative peak (Luck, 2005).
Mean peak amplitudes were entered into data analysis procedures, since they avoid arbitrary identification of a physiological response that does not necessarily directly reflect a peak in neural activation (Luck, 2005). Behavioral and neurophysiological data were analyzed within and between groups to evaluate differences in behavioral responses as well as in the N400 mean amplitude and/or latency associated with the two training conditions and the two types of stimuli. Correlations between brain and behavioral responses were also investigated.

4.3. Stimuli

The stimuli were composed of four consonants and three vowels. These were combined to form an array of 12 syllables, from which 11 were used in order to provide a probability range similar to that used by Saffran et al. (1996). By combining these syllables, six tri-syllabic word-like stimuli were produced: [gimysy], [mimosi], [pogysi], [pymiso], [sipygy], [sysipi] (table 1); in turn, this array yielded six part-words, combinations that cross the word boundaries of the six “words”: [gysimi], [mosigi], [pisipy], [pygymi], [sogimy], [sypogy] (as used by Schön et al., 2008).
Table 1. Pseudo words and frequency
Specific IPA notation for the stimuli, their tone sequence and specific frequency for each phoneme

<table>
<thead>
<tr>
<th>Word (IPA)</th>
<th>Tone sequence</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[mimosi]</td>
<td>Db4Bb3E3</td>
<td>277 – 233 – 165</td>
</tr>
<tr>
<td>[pogysi]</td>
<td>B3A3E3</td>
<td>247 – 220 – 165</td>
</tr>
<tr>
<td>[pymiso]</td>
<td>Ab3Db4D4</td>
<td>208 – 277 – 293</td>
</tr>
<tr>
<td>[sipgy]</td>
<td>E3Ab3A3</td>
<td>165 – 208 – 220</td>
</tr>
<tr>
<td>[sysipi]</td>
<td>D3E3Gb3</td>
<td>147 – 165 – 185</td>
</tr>
</tbody>
</table>

Transitional probabilities between syllabic pairs were maintained across the training set, ranging from 0.31 to 1.0 within each word and from 0.1 to 0.2 between words. These were calculated over all the possible combinations that could result from 216 random repetitions each for each of the six words, using Saffran et al.’s (1996) formula:

\[ TP = P(Y|X) = \frac{\text{frequency}(XY)}{\text{frequency}(X)}. \]

This approach permitted calculation of the frequency of one syllabic pair given the frequency of the first syllable. Bisyllable frequency is the probability of two particular syllables appearing together during the speech stream. For example, in order to calculate the probability of the combination [sipy] (which is actually part of a “word” in this stimulus set), we had to calculate the frequency of [sipy] (how many times did [sipy] appear in the whole speech stream)
and divide by the frequency of [si] (how many times did [si] appear in the whole speech stream). This was true for all syllable pairs, even those that were not part of a word. For example, the syllabic pair [sipo], which is only a part-word, appeared in the speech stream, because of the nature of the continuous speech training – no pauses or breaks. This grouping of syllables results from mixing the end of one word-like stimulus and the beginning of another one. This means, therefore, that the frequency of a pair of syllables from a word-like stimulus was higher than that of part-word syllable pairs. This statistical information was used to find word boundaries (more probable vs less probable syllable pairs) and hence to determine whether a combination was part of the six word corpus to which participants were exposed during training, or if it was a part-word that formed because of random properties of the speech stream.

The speech stimuli were synthesized using Mbrola (speech synthesizer) using a female French voice (Polytechnique de Mons - TCTS lab, 1998). During generation of the speech stream, no acoustic cues were inserted at word boundaries. First a monotone (200Hz per word) continuous stream of syllables was generated, and secondly tones were added to each syllable to obtain the training set for the musical tones condition. The monotone speech stream applied a fundamental frequency of 200Hz per phoneme, the average of the normal fundamental frequency per phoneme (following Schön, Boyer and Moreno et al., 2008). The second speech stream was created by using syllable-pitch mapping to match each syllable to a different musical tone (Table 1 and 2) while maintaining the statistical structure of syllables and tones, following procedures applied by François and Schön (2010). All stimuli were matched for duration (720 ms per word) and intensity (70 dB SPL); all manipulations of the speech sounds were carried out using Praat v5.3.2 (Boersma & Weenink, 2005; Wood, 2005).
Table 2. Frequencies syllables in each condition
Specific frequencies (in Hertz, Hz) for all syllables in each condition

4.4. Training Methods

Participants were trained by listening to a concatenated speech stream. This speech stream was identical to that used by Schön, Boyer and Moreno (2008), with one difference: in this case there were vowel contrasts not found in the native language of the participants, increasing the difficulty (specifically the high front rounded vowel /y/, which is not found in Standard American English) (see Figure 7 below).

