ABSTRACT

This study tested the hypothesis that implicit meter can serve as a way to measure oral reading fluency which in turn could predict reading comprehension. By investigating the neural mechanisms that inform implicit meter, it should be possible to measure oral reading fluency. Therefore, by studying the neural and behavioral underpinnings of implicit meter in children during development (6–10-year-olds), it should be possible to map specific neural mechanisms to reading comprehension. Additionally, this study investigated whether one of the specific mechanisms, the LMN component, had a direct correlation on childrens' reading comprehension KTEA scores. This study found that adolescents are more so successful in sustaining a quadruple implicit meter than a triple meter and are more successful sustaining a slow tempo meter than a fast tempo meter. This study also found that the N1 component matures over the course of adolescence, with it being unobservable in the youngest cohort and fully emergent by the oldest cohort. The LMN component was found to be observable throughout all cohorts without showing visual signs of maturation. Additionally, this study found that the LMN component does not exhibit direct correlation to reading comprehension. Therefore, the LMN component is not a neural component that can be used to measure oral fluency, and therefore is not suitable to measure reading comprehension. This deeper understanding of the neural underpinnings of reading comprehension can then inform future educational policies that aim to promote childhood reading comprehension and provide clear targets that policies can focus on when designing interventions for the vast population of children who are struggling to meet the threshold to be considering as being able to read to their grade level.

Neural Representation of Implicit Meter

in Typical Early Readers

By

Madeline Marie Frances Rose

A Thesis

Presented to the Faculty of Mount Holyoke College

In partial fulfillment of the requirements

For the degree of Bachelor of Arts

With Honors

Department of Neuroscience and Behavior

Mount Holyoke College

South Hadley, Massachusetts

ACKNOWLEDGMENTS

First and foremost, I thank my thesis advisor, Mara Breen, for her continued enthusiasm, dedication, and overwhelming support to seeing this project to completion. Thank you so much for your guidance and patience throughout the course of this project and for allowing me the opportunity to pursue such a rewarding experience.

I also thank the members of my thesis committee, Marta Sabariego, and Laura Sizer, for their helpful comments and revision suggestions. I thank Katerina Drakoulaki for the various studies she shared with me, and the other CAPSlab members for helping collect and process the data.

Thank you to Professor Janelle Gagnon and my Psych-399 classmates for continuously giving me feedback on my thesis, and for helping me see my research from an outsider's perspective. I thank Mr. and Mrs. Harap for generously providing financial support for my research.

I would not have been able to complete this thesis without the support of my family and friends. To my friends, thank you for your constant encouragement, listening to my rants, and always grabbing me bougie water. To my family, thank you for supporting me in all of my endeavors and whims and for reminding me that a world exists outside of the madness. Everything here is the best thing ever, nada aqui es casi el cielo.

TABLE OF CONTENTS

Introduction .... ***Error! Bookmark not defined.***
Reading Comprehension as a Socioeconomic Predictor 2
Reading Curriculum and the Science Behind Reading 3
Reading Comprehension and Prosody 4
Prosody and Reading Comprehension 5
Inner Speech 6
Event-related Potentials 9
Investigating Implicit Meter with ERP 11
Definitions of Relevant Neural Representations 11
Intent of Study 12
Methods *13*
Participants 13
Stimuli 14
Procedure 14
Day 1 15
Day 2 18
EEG Recording and Processing 19
Behavioral Data Analysis 19
EEG Data Analysis 20
Combined Behavioral 21
Results *23*
Behavioral 23
ERPS 25
Cohort 1 25
Cohort 2 27
Cohort 3 30
Cohort 4 32
Combined Behavioral and ERP Effects 35

Discussion.......... *39*
Grand Average ERPs.......... 41
Combined Behavioral and ERP Analysis.......... 42
Limitations & Future Directions.......... 43
Conclusion.......... 44
References.......... *1*

LIST OF TABLES

Table 1. Fixed effects in the logistic regression model predicting accuracy from Meter, Tempo, and Cohort 24
Table 2. Mixed Repeated Measure ANOVA for Cohort 1 Quadruple Fast versus Slow Trials 27
Table 3. Mixed Repeated Measure ANOVA for Cohort 1 Triple Fast versus Slow Trials 27
Table 4. Mixed Repeated Measure ANOVA for Cohort 2 Quadruple Fast versus Slow Trials 29
Table 5. Mixed Repeated Measure ANOVA for Cohort 2 Triple Fast versus Slow Trials 30
Table 6. Mixed Repeated Measure ANOVA for Cohort 3 Quadruple Fast versus Slow Trials 32
Table 7. Mixed Repeated Measure ANOVA for Cohort 3 Triple Fast versus Slow Trials 32
Table 8. Mixed Repeated Measure ANOVA for Cohort 4 Quadruple Fast versus Slow Trials 34
Table 9. Mixed Repeated Measure ANOVA for Cohort 4 Triple Fast versus Slow Trials 35

LIST OF FIGURES

Figure 1. A map of various pathways neural substrates support that facilitate reading comprehension 5
Figure 2. Pie chart illustrating the diversity of ethnicity within the participants collected. 13
Figure 3. Presentation times in milliseconds of each event in an experimental trial. The word 'four' indicates the intended meter of the trial. The trial consisted of Entrainment, Listening, and Response portions. See text above for details. 17
Figure 4. Picture of an example KTEA question during the beginning part of the assessment. The participant in this instance is asked to match a one-word reading with the corresponding image 18
Figure 5. Map of the positions of a 64 electrode system including the number and designation. The schemata are based on the internationally established 10-20 system 20
Figure 6. Trial accuracy for all participants aggregated by tempo, meter, and cohort. Error bars indicate standard error. An accurate trial is one where all three participant responses were aligned with Beat 1 24
Figure 7. Grand Average ERP of Cohort 1 for all combinations of Tempo, Meter, and Cohort. A represents the average Quadruple Fast trial, B represents the averaged Triple Fast, C represents the averaged Quadruple Slow trial, and D represents Ternary Slow trial. 26
Figure 8. Grand Average ERP of Cohort 2 for all combinations of Tempo, Meter, and Cohort. A represents the average Quadruple Fast trial, B represents the averaged Triple Fast, C represents the averaged Quadruple Slow trial, and D represents Ternary Slow trial. T 29
Figure 9. Grand Average ERP of Cohort 3 for all combinations of Tempo, Meter, and Cohort. A represents the average Quadruple Fast trial, B represents the averaged Triple Fast, C represents the averaged Quadruple Slow trial, and D represents Ternary Slow trial. 31
Figure 10. Grand Average ERP of Cohort 1 for all combinations of Tempo, Meter, and Cohort. A represents the average Quadruple Fast trial, B represents the averaged Triple Fast, C represents the averaged Quadruple Slow trial, and D represents Triple Slow trial. 33
Figure 11. Measured amplitude (mV) in all, fast, and slow triple trials, separated by beat, from Cohort 4 36
Figure 12. Measured amplitude (mV) in all, fast, and slow quadruple trials, separated by beat, from Cohort 4 37
Figure 13. Beat 1 minus Beat 2 as in the average LMN associated with beat 1 stimuli minus the average LMN associated with beat 2 stimuli therefore showing 'Individual metric negativity'. Individual metric negativity was then correlated against against participant's KTEA standard score. 38

INTRODUCTION

In 2022, the National Assessment of Educational Progress (NAEP) administered a reading assessment to representative samples of fourth and eighth grade students throughout the United States. These assessments were administered throughout public, private, Bureau of Indian Education, and Department of Defense schools in order to acquire a more complete reflection of students throughout the country, with a total of 108,200 fourth grade students completing the assessment. The NAEP found that the average reading score for fourth grade students was 217 out of 500 points (*NAEP Reading: Reading Highlights 2022*, n.d.). The score needed to be considered proficient in grade-based reading is 238 points (*NAEP Nations Report Card - The NAEP Reading Achievement Levels by Grade*, n.d.). This means that 67% of fourth grade students score below NAEP proficiency; in other words, below grade level. NAEP's 2019 report found that the average reading score was 220 and that 65% of students score below proficiency. This shows that since 2019 there has been a significant decrease in average reading comprehension score and an increase in percentage of fourth graders who do not meet the criteria to be considered proficient readers. The last time the average score in reading comprehension was 217 points was in 1992, almost thirty years ago (*NAEP Reading: Reading Highlights 2022*, n.d.). It is also equally important to note that the highest score obtained by 4th graders within the time that the NAEP has been administered was in 2013, where they scored an average of 222 points (*NAEP Reading: Reading Highlights 2022*, n.d.). This shows that although there was an increase from the 1990’s to the mid 2010’s, the average score only increased by 5 points. Evidently, there are still significant improvements to be made to our reading curriculum. Overall, these statistics illustrate a troubling trend of not only how the covid-19 pandemic has affected early childhood education, but also perhaps more important how ineffective the past thirty years of educational reforms have been, given that there has not been a single year where the average score is at least 238 points. This means that, according to the NAEP, despite the past thirty years of educational reforms, there has not been a single year where children have been giving enough support to where at least half of the population reaches reading proficiency by fourth grade standards.

Reading Comprehension as a Socioeconomic Predictor

Reading comprehension is considered to be one of the most significant development al achievements that relate to an individual's growth and development (Cheng & Wu, 2017; Ferrer et al., 2007). Various studies have demonstrated that socioeconomic status is a strong predictor of adolescent reading development (Bradley & Corwyn, 2001; Kieffer, 2012; Romeo et al., 2022). For instance, it has been found that even after children enter school, children from lower socioeconomic status background experience a decrease in reading development during summer break due to the decreased access to reading resources and materials (Alexander et al., 2007). The same effect has also been observed during the pandemic (Domingue et al., 2022; Kuhfeld et al., 2023). In addition, not only is socioeconomic status considered a strong predictor of reading development, but reading proficiency is a strong predictor of future socioeconomic status. Various studies have shown that reaching proficiency by third grade is a reliable indicator of the likelihood of graduating high school on time (Hernandez, 2011; Lloyd, 1978), but it is also linked to ongoing difficulties in schools and likelihood of becoming economically successful later in life (Cappella & Weinstein, 2001; Foundation, 2010). A 2013 report by Annie E. Casey foundation, a prominent charity organization focused on improving child welfare, found that not only does that gap between struggling and proficient increase over time, but there is also significant interactions between poverty, place, minority populations, reading proficiency, and academic achievement (Foundation, 2010). This shows that it is crucial to work on improving early childhood reading skills, especially in underserved communities, in order to help close the achievement gap between populations and set children up for socioeconomic success during their foundational development years.

Reading Curriculum and the Science Behind Reading

The question therefore becomes how to best develop an effective reading curriculum in order to increase the amount of fourth graders who can, at the very least, be considered proficient readers. Curriculum for elementary schools seems to be in constant flux and notably, science backed curricula has only recently begun being mandated (Schwartz, 2022). However, these science backed curricula have only been found effective in varying degrees (Allor et al., 2014; Cummins, 2007; Solari et al., 2020).

