THE INVESTIGATION OF TEMPORAL ORDER IN LANGUAGE LEARNING USING BEHAVIOURAL TASKS AND MMN

THE INVESTIGATION OF TEMPORAL ORDER IN LANGUAGE LEARNING USING BEHAVIOURAL TASKS AND MMN

By ERIN ELIZABETH DEBORBA, B.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

MASTER OF SCIENCE (2020)

Hamilton, Ontario (Cognitive Science of Language)

TITLE: THE INVESTIGATION OF TEMPORAL ORDER IN LANGUAGE LEARNING USING BEHAVIOURAL TASKS AND MMN

AUTHOR: Erin Elizabeth DeBorba, B.A. (Hons.) (McMaster University)

SUPERVISOR: Dr. Elisabet Service

NUMBER OF PAGES: xi, 122

Lay Abstract

Spoken language is driven by rhythm and keeping track of this rhythm allows us to keep track of the order in which sounds in language are presented. Remembering the order of items requires the use of short-term memory. The better one is at repeating back the order of items, the better they are at learning new words. This thesis investigates the relationship between various short-term memory tasks (English nonword sentence repetition task, foreign sentence repetition task, temporal rhythm accuracy task, auditory judgment task, visual judgment task) and foreign-word learning. This thesis also explores whether there is a correlation between one's brain responses to differing stimuli and a person's ability to track the timing and order of items, as well as a person's ability to learn new words. The results reveal that only the foreign sentence repetition task, using the same foreign language as the word learning task, significantly predicts one's ability to accurately learn foreign words. The results did not show any significant interaction between one's neural responses and rhythm or word learning. These results suggest that the ability to maintain the order of items in memory aids word learning, but further exploration is required with regards to non-verbal stimuli and neural responses. It is important to investigate individual differences in repetition tasks that require short-term memory, as this will aid in understanding normal language development and language acquisition.

Abstract

Short-term memory (STM) has demonstrated to be affected by serial order, involving the use of rhythm and entrainment to stimuli. However, less is known of the extent of this relationship and language learning, and the literature focuses on words rather than sentences. Moreover, the literature lacks an exploration of whether this relationship has a correlation with MMN responses.

We had 30 participants (21 female) complete two sentence repetition tasks, a temporal rhythm accuracy task, and two temporal order judgment tasks. We also recorded the electroencephalograms (EEG) from 24 of the participants (17 female) while they listened to syllables differing by time of presentation and differing by consonant and vowel. We then correlated performance on these tasks to performance on a foreign-word learning (FWL) task. We hypothesized that the STM tasks would predict performance in the FWL task, and we explored whether temporal accuracy and word learning correlated with MMN responses to early stimuli. We found that only the foreign sentence repetition task significantly predicted performance in the FWL task. We also did not find any significant correlations with MMN responses and temporal accuracy and word learning abilities. Findings show that with previous exposure to a novel language, the prosodic pattern of the foreign language is stored temporarily in STM, which enhances learning of the foreign words. Further exploration is needed to understand the relationship of temporal order and language learning with cortical responses.

Acknowledgements

I would like to thank my supervisor Elisabet Service for her guidance throughout the thesis and various other projects I completed in my Graduate career. I would also like to thank my committee members Daniel Pape and John Connolly for their helpful guidance and comments throughout the process. I would like to thank Chia-Yu Lin and ARiEAL for providing me the opportunity and space to learn and grow in my Graduate career. I thank my colleagues Fareeha Rana, Nathalee Renée and Melda Coşkun for their contribution to project ideas, guidance, and support. Additionally, I would like to thank my lab members and the undergraduate students in the lab who assisted with running participants and data input. I would like to thank my colleague Tess Hudson, and colleague and collaborator Angie Lopez Ricote for all of their amazing support and advice throughout the thesis and Graduate studies. Last, but not least, I thank my family, friends, and my cats for their emotional support throughout my Graduate journey.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

ANOVA/s = analysis/analyses of variance

AVSTM = auditory-verbal short-term memory

BVATM = brain vision analyzer TM

EEG = electroencephalography

ERP = event-related potential

FFR = fast-frequency following response

FWL = foreign-word learning

HiREB = Hamilton integrated research ethics board

ICA = independent component analysis

IRR = inter-rater reliability

ISI = interstimulus interval

ITL = iambic-trochaic law

LTM = long-term memory

LC = left central

LF = left frontal

LP = left parietal

MC = middle central

MF = middle frontal

MMN = mismatch negativity

MP = middle parietal

PHPH = phonological phrase

PNB = psychology, neuroscience, and behaviour

RC = right central

RF = right frontal

ROI = regions of interest

RP = right parietal

SLI = specific language impairment

SOA = stimulus onset asynchrony

SOV = subject-object-verb

STM = short-term memory

SVO = subject-verb-object

TD = typically developing

TOJ = temporal order judgment

TRA = temporal rhythm accuracy

WM = working memory

Declaration of Academic Achievements

The conceptualization of this thesis was conducted by myself, in collaboration with Dr. Elisabet Service. I explored and analyzed the data under the supervision of Dr. Elisabet Service. The statistical analysis was completed by myself and Dr. Elisabet Service. The present thesis was written by myself, with comments and feedback from Dr. Elisabet Service and Dr. Daniel Pape.

Introduction

This thesis explores temporal order in language learning. This topic will be explored through verbal and non-verbal short-term memory tasks and a mismatch negativity oddball paradigm for the detection of event-related brain responses. Short-term memory (STM) has been demonstrated to be able to record serial and temporal order in research involving the use of rhythm and entrainment to stimuli (e.g., Gilbert et al., 2014; Henson et al., 2003; Majerus & Cowan, 2016). This STM for the ordering of elements is not restricted to phonological material (e.g., Laasonen et al., 2012; Richie & Aten, 1976; Szmalec, Loncke, Page, & Duyck, 2011). Experiments have focused on the ability to correctly repeat back the phoneme order of nonwords and its potential to predict foreign-word learning (e.g., Schraeyen et al., 2018; Service & Craik, 1993). However, less is known about the predictive effects of sentence repetition.

The present study consisted of short-term memory tasks, including a temporal rhythm accuracy task, two sentence repetition tasks (an English nonword sentence repetition task and a foreign sentence repetition task), and two temporal order judgment (TOJ) tasks. The experiment also included registration of event-related brain potentials in a mismatch negativity (MMN) oddball paradigm that required participants to passively listen to a train of syllables. There is evidence to suggest a relationship between this negative MMN response and short-term memory (e.g., Bonetti et al., 2017). The STM, TOJ, and MMN tasks were used to predict performance on a foreign-word learning task.

The experiments aimed to understand the relationship between individual differences in STM tasks and language learning, and if that relationship could be predicted by one's ability to perceive variation in speech. This chapter begins with a general overview of the field, followed by a literature review of relevant studies and theories, and concludes with the experimental design and hypotheses. The second chapter consists of the methodology, the third chapter the results, the fourth chapter the discussion of results and limitations of the study, and the final chapter the conclusion.

Working Memory

There are various models regarding working memory (WM), but this thesis will focus on that proposed by Baddeley and Hitch (for the original version, see 1974). Baddeley and Hitch refer to WM as "the system or systems that are assumed to be necessary in order to keep things in mind while performing complex tasks such as reasoning, comprehension, and learning" (Baddeley, 2010). This concept of WM evolved from STM, which refers to the "temporary storage of small amounts of material over brief periods of time" (Baddeley, 2010). Therefore, STM is seen as a subsystem of WM in that it is the temporary storage of information, whereas WM is the combination of storage and manipulation (Baddeley, 2012).

The conceptualization of the complex memory system containing storage systems and buffers feeding into a separate long-term memory (LTM), came to be through the study of brain impaired patients (Baddeley, 2010). It was found that maintaining information in STM does not guarantee long-term learning. The encoding of a word based on its meaning was found to be much more effective for LTM storage than processing a word with regards to its perceptual appearance, which dominates verbal STM (Baddeley, 2010; Craik & Lockhart, 1972). In addition, keeping multiple information sets in STM is possible, but performance declines as the length of sequences increase and parallel tasks begin disrupting STM (Baddeley & Hitch, 1974). The finding that information from different modalities could be stored in parallel, as long as the task was not too complex, led to the proposal of a three-component model of working memory that involves the *central executive*, the *visuo-spatial sketchpad*, and the *phonological loop*. The visuo-spatial sketchpad was proposed to store visual material, and the phonological loop was proposed to store verbal-acoustic material and the central executive was proposed to use attention to manage the different types of material (Baddeley, 2010).

The phonological loop is suggested to contain both a phonological store and a rehearsal process (Baddeley, 1986). Baddeley and Hitch assume a process of trace decay for items if they are not actively maintained by rehearsal in the phonological loop (Baddeley, 2012). Rehearsal involves a process of using internal speech to reencode the memory traces into the phonological store before the traces have decayed past the point of identification (Roodenrys et al., 2002). Memory traces can be refreshed by open verbal or subvocal rehearsal, i.e., silent articulation (Baddeley, 2010), however immediate recall declines as the length of the to-be-remembered words increases. This is referred to as the word length effect (Baddeley et al., 1975; Baddeley, 2010). Spoken material is assumed to gain direct access to the phonological store, whereas, in contrast, written material needs to be subvocalized (Baddeley, 2010; Baddeley, 2012; Baddeley et al., 1984; Baddeley et al., 1975). Uttering aloud a string of irrelevant sounds prevents rehearsal of words and suppresses the word length effect (Baddeley, 2010), as it is thought to prevent rehearsal. The main function of the phonological loop was demonstrated with an Italian patient, PV, who had an auditory digit span of only two items, but her intellect was preserved. Baddeley and colleagues (1988) discovered that her capacity to learn lists of native language word pairs was normal, but she failed to learn native–foreign-word pairs. Thus, the phonological loop appears to facilitate new phonological learning, which implies a direct link from the loop to LTM (Baddeley, 2012). The extent to which one can remember and repeat a string of unrelated words is about five items, but if the items form a meaningful sentence, the span is around 15 words (Baddeley, 2010). This reflects a contribution from grammar and meaning, both relying on different features of LTM (Baddeley, 2010). Existing language knowledge influences immediate nonword recall (Gathercole, 1995), suggesting that information flows back and forth between the phonological loop and LTM (Baddeley, 2012). This has been argued to demonstrate the distinction between WM and LTM: the former is constructed of various fluid systems that need temporary activation, whereas the latter is thought to represent more permanent definite skills and knowledge (Baddeley, 2012).

The visuo-spatial sketchpad was studied through testing the rehearsal of information with material that was not easily named, such as a matrix of squares filling up in a pattern (Phillips & Baddeley, 1971). A second matrix of squares was presented at varying time delays and participants were required to make a same/different judgment. The researchers found a steady decline over time through measuring performance on both accuracy and reaction time (Baddeley, 2012). Later on, Phillips (1974) demonstrated that the accuracy of recall for visual memory declines with the number of cells to be remembered. This led to the conclusion of there being a limited capacity for visual STM, and thus to the concept of the visuo-spatial sketchpad (Baddeley, 2012).

The proposed WM model does not include a mechanism to properly account for the serial order of temporal or spatial sequences within the phonological loop and the visuo-spatial sketchpad (Hurlstone & Hitch, 2014), which Baddeley (2012) and Baddeley and Hitch (2019) note as a shortcoming in the original model. A review of studies of spatial and verbal STM for serial order suggested that they rely on similar mechanisms (Hurlstone et al., 2013; Pickering et al., 1998), which suggests that the principles of serial order in STM could be shared in the verbal and the spatial domain. Transposition errors refer to when an item is recalled in the wrong position (Hartley et al., 2016). Transposition errors in spatial serial recall have been found to be comparable to those seen in serial recall of verbal stimuli (Farrell & Lewandowsky, 2004; Hurlstone & Hitch, 2014), providing supporting evidence for the hypothesis that spatial and verbal STM represent serial order by shared mechanisms (Hurlstone & Hitch, 2014). A competitive queuing mechanism has been proposed for recall of serially ordered items along an activation gradient from first to last, so that during recall, the most activated item would be recalled, then suppressed to allow the following item to stand out as the most activated (Page & Norris, 1998). According to the review by Hurlstone and Hitch, serial order may be represented by three mechanisms: a primacy gradient of activation, position marking, and response suppression after each recalled item (Hurlstone & Hitch, 2014). This raises further questions as to whether serial order is processed separately in each domain, or whether there is a central, multi-modal serial ordering mechanism (Baddeley & Hitch, 2019). This thesis uses the WM framework by Baddeley and Hitch as well as Hurlstone and Hitch's (2014) review of mechanisms that may account for representation of serial order in STM to explore and understand temporal order in language learning.

Temporal Order in STM

Being able to represent temporal order in addition to item information in STM has lately been linked to language, especially, vocabulary acquisition in L1 (Attout et al., 2020; Leclercq et al., 2010; Majerus & Boukebza, 2013; Majerus et al., 2006). Temporal order refers to the order of items in time (Laasonen et al., 2012). With regards to verbal recall, it is the order in which items are presented (Schraeyen et al., 2018). In visuo-spatial recall, temporal order is less well understood. It is important to investigate the difference between memory for items and memory for their order, whether it be serial or temporal, as this will aid in understanding normal language development.

In a number of theoretical proposals (Burgess & Hitch, 1999; Brown et al., 2000; Henson et al., 2003), verbal STM for serial order is supported by a timing signal. The timing signal derives from a set of internal oscillators, working as brain clocks at different time scales, similar to second, minute, and hour hands on real clocks. The time signal enables the coding of the positions of items within a sequence (Burgess & Hitch, 1999; Brown et al., 2000). The involvement of these oscillators in encoding and retrieving verbal material could help to explain errors found in phonological output tasks (Hartley, 1995, 2002; Hartley & Houghton, 1996; Vousden et al., 2000). Conrad and Hull (1964) first showed that serial recall for similar sounding visually presented letters is less accurate than that for dissimilar sounding letters (Baddeley & Hitch, 2019). It has also been shown that background speech can be disruptive to tasks that require maintenance of serial order of visually presented verbal material (Salamé & Baddeley, 1982). Henson and colleagues (2003) used an item probe task and a list probe task that differed in the degree to which they required preservation of serial order. In the item probe task, participants saw a list of items presented sequentially, followed by a single probe item, and they were asked to judge whether or not that probe was in the list. This item probe task was likely to indicate STM for item information in the absence of serial scanning or rehearsal. In the list probe task, on the other hand, a list was presented sequentially, and the probe was a second list that the participants had to judge as being the same or different from the first list. The probe list contained the same items as the first list, but in the different-trials, two items differed in their placement. This latter task was described as encouraging forward serial processing, since the participants are comparing successive items in the probe against their memory for the original list. Background speech resulted in greater disruption of the list probe task. Both tasks were performed less accurately when the probes included phonologically similar material to the sequence items, which indicates that the tasks were using phonological STM. The overall results indicated that both irrelevant speech and articulatory suppression, i.e., speaking out loud to restrict silent rehearsal of items, degrade memory for serial order. As articulatory suppression captures the articulatory system it prevents participants from using subvocalization (Henson et al., 2003).

Temporal grouping, inserting a pause every few items in a sequence, is known to improve serial recall (e.g., Ryan, 1969a) and is independent of word length and phonological similarity (Hitch et al., 1996). Temporal grouping is independent of articulatory suppression when the list items are presented auditorily, but not visually (Hitch et al., 1996). The reasoning behind temporal grouping is that it results in a contrast of the timing signal into two components: one tracking the timing of the items within groups, and one tracking the timing of items within lists, therefore assisting with serial recall (Henson et al., 2003). This temporal grouping, or perceptual chunking, has also been reported to be supported by the N400 and P300 event-related brain responses (Gilbert et al., 2014). Perceptual chunks influence immediate memory of monosyllabic words, and there are better memory traces for temporal groups of three than there are of four (Gilbert et al., 2014). Utterance-related information is proposed to be stored in serially-ordered chunks and the scanning of this information in WM follows on a chunk-by-chunk basis (Gilbert et al., 2014).

The link between memory for order and language has been studied in reading difficulties. Dyslexia or developmental reading disorder is a label for individuals whose reading skills do not develop at typical pace. Although the dominant account for reading difficulties assumes problems with awareness and processing of phonemes to be the proximate cause, persons with dyslexia also show deficits with non-phonological tasks, and their STM impairment is not necessarily a consequence of the phonological processing deficit (Laasonen et al., 2012; Szmalec et al., 2011). Previous studies have shown STM impairments in individuals with dyslexia for recall of nonverbal material, such as rhythm and duration (Richie & Aten, 1976), and spatiotemporal flash and tone sequences (Jones, 1974). However, many have failed to show differences between dyslexic and typically developing (TD) individuals in the visuospatial STM domain (e.g., Smith-Spark & Fisk, 2007; Smith-Spark et al., 2003, as cited in Laasonen et al., 2012). Laasonen and colleagues (2012) demonstrated deficits in tasks that require updating of temporal patterns, i.e., tasks which require the binding together of a temporal sequence of stimuli to a single representation in STM. The researchers used two types of STM tasks, Item and Time spans for simple perceptual stimuli, i.e., light flashes, simple tones, finger touches (Laasonen et al., 2012). They believed that the Item STM tasks should be sensitive to the ability to incrementally update an increasingly complex temporal pattern in a comparable way that a verbal sequence would be updated with additional words presented in a list (Laasonen et al., 2012). For the Time STM tasks, the researchers expected that if the durability of traces is an issue for individuals with dyslexia, then they should show poorer performance as the pauses between the items lengthen (Laasonen et al., 2012). The results indicated a correlation between STM performance with phonological sequences, simple sensory stimuli, temporal acuity, and reading abilities (Laasonen et al., 2012). Reading disability was related to an impairment in sensory STM for temporal order patterns, and this generalized to the different modalities and their combinations in the Item STM tasks (Laasonen et al., 2012). Perhaps the difficulties found in nonverbal tasks are a result of a more general difficulty in binding together multiple elements into a simultaneous representation in STM (Laasonen et al., 2012).

As reviewed above, Hurlstone, Hitch, and Baddeley (2013) suggest that there are functional similarities across domains that support the idea that the verbal, visual, and spatial domains are controlled by a competitive queuing mechanism, where the items are simultaneously active and the strongest one is selected for output. With regards to the verbal STM competitive queuing system, results suggest that serial order is represented by a primacy gradient, position marking, response suppression, and cumulative matching and that item similarity effects are seen during serial order encoding and retrieval (Hurlstone et al., 2013). On the other hand, the underlying representation of serial order in visual and spatial competitive queuing systems are not as clear (Hurlstone, et al., 2013). In the spatial domain, there is some evidence suggesting that there is a primacy gradient and position marking like the verbal domain, but there is no direct evidence for these in the visual domain (Hurlstone et al., 2013). For example, there has been evidence of temporal grouping effects in the visual domain (e.g., Henson et al., 2003; Hitch et al., 1996; Ryan, 1969a), in the auditory-spatial domain (Parmentier et al., 2004), and the visual-spatial domain (Parmentier, Andres, et al., 2006, Experiment 3), but not in visual STM (Hurlstone et al., 2013). As previously mentioned, Hurlstone and Hitch (2014) performed experiments that sought to find evidence that spatial and verbal STM rely on a common mechanism for the representation of serial order. They examined transpositions in spatial serial recall tasks to test for error latency predictions of various models in representing serial order, showing that temporally grouped spatial sequences recruit a primacy gradient, position marking, and response suppression (Hurlstone & Hitch, 2014). Additional studies have demonstrated that serial order STM impairment occurs for the retention of both verbal and visuo-spatial sequence information (for a review, see Majerus & Cowan, 2016). This indicates that spatial and verbal STM may rely on common mechanisms for the representation of serial order (Hurlstone & Hitch, 2014).

In addition to serial recall, nonword repetition tasks have been used to investigate problems with coding and retrieval of phonological order information in verbal STM, as this task requires one to temporarily store the phonemes and their serial order for later recall (Schraeyen et al., 2018). Individuals with dyslexia perform as well as controls in recalling phonemes' identity when repeating a nonword (Schraeyen et al., 2018). However, controls perform better than adults with dyslexia when recalling phonemes' serial order, regardless of the number of syllables in the nonword (Schraeyen et al., 2018). Therefore, there is some evidence that a specialized order mechanism in verbal STM also operates at the level of phonemes (Schraeyen et al., 2018). The issue, with correctly recalling serial order in impaired readers, points to a difference between these readers' domain general sequence memory mechanisms in STM and that of individuals with typical language development (Schraeyen et al., 2018). However, it is unclear whether these findings in dyslexia reflect a more general problem in language processing or especially relate to literacy.