Figure 7. Vowel space English and French
Differences in vowel space between French (left) and English (right). Reprinted from “Illustrations of the IPA: French” by Fougeron and Smith, 1993, *Journal of International Phonetic Association*, 23, p. 73. Copyright 1993 by Creative Commons Attribution License. Reprinted with permission.
Each word-like stimulus for the training set was selected from the word-like and part-word sets described above. Each word-like stimulus was repeated 72 times in a pseudo-random order during the training, with the stimulus presentation constrained by the requirement that the same word can never be repeated twice in a row. Auditory stimuli were presented using Eprime v 2.0.8.90 (Psychology Software Tools Incorporated, Pittsburgh, Pennsylvania), in free field, via a Tannoy OCV 6 full-bandwidth speaker suspended 193 cm over the participant’s head in a sound-attenuated chamber. The total duration of training per participant was 7 minutes and 3 seconds; the training condition for each participant was randomly assigned until there were equal numbers of participants in each group.

4.5. Behavioral evaluations.

4.5.1. Background Questionnaire.

A background questionnaire was administered to obtain general demographic information, as well as information about participants’ language exposure, educational background, and musical training (see Appendix B).

4.5.2. Working Memory (WAIS – IV).

The Digit Span and Arithmetic subtests of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS–IV; Wechsler, 2008a/b) was used as a measure of working memory. A working memory evaluation is important for the experimental protocol because working memory has been correlated with the ability to learn new vocabulary (Papagno et al., 1991; Service, 1992; Papagno and Vallar 1995; Gathercole et al. 1997; Baddeley et al. 1998; Gathercole et al. 2003). In order to avoid one potential source of confound, this study attempted to maintain consistency
between the groups with respect to working memory span. The Digit Span subtest of the WAIS-IV consists of three tasks: Digit Span Forwards, Digit Span Backwards and Digit Span Sequencing. All three subtests require participants to repeat a certain set of numbers that are read aloud by the experimenter. In the first task the digit sequences are repeated forward; in the second task the sequences are repeated in reverse order; and in the third task participants sequence the digits in numerical order. The Arithmetic subtest of the WAIS-IV asks the examinee to mentally solve a series of arithmetic problems to measure basic mental manipulation, concentration, attention, short- and long-term memory, numerical reasoning ability, and mental alertness. This last task was included so that a standard score could be obtained for the WAIS-IV Working Memory composite score.

4.5.3. Tone Test.

In order to ensure that participants were able to discriminate differences between tones prior to including them in the study, all participants carried out a tone test using only the tones associated with the training. This test presented pairs of tones of the same frequencies as the musical tones used in the task stimuli (see Table 1 above), and asked participants to indicate (via button press) whether the two tones in each pair were the same or different. The corresponding 11 tones were generated for this task (and matched to each syllable), using an online function generator: http://www.wavtones.com/functiongenerator.php. Each tone was repeated as part of a pair 6 times, with an ISI of 1000 milliseconds. Tones for the tone test had a duration of 50 milliseconds, were downloaded with a sampling rate of 44.1kHz, and were all amplitude normalized in Praat to have a sound pressure level of 70dB. The tone test was administered under the same testing conditions, and in the same room, as the EEG recordings (see below).
4.6. Neurophysiological measures

This study used electroencephalography (EEG) recordings in order to derive the primary outcome measure, the N400 event-related potential. EEG was recorded during a lexical decision task. EEG recording equipment and parameters, and the nature of the lexical decision task, are described here.

4.6.1. Lexical Decision task

Continuous EEG was recorded using NetStation v4.5.6 (Electrical Geodesics Inc., Eugene, OR) after training, during administration of a lexical decision task, requiring participants to make lexical decisions (via button press) in response to word-like and part-word stimuli as described above. The inter-stimulus interval (ISI) and the inter-trial interval (ITI) were set at 1000ms each. Stimuli for the lexical decision task were presented under the same conditions, and in the same room, as the training was administered. Stimulus presentation and collection of button press data (response and reaction times) were controlled via Eprime v 2.0.8.90 (Psychology Software Tools Incorporated, Pittsburgh, Pennsylvania).

Participants were asked to determine (by pressing one of two buttons) whether each auditorily presented stimulus was heard or not heard previously as part of the new “language” training. The task took approximately 14 minutes in total (two blocks, each 7 minutes long). Each word-like and part-word stimulus was repeated 12 times each during each block, in a pseudo-random order avoiding consecutive presentations of the same stimulus, yielding 72 trials per condition per block (word vs. part-word). After each block there was a break, to make sure
participants were comfortable and to maintain electrode impedances (see below) before continuing with the second block.