How to best teach reading to children has a surprisingly fraught historical past. Since the 1960's there has been an ongoing debate over how to best implement early reading instruction in schools (Kim, 2008). In the 2000's, the federal government stepped in and hired a panel of experts in psychology and education to conduct multiple longitudinal studies. The meta-analysis showed that "instruction in phonemic awareness, phonics, and guided oral reading fluency improved children's ability to read words, to read connected text with speed and accuracy, and to comprehend text (Kim, 2008. p.101). Therefore, the federal government encouraged "practitioners that instruction should 'integrate attention to the alphabetic principle with attention to the construction of meaning and opportunities to develop fluency'" (Kim, 2008, p.101). However, given the marginal improvements made over the past decades, as can be seen in the NAEP reports, it is clear that this suggested curriculum was not entirely effective.

In the decades since there has been research demonstrating that the majority of reading curriculums that are actively implemented in schools are not scientifically based nor valid (Suárez et al., 2018). Since around 2020, there have been numerous calls for the standardized implementation of science-based reading curricula. However, in a meta-study conducted by Petscher et al., it was found that there was a lack of compelling evidence regarding the benefits to the science of reading as used in classroom curricula (Petscher et al., 2020). However, the same study also noted that an increase in the body of knowledge from contributions from fields such as linguistics, neuroscience, and computational science will inform and shape the science of reading and therefore the curricula for how reading should be effectively taught. Therefore, the primary aim of this study is to investigate reading comprehension from a neurocognitive perspective.

Reading Comprehension and Prosody

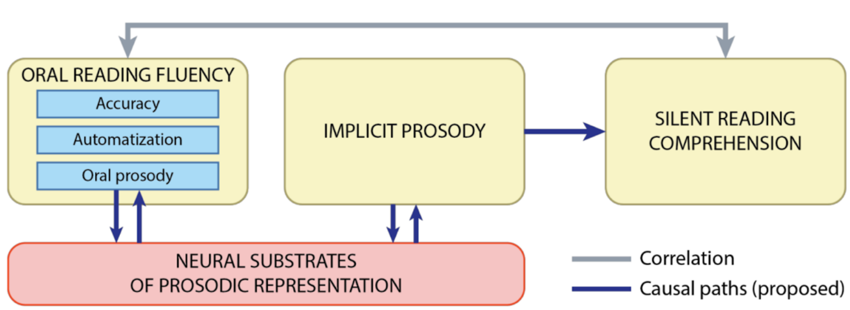
The goal for this study is to identify a neurocognitive correlates of effective reading comprehension, with the ultimate goal of improving reading acquisition through targeted training of these correlates. In order to discover this correlate candidate, this study is focused on discerning the neural processes that best predict reading comprehension. Currently, several predictors of reading comprehensions have been identified. Notably, studies have shown a positive association between children who exhibit more adult-like prosody, encompassing elements such as phrasing, stress, and their reading comprehension skills (Breen et al., 2016; Miller & Schwanenflugel, 2006, 2008). Additionally, it has been shown that facility with rhythm and meter can be linked to linguistic and literacy skills in children (Huss et al., 2011; Ozernov-Palchik et al., 2018; Tierney et al., 2021; Woodruff Carr et al., 2014). Furthermore, disruptions in stress perception have been observed in children diagnosed with developmental learning disorders such as developmental dyslexia, underscoring the importance of prosodic processing in reading comprehension (Goswami et al., 2013). Building upon these findings, recent work by Breen, van Dyke, Krivokapić and Landi (2024) has highlighted the significance of prosodic features in production as reflective of reading comprehension skill in high school students. By examining these predictors and their neural underpinnings, we aim to uncover a robust candidate for a neurocognitive correlate of reading comprehension, which could then inform future interventions aimed at enhancing literacy skills in children and adolescents.

Prosody and Reading Comprehension

Given that there is no direct way to measure reading comprehension through neurocognitive means, it is important to trace the different ways in which these specific neurocognitive mechanisms quantitatively give rise to various factors that inform reading comprehension.

Figure 1.

A map of various pathways neural substrates support that facilitate reading comprehension



Oral reading fluency, the ability read quickly, accurately, and with a natural intonation, has been demonstrated to predict reading comprehension (Cutting et al., 2009), as seen by the correlative path displayed in the figure. As seen in Figure 1, oral reading fluency directly influences silent reading comprehension, and is indirectly influenced by neural substrates of prosodic representation. There are three main components of oral reading fluency: accuracy, automatization, and most importantly, oral prosody. Oral reading prosody is when text is read with “appropriate phrasing, pause structures, stress, rise and fall patterns, and general expressiveness” (p.119, Schwanenflugel et al., 2004). Therefore, by focusing on oral reading prosody we gain insight into oral reading fluency which, following the correlative arrows in the figure, then gives us insight into silent reading comprehension.

A component of oral reading prosody is meter, which facilitates the organization of stress when someone is speaking. Sensitivity to meter also serves as a predictor of reading skill during childhood (Wade-Woolley, 2016). There are various studies which investigate the role of the implicit meter in adulthood and its influence on reading (Breen et al., 2019). However, there is little known about the impact and development of implicit meter in childhood.

As pathways have been shown connecting neural substrates of prosodic presentation to meter, to oral prosody, to oral reading fluency, and ultimately to silent reading comprehension, it should therefore be possible to investigate specific neural substrates and determine if they are able to directly influence reading comprehension.

Inner Speech

These proposed pathways assume that oral reading fluency and implicit prosody can both be studied in the same way, given that not only is there external prosody when a person speaks out loud, there is also implicit prosody, where a person only reads inside their head. This is often referred to as someone’s ‘inner voice’ or ‘inner speech’ (Vissers et al., 2020). Inner speech has been suggested to have an important role in the self-management of both cognition and behavior and therefore contains implications for psychiatric conditions in which there is inner speech dysfunction, and developmental disorders that exhibit atypical language skills or decreased self-regulation ability (Alderson-Day & Fernyhough, 2015). This assumption, that inner speech can be directly measured as a correlate to external speech, or rather that implicit prosody can directly map onto external prosody, as both are components of their respective speech type, relies on the assumption that our inner speech consists of natural language sentences. This means that our proposed pathways contain an underlying assumption that when we use internal speech, for introspective or reading purposes, the mental representations that arise in these cognitive processes constitute a natural language. According to the cognitive philosopher Edouard Machery, when "you, English speaker, entertain a thought that *p*, you token a sentence of a language that does not possess the characteristics properties of natural languages" (p.470, Machery, 2005). For example, when a person thinks "I love San Diego" in English, that thought sentence can be represented by token *p*, however if a person were to say "I love San Diego" that spoken sentence cannot be represented by token *p* and instead will be represented by token *q*, as the thought sentence does not possess the same characteristics as spoken English, and therefore the thought sentence and the spoken sentence cannot both be represented by token *p.* If this hypothesis were true, then that means that it would be impossible to examine the prosody of inner speech with the aim of discovering something true about external speech, or anything else related to the natural language in which we read and talk, as the internal language of inner speech and external language of our world are in fact two different languages.

However, Peter Carruthers, a cognitive scientist and philosopher, instead proposes a single-state mixed contents view, and proposes that inner speech involves the generation of a presentation of word sounds which, in a process similar to what occurs when we perceive external speech, is then interpreted by auditory comprehension mechanisms so that semantic content can then be assigned to the mentally representing speech sound (Rescorla, 2023). This will be the framework that this study operates from, as Carruther's model emphasized the importance of the neural processes which are used to interpret mental sentences, regardless of whether they are in a natural or non-natural language. This means, that although the thought sentence *p* and speech sentence *q* are distinct from one another due to differing characteristics, we can use *p* to learn something about *q*, and vice versa, by looking at the mechanisms we used to understand *p* or *q*. Therefore, we are able to translate the neural processes of inner speech onto the neural processes of comprehension of external speech. In addition, this particular view is supported by Fodor's Implicit Prosody hypothesis, which states that during silent reading, the reader is able to generate a representation of sentence intonation, phrasing, stress and rhythm, and therefore that generated representation is able to influence the reader's interpretation of that text (Bishop, 2021; Fodor, 2002). This therefore means that for our study, which partially rests on the assumption that a strong oral reading fluency implies strong implicit prosody, our methodology will be valid given that we are examining how children process components of prosody.

Under models of intonational phonology, where sound features of speech are conceptualized as categories of acoustic variation (Ladd, 2008), implicit prosody consists of 4 components: intonational contour, phrasing, meter, and stress. Intonational contour is the pattern of significant pitch patterns that are distributed over vocalized utterances (Kapatsinski et al., 2017). Phrasing is defined as the "domain of a perceptually coherent intonational contour"

(Skarnitzl & Hledíková, 2022). Stress is the relative emphasis a speaker may give to certain syllables or words. And finally, meter facilitates the organization of said stress (Breen, 2014). Sensitivity to meter also serves as a predictor towards reading skill during childhood (Wade-Woolley, 2016). There are various studies which investigate the role of implicit meter in adulthood and its influence on reading (Breen et al., 2016; Huss et al., 2011; McCurdy et al., 2013). However, there is little known about the impact and development of implicit meter in childhood. Currently, the existing body of research shows that infants demonstrate metric perception, that this neural response differs from that of an adults, and that there is behavioral evidence of metric stress perception that emerges between preschool and adolescence(Cirelli et al., 2016; Fitzroy & Sanders, 2020; Nave-Blodgett et al., 2021). Given this, by investigating implicit meter, we can gain insight into prosody which mediates the relationship between oral fluency and reading comprehension, which directly impacts the skill level of a person's reading comprehension. This proposed pathway can also be observed in Figure 1.

Event-related Potentials

For this study we will be using Event Resting Potential to investigate implicit meter. ERPs are a subset of a wider research technique known as electroencephalography (EEG). This method was invented in 1929 by German psychiatrist Hans Berger. Interestingly, Berger invented EEG during his attempt to find the scientific basis for telepathy. By using EEG, a scientist is able to read the brain's electrical activity in real time by observing the electrical waves on a human scalp. EEG is able to capture the electrical changes on a human scalp because of the way the brain's neurons communicate between them. These neurons are excitable cells with intrinsic electrical properties which therefore leads to magnetic and electrical fields (Beres, 2017). This electrical activity is a result from two types of neuronal activity: action potentials and postsynaptic action potentials. Action potentials result from when a neuron rapidly depolarizes. Post-synaptic potentials are instead mediated by a variety of neurotransmitter systems and therefore generally results in slower changes to the membrane potentials (Beres, 2017). EEG measures the variation over time of electrical activity of a specific type of neurons — pyramidal cells — through electrodes on the scalp (Beres, 2017). The aggregation of these signals creates an alternating current that can be amplified and recorded, resulting in scientists being able to observe both positive and negative current due to factors that may be happening within the brain. Although there have been more recent scientific innovations that also allow scientists to examine the brain in great depth, EEG is still a popular research technique due to being able to measure the brain's electrical activity in real time, despite having poor spatial resolution which limits the ability to recognize which specific areas of the brain are active.