As is well known, the capacity of serially-ordered auditory-verbal STM (AVSTM) is generally measured using immediate serial recall tasks, where a sequence of spoken items must be remembered in the correct order (Gilbert et al., 2017). Order errors are extremely sensitive to temporal characteristics of spoken sequences, such as their rhythmic structure (Frankish, 1985; Gilbert et al., 2017; Hurlstone & Hitch, 2016; Hitch et al., 1996; Ryan, 1969a, 1969b). However, a major limitation of previous work done on the link between AVSTM for serial order and temporal properties of verbal materials focuses on memory for item *order*, not item *timing* (Gilbert et al., 2017). Studies have shown deficits in AVSTM for serial order in both children and adults with atypical language development (Corkin, 1974; Perez et al., 2012). Using a rehearsal probe task and a STM preload paradigm (e.g., Baddeley & Hitch, 1974; Cocchini et al., 2002; FitzGerald & Broadbent, 1985), Gilbert, Hitch, and Hartley (2017) demonstrated that common resources constrain both span and temporal precision in AVSTM (Gilbert et al., 2017). AVSTM is a more flexible system than the fixed time-based storage capacity in the early models of the phonological loop (Baddeley et al., 1975), as resources are allocated to temporal precision with implications for amount stored (Gilbert et al., 2017). The capacity of serially-ordered AVSTM has to do with the timing of the material to be stored, and both temporal processing and AVSTM capacity are proposed to be involved in language development (Gilbert et al., 2017; Laasonen et al., 2012).

Some advancements have been made in understanding the underlying neural correlates for item and order information in STM. A study by Papagno and colleagues (2017) investigated this by observing participants performing a digit span task, while awake during brain surgery. Stimulation of Broca’s area in the left inferior frontal gyrus interfered with span, producing more item than order errors; whereas stimulation of the left supramarginal gyrus produced more order errors (Papagno et al., 2017). The anterior segment of the arcuate fascicle (the nerve tract that connects Broca’s and Wernicke’s area) also produced more order than item errors (Papagno et al., 2017). There is evidence to suggest that there is atypical development of white matter in the arcuate fasciculus in children with reading disabilities (Wang et al., 2017). Although the study used digits, for which it is harder to dissociate item and order errors, it demonstrates that order information might be stored in the supramarginal gyrus (Papagno et al., 2017). These studies support the assumption of a difference between STM for verbal items and temporal and serial order.

Rhythm and STM

Spoken language is driven by rhythm (Langus et al., 2016). It was suggested by Burgess and Hitch (1999) that rhythmic production tasks might impair the encoding of serial order in STM, as they could compete for a common timing signal (Henson et al., 2003). This was compatible with a study by Henson and colleagues (2000) that showed a dorsal premotor brain region was differentially activated as part of serial rehearsal and temporal grouping in STM, the same region has been reported as involved in rhythmic motor finger movements in neuro-imaging studies (Catalan et al., 1998; Halsband et al., 1993). In a behavioural study, Henson and colleagues (2003) examined the interference of a concurrent tapping task on item probe and list probe STM tasks. The researchers predicted that if the same timing signal is responsible for serial rehearsal and rhythmic tapping, the list probe task should show poorer performance when coupled with concurrent tapping (Henson et al., 2003). Concurrent tapping impaired performance on both list probe and item probe tasks, with poorer performance when tapping a complex rhythm (Henson et al., 2003). However, the type of tapping did not have a differential effect on performance in the item and list probe tasks (Henson et al., 2003). Therefore, the results did not warrant the conclusion that serial order and rhythm compete for a common timing signal, although they did show that regular tapping can interfere with STM when externally paced at a different rate (Henson et al., 2003). Part of the observed impairment could express general rather than task-specific demands of dual-versus single-tasking (Henson et al., 2003).

Rhythm, in speech and music, cues the location of structural boundaries, as the ends of phrases are marked by lengthened durations and longer pauses, among other cues (Tierney et al., 2017). Adults with dyslexia show greater inter-tap interval variability compared to controls when tapping to a metronome beat (Thomson et al., 2006). Children with developmental dyslexia show motor sequencing difficulties when asked to tap to a metronome beat (Wolff, 2002), as do children diagnosed with Specific Language Impairment (SLI) (Corriveau & Goswami, 2008). Children with SLI might have subtle neural impairments that affect both motor and language development, and there might be a mechanism that serves the perception and expression of rhythm and timing (Corriveau & Goswami, 2008).

Various theoretical proposals have been suggested for why motor tapping would correlate with measures of literacy and language. One proposal is that a timing mechanism mediates between central nervous system function and coordinated behaviour, suggesting that children with dyslexia might have a dysfunction of this mechanism within a specific time window (Llinas, 1993). On the other hand, it has been proposed that children with dyslexia experience a temporal information processing deficit in reaction-anticipation transitions (Wolff, 2002). There is an additional hypothesis stating there is a developmental connection between auditory rhythmic timing, the accuracy of motor tapping, and language and literacy as the perception and production of exact rhythmic and temporal patterns is an essential part of language acquisition (Thomson & Goswami, 2008). However, an explicit connection between rhythm and language learning has not been made.

In studying early language development, Nittrouer (2006) notes the developmental importance of acoustic changes that emerge from the slow modulations of the vocal tract, which are first produced by infants. It is argued that signal amplitude envelope information plays an important role in language learning (Nittrouer, 2006). Temporal segmentation of the continuous acoustic signal, at the syllabic level, is facilitated by tracking the amplitude envelope rise time (Thomson & Goswami, 2008). Rise time is important for phonological representation as it carries information regarding syllable structure and vowel-like structures in the syllables (Scott, 1998 as cited in Thomson & Goswami, 2008). When we intentionally produce rhythmic speech, we time our motor production of the time of the vowels in stressed syllables (Scott, 1998). It has also been proposed that the “phonological grammar” of a certain language is built upon the distinctive rhythms of that language at the time scale of syllables (Thomson & Goswami, 2008; Port, 2003). Children with dyslexia are significantly less sensitive to amplitude envelope rise time, compared to age-matched controls (Goswami et al., 2002).

Based on the previous studies mentioned above, relationships between motor skills and language tasks might be stronger earlier in development (Thomson & Goswami, 2008). Children with dyslexia have been found to be impaired in tasks involving the discrimination of rise time, but not intensity, which leads to the conclusion that these individuals experience specific difficulties in hearing the rate of change of amplitude envelope at onset, i.e., rise times (Thomson & Goswami, 2008). In general, children who are inconsistent in synchronizing their tapping to a particular rhythm show the poorest literacy and phonological development (Thomson & Goswami, 2008). This indicates a specific difficulty for children with dyslexia in synchronizing motor behaviour to an external auditory rhythm (Thomson & Goswami, 2008). A special link could exist between motor rhythm skills and written language skills, which would possibly be explained by neural impairments that affect both motor and language development (Thomson & Goswami, 2008).

Difficulties in perceiving rise times are related to less efficient processing of prosodic and sub-lexical phonology, and individuals with dyslexia show deficits in prosodic awareness (Goswami et al., 2013; Goswami et al., 2015). Individuals with dyslexia show equivalent phonological similarity effects in STM to those of TD individuals, even though the capacity of their phonological STM is impaired (Goswami et al., 2016; Ramus & Szenkovits, 2008). This phonological deficit in dyslexia involves prosodic levels of phonology, as well as phonemic levels (Goswami et al., 2016). A study by Goswami and colleagues (2016) used a prosodic STM task that was adopted from a phonological STM task (Thomson et al., 2005). Participants were required to recall three or four three-syllable words with either the same or different prosodic stress pattern (Goswami et al., 2016). Participants with dyslexia recalled significantly fewer items correctly than the control participants, even though they exhibited intact prosodic similarity effects, which suggests impaired STM capacity (Goswami et al., 2016). In addition, memory for the most frequent rhythmic stress template was significantly poorer than memory for the least frequent, which suggests top-down redintegration effects

(Goswami et al., 2016). In turn, this suggests that, for prosodic recall, redintegration based on long-term knowledge is more successful for smaller word sets (Goswami et al., 2016). The strongest predictor of prosodic STM was an individual's earlier prosodic awareness (Goswami et al., 2016). This implies that the prosodic patterns of individual words are stored in the mental lexicon and, with more accurate rhythmic stress templates, there is a recall advantage (Goswami et al., 2016). Individuals with dyslexia have atypical prosodic encoding of lexical forms in LTM, which hinders reconstruction processes and leads to poorer short-term recall of multi-syllable and single-syllable words (Goswami et al., 2016).

Rhythm in speech can be seen as alternation of weaker and stronger elements at different levels of the prosodic hierarchy (Langus et al., 2016; Nespor and Vogel, 1986; 2007; Nespor, 1990). There are three universal levels of linguistic rhythm: the segmental level, the level of the metrical feet, and the phonological phrase level (Langus et al., 2016). At the segmental level, rhythm is signalled through the alternation between consonants and vowel-like structures (Langus et al., 2016). At the level of the metrical feet, rhythm is created in the alternation of stressed elements as defined by the Iambic-Trochaic Law (ITL) (Langus et al., 2016). The ITL states that rhythmic alternations of duration are perceived iambically (weak-strong), while alternations of intensity are perceived trochacially (strong-weak) (Henny Yeung et al., 2018; Langus et al., 2016). Lastly, at the level of the phonological phrase (PHPH), words with PHPH stress vary with words that have lexical stress but are weak at the PHPH level (Langus et al., 2016). Rhythm at the segmental level is thought to contribute to word segmentation, therefore helping in word acquisition, and the rhythm at the PHPH is thought to be crucial for word order acquisition (Langus et al., 2016). Lexical stress cannot be used to group syllables into words according to the ITL because lexical stress does not alternate (Langus et al., 2016). Therefore, lexical stress is not rhythmic (Langus et al., 2016).

Iambic grouping is characterized by variation in duration and trochaic grouping is characterized by variation in pitch and intensity (Langus et al., 2016; Nespor et al., 2008). Therefore, in SVO, iambic, languages, duration is the main correlate of PHPH prominence, while pitch and intensity are the main correlates of PHPH prominence in SOV, trochaic, languages (Langus et al., 2016; Nespor et al., 2008). When iambically grouping sequences of speech sounds varying in duration and trochaically grouping speech sounds varying in intensity or pitch, there are discrepancies between speakers of SVO languages and SOV languages (Langus et al., 2016). Italian, SVO, speakers adhere to the ITL, whereas speakers of SOV languages, Turkish and Persian, group syllables alternating in duration trochaically, as they do with syllables alternating in pitch (Langus et al., 2016). However, speakers of SOV languages do not show any deficit in grouping tones of sine wave speech iambically if they alternate in duration (Marino, 2014). Therefore, while speakers of SVO languages use the ITL for grouping linguistic stimuli, the absence of duration as a cue for PHPH prominence in SOV languages causes participants to rely on their native language trochaic prominence (Langus et al., 2016). This implies that perception of rhythm conforms to the knowledge of our native language (Langus et al., 2016).

The ITL might explain how universal aspects of rhythmic perception can bootstrap infants' learning of their native language (e.g., Hay & Saffran, 2012, as cited in Henny Yeung et al., 2018). However, as noted by Langus and colleagues (2016), part of the ITL-like groupings are modulated by one's native language (Henny Yeung et al., 2018). In acquiring a different language-like-phonology, participants who are better at distinguishing between iambic and trochaic, defined by duration, are also better at perceiving more ITL-like rhythmic grouping of duration (Henny Yeung et al., 2018). This again indicates that rhythm plays an important role in temporal grouping.

It has been established that participants who are good at segmenting visual sequences, are also good at remembering (Swallow et al., 2009). Participants are better at remembering sequences of shapes consistent with the visual information of the ITL than sequences that violate the ITL (Pena et al., 2011). Participants are also better at discriminating nonsense sentences that are iambic or trochaic at the PHPH, if they can transfer representations of rhythm between modalities (Pena, submitted, as cited in Langus et al., 2016). While the representations of iambic rhythm acquired through the visual modality of speech are amodal, they are modality specific when perceived in the auditory modality (Langus et al., 2016). This highlights the use of ITL in studying grouping of stimuli across domains, not restricted to linguistic stimuli.

Oscillation and Entrainment of Rhythm. The dynamic attending theory proposed by Large and Jones (1999) was set to explain how listeners respond to systematic changes in everyday events, such as speech, while retaining a general idea of the rhythmic structure. This theory supposes attending as the behaviour of internal oscillators that entrain to external rhythmicity in an auditory signal (Kotz et al., 2018; Large & Jones, 1999). Entrainment to rhythmic stimuli is predicted to involve phase coupling of oscillations in the theta range (3-8 Hz) with gramma-range oscillations around 40 Hz, directing attention to the rhythmic properties of a stimulus (Schoeder & Lakatos, 2009; Giraud & Poeppel, 2012). This coupling of neural oscillations may provide the underlying neural basis of dynamic attending (Henry & Hermann, 2014). When neural rhythms can synchronize to acoustic input, this leads to increased processing and aids speech comprehension (Kotz et al., 2018).

Much of the research relating neural entrainment to rhythm with language abilities has targeted literacy rather than oral language, especially individuals with dyslexia. In neural entrainment to syllables, there is a distinction between children with dyslexia and TD children (Power et al., 2013). When children are asked to press a button when they recognize a delay in the isochronous auditory syllable stimulus, i.e., the sequence where the stimulus is not presented at the same time as the previous stimuli, there is a significant group difference in the preferred phase of entrainment; children with dyslexia show an earlier preferred phase in the delta band to the auditory and audio-visual stimulus streams (Power et al., 2013) This indicates a difference in speech encoding between TD children and children with dyslexia (Power et al., 2013). Therefore, the auditory temporal reference frame for speech processing appears to be atypical in persons with dyslexia, as their low frequency (delta) oscillations entrain to a different phase of the rhythmic syllabic input (Power et al., 2013).

Sensory/neural difficulties in recovering rhythmic structure from speech input could be the main contributor to the phonological deficit that is seen in dyslexia (Goswami, 2016). A temporal sampling theory suggests that auditory sensory impairments present from birth, affect language acquisition from the start, impairing the accuracy of oscillatory entrainment to amplitude modulations slower than 10 Hz, e.g., delta bands, therefore contributing to the impairments found in children with dyslexia (Goswami, 2011; Goswami, 2015; Goswami, 2016, Power et al., 2013). Adults and children with dyslexia have been found to show impaired oscillatory entrainment in the delta band. These effects originated in the primary auditory cortex (Goswami, 2016; Molinaro et al., 2016). Cell networks in the auditory cortex form an oscillatory hierarchy, which mimics the hierarchy of amplitude modulations found in rhythmic speech (Goswami, 2018). It could be that oscillations at ~2 Hz may help identify stressed syllables and those at ~5 Hz may help to identify syllables (Goswami, 2018). Rhythmic patterning of stressed syllables could provide an outline for phonological development (e.g., Goswami, 2015). This helps to explain why the phonological development of individuals with dyslexia is often poor. Impairments in oscillatory entrainment is thought to impair the ability to entrain to rhythm in speech. This provides support for the hypothesis of impaired oscillatory entrainment through development in individuals with dyslexia (Goswami, 2016).

In STM research, a new model to account for serial order was suggested by Hartley and colleagues (2016): the bottom-up multi-scale population oscillator model, i.e., the BUMP model. The BUMP model assumes that “during presentation, each item is associated with the current state of a temporal context signal, [and] at recall, successive states of the temporal context are replayed, and the learned associations reactivate the items" (Hartley et al., 2016). To test the model, the researchers studied STM for predictable temporally grouped lists, and unpredictably grouped structures. They included an articulatory suppression condition to test the role of subvocal rehearsal. In the no suppression condition, for both predictably and unpredictably grouped lists, grouping increased recall accuracy, primacy, and recency effects within groups, reduced transpositions between temporal groups, but increased interposition errors, i.e., exchanges between items in similar positions within different groups. Grouping effects were found also for the unpredictable list, which the researchers suggest as highlighting that advance knowledge of grouping structure is not necessary for grouping. The researchers concluded that this favoured the view that temporal grouping effects are attributable to a bottom-up mechanism (Hartley et al., 2016).

In a second experiment, the researchers examined the effect of foreknowledge of the grouping pattern for a range of irregular patterns beyond groupings of threes (Hartley et al., 2016). Neither the main effect of list-type, nor the interaction between list-type and serial position were significant, which suggests that foreknowledge of the grouping structure did not benefit serial recall. The researchers state that BUMP provides an account of the mechanisms underlying grouping effects and explains that by exploiting temporal structure in the input sequence, the capacity of auditory-verbal memory can be extended. This was interpreted as suggestive of a competitive queuing mechanism. However, Baddeley and Hitch (2019) note that a shortcoming of the BUMP model is that it can only be applied to auditory sequences. Nevertheless, this study demonstrated how immediate memory for spoken sequences depends on their rhythm through temporal grouping effects.

Precise synchronization to a rhythm requires the detection and correction of small auditory-motor asynchronies, relying upon temporal processing on a fast time scale (Tierney et al., 2017). On the other hand, perception of entire rhythmic sequences requires the integration of rhythmic information across time, relying upon temporal processing on a slow time scale (Tierney et al., 2017). It is hypothesized that system-wide pathways exist to specialize for faster and slower auditory processing (Tierney et al., 2017). In addition, in humans, the fast frequency-following response (FFR), which indicates precise temporal encoding across the auditory system, reflects faster processing whereas the cortical evoked response to sound reflects slower processing (Tierney et al., 2017). As noted in previous studies (e.g., Goswami et al., 2013; Goswami et al., 2016; Thomson & Goswami, 2008), deficits in the perception of nonverbal rhythms appear to have an effect on language skills (Tierney et al., 2017).

Tierney and colleagues (2017) sought to investigate how individual differences in rhythm skills could explain variation in language abilities. Participants completed a perceptual test, various cognitive and language tests, as well as multiple rhythm tasks: a synchronization task; a tempo adaptation task; a timing adaptation task; a beat synchronization task; drumming along to sequences; as well as a sequence memory task to measure participants' ability to remember metrical sequences. Participants also underwent an electrophysiological recording to investigate their perception of sound through elicited FFRs and evoked cortical responses. Among the behavioural measures, there was a correlation between the synchronization, tempo adaptation, and timing adaptation tasks, and a correlation between sequence memory and drumming to sequences. The performance on the beat synchronization task was also significantly correlated with performance on all rhythm tests, except for the sequence memory task. This indicates two separate factors: a synchronization factor consisting of timing adaptation, synchronization, beat synchronization, and tempo adaptation, and a sequencing factor, consisting of drumming to sequences and sequence memory. It was also found that participants who were better at synchronizing had more consistent FFRs but not more consistent evoked cortical responses. Participants who were better at perceiving and remembering rhythms, had evoked cortical responses that were more consistent across trials. However, the FFRs did not correlate with linguistic or cognitive skills. Also, although rhythm sequencing was linked to the consistency of slow cortical responses, the researchers did not find this cortical consistency to be directly linked to verbal memory and reading. The behavioural link between rhythm sequencing and verbal memory could suggest that music training, which incorporates the memorization of rhythms, could benefit memory for verbal material as well (Tierney et al., 2017).

Music and Speech Perception. There are various components underlying rhythmic behaviour in humans, which include periodic motor pattern generation, beat extraction from complex auditory patterns, entrainment of motor output to this beat, and meter perception (Kotz et al., 2018). Healthy adults perceiving music and speech engage partly overlapping neural regions (e.g., Koelsch et al., 2002; Tillmann et al., 2003) and syntactic violations in speech and music elicit similar P600 responses in the brain (Hausen et al., 2013; Patel et al., 1998). Amusia and aphasia, disorders of music and speech perception/expression, respectively, have been generally seen as independent deficits as there are cases of amusia without aphasia and vice versa (see Peretz and Coltheart, 2003 for a review). In patients with non-fluent aphasia, music has been shown to be effective in language rehabilitation (Racette et al., 2006; Schlaug et al., 2010; Stahl et al., 2011), which suggests a link between the processing of speech and music in an impaired brain (Hausen et al., 2003). As previously mentioned, patterns of stressed and unstressed tones or syllables, i.e., rhythm, build up the hierarchical structure of both music and speech (Hausen et al., 2003; Scott, 1998; Tierney et al., 2017). A study by Hausen and colleagues (2003) focused on the role of acoustic differences in the perception of word stress. The results showed that rhythmic cues, particularly time and stress, provide a link between music and speech perception (Hausen et al., 2003).

Previous research has demonstrated that cortical activity in music and speech domains phase-locks (Doelling & Poeppel, 2015; Doelling et al., 2014; Luo & Poeppel, 2007), or entrains (Buiatti et al., 2009; Nozaraden et al., 2011; Nozaraden et al., 2012) to rhythm, but only one study compared these domains in the cortical tracking of rhythm (Harding et al., 2018). A number of studies have demonstrated that musical training improves auditory perception of temporal features in music (e.g., Geiser et al., 2010 as cited in Harding et al., 2018) and speech (e.g., Marie et al., 2011, as cited in Harding et al., 2018). However, there is evidence suggesting that cortical responses to music rhythm, but not for speech rhythm, increase with the years of music training (Harding et al., 2018). A study by Harding and colleagues (2018) measured this by examining cortical responses to stimulus envelopes in participants with a range of music training. The speech rhythm could have been easier than the music rhythm to detect, as it was not as complex as natural speech (Harding et al., 2018; Meninghuas et al., 2015; Patel, 2011). The asymmetric response associated with music training indicates active top-down adjustment of entrainment to rhythm based on listening conditions and task demands (Harding et al., 2018).