4.7. EEG Nets

All EEG recordings took place in the Neurocognition of Language Lab, in the Department of Biobehavioral Studies at Teachers College, Columbia University. The lab uses a 128 electrode, high density HydroCel EEG recording system (Electrical Geodesics; Tucker, 1993) manufactured by Electrical Geodesics, Inc. The electrodes themselves are made of carbon fiber silver chloride embedded in small sponges that before use are soaked in an electrolyte solution (potassium chloride) to ensure optimal conductivity. This system permits the rapid and accurate application of large numbers of electrodes in high-density arrays with minimal time while maximizing participant comfort and safety. The circumference of each participant’s head was measured to ensure the correct size sensor net was selected, and the vertex (center of the head) marked to ensure accurate placement of the net. The net was then soaked in a potassium chloride solution for five minutes, which consists of 1 liter of water purified via reverse osmosis, two teaspoons of potassium chloride, and 3 cc’s of Johnson & Johnson baby shampoo to break down oils on the scalp. The sensor net was then applied, and each electrode appropriately seated. Next, the participant was seated in a chair in a sound attenuated room inside the laboratory containing the computer monitors that delivered the task instructions and stimuli. The sensor net was connected to an amplifier (EGI Net Amps 300 system) that was checked and calibrated prior to each experimental run. Impedances (loss of signal between scalp and sensor) were measured by feeding a minute (400 microvolt) electrical field through each electrode, which was then ‘read back’ by the acquisition system so that the amount of signal loss could be calculated.
4.8. EEG Recording

Continuous EEG data was recorded using EGI’s Netstation (v4.5.6) data acquisition software (Electrical Geodesics Inc., Eugene, OR) with a sampling rate of 500 Hz, or a sample taken every 2 ms (Luck, 2005). The high density hydrocel nets and associated high impedance amplifiers have been designed to accept impedance values ranging as high as 100kΩ, which permits the sensor nets to be used without scalp abrasion, recording paste, or gel (e.g., Ferree, Luu, Russell, & Tucker, 2001; Pizzagalli, 2007); however, we maintained electrode impedances below 40kΩ during EEG recording (see below). Electrode impedance was reassessed and electrodes were rehydrated with potassium chloride solution as needed. The data recording was monitored in real time and bad channels and artifacts were noted and marked so they could be addressed using offline processing techniques. During recording, the vertex channel (Cz) was used as the reference electrode; later offline processing included re-referencing of recorded data to the average of all electrodes (average reference: Luck, 2005).

4.9. Experimental Procedures

This section details a step-by-step of the procedures that experienced by each participant.

1. Prior to the participant arriving at the lab, participants were asked to answer a short background questionnaire about their education, language and musical training. This was administered through RedCap. (5min approx., see appendix B)

2. The participant were presented with a consent form and asked to read it carefully. Risks were explained fully and any questions were answered before the participant signed the form. They were also reminded that they could withdraw consent at any time.
3. A hearing screening was administered (500 Hz, 1000 Hz, 2000 Hz at 20 dB).

4. The head circumference of each participant was measured and the net size corresponding with that measurement was determined. The researcher then located the correct position for the vertex electrode, at the center of the head, according to lab protocol in order to ensure proper placement of net electrodes. The location for the vertex electrode was marked on the participant’s head using a soft wax pencil for later reference.

5. Participants were introduced to the experimental tasks and equipment. Then they were instructed to “listen to the sounds carefully” for the training, lasting about 14 minutes and 6 seconds.

6. Once training finishes, participants were fitted with an appropriate 128-channel HydroCel Geodesic Sensor Net (HCGSN) (Net Amps 300, Electric Geodesics Inc., Eugene, OR). Electrodes were positioned with reference to the vertex marking made previously on the participant.

7. Participants were seated in a chair in a sound attenuated chamber within the Neurocognition of Language Lab. A video camera gave the researcher visual information about the participant during the experiment. The amplifier was checked and calibrated, the EEG sensor net was connected, and impedances (loss of signal between scalp and sensor) were measured. In order to improve impedances the electrodes were adjusted as necessary so that they were making good contact with the participant’s scalp. EEG recording started.

8. Participants were instructed to decide whether the words being played to them were heard or not heard previously, as part of the “language” they were just exposed to. Sounds were played through free field speakers, and participants indicated their decisions about the “wordness” of each stimulus via button press.
9. Once the experiment ended, participants were asked to listen to a series of pure tones and to decide, when presented with two sounds, if they were different or the same. Participants indicated their decisions (same / different) via button press.

10. Following the final tone test, the sensor net was removed.

11. Finally, participants were asked to complete the verbal working memory assessment, including Digit Span and Arithmetic (WAIS-IV). After this the experiment was completed, and participants were given time for questions and debriefing.

4.10. Data Processing

In this section the initial procedures for processing data and deriving the ERPs from the continuous EEG recordings are described.

4.10.1. Behavioral data

Reaction times (RT) to the word-like and part-word stimuli in the lexical decision task, and accuracy (ACC) of responses, were recorded as behavioral measures for later correlation with the neurophysiological data. Mean RT and ACC per group (tone-trained vs monotone-trained) were used for statistical comparisons.