Event-related potentials (ERPs) are time-locked to a stimulus and are the averages of the continuous electroencephalogram. Early ERPs are referred to as exogenous as they are directly related to the stimulus, such as an auditory sound. Examples of early ERP’s would be the N1 component, which is observed when a stimulus is presented, and the P2, which is observed when there is a sensation-seeking response (Sur & Sinha, 2009). ERP’s that occur later are known as endogenous, as they’re internally generated and more so reflect when the subject evaluates the stimulus (Sur & Sinha, 2009). Examples of these endogenous ERPs would be the P3, which demonstrates the speed of stimulus classification, and the N400 which is linked to meaning processing (Kutas & Federmeier, 2011).

The advantage of using ERPs as a research tool is that they are implicit- they don’t require an overt response from the participant, and they can elucidate unconscious cognitive processing. This last point about ERPs is especially relevant for the current study, as we will be using ERPs to measure the neural substrate of implicit metric processing.

Investigating Implicit Meter with ERP

A recent study by Ahren Fitzroy and Lisa Sanders, on which the current study is based , investigated the neural indices of implicit metric processing (Fitzroy & Sanders, 2020). This study led to them studying how metric structure directs attention across time. Fitzroy & Sanders used Event-Related Potentials (ERPs) to assess how the brain functions in response to one of the senses being stimulated, which in this study was sound. They implemented this by having adult participants listen to an isochronous stream of beats, that did not vary in terms of pitch nor intensity, and then impose a metrical structure onto them based on if they were told to group the beats in sections of three or four. The authors found that in adults, there are four specific neuro-cognitive mechanisms that aid in processing subjecting metric strength: N1, LMN, CNV, and P300. However, there is not yet any research investigating these mechanisms relative to subjective metric strength during childhood development.

Given the prior existing body of knowledge, it is clear that reading comprehension can be predicted by oral reading fluency. Meter, a specific component of prosody, has not yet been studied in children although it has previously been studied in adults. Therefore, this study aims to investigate implicit prosody in children by characterizing the neural representation of implicit meter in typical early readers and how they mature over childhood development.

Definitions of Relevant Neural Representations

Given that the N1, LMN, CNV, and P300 have previously been shown to process metric strength, this study will study two of these four components using EEG methods.

Auditory N1 is the primary component of interest in this study, as it is part of auditory evoked responses in the brain and is a negative going potential that, within adults, peaks between 80 and 120 ms after the onset of an auditory stimulus and is mostly distributed over the front-central regions of the brain. This means that the N1 will occur when a person is specifically allocating attention to a strong tone. A 2016 longitudinal study that investigated the neural correlates of accelerated auditory processing in children showed that the N1 increases in amplitude until young adulthood (Habibi et al., 2016). Over the years, auditory stimuli have been used as suitable research stimuli in order to investigate neural activities.

The Late Metric Negativity (LMN) has been shown to be larger over the left hemisphere of the brain (Potter et al., 2009) and is evoked when there are metric accents, either perceptible or imagined. This means that the LMN occurs when a person anticipates an emphasized sound or is preparing to process the tone, even when they are only mentally assigning emphasis on sounds. According to Fitzroy & Sanders (2020), the LMN was exhibited from between 250 to 450 ms post stimuli, however that study only examined the LMN components in adults.

Intent of Study

Given that the N1 and LMN are both neural components that are influenced by metric structure, which is itself a component of oral prosody, this study investigated if the N1 and LMN components would be suitable candidates to be studied as strong neurocognitive correlates of reading comprehension. In order to investigate this we looked at three different things, the N1 and the LMN components of cohort's average neural signatures, and if a participant's LMN component can directly predict their level of reading comprehension. To test this, we had the participants also take a KTEA assessment which evaluates essential reading skills from children to young adults. We hypothesized that as the participant's age increases, so would their accuracy in completing the task. In addition, we hypothesized that fast tempo trials would have greater accuracy than slow tempo trials. We hypothesized that the subjectively stronger tone would elicit a stronger N1 component regardless of it being a slow or fast trial and would also grow stronger as age increases. We also hypothesized that the subjectively stronger tone will elicit an LMN with the LMN being larger for fast trials compared to slow trials.

Methods

Participants

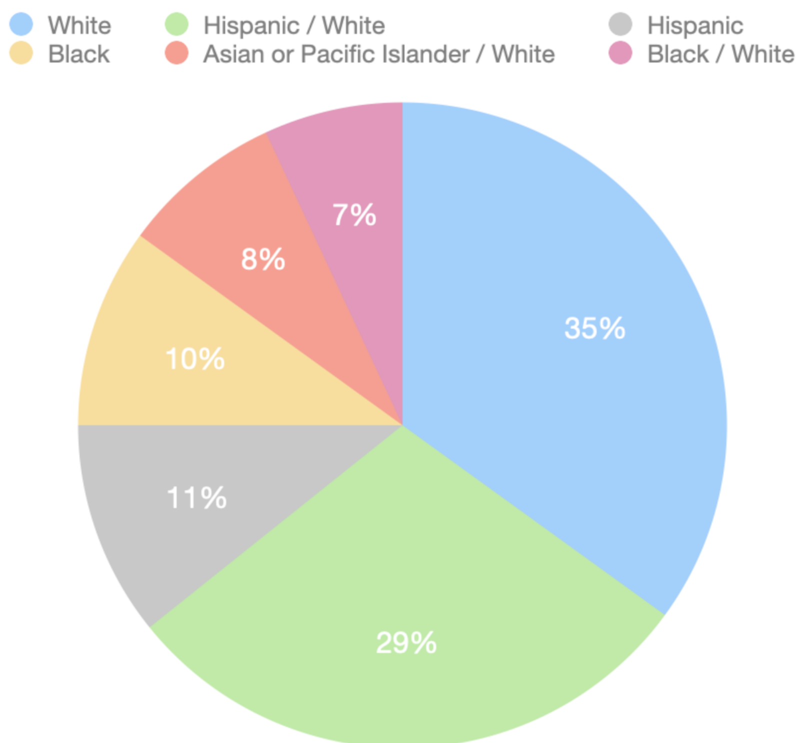
This study recruited children from the ages of 6-10 and grouped participants into four different age groups. There are four age cohorts: 6-7.25 years old, 7.25-8.5 years old, 8.5-9.75 years old, 9.75-11 years old. There were 10 participants in Cohort 1, 12 participants in Cohort 2, 8 participants in Cohort 3, and 9 participants in Cohort 4.

We attempted to recruit evenly from the four age groups. Participants were primarily recruited by tabling at the Springfield Museums. They were additionally recruited from elementary schools in the Springfield, MA metropolitan area. This was accomplished by using the connections CAPSlab has with Springfield museums and by sending flyers home with each child from Sumner Avenue Elementary School and White Elementary School. Springfield Museums is heavily involved with their local elementary schools and after-school programs and allowed CAPSlab to recruit on their premises. Additionally, CAPSlab has a website that advertises the study and the lab created posts on social media such as reddit. As the overarching study takes place over two days, the participant was compensated $20 for the first day, and $30 for the second, adding up to a total of $50.

Participants reported a variety of ethnic backgrounds. The distribution of ethnicities is presented in Figure 2.

Figure 2.

Pie chart illustrating the diversity of ethnicity within the participants collected.



Stimuli

The stimuli consisted of isochronous beats of a 50 ms, 1000 Hz pure tone with 5 ms onset and offset ramps. This tone was stored as a signal-channel 16-bit PCM format WAV audio file. On half of the trials, the beats were presented at a fast tempo (450 inter-onset interval); on half, they were presented at a slow tempo (625 inter-onset interval).

Procedure

The data were collected at Dr. Seuss' childhood home in the Forest Park neighborhood of Springfield, MA. This location is situated close to public transportation and is within walking distance of two elementary schools. Data collection days were scheduled on Tuesdays, Thursdays, and Saturdays with the intention of minimizing disruption to participants' school and extracurricular activities. Due to the high amount (45%) of linguistic diversity in the surrounding metropolitan area, it is imperative that the data reflects this diversity. By making the lab location and time easily accessible, we hoped to lower the barrier of accessibility so that underrepresented populations were well reflected within the dataset. All research procedures were approved by the Mount Holyoke College Institutional Review Board. The room in which the data collection was collected is attenuated so that there is no external and extraneous sound interference.

Children participants were prescreened to ensure they met all eligibility requirements: The child has normal vision, with or without glasses, the child has normal hearing, the child has not been diagnosed with any learning disabilities, developmental disorders, or neurological issues, the child is not on any psychoactive medications. Guardian consent and child assent were obtained, and the parent had the option to wait in an adjacent room while their child completed the experiment or quietly sit in the room as the child completed the experiment. The adult completed questionnaires regarding the child's demographics, developmental history, language, musical experience. On Day 1 the participant would undergo the EEG data collection, and on Day 2 the participant would complete the KTEA assessment.

Day 1

The participant would ideally complete 40 trials, each with a duration of 1 minute. The 40 trials consisted of four subtypes, with ten trials for each combination of listening pattern (triple, quadruple) and tempo (450 ms, 625 ms inter-onset intervals). On each trial, the participant listened to an isochronous pattern of undifferentiated tones and was instructed to interpret this pattern as groups of three or four. The child participant was instructed to ‘feel’ the beat of the tones without counting, moving, or attempting to maintain the beat physically. An experimenter sat beside the child to communicate to them which beat group they were supposed to be listening for, and to help them understand how to count a triple beat vs quadruple. They gave additional instructions at the start of the experiment to explain to the participant what is expected of them, and to monitor the participant during the experiment. In addition, a secondary experimenter sat in the back of the room observing and taking notes on the participants performance and noting any issues with the EEG data. As an aid to increase task engagement, the child participant picked 8 prizes before beginning the experiment and was able to select 1 of the prizes for their goody bag during each of the breaks that occurred during the study. Breaks occurred after six sets: approximately every 15 minutes.

Each trial consisted of three stages: Entrainment, Listening and Response. These are depicted in Figure 3.