ERP Studies

Event-related potentials (ERPs) are the neural responses to specific events recorded through an electroencephalogram (EEG) (Luck, 2014). ERPs can be used to study brain responses to various types of stimuli, such as music and speech. ERP markers of sensitivity to rhythm in speech perception has been studied using sequences of bisyllabic words, where stress of the final word either matched or mismatched the stress pattern of the previous words (Magne et al., 2016). Words with unexpected stress patterns elicited an increased frontal-central mid-latency negativity, ranging from 288 to 576 ms. In addition, rhythm aptitude correlated with the size of the negative effect elicited by words with unexpected iambic rhythm. This supports the idea that rhythm in speech and music share neurocognitive resources.

One well-studied ERP response is the mismatch negativity (MMN), which is a fairly automatic response to an auditory stimulus that differs from the preceding train of stimuli (Luck, 2014). It peaks around 100-250 ms (Brückmann & Garcia, 2020; Luck, 2014; Näätänen et al., 2007; Schaadt et al., 2014; Schwade et al., 2017), particularly at the fronto-central midline of the scalp (Luck, 2014). The MMN relies on the presence of a STM trace in the auditory cortex of the preceding auditory events (Näätänen et al., 2007). Many studies focus on MMN responses as markers of stimulus discrimination (see Näätänen et al., 2007, for a review). For example, a study done by Brückmann and Garcia (2020) highlighted a larger MMN response to a verbal contrast (syllables /da/ and /ta/) than to a tonal contrast (tones of 750 Hz and 1,000 Hz). There are also studies focusing on the temporal grouping of auditory stimuli, demonstrating that in the presence of multiple stimulus streams, the grouping and integration processes to perceptual units take place modularly within the already formed streams (Sussman, 2005). This signifies that speech intelligibility when listening to two or more speakers is supported by automatic within-channel grouping (Näätänen et al., 2007). When there are paired and unpaired tones, the second deviant of the unpaired tones evokes a larger MMN amplitude than that in the unpaired condition (Müller & Schröger, 2007).

Literacy training has been demonstrated to elicit a larger MMN response in individuals with reading disabilities discriminating auditorily presented phonemes (Schaadt et al., 2014). This reinforces the importance of the relationship between phonological awareness and reading skills (Ziegler & Goswami, 2005; Goswami, 2018). WM load and WM performance scores have also been shown to affect the MMN response. When the WM load increases, so does the amplitude of the MMN, which suggests that a more engaged change detection process is engaged under higher WM load conditions (Lv et al., 2010). As well, by inserting acoustic features into a tone sequence, a significant correlation was found between frontal MMN amplitudes and WM performance (Bonetti et al., 2017). This is suggestive of a link between conscious WM abilities and the STM storage of environmental consistencies and pre-attentive sensory-specific neural skills of prediction (Bonetti et al., 2017). To our knowledge, no studies to date have investigated the relationship between the amplitude of the MMN response and STM tasks involving temporal and serial order.

Purpose

This thesis describes an experiment exploring temporal order in language learning. The ultimate goal is to explore the relationship between rhythm and order coding in STM and how that predicts language learning. We investigated this topic through verbal and non-verbal STM tasks, as well as through an MMN oddball paradigm.

We used tasks that measured verbal and non-verbal STM: a temporal rhythm accuracy (TRA) task, two sentence repetition tasks (an English nonword sentence repetition task and a foreign sentence repetition task), and two TOJ tasks. We used these tasks to predict performance in a foreign-word learning (FWL) task. We included the TRA task to study how well one replicates a series of tones differing in length by tapping. This was included because Tierney and colleagues (2017) noted that rhythm sequencing performance was linked to language skills such as reading and verbal memory. Therefore, we predicted a correlation between the TRA task and language learning.

Pseudoword repetition tasks have been shown in many studies to correlate with word learning in children, but they are often too easy for adults. We created the English nonword sentence repetition task to examine participants' ability to code and retrieve familiar phonological information in verbal STM (Schraeyen et al., 2018; Szewczyk et al., 2018). The foreign sentence repetition task is a new task that is comparable to the English nonword sentence repetition task. We used Turkish as the foreign language. In SVO languages, Head-Complement languages, such as English, the PHPH is characterized by PHPH stress on the final object with a durational contrast (Gervain, 2018; Langus et al., 2016). In SOV languages, Complement-Head languages, such as Turkish, the PHPH stress is on the initial object (Gervain, 2018) and is realized as increased pitch or intensity (Langus et al., 2016). It has been shown that 6-12-week-old infants can discriminate between phrasal stress in SVO and SOV languages, even if the two languages are similar with regards to syllabic structure and word primary stress (Langus et al., 2016). Therefore, due to Turkish being a foreign language to participants and having a different rhythmic pattern to that of English, it allows us to see how well participants are able to repeat back syllables of words they have not heard before in an unfamiliar prosodic context.

How well one can repeat back phonologically possible words is a good indicator of one's language-related abilities, and performance on nonword repetition tasks has shown to be an important predictor of novel word learning in native and foreign language acquisition (Baddeley et al., 1998; Gathercole & Baddeley, 1990; Gathercole et al., 1997; Service, 1992; Service & Craik, 1993; Service & Kohonen, 1995; Szewczyk et al., 2018). Whereas previous studies have used single nonwords in repetition tasks (Gathercole & Baddeley, 1990; Schraeyen et al., 2018; Service & Craik, 1993; Szewczyk et al., 2018), this study used nonword sentences. The rationale behind using sentences for the repetition tasks is that the sentences contain more of the language's prosody, than do individuals words, and thus should better support learning of longer phonological sequences with more phonological information than contained in single words. The English nonword repetition task should test phonological memory with a familiar prosodic context whereas the Turkish sentences should test phonological memory in the absence of familiar prosody, but still structured prosody. The foreign sentence repetition task therefore is designed to be more taxing on phonological memory. Both sentence repetition tasks allow us to investigate how one codes order with and without the use of familiar prosody in verbal STM.

The TOJ tasks were taken from a study by Lahti-Nuuttila and colleagues (submitted) to investigate nonverbal STM for item serial order. The stimuli were not words but sequences of visual fantasy animals or auditory animal sounds. This task was similar to the Item STM task used by Laasonen and colleagues (2012) where participants were asked to match pairs of sequences of increasing stimulus lengths. This task was designed to "access the ability to incrementally update an increasingly complex temporal pattern in a similar manner that a verbal sequence would be incremented by the presentation of additional words in a list" (Laasonen et al., 2012). The TOJ tasks were also similar to the list probe task used by Henson and colleagues (2003) that encourages forward serial processing, since the participants are comparing serial items in the probe against their memory for the original list. The auditory and visual TOJ tasks were included due to research showing that memory for serial order and memory for items involve separate processes (Henson et al., 2003). In addition, Hurlstone and Hitch (2014) found that spatial and verbal STM rely on similar, possibly common, mechanisms for the representation of serial order. These tasks also included articulatory suppression, as it has been shown to prevent subvocal rehearsal and have an adverse effect on STM (Henson et al., 2003; Service & Craik, 1993). Therefore, the TOJ tasks presented here hope to show correlations with the verbal tasks.

The FWL task was adopted from Service and Craik (1993), but instead of Finnish words and English pseudowords, we used Turkish words and their English counterparts. We used real words in an unfamiliar language, as it has been found that nonwords or unfamiliar words are significantly harder to recall than familiar words (Goswami et al., 2016; Roodenrys et al., 2002). If we had used Turkish nonwords, they may have been too unnatural and difficult for participants to recall, or the phonology could have been too similar to that of English. Therefore, Turkish words were employed to limit support from native language phonological representations. Together, these tasks allow us to investigate the effect of verbal and non-verbal STM on foreign-language learning.

Lastly, we investigated MMN responses to deviant syllables in an oddball paradigm with a train of evenly spaced standard syllables. The deviant was either presented earlier than expected based on the standards or the initial consonant and the vowel differed from the standard syllables. We used syllables instead of tones because we were examining language learning and therefore preferred to use language stimuli. We correlated the amplitude of the MMN responses, as measures of timing expectancy, with the STM tasks. MMN amplitudes have previously shown a correlation with STM measures (e.g., Bonetti et al., 2017).

The hypothesis was that one's ability to reproduce verbal and non-verbal temporal order would translate to their ability to form mental representations of variation in speech. Specifically, we first predicted that the better a person is at serial order tasks, the better they will perform in the FWL task. Second, we predicted that the better one performs on the STM tasks, the larger their MMN response will be to the deviant stimuli. Additionally, we explored whether the larger a person's MMN response to the deviant stimuli, the better they would be at learning foreign words and the better their temporal accuracy. Lastly, we explored whether music experience has an effect on STM task performance.

METHODS

Participants

Thirty participants aged 18-35 years-old participated in this study (21 females, 19 right-handed, 1 left-handed, 1 ambidextrous; M_{age} = 20.67 years, SD = 5.72 years, and 9 males, 8 right-handed, 1 left-handed; M_{age} = 20.04 years, SD = 7.32 years). Six participants (4 female, 2 male) were removed from the final EEG analysis due to noise in the data, resulting in thirty participants for the behavioural analyses, and twenty-four participants for the EEG analyses (17 females, 15 right-handed, 1 left-handed, 1 ambidextrous; M_{age} = 20.49 years, SD = 6.3 years, and 7 males, 6 right-handed, 1 left-handed; M_{age} = 19.06 years, SD = 7.94 years). At McMaster University, we recruited participants through the Linguistics and Languages Research Participant Pool and the Psychology, Neuroscience, and Behaviour (PNB) Research Participant Pool. Our participants consisted of native English speakers with no knowledge of Turkish, normal or corrected-to-normal vision, and normal hearing. Eligibility criteria also included no history of neurological disorders/disease (e.g., head injury/concussions, tumours, etc.), and no consumption of any medication that affects the nervous system (i.e., anxiety, depression medication, etc.). The mentioned eligibility criteria were in place to maintain the quality of the EEG data recorded; to prevent any confounds. Participants were also screened for their music experience: whether they played or currently play an instrument, type of instrument(s), years of playing, whether they were self-taught or professionally taught, and they were asked to rate themselves as either beginner, intermediate, or advanced (**Table 1**). All participants provided informed consent before beginning the experiment and were compensated for their participation via course credit or cash. This study's protocol was approved by the Hamilton Integrated Research Ethics Board (HiREB).

Table 1: Mean and standard deviants of participant's music experience, i.e., years of playing. Participants were grouped into instrument type, level of music ability, and whether they were self-taught of professionally taught. *N = 29*.

Instrument	**Beginner Self-Taught**	**Beginner Professionally Taught**	**Intermediate Self-Taught**	**Intermediate Professionally Taught**	**Advanced Professionally Taught**
Piano (N=14)	M=2, SD=2.64	M=2.33, SD=2.08	N/A	M=8.1, SD=5.19	M=14, SD=0
Strings (N=16)	M=1.75, SD=1.95	M=3, SD=2.34	N/A	M=8, SD=5.37	M=14, SD=0
Wind (N=8)	M=1.94, SD=2.20	N/A	N/A	M=8.2, SD=6.87	M=14, SD=0
Brass (N=3)	N/A	M=0.66, SD=0.577	N/A	N/A	M=14, SD=0

Percussion (N=3)	N/A	M=2, SD=1.73	M=9, SD=0	N/A	N/A

*Note: one participant did not answer whether they were self-taught or professionally taught, so they are excluded from this table and from analyses involving music measures.

Tasks and Stimuli

Working Memory Tasks. For the behavioural component of the experiment, there were four STM tasks performed on a lab computer and two non-verbal TOJ tasks performed on a lab tablet. Participants listened to the stimuli through Sony MDRZX110NC noise-cancelling headphones. For scoring purposes, we recorded all participant verbal responses to all the phonological STM tasks. Two participants were recorded with a Yeti microphone with a frequency response of 20 Hz-20 kHz, 16-bit range, maximum SPL of 120dB, and a sample rate of 48 kHz. Due to technical difficulties with the Yeti microphone, we switched to using a NT-USB RODE microphone for the remainder of the participants. The NT-USB RODE microphone had a frequency range of 20 Hz-20 kHz, a dynamic range of 96dB, a maximum SPL of 110dB, and a cardioid directional pattern. This microphone had a 3.5 mm stereo jack that connected directly to the lab computer.

Phonological STM Tasks. There were two auditory phonological STM serial order tasks: an English nonword sentence repetition task and a foreign sentence repetition task. In the English nonword sentence repetition task, there were 10 sentences. Each sentence consisted of 5 to 6 word-like stimuli and ranged from 6 to 8 syllables. Function words were real English words, e.g., “the”, “a”, and “some”. The nouns and verbs had been replaced by nonwords that consisted of various consonant clusters and vowels

(**Table 2**). In the foreign sentence repetition task, there were also 10 sentences. The sentences consisted of 3 to 4 Turkish words and ranged from 6 to 7 syllables (**Table 3**).

Table 2: English nonword sentence repetition task stimuli.

Sentence	**Number of Syllables**	**Number of Words**
The muthin faned a grasphit.	7	5
A dreg pribed in the blantis.	7	6
These smorkets nooled a guzdin.	7	5
The flogin janed to a kinto.	8	6
Some flotters peft their laumses.	7	5
A pondle boyed to the trompers.	8	6
The nolk winted the priles.	6	5
Some stordins rimple with mipters.	8	5
These gilperns karnayed the wups.	7	5
A jowler frined with a vulkit.	8	6

Table 3: Foreign sentence repetition task stimuli.

Sentence	Number of Syllables	Number of Words
Sana cicek aldim.	6	3
Sende bu topla oyna.	7	4
Kopek eve kacti.	6	3
Ugur bahceye atladi.	7	3
Kediyi guzel yedir.	7	3
O ati bana al.	6	4
Can kiza yemek verdi.	7	4
Bu kelebek cok guzel.	7	4
Siz arabaya binin.	7	3
Hepsi et yediler.	6	3

Rhythm STM Task. The stimuli in the rhythm task consisted of tones, presented for a TRA task. There were 10 tone sequences, each consisting of the same 7 tones. Each sequence consisted of varying short and long tones. The tones were created in the software program Audacity. The orders of short and long stimuli in the sequences were first randomized to be all different from each other. In Audacity, the tones were generated by selecting one default frequency of 525 Hz. The short tones were 200 ms in length and the long tones were 800 ms in length. The interstimulus interval (ISI) duration was kept constant at 200 ms. The tones were then pasted into the order created previously to produce 10 audio tracks. The participants' task was to listen to the sequences and tap the order of short and long stimuli on a computer keyboard. They were given 8000 ms to repeat all the tones in each sequence before moving on to the next tone. This task was then programmed on the same Superlab 5.0 experiment creator program and Apple iMac computer as the phonological tasks mentioned above.

Foreign-Word Learning Task. Language learning was probed with an auditory FWL task consisting of English-Turkish word pairs. This task was adopted from Service and Craik (1993), but instead of English words being paired with Finnish words and English pseudowords, we presented English words with Turkish words. The pairs were not real translation equivalents.

The tasks were programmed in SuperLab 5.0 (Cedrus Corporation) on an Apple iMac computer with a 21.5-inch screen with a resolution of 1920x1080 pixels. The English stimuli were recorded by a female native English speaker. The foreign stimuli were recorded by a female native Turkish speaker. The English and foreign stimuli were recorded in a soundproof room with a linear response Sennheiser ME62 omnidirectional capsule mounted on a Sennheiser K6 body that was connected directly to a Focusrite Scarlett microphone amplifier, which connects digitally to the computer. The microphone was placed 10 cm in front of the mouth in the center of the sound-treated room, slightly off-center to avoid plosive noises. The software used for recording was Soundforge Pro. The signal was high-pass-filtered at 150 Hz (which filters out unwanted ambient noises) and recorded as standard 44100 Hz, 16 bits mono wav files.

The foreign-word learning task in this study consisted of monosyllabic English words, with the paired Turkish words being 3 syllables. The lexical stress fell either on the second or third syllable of the Turkish word. There were two lists, list A and B, each consisting of six-word pairs (**Table 4**).

Table 4: Foreign-Word Learning task stimuli. The bolded syllables indicate stress.

List A	**List B**
Hand - Lo**kan**ta	Crab - Dondur**ma**
Girl - E**niş**te	Door - Kalem**lik**
Plum - Sepet**ler**	Feet - Sal**yan**goz
Nose - Ka**ra**yel	Rock - Mayda**noz**
Scar - Ta**ban**ca	Silk - A**nah**tar
Boat - Arka**daş**	Pond - Çe**kir**ge

Temporal Order Judgment Tasks. This task was a tablet program taken from the study by our Finnish collaborators, Lahti-Nuuttila et al. (submitted). This task consisted of two non-verbal STM tasks in the visual and auditory domains, where lengthening pairs of stimuli were presented for comparison. The participant had to indicate on a touch screen whether two sequences had the stimuli in the same or a different order. In the original study with child participants, the task was a custom-created application program based on the Unity game engine (Lahti-Nuuttila et al., submitted). For the present study, the task was downloaded onto a Samsung Galaxy Tab A (2016) Tablet, model SM-T585 version with a 10.1-inch WXGA display, and a resolution of 1200x1920 pixels.

Both modalities, visual and auditory, were presented as a game to the participants. For the visual modality, two pairs of barns facing each other were pictured on the screen in daylight. A sequence of animated animals travelled from the top left barn to the top right barn (**Figure 1**). Then, another sequence of animated animals travelled from the bottom left barn to the right barn (**Figure 2**). The animals were not based on real animals. All animals were created from the same thirteen shapes but differed in the position and proportions of the features. The animals also differed in the way they moved across the screen. Only binary sequences were presented so that each sequence consisted of two different animals from the pool of five possible stimuli. The animals were equally spaced from one another and moved at the same speed. Only one animal was seen at a time for approximately 1500 ms (Lahti-Nuuttila et al., submitted).

Figure 1: An example of the visual domain of the tablet when the first sequence is presented.

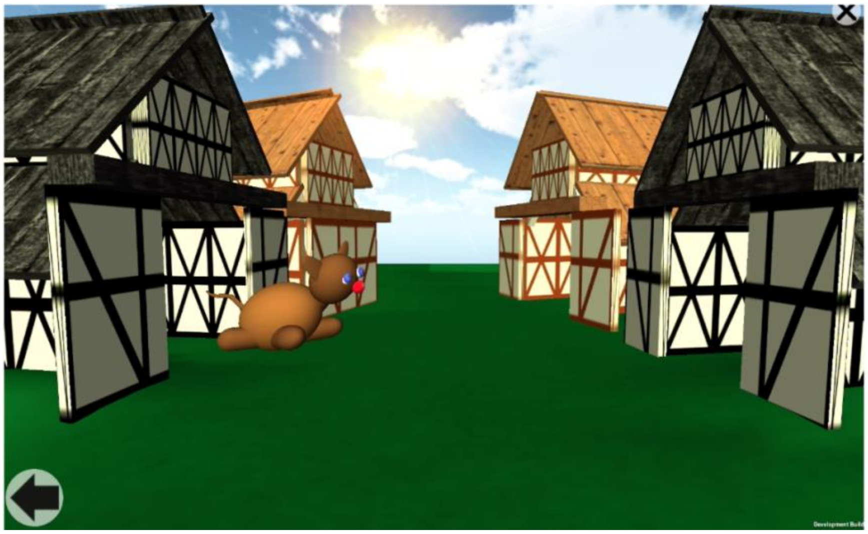


Figure 2: An example of the visual domain of the tablet when the second sequence is presented.

In the auditory modality, the same two pairs of barns as seen in the visual modality were pictured on the screen, but with the picture dimmed to create a nighttime scene. The auditory stimuli were chosen from five different sound files that consisted of animal sounds played backwards. The top right barn lit up and the first sequence of animal sounds was presented (**Figure 3**). Then, the bottom right barn lit up and the second sequence of animal sounds was presented (**Figure 4**). The sounds lasted approximately 1500 ms and were spaced equally between one another (Lahti-Nuuttila et al., submitted).

Figure 3: An example of the auditory domain of the tablet when the first sound sequence is presented.

Figure 4: An example of the auditory domain of the tablet when the second sound sequence is presented.

Both modalities consisted of several trials. There were 6 sequence lengths (2–7), with 6 sequences of animals per sequence length. Before the first test trial of each modality, participants were presented with five practice trials (see **Figures 5 & 6**).

Figure 5: An example of the practice trials for the visual domain.

Figure 6: An example of the practice trials for the auditory domain.