4.10.2. EEG data

The recorded EEG data were pre-processed using a standard ERP analysis protocol (Handy, 2005; Luck, 2005; Picton et al., 2000). See Figure 8 below. The pre-processing procedures were carried out within the EGI NetStation (v5.4) software.
Following pre-processing and removal of artifact and error trials from analysis, segments of the continuous EEG recordings were averaged together to increase the signal-to-noise ratio and to identify the time-locked event-related responses associated with the onset of the stimuli. The EEG epochs were 700 milliseconds long, taken from 100 milliseconds prior to each stimulus onset to 600 milliseconds post. Epochs from each experimental condition were averaged together for each participant, and the averaged waveforms were baseline-corrected to control for drift. Baseline correction procedures involve subtracting the average electrical potential during the 100ms baseline period from the epoch of interest in order to bring the recording closer to zero, further increasing the signal-to-noise ratio by removing baseline activity that occurred prior to stimulus presentation. Grand averages for each condition were derived by averaging together all the processed data for each participant in each group (see figure 9 below).
Analysis was constrained to a montage of electrodes of interest in the current study. The N400 ERP is typically found at frontocentral electrodes, for experiments using pseudo language and transitional probability (e.g., François & Schön, 2010). We used a frontocentral montage composed of electrodes: 20, 12, 5, 118, 13, 6, 112, 7, 106 (see figure 10 below).

Figure 10. N400 Montage
Fronto-central montage for identification of the N400 Event-Related Potential
5. RESULTS

This chapter will describe the results for electrophysiological and behavioral data, as described in previous chapters.

5.1. Participants

Twenty-one English monolingual individuals between 19 and 39 years of age (mean = 26.91, SD = 5.48) participated in the study. Three participants could not be included in the analysis because of artifacts in their recordings or issues with the equipment. Therefore, a total of eighteen participants were included in the analysis, nine in each group.

Fourteen of the participants were right handed, per self-report, and one reported to be left-handed. They all passed a hearing screening (500 Hz, 1000 Hz, 2000 Hz at 20 dB) on the day of the EEG recording.

Written consent was signed by all participants on the day of data collection and all were assured that they could withdraw consent at any time during the duration of the study without penalty. Participants were reimbursed $15 dollars for their time. Compensation was paid for by a Teachers College, Columbia University Dean’s Grant that was awarded to this study in 2015. All procedures were carried out under approval from the Teachers College Institutional Review Board (see Appendix A).
5.2. Standardized Testing

Research has shown that there are many factors contributing to learning a language successfully. Among these are intelligence (Skehan, 1986), foreign language aptitude (Carroll & Sapon, 1959; Carroll, 1981) and working memory capacity (Sawyer & Ranta, 2001). Because we were looking at initial states of language acquisition, we used the working memory subtest of the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS–IV; Wechsler, 2008a, b) to provide a measure of non-linguistic processing ability in the participant group.

All participants scored within average range, which was congruent with the inclusion criteria. See table 3 below.
Table 3. Demographics
Demographic characteristics of the 18 participants included. Self reported language knowledge and musical exposure and scores for Working Memory subtest in WAIS. (T= music tone training; M= monotone training)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Gender</th>
<th>Language</th>
<th>Music</th>
<th>Handedness</th>
<th>WAIS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>F</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>SUPERIOR</td>
</tr>
<tr>
<td>T2</td>
<td>F</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>VERY SUPERIOR</td>
</tr>
<tr>
<td>T3</td>
<td>F</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>AVERAGE</td>
</tr>
<tr>
<td>T4</td>
<td>F</td>
<td>Monolingual</td>
<td>None</td>
<td>L</td>
<td>AVERAGE</td>
</tr>
<tr>
<td>T5</td>
<td>M</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>HIGH AVERAGE</td>
</tr>
<tr>
<td>T6</td>
<td>M</td>
<td>Monolingual</td>
<td>Played/sing for fun once in a while for more than 5 years</td>
<td>R</td>
<td>HIGH AVERAGE</td>
</tr>
<tr>
<td>T7</td>
<td>F</td>
<td>Monolingual</td>
<td>Played/sing for fun once in a while for more than 5 years</td>
<td>R</td>
<td>AVERAGE</td>
</tr>
<tr>
<td>T8</td>
<td>F</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>AVERAGE</td>
</tr>
<tr>
<td>T9</td>
<td>F</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>HIGH AVERAGE</td>
</tr>
<tr>
<td>M1</td>
<td>F</td>
<td>Monolingual</td>
<td>Played/sing for fun once in a while for more than 5 years</td>
<td>R</td>
<td>SUPERIOR</td>
</tr>
<tr>
<td>M2</td>
<td>M</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>AVERAGE</td>
</tr>
<tr>
<td>M3</td>
<td>F</td>
<td>Monolingual</td>
<td>None</td>
<td>R</td>
<td>SUPERIOR</td>
</tr>
<tr>
<td>M4</td>
<td>F</td>
<td>Monolingual</td>
<td>Played/sing for fun once in a while for more than 5 years</td>
<td>R</td>
<td>HIGH AVERAGE</td>
</tr>
</tbody>
</table>
### 5.3. Behavioral Results

The behavioral measures obtained from the lexical decision task were analyzed for each group (tone-trained and monotone-trained). Accuracy and reaction times were submitted to arcsine and log transformations respectively, to obtain data sets that were more normally distributed, hence minimizing the probability of type I or II errors (Cohen & Cohen, 1983). There was a technical glitch for three participants (one in each group), resulting in partial data sets; these participants were therefore excluded from only this analysis. Arcsine transformed accuracy and log-transformed reaction times for both groups (tone-trained and monotone-trained) were analyzed through a two-way t-test including all participants whose data was collected (n = 16). Arcsine and log-transformed values are presented below in figure 11.