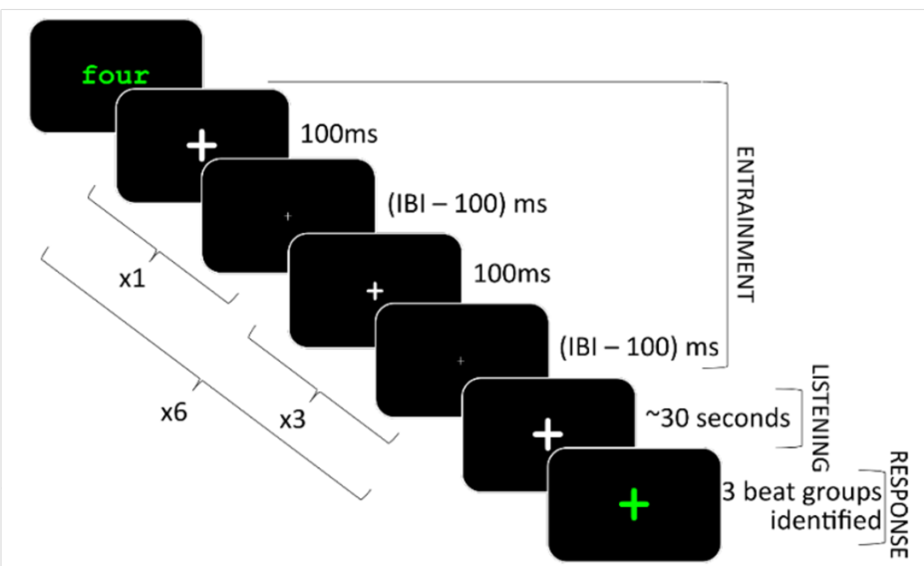
1. Entrainment: The purpose of this stage was to train the participant to recognize the intended metric structure of the trial. On the first beat of every three- or four-beat group, the fixation cross increased in size temporarily to indicate the first tone for the intended listening group for that trial (triple or quadruple). After the seventh group, the size of the cross would stay increased permanently, therefore indicating to the participant that the Listening portion was about to begin.
2. Listening: The child would then attempt to internally maintain the entrained metric pattern for 30 seconds while the tones continued, and the fixation cross stayed unchanged. On one of the five groupings that followed the beginning of the listening portion, which would be randomly chosen for each trial, the fixation marker would then permanently turn green, indicating to the participant that the response portion was about to begin. Additionally, there was a random number of ‘Filler’ events between the

Listening and Response sections of the experiment with the intention of preventing the participants from using the cross changing to green as an indicator for the participant to count as beat 1. Only data from the Listening stage were analyzed.

3. Response: During the Response stage, the child participant counted the beats out loud (either "1, 2, 3" or "1, 2, 3, 4") to indicate their interpretation of the listening pattern; the experimenter pushed a button on a serial response box whenever the participant counted '1' out loud. Each trial concluded after three button presses.

Figure 3.

Presentation times in milliseconds of each event in an experimental trial. The word 'four' indicates the intended meter of the trial. The trial consisted of Entrainment, Listening, and Response portions. See text above for details.

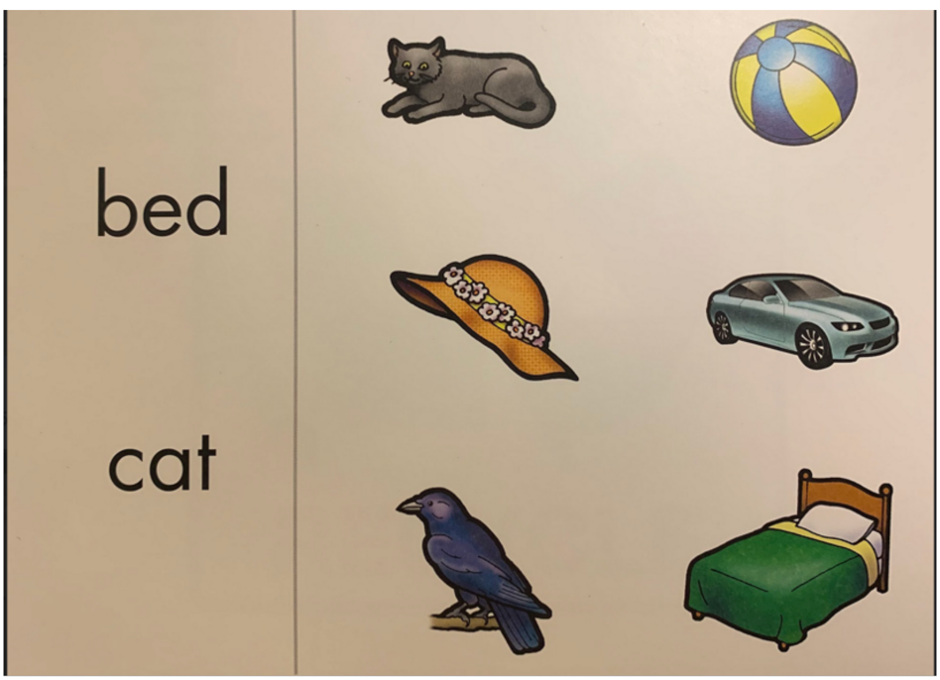


Day 2

On Day Two, each participant completed the KTEA assessment (KTEA-3). The participant was presented with a list of standardized questions consisting of reading scenarios and respective reading comprehension questions. They then received a standardized score that reflected their performance and their age. In all the process took between 10 and 20 minutes. See figure 4 below for an example KTEA task.

Figure 4.

Picture of an example KTEA question during the beginning part of the assessment. The participant in this instance is asked to match a one-word reading with the corresponding image



EEG Recording and Processing

Reference-free electroencephalography (EEG) data were collected using a high density 64-channel electrode cap connected to a BioSemi Active-Two system. Data were collected at a sampling rate of 2048 Hz with a bandpass of 0.1–200 Hz. EEG processing and quantification were completed using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) within MATLAB (The Mathworks, Inc.). The collected EEG data was down sampled to 250 Hz. A high pass filter was then used at 1 Hz. Algorithms were run to clean the data of physical or electrical artifacts. Afterwards, the data were decomposed through independent components analysis (ICA) with a matrix corresponding to how many channels were in the dataset. This process identifies separate sources of electrical activity in the overall EEG signal, allowing for the removal of consistent sources of noise, like eye blinks, muscle tension, and other electrical interference. Components were then rejected or accepted based on visual inspection - components corresponding to neural activity were retained while those corresponding to noise were removed from the data. In the final cleaning step, the EEG from individual electrodes which had been removed due to poor signal was recreated through mathematical interpolation (i.e., by taking the spherically corrected average of adjacent electrodes) so that each dataset had the same number of electrodes. Individual subjects' averages were created for each electrode for each condition of the experiment. Finally, Grand Average ERPs were created for each cohort by averaging the individual subject averages within each cohort.

Behavioral Data Analysis

An accurate trial was one in which the participant’s response coincided with the strong beat three times (i.e., the experimenter made three accurate keypresses on the trial). A mixed-effects logistic regression model was fitted with a binomial family and a logit link function in order to reflect the binary nature of the outcome variable, Accuracy. The model formula is as follows:

$$Accuracy \sim Meter + Tempo + Cohort + Meter{:}Tempo{:}Cohort + (1 \mid participant)$$

This formula indicates that accuracy is predicted by main effects of Meter, Tempo, and Cohort, as well as the three-way interaction between Meter and Tempo and Cohort, and accounts for random intercepts for each person in order to accommodate participant specific differences. This model was used on a set (N = 39) of participants who completed at least twenty out of the forty total possible EEG recording trials.

EEG Data Analysis

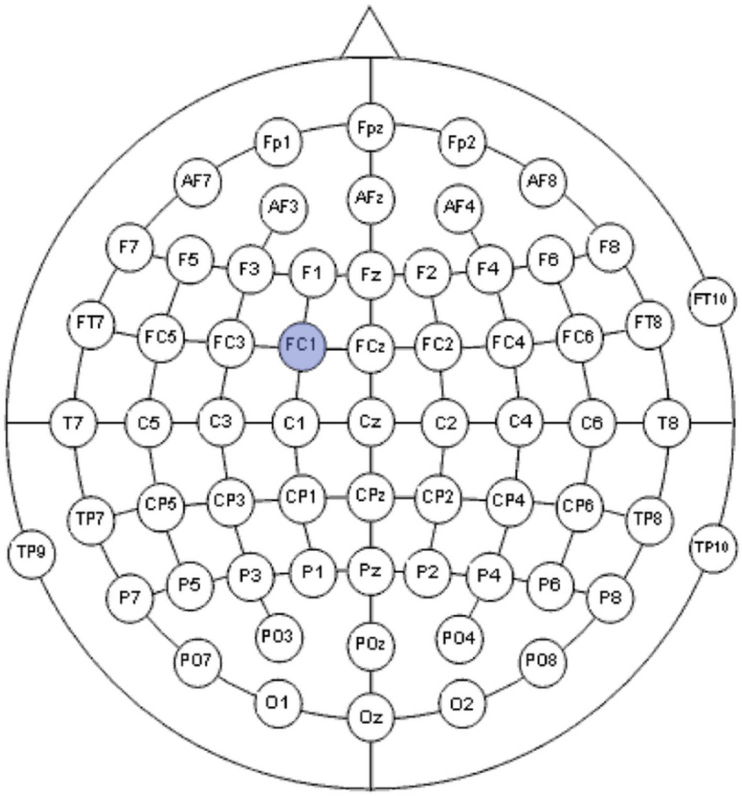
Commented [MR1]: Why such big gap above

For each Cohort, a mixed ANOVA was conducted with two within-subject fixed effects (Beat, Tempo), and a between-subjects (Cohort). For this study, Beat refers to the specific beat within a group (i.e. Beat 1, Beat 2 …) and Tempo refers to if the grouping was presented at a fast or slow speed. An ANOVA was performed on Quadruple Meter data and Triple meter conditions separately. Quadruple meter refers to a grouping of four beets, and Triple meter refers to a grouping of three beats. The effects were considered statistically significant at the 0.05 significance level.

Based on prior investigations of the time course and topography of the N1 and LMN effects, ERP measures in the current study were taken from electrode FC1, which is in the frontal-central area of the scalp and slightly leftward (Figure 5). Figure 5 is based on another figure found in a different EEG study (Demin et al., 2016).

Figure 5.

Map of the positions of a 64-electrode system including the number and designation. The schemata are based on the internationally established 10-20 system



The N1 was determined impressionistically from the waveforms. The LMN effect was operationalized as the mean amplitude of the waveform between 100 to 300 ms after the onset of the sound, averaged across each beat separately for both fast and slow conditions, and triple and quadruple meter.

Combined Behavioral

To investigate the relationship between LMN and KTEA scores, we ran a correlation analysis on the LMN and KTEA scores for each participant. The LMN was determined by taking the average LMN amplitude in mV between Beat 1 and Beat 2. The participant's LMN score was then plotted against their KTEA score. To determine correlation, or lack thereof, a Pearson's correlation analysis was conducted.

RESULTS

Behavioral

Overall, performance in terms of accuracy was poor. For Cohort 1, participants scored an average accuracy of 19.35% (SD = 0.13). For Cohort 2, participants scored an average accuracy of 28.96% (SD = 0.15). For Cohort 3, participants scored an average accuracy of 23.61% (SD = 0.13). For Cohort 4, participants scored an average accuracy of 35.36 % (SD = 0.18).

The model predicting accuracy from meter, tempo, and cohort was shown to accurately describe the data (AIC = 1557.7, BIC - 1583.8) and revealed significant main effects of Meter and Tempo (Table 1). Quadruple meter is associated with a 0.43471 increase in the logs-odd of accuracy ($p<0.001$) compared to Triple meter. Fast Tempo trials lead to a 0.30038 increase in the logs-odd of accuracy ($p = 0.034$) over Slow Tempo Trials. The estimate of the interaction between Meter, Tempo, and Cohort is -0.16706, which is not significant ($p = 0.070$).