The first three pairs of sequences in the practice trial consisted of two stimuli and the two remaining pairs of sequences consisted of three stimuli. Depending on the modality, the tasks started with two animal-like characters or animal-like sounds in the sequence and increased in length, to a maximum of seven, as the participant continued. In each round, there were six pairs of sequences; three comparisons where the stimuli in the sequence were in the same order and three comparisons where the stimuli were in a different order. The order of same and different sequences was pseudorandomized in the game and was the same for all participants. For the trials that were the same, the sequences were the same. For the trials that were different, two different consecutive stimuli had changed places in the second sequence. When the blocks consisted of four or more stimuli, the switched stimuli were always in the n-2 middle position, never including the first or last stimulus. Before the first stimulus, there was a two-second onset asynchrony. The two sequences were presented one after another with a 3-second interval between the last stimulus of the first sequence and the first stimulus of the second (Lahti-Nuuttila et al., submitted).

Once both sequences had been presented, participants were to indicate whether the two sequences were the same or different, that is, to indicate whether the animal-like characters (visual), or the animal-like sounds (auditory) in the sequences had been in the same order. Participants were asked to press the green circle with ✓for YES and the red circle with ✕ for NO (see **Figures 7** & **8**). If the participant answered correctly to at least four out of six pairs, they moved onto the next sequence length with an additional stimulus per sequence. If the participant got three or more comparisons wrong in a sequence length, the game was terminated (Lahti-Nuuttila et al., submitted). Participants were unaware of how far along they got in the game.

Figure 7: An example of the visual domain where the participant presses the green or red circle to indicate whether the animals were in the same order or not.

Figure 8: An example of the auditory domain where the participant presses the green or red circle to indicate whether the animal sounds were in the same order or not.

Event-related potential (ERP) task. The EEG paradigm for eliciting the MMN was an MMN oddball paradigm with auditory syllable stimuli adopted from Rana et al. (in preparation). The paradigm is referred to as an "oddball" paradigm because there are infrequent auditory stimuli (deviants) that are presented periodically among other frequent stimuli (standards) (Yue et al., 2014). The syllables were generated in the software VocaltractLab 2.1 using the new glottal model (the website for this software can be found at https://vocaltractlab.de/). All syllables were plosive-vowel sequences and were created to have ~50 ms VOT. Then, they were adjusted on Praat to a total length of 150 ms (~50 ms per consonant, ~100 ms per vowel, with 5 ms rise and fall times) and normalized using SoundForgePro, version 3.0 for Mac OsX. There were a total of 1000 tokens presented in total per block, with 900 frequently repeating standard stimuli and 100 infrequent deviant stimuli. There were two blocks: in one block there were 3 variants of standard syllables presented with a syllable onset asynchrony (SOA) of 650 ms: /ti/, /tu/, /tə/, and a syllable deviant beginning with a different consonant, /ka/, with SOA of 650 ms; the second block consisted of the same standard syllables with early deviants (one of /ti/, /tu/, or /tə/) occurring after a SOA of 250 ms instead of 650 ms. Therefore, the deviant of interest was an early response. The stimulus order was pseudorandomized before presentation with the restriction that a minimum of 15 standards be presented at the start of the block and a minimum of 2 standards be presented in a row throughout the block.

Procedure

We conducted the behavioural and EEG components of the study during a single session. All testing took place at the ARiEAL Research Centre at McMaster University. The entirety of the experiment took approximately 2 hours to complete: ~1 hour for the behavioural component (~15 minutes for the sentence repetition tasks, ~5 minutes for the TRA task, ~10 minutes for the FWL task, ~30 minutes for the TOJ task), and ~1 hour for the MMN component (~20-minute paradigm, plus set-up). All participants completed the behavioural tasks of the experiment prior to the MMN experiment.

STM and Word Learning Tasks. We decided to counterbalance the order of the sentence repetition tasks, the TOJ modalities, as well as the two lists of the FWL. All participants completed the behavioural tasks in the same order. The only variant was the order of the word lists. Participants started with the English nonword sentences, then the TRA task, then the foreign sentence repetition task, and finally the English–foreign word pair lists. For each task, participants began with three practice trials. Participants were able to take breaks between each task.

Phonological STM Tasks. In the English nonword sentence repetition task, a fixation cross was presented on the screen for 500 ms followed by the word “LISTEN” for an average of 2450 ms and then the recorded sentence was auditorily presented. After hearing each sentence, the word “REPEAT” would appear on the screen and the participants would repeat aloud what they had heard. If they could not remember a word, they were to say “blank”. The word “REPEAT” remained on the screen as long as was needed for each participant, as they pressed the spacebar to continue on. There were 10 sentences in total. Participant responses were phonetically transcribed using broad IPA transcription on a scoring sheet.

In the foreign sentence repetition task, a fixation cross was presented on the screen for 500 ms followed by the word “LISTEN” for an average of 2682 ms and then the recorded sentence was auditorily presented. When the word “REPEAT” appeared on the screen, the participants were asked to repeat what they had heard. If they could not remember a word, they were asked to say “blank”. The word “REPEAT” remained on the screen as long as was needed for each participant, as they pressed the spacebar to continue on. There was a total of 10 sentences. Participants were told that these sentences were from a foreign language, this was to be consistent with the English nonword sentence repetition task. Participant responses were phonetically transcribed using broad IPA transcription on a scoring sheet.

Rhythm STM Task. For the TRA task, again, a fixation cross was presented on the screen for 500 ms followed by the word “LISTEN” for an average of 5152 ms and then a sequence of short and long tones was auditorily presented. Once the word “REPEAT” appeared on the screen for 8000 ms, the participants were to repeat the sequence by tapping a short or long press on the spacebar for each short and long tone, respectively. There were 10 tone sequences in total. The computer program registered the down and up times of the space bar presses, allowing the recording of press durations.

Foreign-Word Learning Task. English words were presented with the foreign words. Participants were told that they would hear English words paired with foreign words. There were 6 words in each block, and the participant heard each word 4 times. They would hear the first list of paired words and then right after there would be a test phase. An ‘X’ would appear on the screen for an average of 3328 ms, and participants would hear the English word, then when a ‘#’ appeared on the screen, participants were asked to say the foreign word equivalent. The “#” remained on the screen as long as was needed for each participant, as they pressed the spacebar to continue on. After four repetitions of the first list, the second list of words was presented, following the same procedure. Participant responses were phonetically transcribed using broad IPA transcription on the scoring sheet.

Temporal Order Judgment Tasks. Participants wore Sony MDRZX110NC noise-cancelling headphones when completing the tasks in both modalities, so as to limit noise distractions. Participants were instructed to perform articulatory suppression by saying "la la la" for the duration of the task. This was to ensure that participants were not using any phonological memory tactics for the visual and auditory modality. This task was taken from a study of children aged 4-6 (Lahti-Nuuttila et al., submitted), so the articulatory suppression was to increase difficulty as well as to prevent adult participants from relying on their phonological STM. Participants were told they could stop saying "la la la" when answering YES or NO to the sequence order, to allow for a break.

ERP task. A 64-electrode cap was used for each participant. Five external electrodes were used: two were placed on the left and right mastoids; one beside the left eye for horizontal eye movements; one above the left eye for vertical eye movements and blink artifacts; and one on the tip of the nose as a reference for all electrodes (Duda-Milloy et al., 2019). Conductive gel was placed between the scalp and each electrode site to maximize signal for recording. Participants were given Etymotic ER-1 Insert Earphones to hear the auditory stimuli. Due to technical issues, only one of the earphones worked for this experiment. Previous literature suggest that ear of presentation has minimal effect on the MMN brain response (Duda-Milloy et al., 2008; Grimm et al., 2008, Schwade, Didoné, and Sleifer, 2017), so we expected that the monoaural presentation of the stimuli would not affect how participants perceive the stimuli or affect the MMN response itself. Participants listened to the auditory stimuli while watching a silent Discovery Channel nature video on the screen. As it was an MMN oddball paradigm, participants were asked not to pay attention to the stimuli and to minimize movement and eye-blinking. The order of the two blocks was counterbalanced, each ~10 minutes long with a break in between if the participant needed one. Data was recorded through Active View. The high-pass filter was set to 0.01 Hz, the low-pass at 100 Hz, and the decimation to ¼.

Data Processing

Phonological Tasks. The data were first pre-processed in a MS Excel spreadsheet. For the sentence repetition tasks, the data were scored both by the proportion of correct words and proportion of correct syllables per sentence. For the FWL task, the word-pairs were pre-processed in an Excel sheet that separated each repetition. The scores in the two lists were averaged together to form one score. Each repetition was scored for the proportion of syllables that were correct and for the correctness of the stressed syllable. A syllable was scored as correct if the syllable and all of its phonemes were recalled in their correct serial order. All scores were turned into z-scores. The data were then passed into the Jamovi 1.1.9 statistical software package (The Jamovi project, 2020) for further analysis.

Rhythm STM Tasks. For the tone sequences, the data were pre-processed in a MS Excel spreadsheet that examined the proportion of correct tone lengths per sequence. Short tones were scored as correct if the participant pressed the key down for less than 500 ms, and the long tones were scored as correct if the participant pressed the key down for longer than 500 ms. All scores were turned into z-scores. The data were then imported into Jamovi (The Jamovi project, 2020) for further analysis.

Temporal Order Judgment Tasks. For both the visual TOJ and the auditory TOJ task, we exported the data from the Samsung tablet into a MS Excel spreadsheet and computed the proportion of the maximum score per trial per participant. All scores were turned into z-scores. The data were then exported to Jamovi (The Jamovi project, 2020) for further processing.

ERP task. Raw EEG data was processed before further analyses were conducted. Using Brain Vision Analyzer (BVA) TM EEG processing software, the raw EEG data was re-referenced, re-sampled to 512 Hz, low pass filtered to 0.1 Hz, and high pass filtered to 30 Hz. The data was visually inspected for artifacts introduced to noise and if a channel contained a high level of noise, the channel was interpolated. If more than four channels needed to be interpolated, the data were rejected (Duda-Milloy et al., 2019). This was the case for one participant. Independent Component Analysis (ICA) was used to identify and remove artifacts introduced by eye movement and blinks. If more than 20% of deviants were removed during artifact rejection, the dataset was rejected. This was the case for five participants. Data were segmented by the relevant condition, early deviant block or syllable block, and then baseline corrected using activity 100 ms before stimulus onset (-100 to 0). The MMN peak detection window was set to 100 ms - 250 ms (Brückmann and Garcia, 2020; Näätänen et al., 2007; Schaadt, Pannekamp, and Meer, 2014; Schwade, Didoné, and Sleifer, 2017). Based on the target syllables and expected time latency of the MMN, the data was further segmented. Epochs were 600 ms long in total, from -100 ms to 500 ms. Once data was pre-processed and segmented for all participants, grand averages were obtained by averaging segments across all participants. ERPs were obtained based for the different types of stimuli (the /ka/ syllable and early deviants), as well as across scalp location. For the early deviant stimuli, the data were visually inspected to identify the MMN response and compared to responses obtained for the /ka/ control syllable deviant and for the frequent standard stimuli obtained from the 64 electrodes. Grand averages and MMN peaks from BVA TM can be viewed in Appendix A. Further quantitative analyses will be conducted on latency and amplitude data exported from BVATM into statistical analyses scripted in R Studio.

Statistical Analyses

Interrater Reliability for Phonological STM Tasks. We ran Shapiro-Wilk tests to check if the data were normally distributed. Although the scores for the English nonword sentences were not normally distributed, a Pearson's r correlation test produced similar results to the non-parametric Spearman's rho test. We computed Pearson's r correlations to assess the inter-rater reliability (IRR) between rater 1 and rater 2 for the English nonword sentences, the foreign sentences, and the foreign-word learning lists for a small sample of 10 participants. The inter-rater analyses were carried out on both the word and syllable level scores. We are reporting the syllable analysis here as both yielded similar significant results. For the syllables in the English nonword sentences, we found a strong positive correlation (see Table 5) between the scoring done by rater 1 and rater 2, which was statistically significant. For the syllables in the Turkish sentences (Table 5), we also found a strong positive correlation between the scoring done by rater 1 and rater 2, which was statistically significant. Lastly, for the syllables in the word pairs lists (Table 5), we also found a strong positive correlation between the scoring done by rater 1 and rater 2, which was statistically significant (**Table 5**). Thus, the scoring was highly reliable.

Table 5: Interrater Reliability Correlations on the score proportions of correct syllables. Spearman's rho is also reported for the English task as the scores were not distributed normally, as shown by the Shapiro-Wilk test. *N =10.*

Task	**Pearson's r**	**Spearman's rho**
English Nonword Sentence Repetition	r = 0.986, p <0.001, CI: 0.942, 0.997	rho = 0.935, p <0.001
Foreign Sentence Repetition	r = 0.959, p <0.001, CI: 0.832, 0.991	
Foreign-Word Learning	r = 0.985, p <0.001, CI: 0.962, 0.994	

CI = Confidence Interval

STM and Word Learning Tasks. For analyses involving the recall tasks, we used the proportion of correct syllables for the English nonword repetition task, the foreign-word sentence repetition task, and for the FWL task. For the TRA task, we took the proportion of correct beats. For the TOJ tasks, we took the proportion of the maximum score per trial. Again, we ran Shapiro-Wilk tests to assess the normality of the data. The English nonword sentences, visual TOJ, and auditory TOJ did not have normal distributions. We first computed correlations between the two sentence repetition tasks and then between the two TOJ tasks. Because of the non-normality of some distributions, we ran both Pearson's r and Spearman's rho correlations, but the values were similar, so we report here the Pearson's r values. However, for the TOJ tasks we only report the Spearman's rho correlations. The *p*-values were adjusted using Holm's corrections. Lastly, we ran a linear regression analysis with the FWL as the outcome variable and the STM tasks as the predictors. Because the TOJ tasks and the English nonword repetition task were not normally distributed, we ran an additional linear regression analysis with the FWL as the outcome variable and only the TRA task and the foreign sentence repetition task as the predictors.

MMN Peak Detection. For statistical analysis in R on the MMN peak data, we used the peak detection results for both standard blocks, as well as for both deviant blocks. We exported the latencies of the peak amplitudes for each electrode in the window between 100 and 250ms, because we wanted to perform analyses of variance (ANOVAs) on the two deviants separately as well as the standards versus deviants in each block. For the analyses in R, we chose 4-6 electrodes at nine pre-defined regions of interest (ROI), which left us using 43 out of the 64 electrodes. This was done in accordance with the lab specifications, to ensure enough power by reducing the number of sites. The nine ROIs were: middle frontal (MF), middle central (MC), middle parietal (MP), left frontal (LF), left central (LC), left parietal (LP), right frontal (RF), right central (RC), and right parietal (RP). For LF we used the electrodes AF3, AF7, F3, F5; for MF we used AFz, F1, F2, Fz; for RF we used AF4, AF8, F4, F6, PO7, PO8, POz; for LC we used C3, C5, FC3, FC5; for MC we used C1, C2, Cz, FC1, FC2, FCz; for RC we used C4, C6, FC4, FC6; for LP we used CP3, CP5, P3, P5; for MP we used CP1, CP2, CPz, P1, P2, Pz; and for RP we used CP4, CP6, P4, and P6. Therefore, there were nine ROIs for the syllable standards, for the early standards, for the syllable deviants, and for the early deviants. As an additional analysis, we exported the peaks in the difference waves between the standard and deviant stimuli averaged inside each ROI for each deviant block. The ROIs were the same as those used in the previously mentioned analysis, but, this time, only nine ROIs for the syllable deviants and the early deviants.

In addition, to check the reliability of the data, we ran analyses including only 200 standards from the first third of each block so we would have the same number of standards and deviants. We split the 200 standards into two groups, odd and even, so the first 100 consisted of the odd numbered standards within that 200 and the latter 100 the even numbered ones. We then computed multiple correlations between odd vs. even datasets on the averaged difference data for amplitude and latency, separately for the early deviant and syllable deviant. After doing Holm’s corrections, the difference wave amplitudes showed a strong positive correlation for the early deviant (**Table 6**). The latency showed a medium positive correlation for the early deviant and but only a small positive correlation for the /ka/ syllable deviant (**Table 7**).

Table 6: Correlations between the amplitude of the odd and even deviants in the early and syllable block. Both Pearson’s r and Spearman’s rho are reported in the table due to the Shapiro-Wilk test showing the data were not normally distributed. *N = 24*.

Block	**Pearson’s r**	**Spearman’s rho**
Early Deviant	r = 0.894, p <0.001, CI: 0.864, 0.918	rho = 0.903, p <0.001
Syllable Deviant	r = 0.734, p <0.001, CI: 0.665, 0.790	rho = 0.693, p < 0.001

CI = Confidence Interval

Table 7: Correlations between the latency of the odd and even deviants in the early and syllable block. Both Pearson’s r and Spearman’s rho are reported in the table due to the Shapiro-Wilk test showing the data were not normally distributed. *N = 24*.

Block	**Pearson’s r**	**Spearman’s rho**
Early Deviant	r = 0.517, p <0.001, CI: 0.412, 0.609	rho = 0.588, p <0.001
Syllable Deviant	r = 0.387, p <0.001, CI: 0.267, 0.495	rho = 0.372, p < 0.001

CI = Confidence Interval

Repeated measures ANOVAs were used to compare the averaged standard and deviant amplitudes and latencies for the two types of deviants separately. The ANOVAs had condition (early standard, syllable standard, early deviant or syllable deviant) and ROI as factors. For the statistical analysis on the MMN peak amplitude difference waves, repeated measures ANOVAs were done to compare the difference waves between the deviants with the ROI as a factor. To check the presence of an MMN response, we also used a series of one-sided one-sample t-tests to compare the means of each difference wave to 0. As the odd and even analysis of standards showed high correlations, all these analyses were performed on the entire sample of standards and deviants for each block.

MMN and Behavioural Tasks. We correlated the peak amplitude and latency difference wave and the peak amplitude for the MF ROI of the early deviant and the peak amplitude for the standard for the MF ROI of the early deviant with the TRA task and the FWL task. The MF ROI was used because it showed the largest response for most participants. We used Pearson's r and Spearman's rho correlations and report Spearman's rho whenever the values varied from the Pearson's r values for correlations that were not normally distributed. The *p*-values were adjusted with Holm's corrections.

Individual Differences. To examine individual differences, we looked at the peak amplitude of the MF ROI. We took an average of each participant's standards and then took all the deviants from the early block. After artifact rejection in BVA TM processing, the number of deviants each participant had ranged from 80-99 deviants. We then performed a series of one-sample t-tests for each participant and calculated the effect size using Cohen's d. We used the Cohen's d value for each participant and correlated it with performance on the TRA task and the FWL task to see if we would find individual differences in MMN related to STM for temporal order, rhythm, phonological form, as well as foreign-word learning. Similar to the other analyses, based on the results of the Shapiro-Wilk tests, both Pearson's r and Spearman's rho correlations were calculated. Again, the *p*-values were adjusted using Holm's corrections.

Music Experience. Lastly, we took the years of music experience and levels of music ability and correlated these music measures with the STM tasks. Because the sentence repetition tasks and the TOJ tasks correlate, respectively, we created a sentence composite score and a TOJ composite. Therefore, we correlated the music measures with the sentence score and the TRA task, and then with the TOJ score. *P*-values were adjusted using Holm's correction.

RESULTS

Correlational Analyses

The following tables show the descriptive statistics of the raw scores for the repetition tasks (**Table 8**), and the TOJ tasks (**Table 9**). For analyses, we turned the raw scores into z-scores to investigate more normally distributed data and to look at individual differences.

Table 8: Descriptive statistics for the repetition tasks.

	English Nonword Sentence Repetition	Foreign Sentence Repetition	TRA
N	30	30	30
Mean	0.712	0.501	0.746
Minimum	0.233	0.151	0.543
Maximum	0.890	0.803	0.929
Shapiro-Wilk W	0.807	0.955	0.971
Shapiro-Wilk p	< .001	0.231	0.562

Table 9: Descriptive statistics for the TOJ tasks.

	Auditory TOJ	**Visual TOJ**
N	30	30
Mean	0.703	0.739
Minimum	0.194	0.0556
Maximum	1.00	1.00
Shapiro-Wilk W	0.777	0.745
Shapiro-Wilk p	< .001	< .001

Sentence Repetition Tasks. For the first correlational analysis, the correlation was done between the proportion of correct syllables in the English nonword sentence repetition task and the proportion of correct syllables in the foreign sentence repetition task (**Table 10**). There was a statistically significant positive correlation between the proportion of correct syllables in the English task and the foreign task, *r*(28) = 0.491, *p* = .006 (**Figure 9**). Both Pearson and Spearman coefficients are reported as the English task was not normally distributed; both were significant. This suggests that performance on the English nonword repetition task relates significantly to the performance on the foreign sentence repetition task.

Table 10: Correlation matrix between the sentence repetition tasks. *N = 30*.

		English Nonword Sentence Repetition	**Foreign Sentence Repetition**
English Nonword Sentence Repetition	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
Foreign Sentence Repetition	Pearson's *r*	0.491 **	—
	p-value	0.006	—
	95% CI Upper	0.723	—
	95% CI Lower	0.159	—
	Spearman's *rho*	0.397 *	—
	p-value	0.030	—

Note. * p < .05, ** p < .01, *** p < .001.