Raw and arcsine transformed accuracy scores (ACC) were measured by calculating the proportion of correct responses per condition out of the total number of trials presented. A two way ANOVA was conducted to compare main effects of group and condition. Participants in the
tone-trained condition were more accurate at distinguishing between words and part words than participants in the monotone training condition. However, this difference did not reach significance; there were no significant differences between groups (F (1, 16) = 0.159, p = 0.696, \( \eta^2 = 0.010 \)) or conditions (F (1, 16) = 0.780, p = 0.390, \( \eta^2 = 0.046 \)) for ACC.

Reaction time (RT) was calculated as the time elapsed from stimulus presentation to the button-press response. A two way ANOVA was conducted to compare main effects of group and condition. Both groups appeared to take approximately equal amounts of time to respond to both stimulus types (words and part-words), and there were no significant differences between groups (F (1, 16) = 0.14, p = 0.714, \( \eta^2 = 0.09 \)) or between conditions (F (1, 16) = 0.065, p = 0.802, \( \eta^2 = 0.004 \)).

![Figure 11. ACC and RT](image)

Average RT in milliseconds after log transformation and average ACC after arcsine transformation (as a proportion of all trials) for each condition and each group (9 for the monotone-trained group and 9 for the tone-trained group). In the average the RT was included for both correct and incorrect trials. WD=word, PW=part word. Error bars indicate standard error of the mean.
As mentioned in previous sections, everyone was tested for tone deafness with a simple tone test. Every participant passed the tone test and an ANOVA was performed to compare the Tone Test ACC and RT responses across groups (see figure 12). This interaction was not significant (F (1, 16) = 0.764, p = 0.395, \( \eta^2 = 0.132 \)) Even though the ACC for the participants included in the tone group was higher (63% compared to 50% in the monotone group), this finding was still not significant.

![Figure 12. ACC and RT Tone Test](image)

*Average RT in milliseconds after log transformation and average ACC after arcsine transformation (as a proportion of all trials) for each condition and each group (9 for each group). In the average the RT was included for both correct and incorrect trials. WD=word, PW=part word. Error bars indicate standard error of the mean.*

5.4. EEG Results

5.4.1. Usable Trials

Following data pre and post processing, numbers of usable trials were documented for each participant and mean usable trials were calculated for each of the two groups. For both groups it was challenging to obtain many trials because of muscle artifact (Luck, 2005). Each block had 144 trials (72 per condition). A comparison was carried out between the numbers of usable trials in the first and second block for all participants, to determine whether participant fatigue could be a factor. It was found that Block 2 contained significantly fewer usable trials for
the whole cohort ($F(1, 32) = 0.014, p = 0.011, \eta^2 = 0.000428$), and the decision was taken to therefore examine only trials from the first block for the EEG data analysis.

Across groups, we wanted to include only participants who were not outliers in terms of the numbers of usable trials. We compared the number of trials between conditions and groups and only included those who were not significantly different from each other on this measure ($F(1, 32) = 0.004, p = 0.948, \eta^2 = 0.000120$). This resulted in the exclusion of 3 participants from further analysis, two monotone trained and one tone trained.

5.4.2. Group N400 Analyses

Prior to analysis, pre-processing procedures were implemented as described in the previous chapter. Data were segmented with respect to stimulus onset in NetStation. Each trial consisted of values recorded from 128 electrodes sampled every 2ms, for epochs starting 100ms prior to stimulus onset and ending 600ms post-stimulus, for every participant. Data were then reduced to include only the electrodes of interest (a montage selected based on Magne, Schön & Besson, 2003; Schön & Francois, 2011). Individual averages were computed by averaging each individual’s epochs associated with responses to words and part words, across electrodes within the montage. In this way individual average ERP waveforms were generated (Luck, 2005; Picton et al., 2000). Individual averaged data were exported to R Studio for further analysis. Mean amplitudes within the N400 time window (200 – 600 ms post-stimulus onset) were computed for each condition and each individual (e.g., Näätänen & Picton, 1987). The N400 response was considered to be present if there was a negative mean over the electrodes of interest during the target time window (Kutas & Federmeier, 2011). Individual average waveforms were then
averaged together within groups (tone-trained vs. monotone-trained), to provide grand averaged waveforms for each condition (word vs part-word).

Inspection of group waveforms derived from the frontocentral montage revealed an N400 response from both tone-trained and monotone-trained participants. Amplitude values (in microvolts, µV) during the time window of interest (200 msec – 600 msec post stimulus onset) were entered into a 2 (Group: Tone Trained vs Monotone Trained) x 2 (Condition: Word vs Part Word) ANOVA, which revealed no significant main effect of Condition (word vs. part word: F (1, 16) = 2.563, p = 1.29, η² = 0.138) or Group (F (1, 16) = 0.492, p = 0.097, η² = 0.163).