When depicting accuracy by means of a graph, as seen in Figure 6, there is a suggestion that participants are more accurate when completing Quadruple meter trials. This graph also suggests that participants are more accurate when completing Slow tempo trials. In addition, it is also important to note the staggering jump in accuracy Cohort 4 displays when completing Quadruple Fast trials.

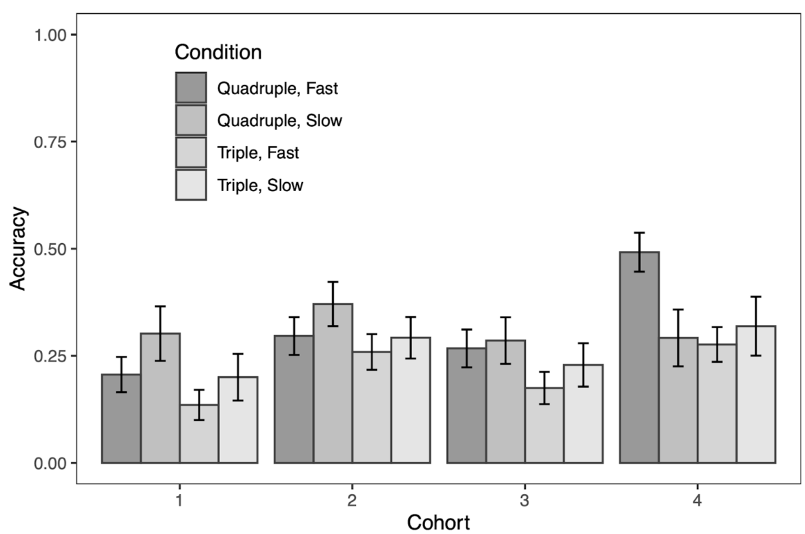
Table 1.

Fixed effects in the logistic regression model predicting accuracy from Meter, Tempo, and Cohort

	Est.	Std. Error	z-value	p
(Intercept)	-1.59881	0.33618	-4.756	<0.001***
Meter	0.43471	0.13054	3.330	<0.001***
Tempo	0.30038	0.14386	2.088	0.034*
Cohort	0.21275	0.12306	1.729	0.084
Meter: Speed: Cohort	-0.16706	0.09235	-1.809	0.070

Figure 6.

Trial accuracy for all participants aggregated by tempo, meter, and cohort. Error bars indicate standard error. An accurate trial is one where all three participant responses were aligned with Beat



ERPS

The following sections describe the results of the ANOVAs comparing results across conditions, as seen in Tables 2-9 as well as the Grand Average ERPs. The Grand average ERPs for all combinations of Tempo, Meter, and Cohort are presented in Figures 7-10.

Cohort 1

The Grand Average ERP for Cohort 1 (Figure 7) displays an immature N1 neural component, as these N1 components in these Grand Average ERPs do not show a typical N1 morphology. The Grand Average ERP shows an LMN neural component in all four types of trials.

As presented in Table 2, for the Quadruple Meter ANOVA, the main effect of Beat yielded an F ratio of $F(1, 8) = 0.895$, $p = 0.372$. This indicates that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of $F(1,8) = 12.99$, $p = 0.006$, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the quadruple meter. The relationship between Beat and Tempo additionally yielded an F ratio of $F(1,8) = 01.442$, $p = 0.264$.

As presented in Table 3, for the Triple Meter ANOVA, the main effect of Beat yielded an F ratio of $F(1, 8) = 1.55$, $p = 0.248$. This indicates that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of $F(1,8) = 11.6$, $p = 0.009$, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the triple meter. The relationship between Beat and Tempo additionally yielded an F ratio of $F(1,8) = 0.650$, $p = 0.443$.

ANOVAs were additionally run within the triple slow and fast tempo groups. For the ANOVA examining statistical difference within the triple fast meter, the main effect of Beat yielded an F ratio of $F(1,8) = 2.52$, $p = 0.151$. For the ANOVA examining statistical significance within a triple slow meter, it yielded an F ratio of $F(1,8) = 0.266$, $p = 0.620$. Neither of these ANOVA's revealed statistical significance.

Figure 7.

Grand Average ERP of Cohort 1 for all combinations of Tempo and Meter. A represents the averaged Quadruple Fast trial, B represents the averaged Triple Fast trial, C represents the averaged Quadruple Slow trial, and D represents Triple Slow trial

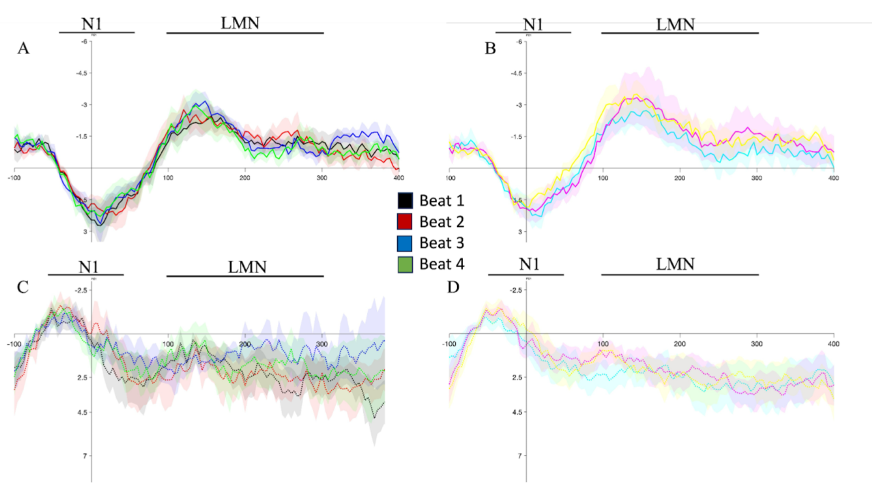


Table 2.

Mixed Repeated Measure ANOVA for Cohort 1 Quadruple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	8	0.8952966	0.372
Tempo	1	8	12.9908284	<0.001***
Beat : Tempo	1	8	1.4420102	0.264

Table 3.

Mixed Repeated Measure ANOVA for Cohort 1 Triple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	8	1.5507364	0.248
Tempo	1	8	11.6070818	<0.001***
Beat : Tempo	1	8	0.6501811	0.264

Cohort 2

The Grand Average ERP for Cohort 2 (Figure 8) displays a developing N1 neural component, as the N1 component in these Grand Average ERPs do not show a typical N1 morphology, however it is beginning to assume it's typical morphology. The Grand Average ERP shows an LMN neural component in all four types of trials.

For the Quadruple ANOVA (Table 4) the main effect of beat yielded an F ratio of F(1, 9) = 1.62, p = 0.235, indicating that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of F(1,9) 22.9, p = 0.0009, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the quadruple meter. The relationship between Beat and Tempo additionally yielded a statistically significant F ratio of F(1,9) = 6.235, p = 0.034.

For the Triple ANOVA (Table 5), the main effect of Beat yielded an F ratio of F(1, 9) = 2.748, p = 0.132, indicating that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of F(1,9) = 22.69, p = 0.001, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the quadruple meter. The relationship between Beat and Tempo additionally yielded an F ratio of F(1,9) = 0.901, p = 0.367.

ANOVAs were additionally run within the quadruple fast and slow tempo groups separately. The ANOVA examining the statistical Beat difference within the quadruple fast meter yielded an F ratio of F(1,7) = 0.193, p = -.674. For the ANOVA examining statistical significance within a quadruple fast meter, it yielded an F ratio of F(1,7) = 0.0927, p = 0.770. Neither of these ANOVA's revealed statistical significance.

Figure 8.

Grand Average ERP of Cohort 2 for all combinations of Tempo and Meter. A represents the average Quadruple Fast trial, B represents the averaged Triple Fast trial, C represents the averaged Quadruple Slow trial, and D represents Triple Slow trial.

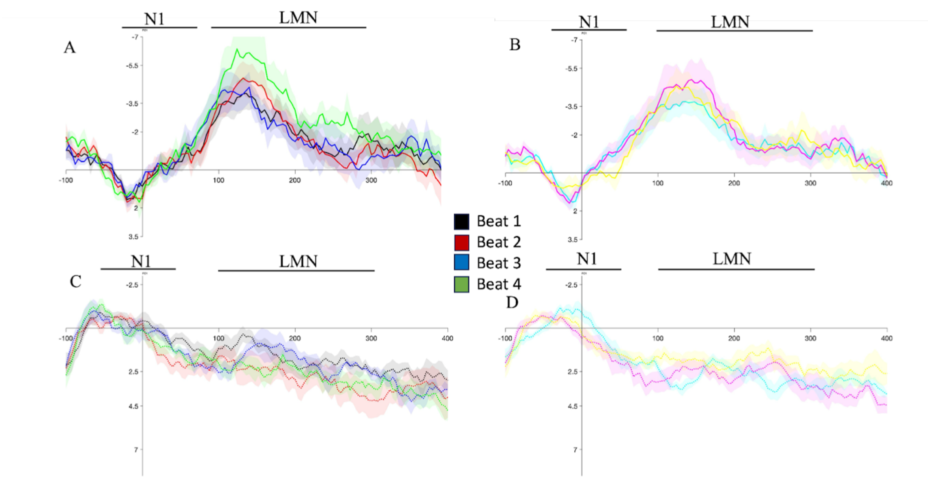


Table 4.

Mixed Repeated Measure ANOVA for Cohort 2 Quadruple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	9	1.620013	0.2349846394
Tempo	1	9	22.906925	<0.001***
Beat : Tempo	1	9	6.234526	*0.034

Table 5.

Mixed Repeated Measure ANOVA for Cohort 2 Triple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	9	2.7475044	0.132
Tempo	1	9	22.6887442	<0.001***
Beat : Tempo	1	9	0.9012213	0.367

Cohort 3

The Grand Average ERP for Cohort 3 (Figure 9) displays a developing N1 neural component, as the N1 component in these Grand Average ERPs do not all show a typical N1 morphology, however N1 component is beginning to assume it's typical morphology. This can be especially seen within Figure 9A and 9B. The Grand Average ERP shows an LMN neural component in all four types of trials.

For the Quadruple ANOVA, the main effect of beat yielded an F ratio of $F(1, 7) = 0.166$, $p = 0.696$, indicating that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of $F(1,7)$ 0.128, $p = 0.009$, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the quadruple meter. The relationship between Beat and Tempo additionally yielded an F ratio of $F(1,7) = 0.125$, $p = 0.734$.

For the Triple ANOVA, the main effect of Beat yielded an F ratio of $F(1, 7) = 0.606$, $p = 0.462$, indicating that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of $F(1,7) =$

26.8, p = 0.00128, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the quadruple meter. The relationship between Beat and Tempo additionally yielded an F ratio of F(1,7) = 0.207, p = 0.663.

Figure 9.