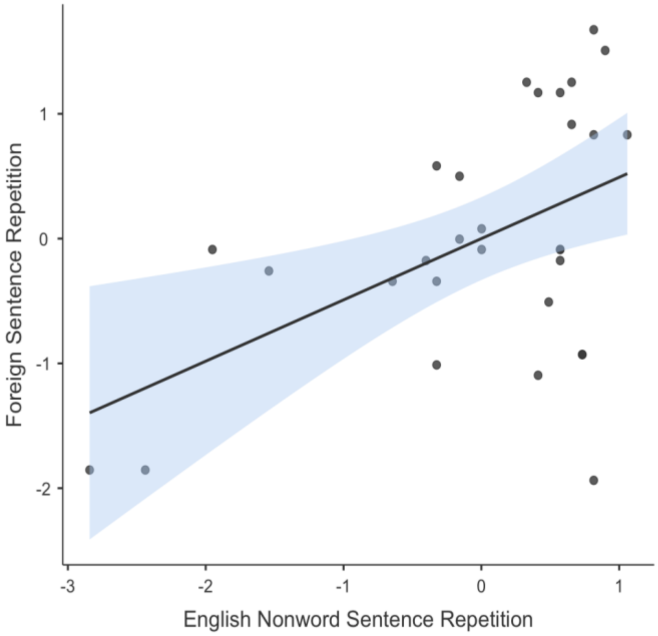


Figure 9: A scatterplot showing the positive correlation between the sentence repetition tasks. *N = 30.*

Temporal Order Judgment Tasks. In the second correlational analysis, the correlation was done between the proportion of maximum scores in the Auditory TOJ task and the proportion of maximum scores in the Visual TOJ task (**Table 11**). The results show a statistically non-significant positive correlation between the proportion of correct responses in the Auditory TOJ task and the proportion of correct responses in the Visual TOJ task, *r*(29) = .295, *p* = .057 (**Figure 10**). However, this is close to significant. Because the data were not normally distributed, Spearman’s rho coefficients were also computed. The rank order correlation *rho*(29) = .45, *p* = .006 was significant. This suggests that performance on the Auditory TOJ task predicts performance on the Visual TOJ task.

Table 11: Correlation matrix for the TOJ tasks. *N = 30*.

		Auditory TOJ		**Visual TOJ**
Auditory TOJ	Pearson's *r*	—		
	p-value	—		
	95% CI Upper	—		
	95% CI Lower	—		
	Spearman's *rho*	—		
	p-value	—		
Visual TOJ	Pearson's *r*	0.295		—
	p-value	0.057		—
	95% CI Upper	1.000		—
	95% CI Lower	-0.013		—
	Spearman's *rho*	0.450	**	—
	p-value	0.006		—

Note. H_a is positive correlation

Note. * p < .05, ** p < .01, *** p < .001, one-tailed

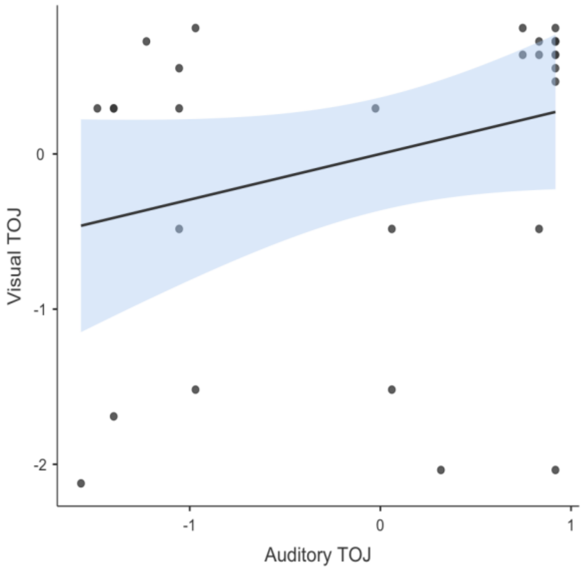


Figure 10: A scatterplot showing the correlation between the proportion of correct responses in the TOJ tasks. *N = 30.*

Predictors for Foreign-Word Learning. We hypothesized that the better a person performs in serial order tasks, the better they will be at word learning. Therefore, we expected to see that the better one performed on these STM tasks, the higher their proportion of correctly remembered syllables. The normalized behavioural task scores were treated as factors to predict performance in the FWL. The linear regression model shows that the five predictors together accounted for 31.8% of the variance in the FWL task. However, the linear regression equation was not significant, $F(5,24) = 2.23$, $p = 0.084$, $R_2 = 0.318$. The only predictor that was significant was the foreign sentence repetition task (**Table 12**). This result suggests that the foreign sentence repetition task significantly and uniquely predicts performance on the FWL task.

Table 12: Linear regression for the predicted variable FWL. *N = 30*.

Predictor	**Estimate**	**SE**	**t**	**p**
Intercept	5.37e-9	0.166	3.24e-8	1.000
English Nonword Sentence Repetition	-0.1767	0.207	-0.856	0.401
Foreign Sentence Repetition	0.5463	0.198	2.752	0.011
Temporal Rhythm Accuracy	0.1311	0.186	0.705	0.488
Auditory TOJ	0.1550	0.196	0.792	0.436
Visual TOJ	0.0280	0.182	0.154	0.879

As the TOJ and English nonword repetition tasks were not normally distributed, performed at ceiling, as shown by the Shapiro-Wilks test (see **Tables 8** and **9**), an additional regression was performed without these variables. The linear regression equation was significant, $F(2,27) = 5.33$, $p = .011$, $R_2 = 0.283$. The linear regression shows that the foreign sentence repetition task and the TRA task, together, account for 28.3% of the variance in the FWL task. Again, only the coefficient for the foreign sentence repetition task was significant (**Table 13**). This demonstrates that the foreign sentence repetition task significantly and uniquely predicts performance on the FWL task, more so than the other repetition tasks and the TOJ tasks.

Table 13: Linear regression for the foreign sentence repetition task and the TRA task on the predicted variable FWL. $N = 30$.

Predictor	**Estimate**	**SE**	**t**	**p**
Intercept	3.85e-9	0.160	2.40e-8	1.000
Foreign Sentence Repetition	0.486	0.166	2.932	0.007
Temporal Rhythm Accuracy	0.147	0.166	0.888	0.382

MMN Peak Detection

The MMN peak was identified within a 100-250 ms time-window. The descriptive statistics of the amplitude and latency of the deviants in the *early* and *syllable* blocks are shown in **Table 14**. The latency of the *early* deviant was not distributed normally.

Table 14: Descriptive statistics for the amplitude and latency of the MMN peak in the *early* and *syllable* deviant blocks, as well as for a test of normality of distribution.

	Early Deviant Amplitude (µV)	**Syllable Deviant Amplitude (µV)**	**Early Deviant Latency(ms)**	**Syllable Deviant Latency (ms)**
N	24	24	24	24
Mean	-5.16	-2.46	152	215
Minimum	-8.17	-6.47	121	180
Maximum	-2.11	2.67	248	248
Shapiro-Wilk *W*	0.957	0.971	0.674	0.918
Shapiro-Wilk *p*	0.377	0.689	< .001	0.052

The descriptive statistics for the amplitude of the difference waves for the *early* and *syllable* deviant blocks are presented in **Table 15**. The amplitudes of both deviants were distributed normally.

Table 15: Descriptive statistics for the amplitude of the difference waves in the *early* and *syllable* deviant blocks, as well as for a test of normality of distribution.

	Early Deviant Amplitude (µV)	**Syllable Deviant Amplitude** (µV)
N	24	24
Mean	-6.96	-2.10
Minimum	-10.4	-5.58
Maximum	-1.95	0.602
Shapiro-Wilk *W*	0.944	0.967
Shapiro-Wilk *p*	0.204	0.587

Standards and Deviants.

ERP Amplitude. For the *early* block, with the deviants occurring at shorter SOAs than the standard stimuli, a 2 (standard vs. deviant) x 9 (scalp ROI) ANOVA showed a significant main effect of condition, $F(1, 23) = 85.72$, $p < 0.001$, the peak ERP amplitude for the deviants, which occurred early, being more negative than that for the stimuli occurring at standard SOAs. There was also a significant main effect of scalp ROI, $F(8, 184) = 33.47$, $p < 0.001$, and an interaction, $F(8, 184) = 20.55$, $p < 0.001$, showing that the strength of the standard versus deviant MMN response depended on the scalp region (**Figure 11**).

For the *syllable* block, in which all stimuli occurred at the same SOAs but the deviant stimuli began with a different consonant than the standard stimuli, the ANOVA showed a significant main effect of condition, *F*(1, 23) = 19.12, *p* <0.001, a significant main effect of *ROI*, F(8, 184) = 3.8, *p* = 0.036, and an interaction between the two, *F*(8, 184) = 2.08, *p* = 0.039 (**Figure 12**). However, after the Huynh-Feldt sphericity correction the interaction was no longer significant, *F*(8, 184) = 2.08, *p* = 0.117. This suggests that there was a significant difference between the standard and deviant, indicative of an MMN, in the *syllable* block in terms of the peak itself. Peak amplitudes also varied depending on the location of the response. However, the greater negativity of the deviant response compared to the standard response did not significantly vary with scalp location.

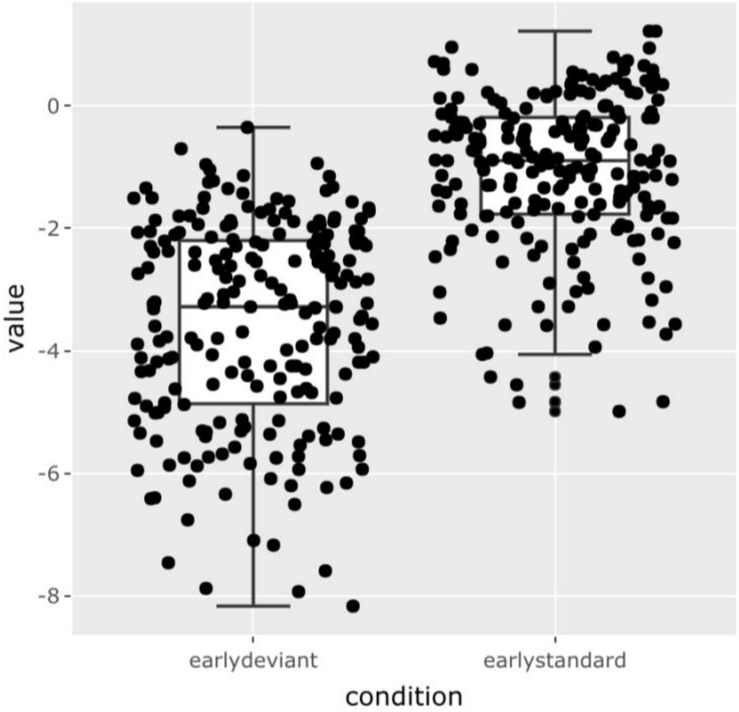


Figure 11: ERP amplitude of the deviants and the standards in the block with *early* deviants. *N = 24.*

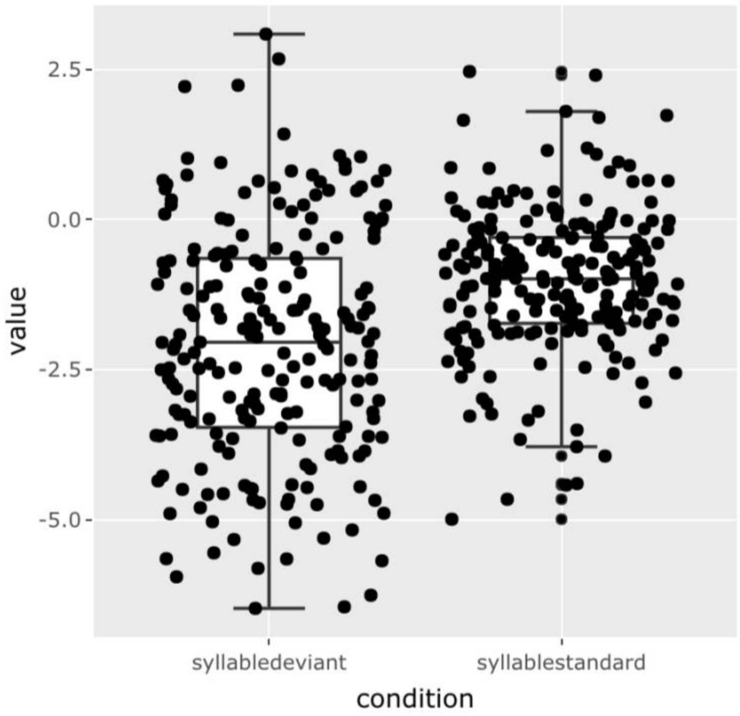


Figure 12: ERP amplitudes of the deviants and the standards in the block with *syllable* deviants. *N = 24.*

Latency. We also analyzed the latency of the ERP peak from the onset of the stimulus. For the *early* block, a 2 x 9 repeated measures ANOVA showed a main effect of condition, $F(1, 23) = 50.23$, $p < 0.001$, no significant main effect of ROI, $F(8, 184) = 0.43$, $p = 0.64$, but an interaction between condition and ROI was significant, $F(8, 184) = 2.88$, $p = 0.01$ (**Figure 13**). The condition effect reflects the finding that the negative response peak occurred later for the standard compared to the deviant stimuli in the *early* block. This difference between the deviant and standard response latencies depended on the region.

For the *syllable* block, a similar latency ANOVA did not reveal a main effect of condition, did show a significant main effect of ROI, $F(8, 184) = 2.68$, $p = 0.049$, but not a significant interaction, $F(8, 184) = 0.7$, $p = 0.69$ (**Figure 14**). This suggests that the location of the response differed significantly between the standard and deviant in the *syllable* block.

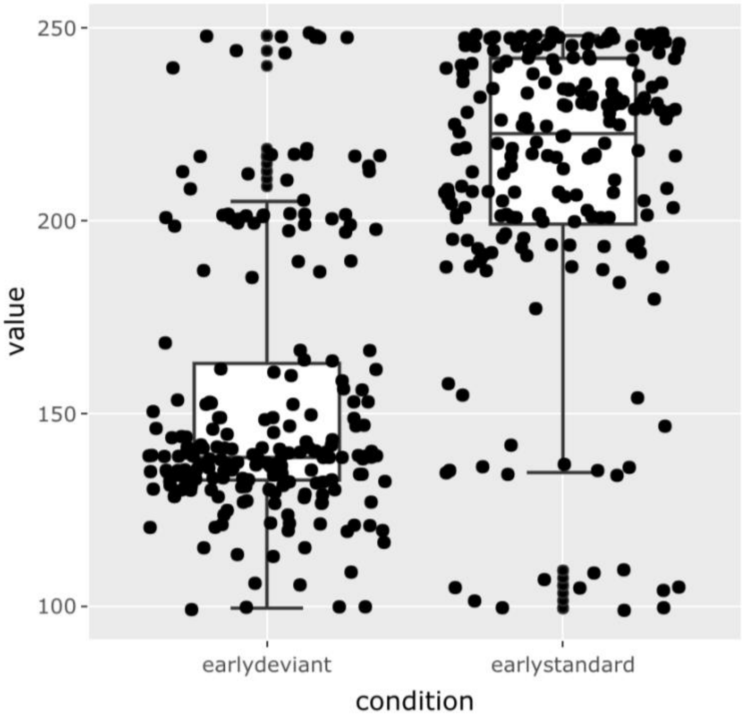


Figure 13: Latencies for ERPs to deviants and standards in the block with *early* deviants. *N = 24.*

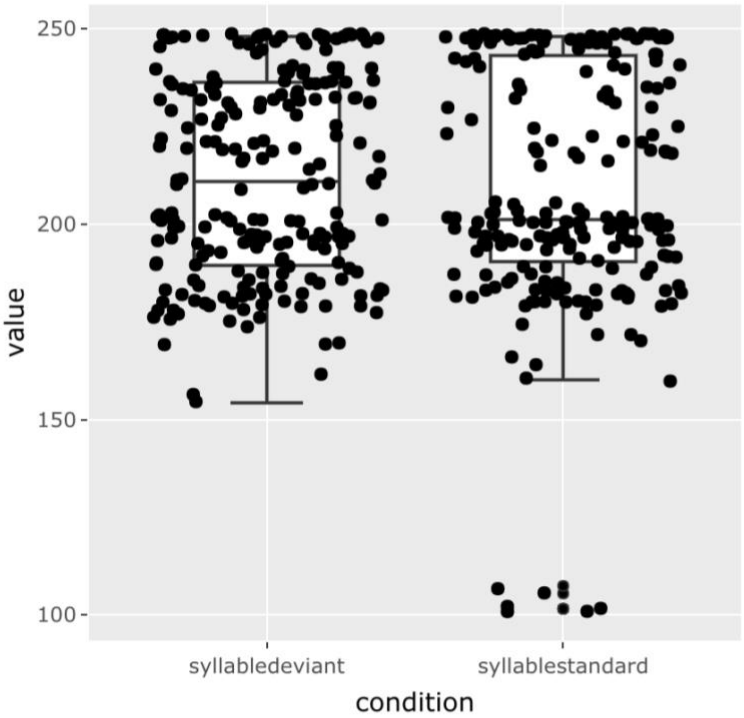


Figure 14: Latencies of deviants and standards in the block with *syllable* deviants. *N = 24.*

Difference Waves.

ERP Amplitude. MMN responses are customarily analyzed based on subtractions between the deviant and standard waveforms. One-way repeated measures ANOVA models were computed separately for the MMN responses (difference wave peak amplitudes) to deviant SOAs and phonologically deviant syllables. ROI was the repeated factor. The ANOVA for the *early* deviant showed there was a significant main effect of ROI, $F(8, 184) = 58.94$, $p < 0.001$. The ANOVA was followed with a one-sample t-test for the middle frontal ROI, a site expected to show an MMN response. The t-test showed a significant difference from zero, $t(23) = -14.83$, $p < 0.001$, 95% CI = -infinity, -5.15. There was also a significant main effect of ROI for the syllable deviant, $F(8, 184) = 4.77$, *p* = 0.003. A one-sample t-test again showed significance for the middle frontal ROI, *t*(23) = -6.26, *p* <0.001, 95% CI = -infinity, -1.52 (**Figure 15**). The amplitude of the MMN for the deviants differed significantly from zero.

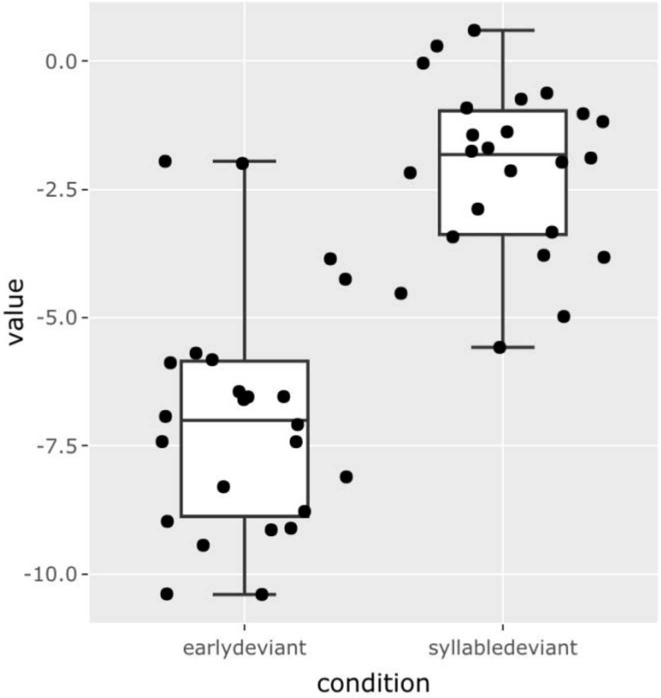


Figure 15: The amplitude of the difference waves in the block with *early* deviants and the block with *syllable* deviants. *N = 24.*

MMN and STM Measures of Temporal Structure.

For further exploration, we aimed to investigate whether the *early* deviant MMN response is sensitive to temporal accuracy, and whether there is a relation with word learning. The rationale to ask questions about the MMN to an *early* deviant was to determine if individual differences in the sensitivity to deviations in timing of a repetitive sequence of auditory stimuli would be related to behavioural measures of temporal structure. Although these analyses were planned, circumstances outside our control prevented us from testing the planned sample size of 30 participants. For this reason, all correlations with our ERP measures must be treated with caution.

We had four MMN measures: difference wave peak amplitude, deviant peak amplitude, standard peak amplitude, and difference wave peak latency. Each of these measures were correlated with the proportion of correct scores in the TRA task to investigate temporal accuracy. The correlations were Holm's corrected and the corrected p-values are reported in the tables. The difference wave peak amplitude showed a non-significant weak positive correlation with the TRA task, *r*(23) = .168, *p* = .432 (**Table 16**). If this correlation were reliable, it would suggest that there is an inverse relationship between one's ability for temporal accuracy in a tapping task and their temporal accuracy in an MMN task. As the correlation is non-significant and in an unexpected direction, we do not interpret it further.

Table 16: Correlation between difference wave peak amplitude of the *early* deviant and the proportion of correct responses in the TRA task. *N = 24*.