Topo plots revealed that there was a frontal right lateralized activation for the tone trained group (Figure 13), that was specially negative for the partword condition. Due to this observation, two new montages were developed, examining right and left hemisphere sensors in the frontal region (following Schön & Francois, 2011) (Figure 14).
Figure 13. Topo Plots
First row shows the electrode 6 from the fronto-central original montage and the waveform graph at 400ms for all conditions and both groups color coded (blue: pw monotone group, red: ww monotone group; green: part-word tone group and black: ww from the tone trained group) The last four rows represent the graph for electrode 124, part of the new right fronto central montage and the graph at 200, 400, 600 and 800ms

Figure 14: Right Fronto-Central and Left Fronto-Central Montage
For the RFC (right fronto-central) montage electrodes 124, 4, 10, 3, 123 were selected and for the LFC (left fronto-central) electrodes 24, 19, 18, 23, 27 were selected.
A two way ANOVA was performed comparing both hemispheres for each group with each condition. The main effect of Group was significant (F (1, 32) = 4.515, p = 0.41, \( \eta^2 = 0.124 \)). A two-tailed t-test was carried out on amplitude values included in each condition for each group, to determine the source of the significant main effect. This revealed a significant difference between responses to part-words from the monotone group and the tone-trained group (t (8) = -2.538, p = 0.035) (see Figure 15). Specifically, in the N400 time window (200 - 600 ms), the amplitude deflection in response to part words (the N400) was greater for participants who were tone trained, but not for those exposed to monotone training stimuli.

![Figure 15. Group waveforms](image)

Group waveforms for the RFC montage for each condition. Words = red, Part words = blue

### 5.5. Results Summary

In summary the main finding of this dissertation study were:

1) No significant differences in behavioral measures (ACC and RT) between groups or between conditions. There was a non-significant difference in RT for monotone-trained participants vs. tone-trained participants (the former
responded more quickly to part-words than the latter); and also a non-significant difference in ACC between groups (tone-trained group was more accurate than the monotone-trained group in identifying words).

2) Participants in the tone-trained group showed a significant enhancement of the N400 amplitude in response to part words compared to words.

3) Participants in the monotone trained group did not show any significant difference in the N400 amplitude between conditions.

4) Only the tone-trained group showed a right hemisphere activation bias (activation greater in the right frontal hemisphere compared to the left).

5) There were no significant differences between groups for the tone test.

These results will be discussed in the next chapter.
6. DISCUSSION

Results from this study provide initial ERP evidence in support of the prediction that musical/tonal association with lexical stimuli can support the initial stages of word segmentation in a novel language, even when the novel language contains phonemes different from those instantiated in the L1. Even though the original hypothesis placed the significant difference in a more central scalp location, the significantly enhanced right fronto-central N4 response seen in the tone-trained group when they were asked to identify novel words vs. part-words, compared to the non-significant effect for the monotone-trained group on the same task, serves as a good indicator that these two groups processed the same speech stream in different ways. The prediction that the N400 response to part-words would be enhanced after tone training is supported by previous research findings showing that adding a musical tone to word stimuli in a novel language facilitates word boundary identification (e.g., Gordon, Schön, Magne, Astésano, Besson, 2010; Francois & Schön, 2010; Schön & François, 2011). Though these research lines suggested that the relevant activation would appear over central electrode locations, the present finding of a bias towards the right hemisphere can be accounted for in several ways. Here, I focus mainly on three explanations related to the fact that the group showing this hemispheric distinction was trained with stimuli associated with musical tones, alongside the fact that this pseudo language can be considered a “second language” (L2).

fMRI studies (Courellis, Mullen, Poizner & Cauwenberghs, 2017) have shown that right frontal activation is linked with the Anterior Cingulate Cortex (ACC) activation. ACC is known to be activated when processing emotion in music (Green, Bærentsena, Stødkilde-Jørgensen, 62
Wallentind, Roepstorfb and Peter Vuustd, 2008). Preira (2011) found a greater activation when participants listened to music that they liked vs. disliked music, as well as when the music presented was familiar (compared to unfamiliar). Bush, Luu and Posner (2000) suggest that the ACC is associated with ventral pathways, linked to affective processing, as well as a dorsal pathway that is connected with higher cognitive processing.

Following these findings, the anterior right frontal activation observed in the present study could be due to the characteristics of the musical tones added to the training. All the musical notes added were part of a major scale. Mizuno and Sugishita (2007) noted a correlation between happy–sad judgments of the major–minor mode in isolated chords presented to musicians with activation in the inferior frontal gyrus (BA 47), the medial thalamus, and the dorsal anterior cingulate cortex. Hence, it is likely that our participants in the tone-trained condition are activating dorsal ACC, resulting in the observed right-lateralized activation in response to the experimental stimuli.

Another explanation for the amplitude-enhanced N400 over right hemisphere sensors could be related to pitch processing. Activation in Broca’s area right-hemisphere homolog has been observed in imaging experiments examining pitch processing in music (Zatorre, Evans & Meyer, 1994). Lesion deficit research, mostly examining patients with acquired amusia, has also revealed an association between frontal anterior activations and the processing of prosody and musical contour (Gingras, Musil, Stewart & Mu, 2014).
The third possible explanation for the anterior activations observed here is associated with second language processing. There has been research linking frontal activation in response to learned phonemic contrasts in a new language after training (Golestani & Zatorre, 2004), though usually this response is observed bilaterally (Fiez, Raichle, Miezin, Petersen, Tallal & Katz, 1995).