Grand Average ERP of Cohort 3 for all combinations of Tempo, Meter, and Cohort. A represents the average Quadruple Fast trial, B represents the averaged Triple Fast, C represents the averaged Quadruple Slow trial, and D represents Triple Slow trial

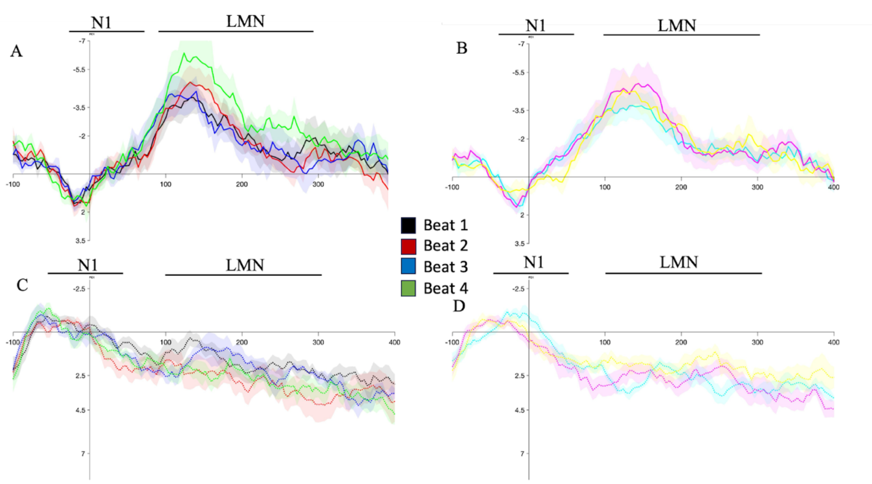


Table 6.

Mixed Repeated Measure ANOVA for Cohort 3 Quadruple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	7	0.1655537	0.696
Tempo	1	7	12.7663424	<0.001***
Beat : Tempo	1	7	0.1247741	0.734

Table 7.

Mixed Repeated Measure ANOVA for Cohort 3 Triple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	7	0.6062795	0.462
Tempo	1	7	26.8129448	<0.001***
Beat : Tempo	1	7	0.2074120	0.663

Cohort 4

The Grand Average ERP for Cohort 4 (Figure 10) displays a significantly more mature N1 neural component, as the N1 component in these Grand Average ERPs do not all show a typical N1 morphology, however N1 component is beginning to assume it's typical morphology. This can be especially seen within Figure 9A and 9B. The Grand Average ERP shows an LMN neural component in all four types of trials.

For the Quadruple ANOVA, the main effect of beat yielded an F ratio of $F(1, 7) = 0.166$, $p = 0.696$, indicating that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of $F(1,7)$ 0.12.8, $p = 0.009$, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the quadruple meter. The relationship between Beat and Tempo additionally yielded an F ratio of $F(1,7) = 0.125$, $p = 0.734$.

For the Triple ANOVA, the main effect of Beat yielded an F ratio of $F(1, 7) = 0.606$, $p = 0.462$, indicating that the mean LMN amplitude was not statistically significant based on differing beats within a quadruple meter. The main effect of Tempo yielded an F ratio of $F(1,7) = 26.8$, $p = 0.00128$, indicating that the mean amplitude was statistically different based on the fast versus slow tempo that was presented with the quadruple meter. The relationship between Beat and Tempo additionally yielded an F ratio of $F(1,7) = 0.207$, $p = 0.663$.

Figure 10.
Grand Average ERP of Cohort 4 for all combinations of Tempo, Meter, and Cohort. A represents the average Quadruple Fast trial, B represents the averaged Triple Fast, C represents the averaged Quadruple Slow trial, and D represents Triple Slow trial

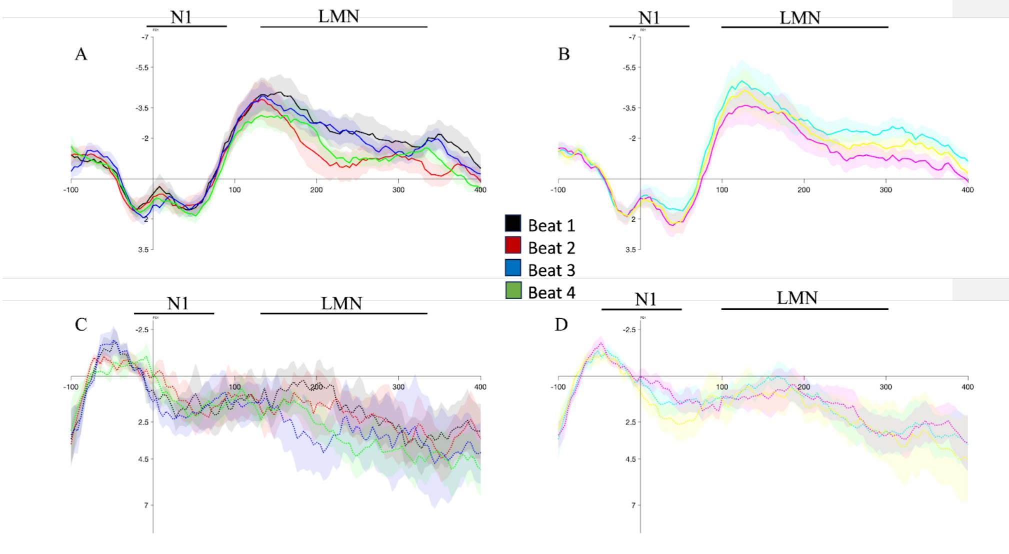


Table 8.

Mixed Repeated Measure ANOVA for Cohort 4 Quadruple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	7	2.99729647	0.127
Tempo	1	7	14.49974972	<0.001***
Beat : Tempo	1	7	0.06871589	0.801

Table 9.

Mixed Repeated Measure ANOVA for Cohort 4 Triple Fast versus Slow Trials

Effect	Dfn	Dfd	F	p
Beat	1	7	1.5270771	0.256
Tempo	1	7	13.6531851	<0.001***
Beat : Tempo	1	7	0.5254219	0.492

Combined Behavioral and ERP Effects

Given that Cohort 4 has the strongest observable LMN component, only EEG data taken from Cohort 4 was taken to find the average LMN amplitude in all, fast, and slow tempo trials. The average LMN for Triple trials is shown in Figure 11, and the average LMN for Quadruple trials is shown in Figure 12. This was done in order to determine the average LMN per Beat in a trial. In addition to calculating the average LMN per Beat, the average LMN was also calculated for each participant, averaging over the four different types of trial.

Figure 11.

Measured average LMN amplitude (mV) in all, fast, and slow triple trials, separated by beat, from Cohort 4

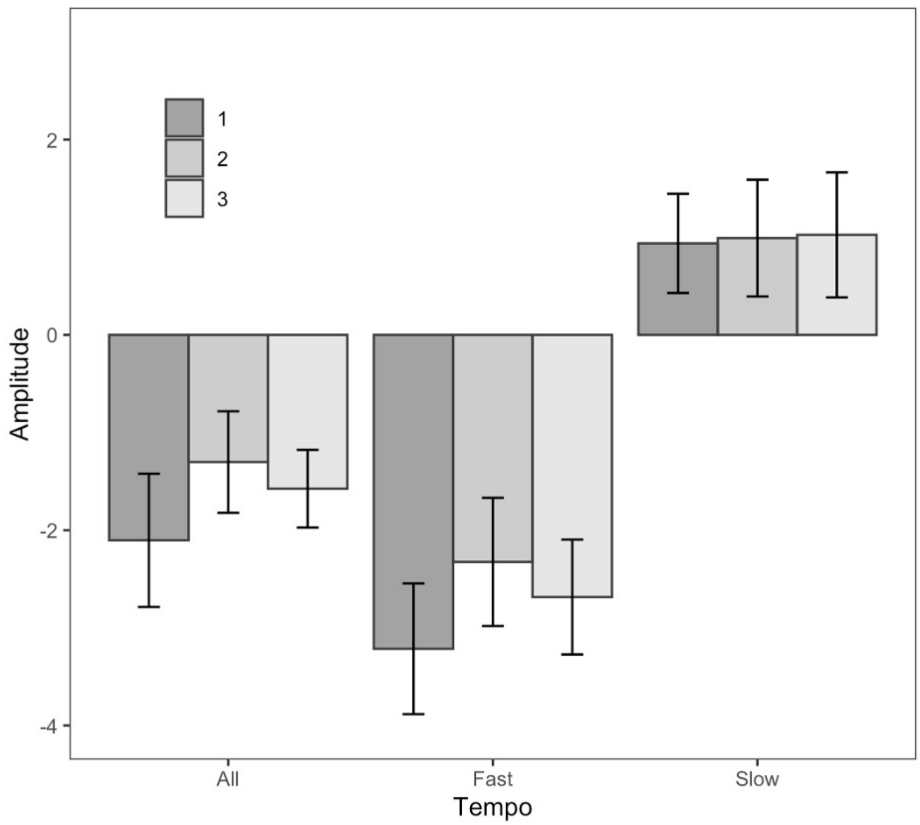


Figure 12.

Measured average LMN amplitude (mV) in all, fast, and slow quadruple trials, separated by beat, from Cohort 4

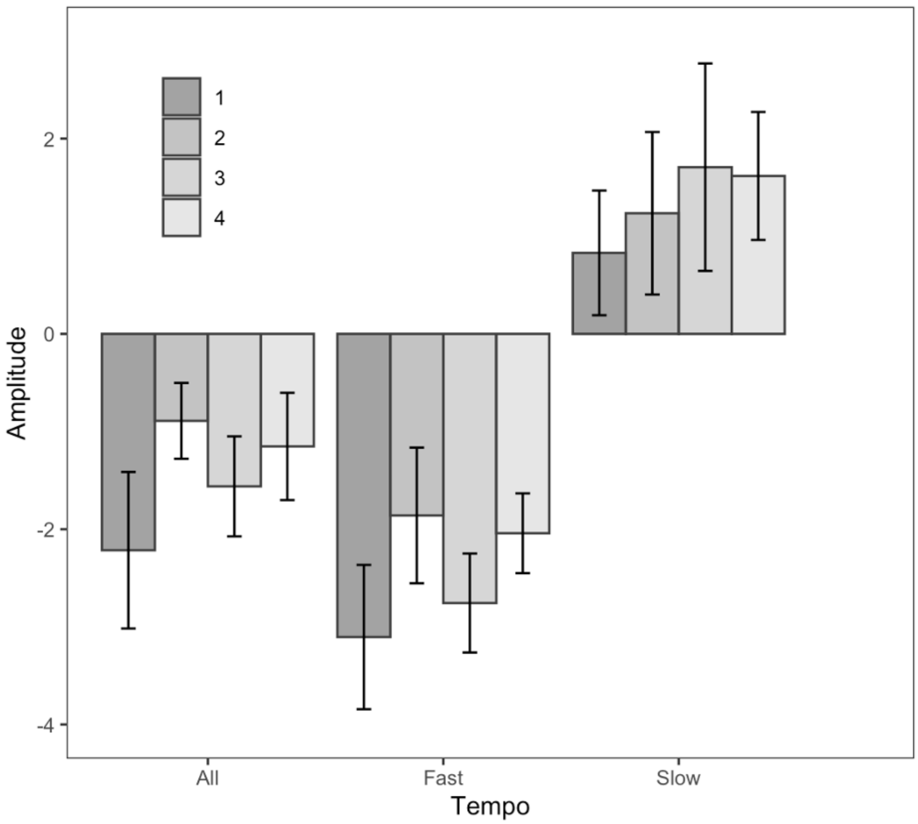
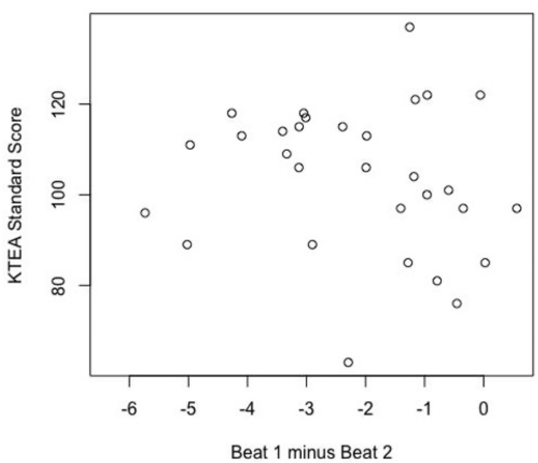


Figure 5 displays the correlation between participant's LMN effect (i.e., the difference in ERP amplitude between Beat 1 and Beat 2) and their KTEA standardized score). There is no significant correlation between LMN effect and KTEA scores, r(28) = -0.14, t = -0.77, p = 0.48.