		Early Deviant Amplitude	**TRA**
Early Deviant Amplitude	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
TRA	Pearson's *r*	0.168	—
	p-value	0.932	—

Table 16: Correlation between difference wave peak amplitude of the *early* deviant and the proportion of correct responses in the TRA task. *N = 24.*

		Early Deviant Amplitude	**TRA**
	95% CI Upper	0.535	—
	95% CI Lower	-0.252	—
	Spearman's *rho*	0.262	—
	p-value	0.864	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

The next correlation with TRA was conducted with the deviant peak amplitude (**Table 17**). The results show that there was no significant correlation between these two variables, *r*(23) = .002, *p* = .993. This suggests that the peak amplitude of the deviant alone does not predict one's score in the TRA task.

Table 17: Correlation between the *early* deviant peak amplitude and the proportion of correct responses in the TRA task. *N = 24.*

		Early Deviant Amplitude	**TRA**
Early Deviant Amplitude	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	

Table 17: Correlation between the *early* deviant peak amplitude and the proportion of correct responses in the TRA task. *N = 24*.

		Early Deviant Amplitude	**TRA**
TRA	Pearson's *r*	0.002	—
	p-value	0.993	—
	95% CI Upper	0.405	—
	95% CI Lower	-0.402	—
	Spearman's *rho*	0.103	—
	p-value	1.00	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

We then correlated the TRA task with the standard peak amplitude (**Table 18**). The results showed a non-significant weak negative correlation, *r*(23) = -.221, *p* = .299. This suggests that there could be a positive relationship between performance in the TRA task and the peak amplitude of the event-related response to the *early* standards in the MMN window, i.e., the larger one's ERP responses to the standards, the better one's performance in the TRA task. Again, this correlation, even if significant, does not have a theoretical interpretation.

Table 18: Correlation between the *early* standard peak amplitude and the proportion of correct responses in the TRA task. *N = 24*.

		Early Standard Amplitude	**TRA**
Early Standard Amplitude	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
TRA	Pearson's *r*	-0.221	—
	p-value	0.932	—
	95% CI Upper	0.200	—
	95% CI Lower	-0.573	—
	Spearman's *rho*	-0.154	—
	p-value	1.00	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

Lastly, we correlated the TRA task with the *early* deviant peak latency (**Table 19**). The results show a non-significant weak negative correlation, *r*(23) = -.253, *p* = .233. This would reveal a positive relationship between how one performs on the TRA task and the latency of their MMN response. Thus, the better one performed on the TRA task, the longer the latency of their MMN response. However, the correlation was not significant and was not expected on the basis of theory.

Table 19: Correlation between the difference wave deviant latency and the proportion of correct responses in the TRA task. *N = 24.*

		Early Deviant Latency	**TRA**
Early Deviant Latency	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
TRA	Pearson's *r*	-0.253	—
	p-value	0.932	—
	95% CI Upper	0.168	—
	95% CI Lower	-0.596	—
	Spearman's *rho*	-0.197	—
	p-value	1.00	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

The same ERP variables were then employed to study correlations with the FWL task. The first correlation investigated the relationship between the difference wave peak amplitude and performance in the FWL task (**Table 20**). The results showed a non-significant weak positive correlation, $r(23) = .245$, $p = .250$. This would indicate that the smaller a person's peak MMN amplitude, the better they performed in the FWL task. As this correlation was not significant or in the predicted direction, we will treat it as spurious.

Table 20: Correlation between the difference wave peak amplitude and the proportion of correct syllables in the FWL task. *N = 24.*

		Early Deviant Amplitude	**FWL**
Early Deviant Amplitude	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
FWL	Pearson's *r*	0.245	—
	p-value	0.756	—
	95% CI Upper	0.590	—
	95% CI Lower	-0.176	—
	Spearman's *rho*	0.274	—
	p-value	0.660	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

The next correlation was performed between the FWL task and the deviant peak amplitude (**Table 21**). The results showed a non-significant weak positive correlation, *r*(23) = .211, *p* = .322. This correlation would suggest an inverse relationship with a person's deviant peak amplitude and how well they perform in the FWL task. Note, again, that this correlation was not significant or in a predicted direction.

Table 21: Correlation between the deviant peak amplitude and the proportion of correct syllables in the FWL task. *N = 24*.

		Early Deviant Amplitude	**FWL**
Early Deviant Amplitude	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
FWL	Pearson's *r*	0.211	—
	p-value	0.756	—
	95% CI Upper	0.566	—
	95% CI Lower	-0.210	—
	Spearman's *rho*	0.293	—
	p-value	0.660	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

Next, a correlation was explored between how well one does on the FWL task and the standard peak amplitude (**Table 22**). The results show no reliable correlation, *r*(23) = -.048, *p* = .823. This suggests that the size of person's peak amplitude ERP in the MMN time window does not relate to how well they are able to learn new words.

Table 22: Correlation between the standard peak amplitude and the proportion of correct syllables in the FWL task. *N = 24*.

		Early Standard Amplitude	**FWL**
Early Standard Amplitude	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
FWL	Pearson's *r*	-0.048	—
	p-value	0.823	—
	95% CI Upper	0.362	—
	95% CI Lower	-0.443	—
	Spearman's *rho*	0.049	—
	p-value	0.820	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

For the final correlation between MMN measures and word learning, we explored if there is a relationship between the deviant latency and the FWL task (**Table 23**). The results demonstrate a non-significant weak negative correlation, *r*(23) = -.278, *p* = .756. This would have suggested that the earlier a person's MMN response, the better they performed in the FWL task. Although this correlation was in the expected direction, it did not approach significance. A larger *N* would be needed to test its reliability.

Table 23: Correlation between the deviant latency and the proportion of correct syllables in the FWL task. *N = 24*.

		Early Deviant Latency	**FWL**
Early Deviant Latency	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
FWL	Pearson's *r*	-0.278	—
	p-value	0.756	—
	95% CI Upper	0.142	—
	95% CI Lower	-0.612	—
	Spearman's *rho*	-0.192	—
	p-value	0.736	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

We had planned regression analyses, but the small *N* and non-significant correlations did not justify them.

Individual Differences. The raw values of ERP amplitude measures can depend on trivial factors, such as scalp conductance. In the next analysis, we inspected individual MMN effect sizes. To further explore possible relations between individual differences in MMN responses and temporal accuracy and word learning measures, we used the difference between the standards and deviants in the *early* deviant block. After calculating Cohen's *d* from a one-sample t-test for each participant, we calculated the effect size, (**Table 24**).

Table 24: Results of t-tests computed for each participant on the average of standards and deviants in the *early* block. Cohen's *d* was also calculated for each participant. *N = 24*.

ID	t	df	p	CI	Effect Size
1	11.3	94	<0.001	5.96, 8.51	1.159
2	10.95	88	<0.001	9.91, 14.31	1.16
3	11.41	97	<0.001	8.46, 12.03	1.16
4	13.74	85	<0.001	12.67, 16.95	1.481
5	11.92	90	<0.001	10.44, 14.62	1.249
6	12.21	92	<0.001	8.82, 12.24	1.266
7	8.38	88	<0.001	8.37, 13.58	0.888
8	12.9	79	<0.001	9.92, 13.55	1.44
9	9.62	94	<0.001	6.49, 9.87	0.987
10	17.68	93	<0.001	12.11, 15.17	1.823
11	7.316	82	<0.001	3.96, 6.91	0.803
12	13.93	98	<0.001	11.89, 15.85	1.4
13	13.14	85	<0.001	10.00, 13.57	1.417
14	14.28	98	<0.001	12.79, 16.93	1.435
15	12.04	89	<0.001	13.59, 18.96	1.269
16	13.43	94	<0.001	9.85, 13.27	1.378

17	9.263	89	<0.001	6.29, 9.74	0.976
18	12.66	87	<0.001	12.06, 16.55	1.349
19	9.428	79	<0.001	8.34, 12.80	1.054
20	6.74	90	<0.001	2.78, 5.10	0.706
21	9.088	86	<0.001	5.52, 8.61	0.974
22	11.45	94	<0.001	7.31, 10.37	1.174
23	14.91	98	<0.001	11.17, 14.60	1.498
24	10.66	83	<0.001	7.15, 10.43	1.163

Note: ID refers to participants, df = degrees of freedom, CI = 95% confidence interval

The correlational analyses were conducted with the MMN effect sizes, the TRA task, and the FWL task (**Table 25**). There were no significant correlations with the TRA task, *r*(23) = .120, *p* = .578 nor with the FWL task, *r*(23) = .171, *p* = .424. Although the correlations were positive, they were very weak and not significant. The effect size *d* appears not to vary much between participants, who all showed a large effect size. The time difference between the early deviant and the standard stimuli may have been too large to be sensitive to individual variability within the tested sample.

Table 25: Correlations between effect size of the difference between standards and deviants in the *early* deviant block and the TRA and FWL tasks. *N = 24*.

		MMN effect Size	**TRA**	**FWL**
MMN effect Size	Pearson's *r*	—		
	p-value	—		
	95% CI Upper	—		
	95% CI Lower	—		
	Spearman's *rho*	—		
	p-value	—		
TRA	Pearson's *r*	0.120	—	
	p-value	0.848	—	
	95% CI Upper	0.499	—	
	95% CI Lower	-0.298	—	
	Spearman's *rho*	-0.018	—	
	p-value	0.996	—	
FWL	Pearson's *r*	0.171	0.280	—
	p-value	0.848	0.555	—
	95% CI Upper	0.537	0.614	—
	95% CI Lower	-0.250	-0.139	—
	Spearman's *rho*	0.202	0.207	—
	p-value	0.996	0.996	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

Music Experience and STM Tasks. As a background variable, we had recorded years of music experience and level of music ability (beginner, intermediate, advanced), measures that were self-judged. We next explored whether these variables predicted performance in the STM tasks. Because the two sentence repetition tasks correlated, and the two TOJ tasks correlated, these tasks were combined to create composite scores, yielding one sentence repetition score and one TOJ score. For the first correlational analysis with musical experience, the correlation was performed between the years of experience and the level of music ability with the sentence repetition composite score (**Table 26**). There was a significant weak negative correlation between the years of experience and the proportion of correct syllables in the sentence repetition composite score, *r*(28) = -.394, *p* = .035, although it became non-significant after Holm's corrections (**Figure 16)**. There was also a non-significant negative correlation between the levels of music ability and the proportion of correct syllables in the sentence repetition composite score, Spearman's *rho*(28) = -0.296, *p* = 0.452 (**Figure 17**). The TRA task did not significantly correlate with years of music experience nor with level of music ability. The values of these correlations were not expected, as our hypotheses differed from that of the findings. Thus, the non-significant results suggest that years and level of musical experience did not reliably predict performance on phonological, temporal, and serial order STM tasks in the present university sample.

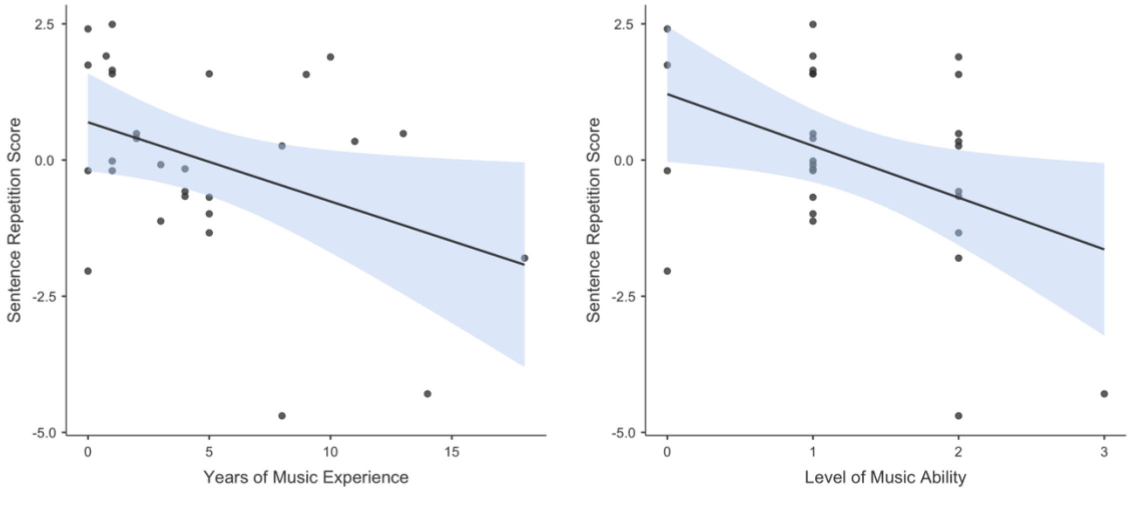
Table 26: Correlation matrix for the sentence repetition score composite, the TRA task, and the years of music experience and level music ability. *N = 29*.

		Sentence Repetition Score	**TRA**	**Years of Music Experience**	**Level of Music Ability**
Sentence Repetition Score	Pearson's *r*	—			
	p-value	—			
	95% CI Upper	—			
	95% CI Lower	—			
	Spearman's *rho*	—			
	p-value	—			
TRA	Pearson's *r*	0.277	—		
	p-value	0.290	—		
	95% CI Upper	0.585	—		
	95% CI Lower	-0.099	—		
	Spearman's *rho*	0.301	—		
	p-value	0.542	—		
Years of Music Experience	Pearson's *r*	-0.394	-0.088	—	
	p-value	0.105	0.648	—	
	95% CI Upper	-0.032	0.287	—	

Table 26: Correlation matrix for the sentence repetition score composite, the TRA task, and the years of music experience and level music ability. *N = 29*.

		Sentence Repetition Score	**TRA**	**Years of Music Experience**	**Level of Music Ability**
	95% CI Lower	-0.664	-0.441	—	
	Spearman's *rho*	-0.337	0.064	—	
	p-value	0.370	1.00	—	
Level of Music Ability	Spearman's *rho*	-0.296	0.079	0.867 ***	—
	p-value	0.452	1.00	< .001	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*



Figures 16 & 17: The left figure is a scatterplot showing the correlation between the years of music experience and the proportion of correct syllables in the sentence repetition composite score. The right figure is showing the correlation between the level of music ability and the proportion of correct syllables in the sentence repetition composite score. *N = 29*.

The second set of correlations was computed on the TOJ composite score and the years of music experience and the level of music ability (**Table 27**). The results show no significant correlations between the music measures and the TOJ composite score. This suggests that music experience and music ability did not reliably predict performance in tasks involving temporal order in the present sample of participants.

Table 27: Correlation matrix for the TOJ composite score and the years of music experience and level of music ability. *N = 29*.

		TOJ Score	**Years of Music Experience**	**Level of Music Ability**
TOJ Score	Pearson's *r*	—		
	p-value	—		
	95% CI Upper	—		
	95% CI Lower	—		
	Spearman's *rho*	—		
	p-value	—		
Years of Music Experience	Pearson's *r*	-0.174	—	
	p-value	0.368	—	
	95% CI Upper	0.206	—	
	95% CI Lower	-0.508	—	
	Spearman's *rho*	-0.111	—	
	p-value	1.00	—	
Level of Music Ability	Spearman's *rho*	-0.107	0.867 ***	—
	p-value	1.00	< .001	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

Music Experience and MMN. We decided to also explore the correlation between the MMN response and music experience. No significant effects of the musical measures were found. This analysis is reported in Appendix B.

Discussion

The described experiment provides an interesting picture of the effects of STM and temporal order on word learning. We originally hypothesized that ability to reproduce verbal and non-verbal temporal order would translate to ability to encode patterns of variation in speech. Specifically, we predicted that the better one is at serial order tasks, the better one will perform in the FWL task. This was found with the foreign language sentence repetition task, but not for the English nonword sentence repetition task, the TRA task, or the TOJ tasks. We also explored whether the MMN response to the *early* deviant was sensitive to temporal accuracy. We found an unexpected correlation between one's temporal accuracy and the difference wave peak amplitude; an expected relationship between one's temporal accuracy and the standard peak amplitude; and an unexpected correlation with temporal accuracy and the deviant peak latency. However, none of the correlations were significant. Additionally, we explored the relationship between MMN responses and foreign-word learning. We found an unexpected correlation with the difference wave peak amplitude and performance on the FWL task; an unexpected correlation with the deviant peak amplitude and performance on the FWL task; and an expected correlation with the deviant peak latency and the FWL task. Again, all of the correlations were not significant. As a further avenue of exploration, we also investigated whether individual differences in the effect size of the MMN *early* deviant difference wave peak amplitude correlated with temporal accuracy and word learning. We did not find significant correlations. Additionally, we looked into the relationship of music experience and level of music ability with sentence repetition tasks and the TRA task and the TOJ tasks. The results showed an unexpected non-significant correlation with the sentence repetition score, but no significant correlation with the TRA task or the TOJ tasks.

Sentence Repetition Tasks

The nonword repetition task consists of temporarily storing a sequence of familiar phonemes in a particular, unfamiliar serial order and then retrieving that representation from STM for recall (Schraeyen et al., 2019). On the other hand, for foreign-word repetition, participants are required to temporarily store a sequence of *unfamiliar* phonemes in a certain unfamiliar serial order. These repetition tasks have been used to measure the individuals' ability to create and maintain phonological representations in STM (Service & Craik, 1993). Where previous studies employed words for the repetition task, this study used sentences. With single nonwords, you are remembering the temporal patterns of the individual word. Sentences, on the other hand, require you to remember the temporal patterns of the words in the context of each other, in addition to the temporal patterns of the words themselves. From the perspective of a learner, every new word is a nonword when first encountered (Gupta et al., 2005). In the experiment employed in this thesis, the English sentence repetition task consisted of nonwords with familiar phonemes, whereas the foreign sentence repetition task consisted of words with some unfamiliar phonemes. Even though the latter were real words in a foreign language, they were novel to the learner, and were, therefore, also perceived as nonwords. This allows us to compare how serial order recall of familiar prosody and phonology influence the serial recall of unfamiliar prosody and phonology. There was a significant correlation between the English nonword sentence repetition task and the foreign language sentence repetition task. This indicates that there is a relationship between STM for material in one's native phonology, partly relying on LTM (Service & Craik, 1993), and one's ability to produce unfamiliar phonology and prosody. The fact that participant scores showed a relationship in these two tasks could suggest that the extent of lexical, phonological, and prosodic knowledge of one's native language influences the learning of foreign phonology and prosody (Szewczyk et al., 2018). On the other hand, this could suggest that perhaps the LTM support for native language nonword sentences relies on a similar STM as the foreign sentences, therefore resulting in correlations on tasks requiring both types of memory for language.

Temporal Order Judgment Tasks

The TOJ tasks were employed to explore serial order in the non-verbal domain. As the sequence length of stimuli increased, the further participants got in the task, serial order had to be accessed. As the data were not normally distributed, Spearman's rho coefficients were computed between the visual and auditory versions of the order comparison task. This showed only a weak expected correlation between the two domains. The correlation could be accounted for by the fact that the tasks, originally designed for pre-school-aged children, were too easy for adult participants, even with the articulatory suppression. In addition, although the sequence length increased throughout the task, a pattern might have been predicted. As the presented sequences were binary, the participants might have recognized this pattern in that the order of only two stimuli would be different, and that the change of order never occurred at the beginning or end of the sequence. Therefore, if participants caught on to this pattern, they may have stopped relying on the temporal aspect of serial order and, instead, focused on the pattern. This may explain why participants performed at ceiling across both tasks. We used articulatory suppression in the TOJ tasks to prevent verbal coding. This should have suppressed subvocal rehearsal as a verbal strategy to support memory for serial order (Baddeley et al., 1965a). Temporal grouping has been found to be independent of articulatory suppression when the list items are presented auditorily, but not visually (Hitch et al., 1996). Studies have shown temporal grouping in auditory spatial sequences, demonstrating serial memory in the audio-spatial domain (Parmentier et al., 2004). However, this serial memory was found when the auditory domain was combined with the spatial domain. In our study, the auditory TOJ was not combined with the spatial domain. Studies have pointed to the possibility of a shared mechanism for serial order in the verbal and spatial domains (Hurlstone & Hitch, 2014; Hurlstone et al., 2013), but no previous studies to our knowledge have explored the order aspect of auditory non-verbal serial memory in the auditory modality alone. The results in the present study suggest the possibility of serial order in the auditory and visual domain sharing mechanisms. Therefore, we expected a strong correlation between the two tasks. However, as the correlation was weak, this indicates that the TOJ tasks were too easy for adult participants. Future studies should adjust the TOJ tasks to be more challenging for adult participants by having sequences with two changing stimuli (as in this task), as well as sequences that unpredictably change the order of stimuli in the sequences. This would better allow the investigation of serial order, as participants would not be able to easily form a pattern of the stimuli based on the strategy mentioned above and, instead, would have to rely on serial order mechanisms for more complex patterns of stimuli.