All these proposed explanations may contribute to the observed experimental findings, considering we used a new vowel contrast in the context of pseudo-L2 exposure, with the deliberate intention of forcing participants to make use of pitch contour (musical tones) as one tool to help them anchor word boundaries in a continuous speech stream. This might be why the present study did not replicate the findings of Schön, et al. (2008), since their participants were all French monolingual speakers and their stimuli used only French phonemes.

The current findings suggest that, even when wordlike stimuli in a novel language are phonetically dissimilar to the native language of the listeners, the addition of tonal information to the speech stream is still advantageous for the perceptual task of word segmentation. The ERP findings from this study are therefore encouraging and have potential relevance for L2 pedagogy.

The behavioral results (reaction times and accuracy rates for the word/part-word discrimination task) indicated that participants were not able to consciously distinguish the novel words from part-words – strings of speech sounds that were not heard in the training stimulus set. Differences in accuracy between groups (tone-trained vs. monotone-trained) and
between conditions (identification of novel words vs. partwords) were not significant. Similarly, there were no differences in reaction times between groups or conditions. These results are contrary to those previously reported in the literature. For example, in a study by Schön, Boyer, Moreno, Besson, Peretz and Kolinsky (2008), participants showed faster reaction times and more accurate responses to words when they received tone training, compared to the monotone-trained group. Since this study showed no significant difference in either reaction time or accuracy between the monotone and tone-trained groups, it is possible that the behavioral measures used here were not sensitive enough to capture differences perceived by participants when they were exposed to wordlike stimuli that are phonetically dissimilar from their native language. Divergence between ERP findings that indicate recognition of words vs. part-words, and behavioral findings that indicate no such recognition, supports the use of brain measures in tasks like these. With only behavioral measures, the erroneous conclusion may be drawn that non-native-like stimuli in a novel speech stream cannot be accurately processed and recognized, suggesting that there is a limitation on the effectiveness of tone training (i.e., that tone training only supports word boundary identification in second language learning when the novel linguistic stimuli are phonetically very similar to the L1 of the listeners). However, the brain measures show that word boundary identification is possible, and is supported by tone training, even when the vowels in the novel linguistic stimuli are distinct from L1 speech sounds; albeit at a level of processing that is not – given the short exposure times provided in this experimental paradigm – reflected in behavioral classification responses.
6.1. **Study Limitations and Delimitations**

The present pilot study has some limitations. It is not known whether participants may have had unreported exposure to the French language (for instance participants may have visited France or heard the French language spoken in school or other settings). Similarly, there is a possibility of unreported musical exposure (for example, participants may have grown up watching or listening to a close relative play an instrument, or may have been peripherally associated with school music activities, and so on). It may not be possible to control for or even identify all such exposures.

There are also delimitations to the study, for example the **number of participants** included in the ERP analysis. Data were collected from 21 participants, but due to excessive movement and other sources of artefact, only 18 were included in the ERP analysis, 9 in each group. Although previous studies used only 36 trials per condition (e.g., Francois & Schön, 2010), and we included 72 trials per condition, it was still problematic to obtain sufficient numbers of trials to achieve statistical power in the analyses. In an attempt to address this, the experiment was repeated twice; but analysis revealed that the 1st block was significantly different from the second one in terms of numbers of usable trials (indicating that participant fatigue played a role), so we were not able to use more trials. To address this issue, larger group sizes are needed.

There is a flip side to this solution, however. Although 72 trials per condition is a higher number of trials than research published in similar studies (e.g., Gordon et al., 2010), it may be insufficient to completely capture the electrophysiological indices of word and part-word
processing in this task. However, raising the number of trials any further would likely increase the amount of movement artefact in the EEG recordings, and would burden participants with a longer run and greater attention load. Still, recruitment remains extremely difficult, and adding greater numbers of participants may not be possible as a measure to increase statistical power in the analyses.

Another limitation of the study is the fact that 7 minutes were used for training time, following Schön et al. (2008). Though initially we thought this would be too short for a language with new phonemes, results suggest that we might need more participants to possibly reveal the effect found by Schön et al. (2008). It could also be that in this model, using cross-linguistic phonemic contrasts, the task is simply too challenging to allow significant changes in participant behavior.

6.2. Conclusions and Future Directions

Much further research is needed to elucidate how language and tones interact and whether it may be beneficial to facilitate this interaction to support the learning of a second language. This study contributes to this question, by providing evidence about relationships between aspects of musical processing (musical contours added to lexical items) and aspects of language learning (the identification of word boundaries in a novel speech stream). A richer understanding of these connections could support development of new educational strategies to facilitate L2 acquisition in different populations.