Figure 13.

Beat 1 minus Beat 2 as in the average LMN associated with beat 1 stimuli minus the average LMN associated with beat 2 stimuli therefore showing 'Individual metric negativity'. Individual metric negativity was then correlated against participant's KTEA standard score.



DISCUSSION

The current study was designed to investigate the development of neural responses to metric attention across development and to determine whether the strength of an individual's neural response predicts their reading comprehension. In order to investigate this, we organized the results into three different stages. First, we analyzed the behavioral data by determining the mean accuracy with which participants completed the experiment task. This data was then divided into cohorts to investigate the differences that emerged within cohorts which would then also give us further insight into behavioral differences across the cohorts. Then we visually analyzed the Grand Average ERPs to see if we could visually determine the maturation of the N1 and the LMN component across cohorts. This would give us insight into seeing how these specific components emerged and developed throughout adolescent development. Finally, we compared the average LMN amplitude of each participant to their respective KTEA score in order to determine if we could find a preliminary direct correlation between a neural component and reading comprehension. There were three overarching hypotheses for this study. The first is that as age increased, so would accuracy in completing the task. The second was that we hypothesized that the subjectively stronger tone (Beat 1) will elicit a stronger N1 component in both types of tempo trials, and that the N1 will grow more pronounced with age.

The final hypothesis was that subjectively stronger tone would elicit an LMN component in all trials, and that the LMN would be larger for fast tempo trials compared to slow tempo trials. As predicted, multiple distinct effects of implicit meter on neural responses were observed, therefore indicating that the implicit metric strength in developing children draws on multiple neurocognitive mechanisms that emerge during childhood development.

Behavioral

We hypothesized that accuracy would increase as age increases. When comparing accuracy, we found that participants, across all cohorts, performed better on slow trials than fast (Figure 6). This is in line with Fitzroy & Sanders who similarly found that adults performed better on slow trials. This suggests that although children may be processing information at a greater speed than adults, there is some neural component at play that allows slower speeds to be processed with more accuracy. This contradicts our hypothesis given that we thought the faster trials would have higher accuracy scores given that a faster tempo provides more scaffolding to keep metric representations in line given that there are more beats that the participant hears. Interestingly, the participants made it clearly known that they preferred the fast tempo trials over the slow tempo trials when completing the experiment.

Additionally, when comparing accuracy, it was also found that participants were better at accurately keeping a quadruple meter than triple meter. This also aligns with what Fitzroy & Sanders found in adults. Not only did the participants make it verbally clear that they preferred quadruple meter over triple meter, but participants also struggled to complete triple trials. Oftentimes, especially in the younger cohorts, the participant would try to complete a triple trial as if it were a quadruple trial. For example, a participant would count three beats out loud, let another beat sound, and then start counting from one again, effectively turning their grouping into a quadruple meter. Or a participant would count to three out loud, and then one, and then count to three again such that they essentially repeated their count of one twice in each grouping, also effectively making their grouping of meter quadruple. Given that most of western music is expressed in either a duple or quadruple meter, and that children songs especially are composed in duple or quadruple meter, it makes sense that children would be strong with triple meter.

Vast improvement in accuracy that Cohort 4 displays in the Fast Quadruple may be due to the pilot participants having only completed fast trials during data collection. This means that there is not an equal distribution of responses between fast and slow trials. This unequal distribution might explain why the fast quadruple trials are seemingly disproportionately improved in terms of accuracy when compared to the other cohorts and the other meters and the other tempo.

The logistic regression model found a suggested negative effect of the interaction between meter and speed. Fitzroy & Sanders found that the “interaction of tempo and listening pattern may reflect that performance was often at a ceiling for both fast and slow quadruple meter trials”. When the population size is increased, it is very likely that this suggested effect will become a significant effect, given that it was modeled at Fitzroy & Sander’s model.

Grand Average ERPs

We hypothesized that participants would elicit a stronger N1 component in both types of tempo trials, and that the N1 will grow more pronounced with age. Given that the N1 did not emerge in Cohort 1, and that Cohort 2 and 3 showed a mixture of emerging and non-emergent N1 components, it was impossible to conduct significant analysis to determine if the N1 was stronger on the first beat of a grouping. It was however found that the component is more distinct in fast tempo trials (Figures 10a,10b) compared to the slow tempo trials (Figures 10c, 10d), when the component does emerge (Figures 10). Given that the N1 component was effectively absent from the youngest cohort, was partially visible in the middle cohorts, and fully visible in the oldest cohort, this suggests that most of the neural development that would facilitate a mature N1 effect develops during the specific development years that encompass Cohort 2 and 3 (Figures 7-

10). The finding that N1 increases over time aligns with the current body of literatures which states that the N1 component matures until early adulthood.

Similarly, we hypothesized that participants would elicit a stronger LMN component for the subjectively stronger tone in a triple meter when compared to the second and third beat. Consistent with this prediction, as seen in Figure 11, beat one elicits a stronger LMN component in both fast and slow tempo trials Interestingly, the average LMN component in the fast tempo trials is much stronger than in the slow tempo trials. In quadruple meter trials the same results were observed where the LMN component is the strongest on the first beat, compared to the other three, and that it is stronger in fast tempo trials than slow. These findings, that beat one has the most pronounced neural signature in both triple and quadruple meters, is in line with Fitzroy & Sander's findings, therefore supporting the notion that there is a hierarchical structure to N1 and LMN components which is depended on beat structure. Additionally, based on visual inspection of the Grand Average ERPs, it seems that the LMN does not matures alongside the maturation of an adolescent.

Given that the N1 component did not appear in all cohorts, and only sometimes in Cohorts 2 and 3, we chose to map the KTEA scores to the LMN component, instead of the preferred N1 component.

Combined Behavioral and ERP Analysis

Simply based on visual inspection, it is clear that there is no direct correlation between LMN and KTEA scores. This indicates that LMN alone does not influence reading comprehension. This makes sense with the data given that based on the Grand Average ERPs, the LMN is present for all cohorts and does not necessarily develop with increasing adolescent age. Ultimately, this shows that the LMN is not a suitable candidate to be used as a neural component that can directly predict reading comprehension. However, we do see the N1 emerges and that the Cohort 2 and Cohort 3 display a significant variance in maturation of the N1. With a large enough sample size where Cohort 2 and 3 have enough N1 components to appear, we may then have enough data to investigate if the N1 can directly predict reading comprehension.

Limitations & Future Directions

There are multiple limitations that impact this study. First, the original aim is designed to accommodate at least 120 participants so that there would be at least 30 participants within each cohort. This would then ensure that all of the ANOVAs are powered enough to show significant results. Given the small population sizes for the ANOVAs, it makes sense that there were not many significant results to be found given that there were often fewer than 10 participants in each cohort. Given that the IRB that this study is a part of calculated that 150 participants, with at least 30 in each cohort, was needed in order to ensure that the study would not be underpowered, it is clear that the current study is underpowered.

Additionally, there is an unequal distribution of fast versus slow trial data, due to the pilot participants only completing fast trials. In future analyses these first 12 participants will most likely be excluded in order to excise the unequal distribution of the quadruple accuracy scores. However, given the already small sample size for this current study, it was determined necessary to keep these 12 participants in, as excluding them would have resulted in losing almost a third of the data.

There is also the added difficulty of explaining to a young participant how they are supposed to 'feel' or 'experience' the beat without being allowed to actually count it. This is either due to experimenter error in terms of how to best describe to the child how to complete the task successfully, and it may also be due to the amount of music experience that specific participants have. Experiencing music means that the person listening feels the beat, or rather the grouping of that music, without having to intentionally count out loud, and rhythm Therefore, given that the younger a child is, the less music they have likely experienced, and therefore the less experience they have with beat rhythm, it makes sense that many of the young participants struggled with understanding how they were supposed to complete the experiment. This is another avenue which may be investigated in the future, as there was information collected about participants' musical experience. It may be interesting to see if there is a correlation between musical experience and level of development that a participant's N1 and LMN exhibits. Additionally, there are two other studies within this overarching study that look at explicit and implicit prosody. Once more data is collected, we can compare the explicit prosody that is captured by investigating their explicit production of Cat in the Hat, and the implicit prosody that is captured via eye tracking of silent reading of Cat in the Hat, against the N1 component. Especially for the study that is looking at implicit prosody, a future study may be conducted that will be correlating these reading comprehension results with the participants silent reading results and investigate if this reveals some additional information about the processes behind implicit prosody.

Conclusion

The purpose of this study was to investigate if the N1 and the LMN could be suitable neural substrates that could be studied as strong neurocognitive correlates of reading comprehension. Therefore, this study investigates these components in three different manners: by looking at the N1 and LMN Grand Average ERPS, and by pairing a participant's LMN component to their KTEA score. In addition, we also looked at the average accuracy score for each cohort to determine if, and by how much, accuracy improves over time, as that would also give us insight into if their neural substrates are maturing.

We discovered that in all, children across all cohorts struggled with accurately completing the task. However, we did find that with an increase in age there was an increase in accuracy, therefore showing that the cohorts are neurally maturing. We discovered that the N1 first emerges during Cohort 2 and Cohort 3, depending on the individual's development speed, and that Cohort 4 displayed prominent N1 signatures. This shows that the N1 emerges over time and suggests that the maturation of the N1 may be involved with the increase in reading comprehension accuracy. Unfortunately, given the lack of N1 in Cohort 1 and the small populations of Cohorts 2, 3, and 4, we were unable to run any tests on the N1 that would yield any significant results.

