UNEXPECTED TIMING IN LINGUISTIC AND NON-LINGUISTIC SEQUENCES

PROCESSING OF UNEXPECTED STIMULUS TIMING IN LINGUISTIC AND NON-LINGUISTIC SEQUENCES

By FAREEHA SHAHID RANA, B.Sc. Hons., M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2022) Hamilton, Ontario (Cognitive Science of Language)

TITLE: Processing of Unexpected Stimulus Timing in Linguistic and Non-Linguistic Sequences AUTHOR: Fareeha Shahid Rana, B.Sc. Hons. (McMaster University), M.Sc. (McMaster University) SUPERVISORS: Dr. Elisabet Service and Dr. John F. Connolly NUMBER OF PAGES: xviii, 167

Lay Abstract

This thesis examines how unexpectedly shorter or longer pauses in speech may affect speech comprehension. Specifically, the research reported here examined how stimuli that are presented unexpectedly early or unexpectedly late in a steady-rate sequence impact a listener's attention and memory. Although the speed at which we speak has been studied previously, work on unexpected changes in its timing has been limited. This research begins to explore this aspect of speech processing. It contributes to our understanding of how speech timing is processed in two important ways. First, we found that unexpected delays in both non-linguistic sounds and syllables were less noticeable than unexpectedly early presentations, when participants were not required to pay attention to them. Second, we found that unexpected delays made recognition memory for stimulus order worse. Overall, the results of these studies indicate that unexpected delays in the rhythm of speech make it more difficult to understand.

Abstract

Timing, and ergo rhythm, are intrinsic features of language that help facilitate real-time speech comprehension. However, work exploring how variable timing is processed in speech is limited. This dissertation addresses this gap in literature by exploring the tenets of how temporal variability is cognitively processed, particularly in the context of real-time stimulus processing. This research is one of the first works to examine temporal variations in linguistic and other acoustically complex contexts.

Using electroencephalography (EEG) and behavioural methods, participants were tested on their perception of temporal variations within a continuous stream of either simple tones, complex waves, or syllables. Two timing deviants were presented that occurred *early* or *late* compared to other stimuli in the sequence. Event-related potentials (ERPs) were recorded for each stimulus type across three experiments. A fourth experiment tested participant recognition memory for syllable order.

Results showed differential processing between the two timing deviants. Unexpectedly earlier tokens elicited larger pre-attentive responses compared to late, suggesting a saliency for the earlier tokens that was not present for the delayed ones. This pattern was observed across all three levels of acoustic and linguistic complexity. Compared to sequences with no timing deviants or an early timing deviant, unexpectedly late tokens were more detrimental to memory, suggesting a negative impact of delays on verbal recognition. Thus, not only were early and late timing variations processed differently, but delays in continuous sequences were also more cognitively taxing for working memory.

The results reported in this dissertation contribute to existing knowledge by enriching our understanding of the fundamentals of how aspects of prosodic timing may affect attention and memory. Additionally, it provides new insights into how speech synthesis can be used in neurolinguistic research by tracking how neurophysiological responses change with increasing acoustic complexity and linguistic familiarity.

Impact of COVID-19

The COVID-19 pandemic led to a substantial delay in the progress of this thesis. There was a direct, measurable 10-month impact on recruitment, data collection, and Ethics clearance processes, as well as a longer, indirect impact in terms of the amendments to documentations and Ethics applications to incorporate new COVID-19 testing protocols, and new programming skills and analyses that had to be acquired to continue the research.

At the beginning of the pandemic, all recruitment for in-person testing was cancelled and delayed until the institution re-instated in-person testing. As this dissertation was using electroencephalography (EEG) as the tool for investigation, testing was halted indefinitely. Recruitment and testing took seven months to get restarted. This was done immediately when the university allowed it.

The original research plan, which would have continued the use of neurophysiological tools (EEG) had to be abandoned amid the uncertainty of in-person testing. A newer research plan was drafted that still incorporated an EEG element, but was complemented with results from behavioural tasks that could be tested online. A delay in the Ethics application process—related directly to the use of EEG in the application and pandemic-related concerns about in-person testing—meant that also this plan had to be abandoned, and only the first originally planned task was run (reported here as Chapter 4). The final thesis consists of therefore, three experiments that make use of EEG and a final experiment that involves a behavioural task.

Acknowledgements

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

This dissertation would not exist in the form it does today without the unwavering support of so many people.

First, to my supervisors Dr. Elisabet Service and Dr. John Connolly, I am extremely thankful that you agreed to take me on as a student all those years ago. Dr. Connolly—your insight and support has been an integral part of my growth as a Ph.D. student, and I am so grateful to have had the chance to work in your lab and learn from you. Dr. Service—I cannot thank you enough for all your knowledge, insight, and support. The pandemic made it harder, but the fact that your office door was always open for a quick question or a chat has always meant a lot to me. I feel very lucky to have learned so much from you and to have worked so closely with you over these years.

To Dr. Daniel Pape, who has been an integral part of my committee and research: thank you. Your friendly words of encouragement have always come at the right time, and I have learned so much working with you. I am very grateful to have had you on my committee.