Here we used experimental parameters based on previous studies (Gordon et al., 2010; Schön et al., 2008; Schön et al., 2004; Magne et al., 2007) to investigate whether the documented
supporting influence of tonally-varied stimuli for word boundary identification in language learning could extend to a language-learning task involving word-like stimuli that are phonetically dissimilar from the native language(s) of the participants. The design and stimuli employed yielded data of value for addressing the research questions. However, some parameters could usefully be altered for future research.

The first suggested modification is to increase either the number of participants included in each group, or the number of trials completed by each participant. A larger number of participants could potentially yield a larger number of trials to enter into analysis without the need to increase the data collection time for each participant. Fatigue appears to be an important factor in this learning paradigm, so greater numbers of participants may be the more ecologically valid approach to enhancing statistical power of the study findings.

Secondly the background questionnaire presented to participants should be altered to be more exhaustive and detailed in terms of language exposure or musical ability. The additional information collected could permit data analysis approaches such as subgrouping, which provides an opportunity for greater nuance in interpretation of study findings. As it happens, in this study some potential participants disclosed their knowledge of other languages and reported that they “were” fluent but not anymore. No participants ho disclosed such information about their language background were included, but a more detailed questionnaire would be of value in eliminating such potential confounds.

A third aspect that could be modified is the instruction during the training phase. Participants in the present study were instructed to just listen to the sound. It is possible that
behavioral measures might change more when participants are given more information – for example, that this is a new pseudo language and that they should “look for the words”. Maybe this would also shift the observed N400 negativity more centro-frontally (as per our initial predictions), possibly indicating that participants are required to allocate fewer cognitive resources to process words and part-words of this new pseudo language.

Therefore, future studies could include these elements to better control some variables and hopefully generate results that could develop even better informed strategies to teach a second language.

Findings suggest the representation of musical tone can be associated with lexical learning in the brain, and that this association can reinforce the identification of statistical regularities in the speech stream. This conclusion is supported by the observation of a significant N400 amplitude enhancement in response to part-words in the lexical decision task, that was seen for the tone-trained group only. This observation suggests that, even when there is limited exposure to a novel speech stream, participants are able to identify statistical regularities in the speech stream with greater facility when there is musical or intonational information present in the input.

On the other hand, the second study hypothesis was not supported by the behavioral responses of the participants (ACC and RT). Even though there were differences between groups, showing a tendency for the monotone group to be faster and the tone trained group more accurate, these differences were not significant. Therefore, the present study does not support the conclusion that the facilitation of word boundary identification by the addition of musical tones
as seen in the brain data will extend to the behavioral domain, allowing for the more accurate and faster identification of words vs pseudowords (part-words) in a novel L2-like speech stream. This suggests one of two things: (1) we might need more participants to obtain a significant difference; or (2) these processes really are separate. In either case, more neurophysiological studies are needed to observe the initial stages of language acquisition.
7 REFERENCES


85


APPENDICES

8.1. Teachers College, Columbia University IRB Approval

To: Dayna Moya Sepulveda  
From: Curt Naser, TC IRB Administrator  
Subject: IRB Modification Approval: 16-131 Protocol  
Date: 07/27/2017

Please be informed that as of the date of this letter, the Institutional Review Board for the Protection of Human Subjects at Teachers College, Columbia University has approved a modification to your study, entitled "N400 evidence for musical facilitation of word boundary identification in second language exposure" on 07/27/2017.

The approval remains effective until 10/10/2017.

The IRB Committee must be contacted if there are any changes to the protocol during this period. Please note: If you are planning to continue your study, a Continuing Review report must be submitted to either close the protocol or request permission to continue for another year. Please submit your report by 09/12/2017 so that the IRB has time to review and approve your report if you wish to continue your study. The IRB number assigned to your protocol is 16-131. Feel free to contact the IRB Office (212-678-4105 or IRB@tc.edu) if you have any questions.

Please note that your Consent form bears an official IRB authorization stamp. Copies of this form with the IRB stamp must be used for your research work. Further, all research recruitment materials must include the study's IRB-approved protocol number. You can retrieve a PDF copy of this approval letter from the Mentor site.

Best wishes for your research work.

Sincerely,

Curt Naser, Ph.D.
TC IRB Administrator
8.2. Background Questionnaire

Confidential

Background Questionnaire

Please complete the survey below.

Thank you!

Name

Date

Age

Gender

☐ Male
☐ Female
☐ Prefer not to disclose

What is your dominant hand?

☐ Right
☐ Left
☐ Both

(Which hand do you write with)

Years of education after high school:

Are you monolingual?

☐ Yes
☐ No

Please list the other language(s) you speak fluently or in which you have some proficiency:

(Specify the degree of proficiency (basic, conversational, fluent, native-like) and years of practice)

Do you play a musical instrument or sing?

☐ Yes
☐ No

What instrument do you play?

Specify how much you practice

☐ I have played/sing for fun once in a while for more than 5 years
☐ I have been playing/singing for fun multiple times a week for more than 5 years
☐ I have been playing/singing professionally for more than 5 years

What age you started practicing?

How many years of formal training have you had?