Additionally, by looking at the N1 and LMN components of the cohort's average neural signatures, we were able to determine that there is a significant difference in neural signatures that is dependent upon the tempo of the tones presented, and that for Cohort 2, there is a significant interaction between Beat and Tempo. Once again, given the small sample size, we suspect more significant results will be revealed once there has been a larger amount of data collected.

It is clear, both by looking at the accuracy of participants, the presence of the LMN in the Grand Average ERPS, and most significantly the lack of correlation between the KTEAs and the LMN, that the LMN should not be considered a suitable candidate for being a neural substrate that can directly predict reading comprehension.

REFERENCES

Alderson-Day, B., & Fernyhough, C. (2015). Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting Consequences of the Summer Learning Gap. *American Sociological Review*, *72*(2), 167–180. https://doi.org/10.1177/000312240707200202

Allor, J. H., Mathes, P. G., Roberts, J. K., Cheatham, J. P., & Otaiba, S. A. (2014). Is Scientifically Based Reading Instruction Effective for Students With Below-Average IQs? *Exceptional Children*, *80*(3), 287–306. https://doi.org/10.1177/0014402914522208

Beres, A. M. (2017). Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Applied Psychophysiology and Biofeedback*, *42*(4), 247–255. https://doi.org/10.1007/s10484-017-9371-3

Bishop, J. (2021). Exploring the Similarity Between Implicit and Explicit Prosody: Prosodic Phrasing and Individual Differences. *Language and Speech*, *64*(4), 873–899. https://doi.org/10.1177/0023830920972732

Bradley, R. H., & Corwyn, R. F. (2001). *Socioeconomic Status and Child Development*.

Breen, M. (2014). Empirical Investigations of the Role of Implicit Prosody in Sentence Processing. *Language and Linguistics Compass*, *8*(2), 37–50. https://doi.org/10.1111/lnc3.12061

Breen, M., Kaswer, L., Van Dyke, J. A., Krivokapić, J., & Landi, N. (2016). Imitated Prosodic Fluency Predicts Reading Comprehension Ability in Good and Poor High School Readers. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01026

Breen, M., Van Dyke, J., Krivokapić, J., & Landi, N. (2024). Prosodic features in production reflect reading comprehension skill in high school students. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001355

Cappella, E., & Weinstein, R. S. (2001). Turning around reading achievement: Predictors of high school students' academic resilience. *Journal of Educational Psychology*, *93*(4), 758–771. https://doi.org/10.1037/0022-0663.93.4.758

Cheng, Y., & Wu, X. (2017). The Relationship between SES and Reading Comprehension in Chinese: A Mediation Model. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00672

Cirelli, L. K., Spinelli, C., Nozaradan, S., & Trainor, L. J. (2016). Measuring Neural Entrainment to Beat and Meter in Infants: Effects of Music Background. *Frontiers in Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnins.2016.00229

Cummins, J. (2007). Pedagogies for the Poor? Realigning Reading Instruction for Low-Income Students With Scientifically Based Reading Research. *Educational Researcher*, *36*(9), 564–572. https://doi.org/10.3102/0013189X07313156

Cutting, L. E., Materek, A., Cole, C. A. S., Levine, T. M., & Mahone, E. M. (2009). Effects of fluency, oral language, and executive function on reading comprehension performance. *Annals of Dyslexia*, *59*(1), 34–54. https://doi.org/10.1007/s11881-009-0022-0

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

Demin, S., Panischev, O., & Demina, N. (2016). Cross MFF—analysis in studying the obsessive-compulsive disorder. *Journal of Physics: Conference Series*, *741*, 012073. https://doi.org/10.1088/1742-6596/741/1/012073

Domingue, B. W., Dell, M., Lang, D., Silverman, R., Yeatman, J., & Hough, H. (2022). The Effect of COVID on Oral Reading Fluency During the 2020–2021 Academic Year. *AERA Open*, *8*, 23328584221120254. https://doi.org/10.1177/23328584221120254

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, *43*(6), 1460–1473. https://doi.org/10.1037/0012-1649.43.6.1460

Fitzroy, A. B., & Sanders, L. D. (2020). Subjective metric organization directs the allocation of attention across time. *Auditory Perception & Cognition*, *3*(4), Article 4. https://doi.org/10.1080/25742442.2021.1898924

Fodor, J. D. (2002). *Prosodic Disambiguation In Silent Reading*. *32*.

Foundation, the A. E. C. (2010, January 2). *Reading by Third Grade*. The Annie E. Casey Foundation. https://www.aecf.org/resources/early-warning-why-reading-by-the-end-of-third-grade-matters

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, *69*(1), Article 1. https://doi.org/10.1016/j.jml.2013.03.001

Habibi, A., Cahn, B. R., Damasio, A., & Damasio, H. (2016). Neural correlates of accelerated auditory processing in children engaged in music training. *Developmental Cognitive Neuroscience*, *21*, 1–14. https://doi.org/10.1016/j.dcn.2016.04.003

Hernandez, D. J. (2011). Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation. In *Annie E. Casey Foundation*. Annie E. https://eric.ed.gov/?id=ED518818

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, *47*(6), 674–689. https://doi.org/10.1016/j.cortex.2010.07.010

Kapatsinski, V., Olejarczuk, P., & Redford, M. A. (2017). Perceptual learning of intonation contour categories in adults and 9 to 11-year-old children: Adults are more narrow-minded. *Cognitive Science*, *41*(2), 383–415. https://doi.org/10.1111/cogs.12345

Kieffer, M. J. (2012). Before and after third grade: Longitudinal evidence for the shifting role of socioeconomic status in reading growth. *Reading and Writing*, *25*(7), 1725–1746. https://doi.org/10.1007/s11145-011-9339-2

Kim, J. S. (2008). Research and the reading wars. In F. M. Hess (Ed.), *When Research Matters: How Scholarship Influences Education Policy* (pp. 89–111). Harvard Education Press. http://hepg.org/hep/book/79

Kuhfeld, M., Lewis, K., & Peltier, T. (2023). Reading achievement declines during the COVID-19 pandemic: Evidence from 5 million U.S. students in grades 3–8. *Reading and Writing*, *36*(2), 245–261. https://doi.org/10.1007/s11145-022-10345-8

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

Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lloyd, D. N. (1978). Prediction of School Failure From Third-Grade Data. *Educational and Psychological Measurement*, *38*(4), 1193–1200. https://doi.org/10.1177/001316447803800442

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Machery, E. (2005). You Don't Know How You Think: Introspection and Language of Thought. *The British Journal for the Philosophy of Science*, *56*(3), 469–485. https://doi.org/10.1093/bjps/axi130

McCurdy, K., Kentner, G., & Vasishth, S. (2013). Implicit prosody and contextual bias in silent reading. *Journal of Eye Movement Research*, *6*(2), Article 2. https://doi.org/10.16910/jemr.6.2.4

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology*, *98*(4), Article 4. https://doi.org/10.1037/0022-0663.98.4.839

Miller, J., & Schwanenflugel, P. J. (2008). A Longitudinal Study of the Development of Reading Prosody as a Dimension of Oral Reading Fluency in Early Elementary School Children. *Reading Research Quarterly*, *43*(4), Article 4. https://doi.org/10.1598/RRQ.43.4.2

NAEP Nations Report Card—The NAEP Reading Achievement Levels by Grade. (n.d.). Retrieved April 15, 2024, from https://nces.ed.gov/nationsreportcard/reading/achieve.aspx

NAEP Reading: Reading Highlights 2022. (n.d.). Retrieved April 26, 2024, from https://www.nationsreportcard.gov/highlights/reading/2022/

Nave-Blodgett, J. E., Snyder, J. S., & Hannon, E. E. (2021). Hierarchical beat perception develops throughout childhood and adolescence and is enhanced in those with musical training. *Journal of Experimental Psychology: General*, *150*(2), Article 2. https://doi.org/10.1037/xge0000903

Ozernov-Palchik, O., Wolf, M., & Patel, A. D. (2018). Relationships between early literacy and nonlinguistic rhythmic processes in kindergarteners. *Journal of Experimental Child Psychology*, *167*, 354–368.

Petscher, Y., Cabell, S. Q., Catts, H. W., Compton, D. L., Foorman, B. R., Hart, S. A., Lonigan, C. J., Phillips, B. M., Schatschneider, C., Steacy, L. M., Terry, N. P., & Wagner, R. K. (2020). How the Science of Reading Informs 21st-Century Education. *Reading Research Quarterly*, *55*(Suppl 1), S267–S282. https://doi.org/10.1002/rrq.352

Potter, D. D., Fenwick, M., Abecasis, D., & Brochard, R. (2009). Perceiving rhythm where none exists: Event-related potential (ERP) correlates of subjective accenting. *Cortex*, *45*(1), Article 1. https://doi.org/10.1016/j.cortex.2008.01.004

Rescorla, M. (2023). The Language of Thought Hypothesis. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/language-thought/

Romeo, R. R., Uchida, L., & Christodoulou, J. A. (2022). Socioeconomic status and reading outcomes: Neurobiological and behavioral correlates. *New Directions for Child and Adolescent Development*, *2022*(183–184), 57–70. https://doi.org/10.1002/cad.20475

Schwartz, S. (2022, July 20). *Which States Have Passed "Science of Reading" Laws? What's in Them?* Education Week. https://www.edweek.org/teaching-learning/which-states-have-passed-science-of-reading-laws-whats-in-them/2022/07

Skarnitzl, R., & Hledíková, H. (2022). Prosodic Phrasing of Good Speakers in English and Czech. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.857647

Solari, E. J., Terry, N. P., Gaab, N., Hogan, T. P., Nelson, N. J., Pentimonti, J. M., Petscher, Y., & Sayko, S. (2020). Translational Science: A Road Map for the Science of Reading. *Reading Research Quarterly*, *55*(S1), S347–S360. https://doi.org/10.1002/rrq.357

Suárez, N., Sánchez, C. R., Jiménez, J. E., & Anguera, M. T. (2018). Is Reading Instruction Evidence-Based? Analyzing Teaching Practices Using T-Patterns. *Frontiers in Psychology*, *9*, 7. https://doi.org/10.3389/fpsyg.2018.00007

Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, *18*(1), 70–73. https://doi.org/10.4103/0972-6748.57865

Tierney, A., Gomez, J. C., Fedele, O., & Kirkham, N. Z. (2021). Reading ability in children relates to rhythm perception across modalities. *Journal of Experimental Child Psychology*, *210*, 105196.

Vissers, C. Th. W. M., Tomas, E., & Law, J. (2020). The Emergence of Inner Speech and Its Measurement in Atypically Developing Children. *Frontiers in Psychology*, *11*, 279. https://doi.org/10.3389/fpsyg.2020.00279

Wade-Woolley, L. (2016). Prosodic and phonemic awareness in children's reading of long and short words. *Reading and Writing*, *29*(3), 371–382. https://doi.org/10.1007/s11145-015-9600-1

Woodruff Carr, K., White-Schwoch, T., Tierney, A. T., Strait, D. L., & Kraus, N. (2014). Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. *Proceedings of the National Academy of Sciences*, *111*(40), 14559–14564. https://doi.org/10.1073/pnas.1406219111