Sentence Repetition Tasks and Foreign-Word Learning

The FWL task gives us an idea of the participants' "efficiency in forming new arbitrary associations between concepts" (Service & Craik, 1993). How well one is able to repeat back phonologically possible words is a good indicator of one's language-related abilities, and performance on the nonword repetition tasks has been shown to be an important predictor of novel word learning in native and foreign language acquisition (Baddeley et al., 1998; Gathercole & Baddeley, 1990; Gathercole et al., 1997; Schraeyen et al., 2019; Service, 1992; Service & Craik, 1993; Service & Kohonen, 1995; Szewczyk et al., 2018). The sentence repetition tasks were anticipated to explore temporal order in the verbal domain, whereas the TOJ tasks were exploring serial order in the non-verbal domain (Laasonen et al., 2012). The foreign sentence repetition task significantly predicted performance in the FWL task, but the English nonword sentence repetition task, the TRA task, and the TOJ tasks did not.

Participants' knowledge of the language used in the verbal repetition task might be a "key factor in determining accuracy of [nonword repetition]" (Szewczyk et al., 2018). The lack of significant correlation between the English nonword sentence repetition task and the FWL task could be because the English task was too easy for the present participants, performing at ceiling. Individuals who are more proficient in a given language are better at repeating nonwords that resemble plausible words in the language (Service, 1992; Szewczyk et al., 2018). Thus, due to the English repetition task consisting of easily plausible words in English, the task was easy for participants.

The novel word forms used in the FWL task were Turkish words, not English nonwords, differing with regards to both phonemes, phonotactics and prosody. As such, the finding that performance on the foreign sentence repetition task predicted performance on the FWL task could be explained by the fact that the language used in both tasks were the same. Therefore, the prosodic pattern of the foreign language (Turkish) was stored temporarily in STM from the sentence task, performed prior to the FWL task, and, as a result, the newly formed phonological and prosodic representations could have aided participants with Turkish word learning. This supports the idea that for tasks in which phonological memory is most important, and consequently more heavily relied upon, you see the highest correlation with learning of new phonological patterns (Service & Craik, 1993). The better one is at reproducing unfamiliar phonology and holding the phonological representations in STM (Service & Craik, 1993), the better one is at learning words in that foreign language.

It has been demonstrated that SVO and SOV languages differ in their correlates of PHPH prominence (Langus et al., 2016). In SVO languages, such as English, the main correlate is duration, whereas in SOV languages, such as Turkish, it is pitch and intensity (Langus et al., 2016). Studies focusing on groupings of non-native rhythms have demonstrated that training of certain rhythmic groupings in an unfamiliar phonology can aid learning (e.g., Henny et al., 2018). The study employed here did not focus on training rhythmic groupings, but the foreign sentence repetition task employing the same language as that in the FWL task could be seen as priming the ability of individuals to hold particular novel phonological patterns in their STM. Thus, the foreign sentence repetition task could be seen as "training" prior to the FWL task, as it was familiarizing participants with phonology they would then need to learn, by asking them to repeat similar material.

In the present study, one cannot rule out the possibility of additional languages playing a role in word learning. The majority of participants were bilingual and multilingual speakers, and therefore many of them could be speakers of SOV languages. A familiar prosodic rhythm (Langus et al., 2016) could have aided them in the learning of novel Turkish words. Future studies should investigate the link between languages in the individuals' repertoire and the prosodic pattern of each language and how that predicts learning words with and without that same prosodic pattern. Additionally, as this study, to our knowledge, is the first to look at nonword repetition in sentences, future studies should look into individual nonwords versus nonword sentences for serial recall, and how each of these independently predict FWL in words and FWL in sentences. This will provide a better understanding of how rhythm and phonology play a role in word learning.

Temporal Accuracy and Foreign-Word Learning

It has been shown that the perception of temporal patterns is an important predictor of language abilities (e.g., Flaugnacco et al., 2014; Thomson & Goswami, 2008). Reproducing rhythm and tapping correlates with pseudoword repetition

(Flaugnacco et al., 2014). Therefore, we expected to see a similar correlation in this experiment, but the TRA task did not significantly predict word learning. The TRA task employed in the current study might have been too easy as the same number of tones was presented in each sequence for repetition. Other tasks using varying sequence lengths of tones, as well as varying musical structures, found a link with serial order representation (Flaugnacco et al., 2014). Therefore, the TRA task may not have been properly assessing serial order as the tone sequences were always the same length, consisting of short and long beats, causing participants to recognize a pattern rather than maintain the serial order of stimuli. Future research should employ the TRA task used here, in addition to a tone task that varies in sequence length to explore the link with language acquisition.

Temporal Order Judgment and Foreign-Word Learning

The TOJ tasks did not significantly predict word learning. As previously mentioned, although the TOJ tasks were assessing serial order as the sequence lengths increased, participants may have caught onto the pattern of the sequences and therefore were no longer using serial order memory, but, instead, were memorizing the pattern. Studies have demonstrated that serial memory across domains predicts language abilities, such as pseudoword repetition abilities (e.g., Laasonen et al., 2012). Therefore, because we did not see a correlation between the TOJ tasks and FWL, we can assume that the TOJ tasks did not properly test for serial memory ability as these tasks were too easy for adult participants.

MMN Responses to Timing and Phonological Stimuli

We investigated the MMN peak amplitudes and latencies in both the *early* block and the *syllable* block with ANOVA models. We were interested in seeing whether the MMN response would differ based on the stimuli differing in temporal and phonological characteristics. The MMN results showed a significantly later latency for the *syllable* deviant than for the *early* deviant. Researchers have noted that "there are differences in latencies of the potentials according to the characteristics of the stimuli" (Brückmann & Garcia, 2019). However, that study used nonverbal stimuli, whereas, in the study employed here, we used verbal stimuli.

The data show that even though the stimuli were both verbal, having a condition which differs from the standard syllables with regards to both the initial consonant and vowel, causes different latencies from one where the deviant is the same as the standards but presented earlier. There was also a smaller amplitude of the MMN response to the *syllable* deviant than the *early* deviant, even though both differed significantly from the standard. This finding could be because the *syllable* deviant /ka/ was perceived as too similar to the standards /ti/, /tu/, /tə/. Therefore, if the main response came from the syllable level, the syllable level may not have established a proper standard. Without a properly established standard, responses to the three different standards would have been too similar to the responses to the true deviant, which would have prevented a strong MMN response. The *early* deviant, on the other hand, was acoustically the same as the standard but presented earlier. Therefore, it could have been easier to establish a standard because the timing changed in deviants but there was no related variability among the standards.

As the MMN responses are generally analyzed based on the difference between the deviants and standards, we also looked at the amplitude of the difference waves in each block. Both the *early* deviant and the *syllable* deviant differed significantly from zero. This result was expected. The *early* deviant did have a larger main effect of ROI than the *syllable* deviant, which, again, suggests that the *early* deviant produced a stronger MMN response. It has been previously shown that a larger MMN amplitude is produced by a duration deviant compared to that of a frequency deviant (Kathmann et al., 1999). However, this result was found with tone stimuli and it has been cautioned against transferring findings of MMN data with tone stimuli to that of speech, as extracting the probability of stimuli has been shown to be slower for verbal stimuli (Bendixen et al., 2015). It would be interesting to recreate this study using tones to compare the MMN responses between the syllable deviant and the early deviant, using syllables, and then to tone stimuli in each block.

Temporal Accuracy and MMN

There are few studies exploring the relationship of STM and MMN responses (e.g., Lv et al., 2010; Bonetti et al., 2018). Our study sought to explore whether the *early* deviant MMN response was sensitive to temporal accuracy. The *early* deviant difference wave peak amplitude and latency showed non-significant weak unexpected correlations with the TRA task. This suggests that the smaller a person’s MMN amplitude response and the later the latency of the response, the more accurate these persons were at detecting temporal changes and reproducing rhythmic patterns. On the other hand, the *early* standard peak amplitude showed a non-significant weak expected correlation with the TRA task. Due to the small sample size, there is a need for further, more powerful studies exploring the relationship between MMN and temporal accuracy. This could be explored, for example, by varying the time differences in the TRA task and by smaller variations between the deviants and standards used in the MMN oddball paradigm. This will allow better investigation into the temporal pattern of both tasks.

The *early* deviant was a better paradigm to measure individuals' MMN response to deviant stimuli, as previously mentioned. The studies that have explored WM and MMN (Lv et al., 2010; Bonetti et al., 2017) used digits and acoustic features inserted into tone sequences. In addition, studies employed the WM tasks while recording the MMN responses (Lv et al., 2010; Bonetti et al., 2017), whereas our study had participants perform the STM tasks prior to the MMN paradigm; the tasks were separate. To our knowledge, there are no studies to date that have explored the link between STM, serial order memory, and MMN. Future research should investigate this phenomenon of serial order affecting MMN responses. Instead of exploring them as separate tasks, experiments should first explore the effect of serial order while recording MMN. This will better relate to previous studies investigating STM and MMN responses and provide a better understanding of the link between neural activation and serial order errors in individuals. In doing so, this can be extended to studies exploring the documented temporal order deficits in dyslexia.

MMN and Foreign-Word Learning

We then investigated if MMN peak amplitude and latency have a relationship with word learning. The *early* deviant difference wave peak amplitude and *early* deviant peak amplitude showed a weak unexpected correlation with the FWL task. This means that the smaller one's MMN peak amplitude, the better they are at word learning. However, these correlations were not significant. The *early* deviant difference wave peak latency showed a weak expected correlation with performance in the FWL task. This indicates that the earlier one's MMN latency, the better they are at learning new words. Again, this correlation was not significant. These correlations could illustrate that the earlier one is able to notice variability in the temporal pattern, the better they are at learning temporal patterns of foreign words. A previous study investigated MMN responses in native Mandarin speakers to novel and familiar word-forms in Mandarin (Yue et al., 2014). Participants' MMN responses were recorded while passively listening to the novel and familiar word-forms during an early exposure phase, and a late exposure phase. Participants demonstrated larger MMN responses to the novel word-forms in the late exposure phase, but the MMN responses did not change with the familiar word-forms (Yue et al., 2014). The researchers suggested that this indicates LTM traces were developed for the novel word-forms through establishing new associations of the stimuli (Yue et al., 2014). In our study, MMN was not recorded while participants underwent the word learning paradigm. Future studies should replicate this FWL task while also recording participants' MMN responses to investigate how word learning affects ERP responses over a learning period. This could provide insight into how individuals learn novel words in a short period of time and how and where the brain represents this learning. We were also interested in exploring individual differences with the *early* deviant MMN response and temporal accuracy and word learning. The effect size of the MMN responses was calculated for the difference between the standards and deviants in the *early* block. The effect sizes did not show any significant correlations with either the TRA task or the FWL task. However, the weak correlations were expected, although we had a small sample size. A larger sample is needed to further explore the possibility of a relationship.

STM Tasks and Music Experience

Lastly, we explored whether there was an association between music experience and ability and the STM tasks. For the STM tasks, we created a sentence repetition composite with the sentence repetition tasks, and a TOJ composite score with the TOJ tasks. The years of music experience and level of music ability showed non-significant unexpected correlations with the sentence repetition score. There was also no significant correlation with the TRA task and the music experience or ability. However, there is little knowledge regarding the relationship between music training and auditory STM (e.g., Kraus et al., 2012; Ramachandra et al., 2012, as cited in Gordon et al., 2015). We also did not find significant correlations between the music factors and the TOJ score. This finding is open to interpretation as there are, as far as we know, no current findings on the relationship of nonverbal serial order and music. Perhaps if multisyllabic words were used, or sentences, as well as music stimuli, this would allow for better insight into the correlation between music training and rhythm. Similarly, the stimuli used in the phonological STM may not have allowed for participants to properly engage in verbal rhythm. The English nonword sentence repetition task was “not natural” to native English speakers (Harding et al., 2018; Meninghuas et al., 2015; Patel, 2011), and therefore the participants may have been more focused on trying to produce the words properly rather than the stress. That being said, the rhythm and prosody of the sentences were not controlled for, as was done in previous studies (e.g., Harding et al., 2018). Future studies could look into controlling for the rhythm of the stimuli and employ an additional music rhythm task to further tease apart the distinction between language and music rhythm processes.

General Discussion. To our knowledge, this is the first study to investigate how serial order embedded in sentences and serial order in the nonverbal domain can help us to understand and predict word learning in adult individuals. This study is also, as far as we know, the first to explore whether MMN responses to various deviant stimuli can predict one’s temporal accuracy and word learning ability. The STM tasks were employed to tap into temporal order, as it has been shown that temporal order predicts word learning (e.g., Majerus & Boukebza, 2013; Majerus et al., 2006; Schraeyen et al., 2019; Service & Craik, 1993). However, some of the tasks were performed at ceiling and did not show correlations with performance in the FWL task. This indicates that our tasks were not necessarily testing what we anticipated them to be testing. In saying this, we did find a significant correlation between the foreign sentence repetition task and the FWL task, which indicates that temporal order in a foreign language can predict how well one will learn words in that foreign language. This study also suggested the possibility that there is a relationship between temporal accuracy and word learning with one's peak MMN response, although this relationship was not significant and needs further exploration and understanding. Future studies should rebuild the STM tasks employed here to be more complex to see if the ease of the tasks is why we did not find significant correlations between these tasks that recruit serial order memory and word learning. Nonetheless, this study contributes intriguing findings to the field of language development and acquisition, as temporal and phonological accuracy of foreign sentence repetition predicted word learning, and neural responses to phonological and temporal deviants showed a relationship with temporal accuracy of tapping rhythms and the learning of foreign words. However, due to the small sample size in this study, interpretation and generalization of the results has to await testing with a larger sample of participants.

Conclusions

STM has been demonstrated to be able to record serial and temporal order, in research involving the use of rhythm and entrainment to stimuli (e.g., Gilbert et al., 2014; Henson et al., 2003; Majerus & Cowan, 2016), and this ordering of elements is not restricted to phonological material (e.g., Laasonen et al., 2012; Richie & Aten, 1976; Szmalec, Loncke, Page, & Duyck, 2011). The processing of serial phoneme order in repetition tasks has been demonstrated to predict individual ability in word learning (e.g., Schraeyen et al., 2019; Service & Craik, 1993). This study sought to understand the role of serial order in word learning and how this related to one's ability to perceive variation in speech. We found that one's ability to repeat English nonword sentences did not predict one's ability to learn foreign words. However, one's ability to repeat foreign language sentences did predict word learning. This suggests that in repetition tasks, exposure to a novel language, the prosodic pattern of the foreign language is stored temporarily in STM and appears to enhance learning of the foreign words later on. Therefore, the better one is at reproducing unfamiliar phonology, and holding the phonological representations in one's STM (Service & Craik, 1993), the better one is likely to be at learning words in that foreign language.

We also investigated if MMN peak amplitude and latency have a relationship with word learning and temporal accuracy. We found an expected relationship between the FWL task and the *early* deviant/difference wave peak amplitude, but an unexpected relationship with the *early* deviant difference wave peak latency. However, none of the correlations were significant. As the relationship between phonological STM with serial order and word learning has, to our knowledge, not yet been investigated with MMN responses, it is difficult to say what this indicates, as serial order needs to be further explored. In addition, temporal accuracy accounted for by the TRA task was found to have an expected relationship with the *early* deviant difference wave peak amplitude and latency. However, the *early* standard peak amplitude showed an unexpected relationship with the TRA task. Again, none of the correlations were significant, and this might have been a result of the TRA task not properly tapping into serial order memory, therefore further exploration is needed. Lastly, we explored whether there was a relationship between music experience and level of music ability and the STM tasks. The results showed no significant correlations. As there is little known regarding this relationship in the literature, further exploration is required for interpretation of results. In conclusion, the results demonstrate that serial order can aid in the understanding of word learning but needs further investigation with regards to non-linguistic stimuli and ERP responses in adult participants. Results from this study, and future studies adopting variations of these tasks, can be used to better understand language impairments and development as these findings have allowed us to explore language and memory abilities in typically developing individuals. Therefore, the results can be widened to individuals with various language impairments, such as those with developmental dyslexia, to further understand the core of atypical language development. Eventually, these tasks could be used as a standardized testing tool for therapists and psychometrists to perform various language and memory assessments.

References

Attout, L., Grégoire, C., & Majerus, S. (2020). How robust is the link between working memory for serial order and lexical skills in children?. *Cognitive Development, 53*, 100854.

Baddeley AD. (1986). *Working Memory*. Oxford, UK: Oxford Univ. Press

Baddeley, A. (2010). Working memory. *Current biology*, *20*(4), R136-R140.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, *63*, 1-29.

Baddeley, A.D., and Hitch, G.J. (1974).Working memory. In G.A. Bower (Ed.), Recent Advances in Learning and Motivation (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex*, *112*, 91-106.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological review*, *105*(1), 158.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586-595.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, *36*(2), 233-252.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, *14*(6), 575-589.

Bendixen, A., Schwartze, M., & Kotz, S. A. (2015). Temporal dynamics of contingency extraction from tonal and verbal auditory sequences. *Brain and language*, *148*, 64-73.

Bonetti, L., Haumann, N. T., Brattico, E., Kliuchko, M., Vuust, P., Särkämö, T., & Näätänen, R. (2018). Auditory sensory memory and working memory skills: Association between frontal MMN and performance scores. *Brain research*, *1700*, 86-98.

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological review*, *107*(1), 127.

Brückmann, M., & Garcia, M. V. (2020). Mismatch Negativity Elicited by Verbal and Nonverbal Stimuli: Comparison with Potential N1. *International Archives of Otorhinolaryngology*, *24*(02), e80-e85.

Buiatti, M., Peña, M., & Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*, *44*(2), 509-519.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological review*, *106*(3), 551.

Catalan, M. J., Honda, M., Weeks, R. A., Cohen, L. G., & Hallett, M. (1998). The functional neuroanatomy of simple and complex sequential finger movements: a PET study. *Brain: a journal of neurology*, *121*(2), 253-264.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*(7), 1086-1095.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British journal of psychology*, *55*(4), 429-432.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *cortex*, *45*(1), 119-130.

Corkin, S. (1974). Serial-ordering deficits in inferior readers. *Neuropsychologia*, *12*(3), 347-354.

Craik, F.I.M., and Lockhart, R.S. (1972). Levels of processing. A framework for memory research. J. Verb. Learn. Verb. Behav. 11, 671–684.

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, *112*(45), E6233-E6242.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage, 85*, 761-768.

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language*, *51*(1), 115-135.

FitzGerald, P., & Broadbent, D. E. (1985). Order of report and the structure of temporary memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(2), 217.

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., ... & Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in human neuroscience*, *8*, 392.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of memory and language*, *24*(2), 200-209.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83-94.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection?. *Journal of memory and language*, *29*(3), 336-360.

Geiser, E., Sandmann, P., Jäncke, L., & Meyer, M. (2010). Refinement of metre perception–training increases hierarchical metre processing. *European Journal of Neuroscience*, *32*(11), 1979-1985.

Gilbert, A. C., Boucher, V. J., & Jemel, B. (2014). Perceptual chunking and its effect on memory in speech processing: ERP and behavioral evidence. *Frontiers in psychology*, *5*, 220.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, *15*(4), 511.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in psychology*, *6*, 1777.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in cognitive sciences*, *15*(1), 3-10.

Goswami, U. (2015). Sensory theories of developmental dyslexia: three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43-54.

Goswami, U. (2016). Educational neuroscience: neural structure-mapping and the promise of oscillations. *Current opinion in behavioral sciences*, *10*, 89-96.

Goswami, U. (2018). A neural basis for phonological awareness? An oscillatory temporal-sampling perspective. *Current directions in psychological science*, *27*(1), 56-63.

Goswami, U., Barnes, L., Mead, N., Power, A. J., & Leong, V. (2016). Prosodic Similarity Effects in Short-Term Memory in Developmental Dyslexia. *Dyslexia*, *22*(4), 287-304.

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), 1-17.

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, *99*(16), 10911-10916.

Halsband, U., Ito, N., Tanji, J., & Freund, H. J. (1993). The role of premotor cortex and the supplementary motor area in the temporal control of movement in man. *Brain*, *116*(1), 243-266.

Hartley, T. (1995). *The role of syllable structure in verbal short-term memory* (Doctoral dissertation, University of London).

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of memory and language*, *35*(1), 1-31.

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135-178.

Harding, E. E., Sammler, D., Henry, M. J., Large, E. W., & Kotz, S. A. (2019). Cortical tracking of rhythm in music and speech. *NeuroImage*, *185*, 96-101.

Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: a common rhythm. *Frontiers in psychology*, *4*, 566.

Hay, J. F., & Saffran, J. R. (2012). Rhythmic grouping biases constrain infant statistical learning. *Infancy*, *17*(6), 610-641.

Henry, M. J., & Herrmann, B. (2014). Low-frequency neural oscillations support dynamic attending in temporal context. *Timing & Time Perception*, *2*(1), 62-86.

Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia*, *38*(4), 426-440.

Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology Section A*, *56*(8), 1307-1334.

Henny Yeung, H., Bhatara, A., & Nazzi, T. (2018). Learning a Phonological Contrast Modulates the Auditory Grouping of Rhythm. *Cognitive Science, 42*(6), 2000-2020.

Hitch, G., Burgess, N., Towse, J., & Culpin, V. (1996). Temporal grouping effects and working memory: The role of the phonological loop. *Quarterly Journal of Experimental Psychology A*, *49*, 116-139.

Hurlstone, M. J., & Hitch, G. J. (2014). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 295.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2013). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological bulletin*, *140*(2), 339.

Jones, B. (1974). Cross-modal matching by retarded and normal readers. *Bulletin of the Psychonomic Society*, *3*(3), 163-165.

Kathmann, N., Frodl-Bauch, T., & Hegerl, U. (1999). Stability of the mismatch negativity under different stimulus and attention conditions. *Clinical Neurophysiology*, *110*(2), 317-323.

Koelsch, S., & Gunter, T. C. v Cramon DY, Zysset S., Lohmann G., Friederici AD (2002). *Bach speaks: a cortical "language-network" serves the processing of music. Neuroimage*, *17*, 956-966.

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in cognitive sciences*, *22*(10), 896-910.

Kraus, N., Strait, D., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, *1252*(1), 100.

Laasonen, M., Virsu, V., Oinonen, S., Sandbacka, M., Salakari, A., & Service, E. (2012). Phonological and sensory short-term memory are correlates and both affected in developmental dyslexia. *Reading and writing*, *25*(9), 2247-2273.

Lahti-Nuuttila, P., Service, E., Smolander, S., Kunnari, S., Arkkila, E., & Laasonen, M. (submitted). Nonverbal serial short-term memory moderates vocabulary growth in children with developmental language disorder.

Langus, A., Mehler, J., & Nespor, M. (2016). Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews*, *81*, 158-166.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, *106*(1), 119.

Leclercq, A. L., & Majerus, S. (2010). Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study. *Developmental Psychology*, *46*(2), 417.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001-1010.

Lv, J. Y., Wang, T., Qiu, J., Feng, S. H., Tu, S., & Wei, D. T. (2010). The electrophysiological effect of working memory load on involuntary attention in an auditory–visual distraction paradigm: an ERP study. *Experimental brain research*, *205*(1), 81-86.

Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and language*, *153*, 13-19.

Majerus, S., & Boukebza, C. (2013). Short-term memory for serial order supports vocabulary development: New evidence from a novel word learning paradigm. *Journal of experimental child psychology*, *116*(4), 811-828.

Majerus, S., & Cowan, N. (2016). The nature of verbal short-term impairment in dyslexia: The importance of serial order. *Frontiers in Psychology*, *7*, 1522.

Majerus, S., Poncelet, M., Elsen, B., & Van der Linden, M. (2006). Exploring the relationship between new word learning and short-term memory for serial order recall, item recall, and item recognition. *European Journal of Cognitive Psychology*, *18*(6), 848-873.

Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, *23*(2), 294-305.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping*, *37*(8), 2767-2783.

Müller, D., & Schröger, E. (2007). Temporal grouping affects the automatic processing of deviant sounds. *Biological psychology*, *74*(3), 358-364.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*, *118*(12), 2544-2590.

Nittrouer, S. (2006). Children hear the forest. *The Journal of the Acoustical Society of America*, *120*(4), 1799-1802.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781. http://dx.doi.org/10.1037//0033-295X.105.4.761-781

Papagno, C., Comi, A., Riva, M., Bizzi, A., Vernice, M., Casarotti, A., ... & Bello, L. (2017). Mapping the brain network of the phonological loop. *Human Brain Mapping*, *38*(6), 3011-3024.

Parmentier, F. B., Andres, P., Elford, G., & Jones, D. M. (2006). Organization of visuo-spatial serial memory: Interaction of temporal order with spatial and temporal grouping. *Psychological research*, *70*(3), 200-217.

Parmentier, F. B., Maybery, M. T., & Jones, D. M. (2004). Temporal grouping in auditory spatial serial memory. *Psychonomic Bulletin & Review*, *11*(3), 501-507.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in psychology*, *2*, 142.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of cognitive neuroscience*, *10*(6), 717-733.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature neuroscience*, *6*(7), 688-691.

Perez, T. M., Majerus, S., Mahot, A., & Poncelet, M. (2012). Evidence for a specific impairment of serial order short-term memory in dyslexic children. *Dyslexia*, *18*(2), 94-109.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283-290.

Phillips, W. A., & Baddeley, A. D. (1971). Reaction time and short-term visual memory. *Psychonomic Science*, *22*(2), 73-74.

Pickering, S. J., Gathercole, S. E., & Peaker, S. M. (1998). Verbal and visuospatial short-term memory in children: Evidence for common and distinct mechanisms. *Memory & cognition*, *26*(6), 1117-1130.

Port, R. F. (2003). Meter and speech. *Journal of phonetics*, *31*(3-4), 599-611.

Power, A. J., Mead, N., Barnes, L., & Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Frontiers in human neuroscience*, *7*, 777.

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: sing along!. *Brain*, *129*(10), 2571-2584.

Ramachandra, V., Meighan, C., & Gradzki, J. (2012). The Impact of Musical Training on the Phonological Memory and the Central Executive: A Brief Report. *North American Journal of Psychology*, *14*(3).

Ramus, F., & Szenkovits, G. (2008). What phonological deficit?. *Quarterly journal of experimental psychology*, *61*(1), 129-141.

Rana, F., and colleagues. Manuscript in preparation.

Richie, D. J., & Aten, J. L. (1976). Auditory retention of nonverbal and verbal sequential stimuli in children with reading disabilities. *Journal of learning disabilities*, *9*(5), 312-318.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019.

Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*(2), 137-147.

Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *The Quarterly Journal of Experimental Psychology*, *21*(2), 148-155.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of verbal learning and verbal behavior, 21*(2), 150-164.

Schaadt, G., Pannekamp, A., & Meer, E. V. D. (2014). Gaining mismatch negativity! Improving auditory phoneme discrimination by literacy training–A pre-post event-related potential study. *Int J Sch Cogn Psychol*, *1*(101), 2.

Schlaug, G., Norton, A., Marchina, S., Zipse, L., & Wan, C. Y. (2010). From singing to speaking: facilitating recovery from nonfluent aphasia. *Future neurology, 5*(5), 657-665.

Schraeyen, K., Van der Elst, W., Geudens, A., Ghesquière, P., & Sandra, D. (2019). Short-term memory problems for phonemes' serial order in adults with dyslexia: evidence from a different analysis of the nonword repetition task *Item and serial order retention in phonological short-term memory representations in relation to reading performance*, 37.

Schwade, L. F., Didoné, D. D., & Sleifer, P. (2017). Auditory evoked potential mismatch negativity in normal-hearing adults. *International Archives of Otorhinolaryngology*, *21*(03), 232-238.

Scott, S. K. (1998). The point of P-centres. *Psychological Research, 61*(1), 4-11.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology Section A*, *45*(1), 21-50.

Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, *32*(5), 608-623.

Service, E, & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition?. *Applied Psycholinguistics*, *16*(02), 155-172.

Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. *Memory*, *15*(1), 34-56.

Smith-Spark, J., Fisk, J., Fawcett, A., & Nicolson, R. (2003). Investigating the central executive in adult dyslexics: Evidence from phonological and visuospatial working memory performance. *European Journal of Cognitive Psychology*, *15*(4), 567-587.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R., & Geyer, S. (2011). Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain*, *134*(10), 3083-3093.

Sussman, E. S. (2005). Integration and segregation in auditory scene analysis. *The Journal of the Acoustical Society of America*, *117*(3), 1285-1298.

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, *179*, 23-36.

Szmalec, A., Loncke, M., Page, M., & Duyck, W. (2011). Order or disorder? Impaired Hebb learning in dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1270.

Thomson, J. M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of research in reading*, *29*(3), 334-348.

Thomson, J. M., Richardson, U., & Goswami, U. (2005). Phonological similarity neighborhoods and children’s short-term memory: Typical development and dyslexia. *Memory & Cognition*, *33*(7), 1210-1219.

Thomson, J. M., & Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *Journal of Physiology-Paris*, *102*(1-3), 120-129.

Tierney, A., White-Schwoch, T., MacLean, J., & Kraus, N. (2017). Individual differences in rhythm skills: Links with neural consistency and linguistic ability. *Journal of Cognitive Neuroscience*, *29*(5), 855-868.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*(2), 145-161.

Vousden, J. I., Brown, G. D., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive psychology*, *41*(2), 101-175.

Wang, Y., Mauer, M. V., Raney, T., Peysakhovich, B., Becker, B. L., Sliva, D. D., & Gaab, N. (2017). Development of tract-specific white matter pathways during early reading development in at-risk children and typical controls. *Cerebral Cortex*, *27*(4), 2469-2485.

Wolff, P. H. (2002). Timing precision and rhythm in developmental dyslexia. *Reading and Writing*, *15*(1-2), 179-206.

Yue, J., Bastiaanse, R., & Alter, K. (2014). Cortical plasticity induced by rapid Hebbian learning of novel tonal word-forms: Evidence from mismatch negativity. *Brain and language*, *139*, 10-22.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, *131*(1), 3.

Appendix A

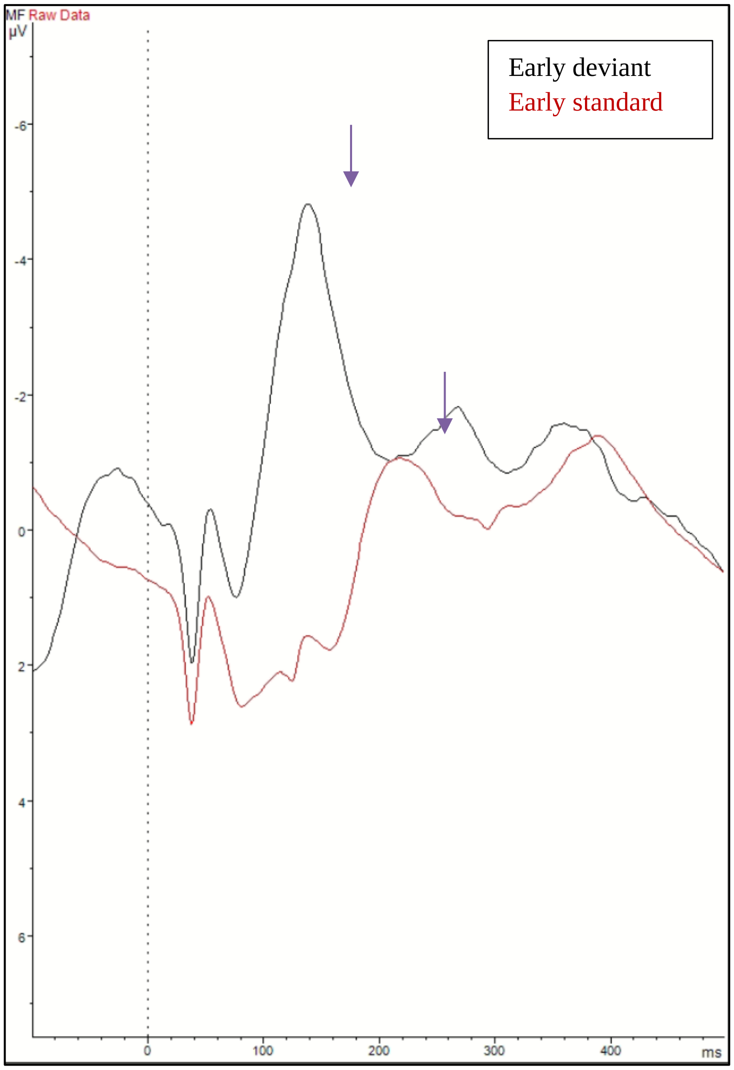


Figure A1: Grand average waveform for MF ROI for the early deviant and early standard. *Note:* the arrow indicates the peak MMN response.

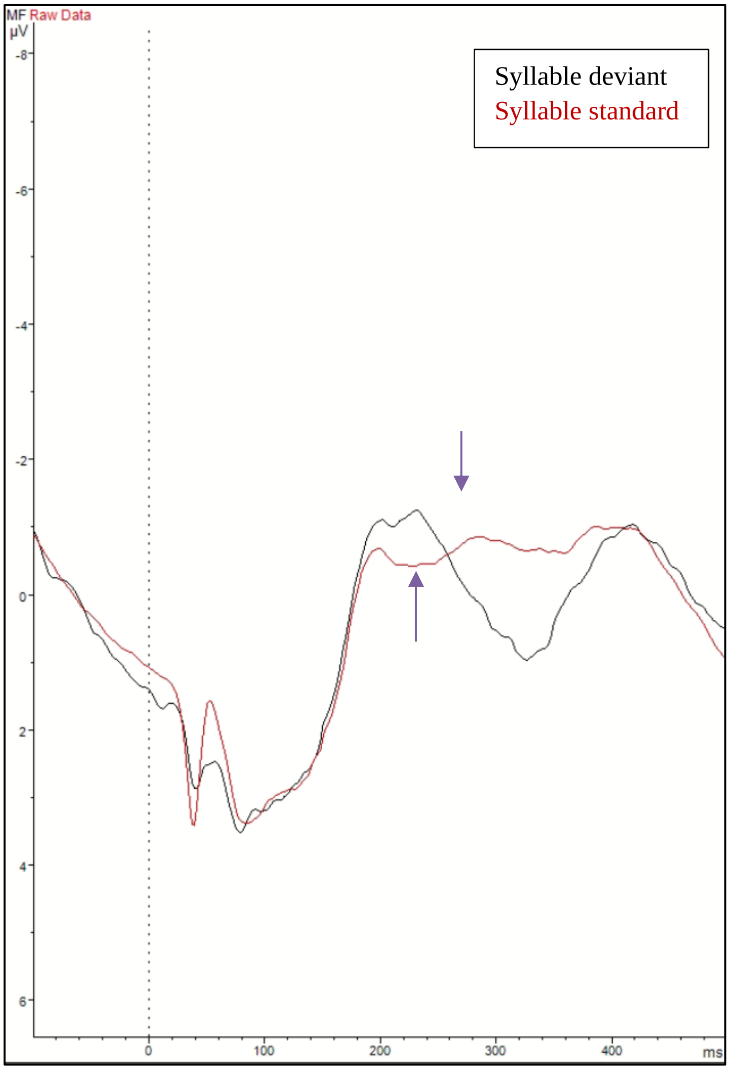


Figure A2: Grand average waveform for MF ROI for the syllable deviant and syllable standard. *Note:* the arrow indicates the peak MMN response.

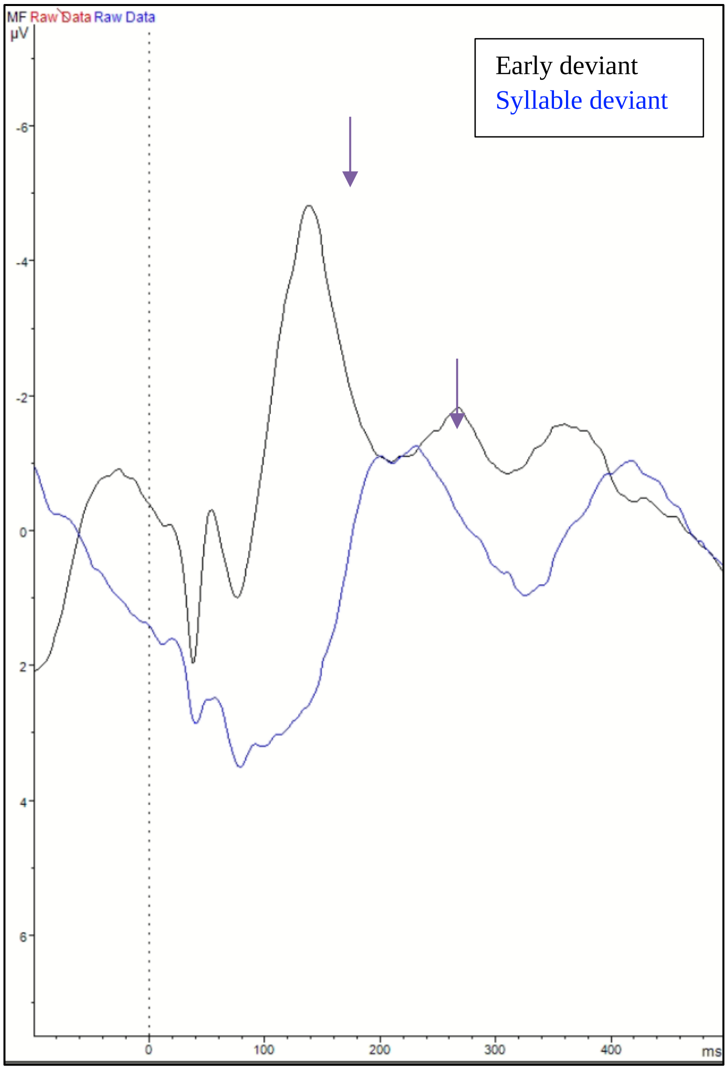


Figure A3: Grand average waveform for MF ROI for the early deviant and syllable deviant. *Note:* the arrow indicates the peak MMN response.

Appendix B

First, we explored the correlation between the *early* deviant amplitude and the *syllable* deviant amplitude to determine if there is a relationship between time expectancy responses (*early* deviant) and phonological difference responses (*syllable* deviant) (**Table B1**). The amplitudes did not significantly correlate, so they were analyzed separately.

Table B1: Correlation matrix for the amplitude of difference waves of each deviant. *N = 24*.

		Early Deviant Amplitude	**Syllable Deviant Amplitude**
Early Deviant Amplitude	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
Syllable Deviant Amplitude	Pearson's *r*	0.088	—
	p-value	0.682	—
	95% CI Upper	0.475	—
	95% CI Lower	-0.327	—
	Spearman's *rho*	0.063	—
	p-value	0.768	—

Note. * p < .05, ** p < .01, *** p < .001

We then correlated the *early* deviant latency and the syllable deviant latency (**Table B2**). The latencies did not significantly correlate, so they will be analyzed separately.

Table B2: Correlation matrix for the latencies of the difference waves of each deviant. *N = 24*.

		Early Deviant Latency	**Syllable Deviant Latency**
Early Deviant Latency	Pearson's *r*	—	
	p-value	—	
	95% CI Upper	—	
	95% CI Lower	—	
	Spearman's *rho*	—	
	p-value	—	
Syllable Deviant Latency	Pearson's *r*	0.291	—
	p-value	0.167	—
	95% CI Upper	0.622	—
	95% CI Lower	-0.127	—
	Spearman's *rho*	0.272	—
	p-value	0.198	—

Note. * p < .05, ** p < .01, *** p < .001

Next, we explored the correlation between the amplitude of the difference waves of each deviant and the years of music experience and the level of music ability (**Table B3**). The results show no significant correlation between music and the amplitude of either deviant. This suggests that the peak of one's MMN response may not rely on one's music training and/or music abilities.

Table B3: Correlation matrix for the amplitude of the difference waves of both deviants and the years of music experience and the level of music ability. *N = 23*.

		Early Deviant Amplitude	**Syllable Deviant Amplitude**	**Years of Music Experience**	**Level of Music Ability**
Early Deviant Amplitude	Spearman's *rho*	—			
	p-value	—			
Syllable Deviant Amplitude	Spearman's *rho*	0.037	—		
	p-value	1.00	—		
Years of Music Experience	Spearman's *rho*	0.071	-0.074	—	
	p-value	0.748	1.00	—	
Level of Music Ability	Spearman's *rho*	-0.019	-0.064	0.784 ***	—
	p-value	1.00	1.00	< .001	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

Then, we correlated the latency of the difference waves of each deviant with the years of music experience and level of music ability (**Table B4**). The results show no significant correlations between the latency of either deviant and music experience and ability. This suggests that music experience and ability may not have an effect on the latency of one's MMN response to the stimuli used in this study.

Table B4: Correlation matrix for the latency of difference waves of both deviants and the years of music experience and the level of music ability. *N = 23*.

		Early Deviant Latency	**Syllable Deviant Latency**	**Years of Music Experience**	**Level of Music Ability**
Early Deviant Latency	Spearman's *rho*	—			
	p-value	—			
Syllable Deviant Latency	Spearman's *rho*	0.265	—		
	p-value	1.00	—		
Years of Music Experience	Spearman's *rho*	0.189	-0.087	—	
	p-value	1.00	1.00	—	
Level of Music Ability	Spearman's *rho*	0.052	0.157	0.784 ***	—
	p-value	1.00	1.00	< .001	—

Note. * p < .05, ** p < .01, *** p < .001. *All p-values are Holm's corrected.*

With regards to the MMN, no significant correlations were found with years of music nor level of music ability with any of the amplitudes and latencies of the deviant difference waves. This is not a surprising finding as music training is seen to increase cortical responses to music rhythm but not speech (e.g., Harding et al., 2018). However, the ERP stimuli employed in this study were single syllables and, therefore, did not allow for an in-depth exploration of rhythm. Studies could incorporate a rhythmic component into the MMN paradigm to gain a better insight into music ability and MMN responses to speech and music rhythm.