There are too many people who have been a part of this journey with me, and it would be a disservice to say that I have named them all. Lab mates, students and friends, past and present—thank you for your help, insight, support, and encouragement. Also a big thank you to the countless RAs who helped me sort stimuli and pilot my experiments at various stages—you are the true backbone of research.

Special thanks to my EEG folks: Dr. Rober Boshra (for being generous with his time and expertise), Dr. Gaisha Oralova, Nathalee Ewers, Dr. Richard Mah, Dr. Kyle Ruiter, Adianes Herrera; it truly takes a village, and I have learned so much from you all.

And to my Memory folks, and particularly my LMBLadies: Laura Beaudin, Chelsea Whitwell, Narcisse Torshizi, Bre-Anna Owusu—thank you for always being encouraging, supportive, and having a sympathetic ear to lend. I truly am lucky to have shared an office with you, and I certainly would not have been able to do this without your friendship.

This section would be incomplete without mentioning Chia-Yu Lin, friend and (former) lab manager at ARiEAL. Your advice, help, and friendship has been an invaluable part of my time at ARiEAL and McMaster University; thank you.

I would also like to acknowledge all the faculty and research colleagues I have had the pleasure of working with in the department over the years. The end of this degree marks the end of my seventh year with this department (and eleventh at McMaster University), which means this number is not small! I am immensely thankful to have been part of such a welcoming community.

Finally, to my family, who did the most to support me in every way possible. Particularly my parents, whose prayers and belief in me got me this far, and my brother, who took the time to read every word I wrote and refused to let me falter. Thank you. Words are not enough to express my gratitude (but they will have to do for now).

Table of Contents

List of Figures

List of tables

List of Abbreviations and Symbols

μV - microvolt

ANOVA – Analysis of Variance

CQ – Competitive Queuing model

C-SOB – Competitive Serial-Order-in-a-Box

CV – Consonant-Vowel

DAT – Dynamic Attending Theory

DLD – Developmental Language Disorders

DRL – Driven Right-Leg

DRN – Deviant-Related Negativity

EEG – Electroencephalography

EOG – Electrooculogram

ERP – Event-Related Potential or Evoked Reaction Potential

GLMM – Generalised Linear Mixed Model

ICA – Independent Component Analysis

ISI – Inter-Stimulus Interval

ITI – Inter-Trial Intervals

LC – Left Central

LDN – Late Discriminative Negativity

LF – Left Frontal

LP – Left Parietal

LPC – Linear Predictive Coding

MC – Middle Central

MF – Middle Frontal

MMN – Mismatch Negativity

MP – Middle Parietal

MREB – McMaster Research Ethics Board

ms - millisecond

NSW – Negative Slow Wave

OS – Operating System

OSCAR – Oscillator-Based Associative Recall model

PC – Predictive Coding

PC-DIM – Predictive Coding/Biased Competition-Divisive Input Modulation model

PMN – Phonological Mapping Negativity

PNB – Psychology, Neuroscience, and Behaviour

PPE – Personal Protective Equipment

PVI – Pairwise Variability Index

RC – Right Central

RF – Right Frontal

ROI – Region of Interest

RP – Right Parietal

RT – Reaction Time

s - second

SOA – Stimulus Onset Asynchrony

SONA – Student Research Participation systems

SPL – Sound Pressure Level

SPS – Syntactic Positive Shift

SSA – Stimulus-Specific Adaptation Model

TBRS – Time-Based Resource Sharing model

TP – Transitional Probability

TWI – Temporal Window of Integration

WM – Working Memory

CHAPTER ONE: INTRODUCTION

Introduction

Discussing speech and all it entails is difficult in a written format. While differences between examples like “This is a sentence?” and “This is a sentence!” are self-evident, no other information beyond the punctuation is available to the reader from the text alone. However, said aloud, it is obvious that the cadence of a question is quite different from that of an exclamation. The change in how these sentences are said has less to do with the punctuation, and more with the subtle variations inherent in the way they are communicated. For example, changing the words that are emphasised, or altering stress placement or intonation. These variations manifest as the *prosody* of the sentence via differences in parameters like loudness and pitch. These features affect the way a sentence is spoken, perceived, and understood. Changes in the underlying timing and rhythm of sentences bring the words to life and give them meaning that is not adequately conveyed through semantics alone. The inherent rhythm present in language is partly what helps the listener to keep track of a conversation and have an idea of what might come next.

Predicting what will follow in a sentence is an innate ability that relies on a combination of knowing what could come next (based on your experience with similarly structured sentences), and the likelihood of those words occurring depending on all types of pragmatic, semantic, morphological, phonological, and prosodic contexts. Rhythm and prosodic cues lend support to both these factors. Acoustic features such as placement of stress, for example, then help favour one prediction over another and ultimately lead to a facilitation of real-time speech processing. The *predictive coding* theory, in its essence, states that predictions are based on the sum of your experiences. On the role of phonology and syntax in prediction, one author notes that, “Selkirk [1986] carefully identifies multiple hierarchical levels ranging from syllables to phrases to sentences (…), but neither rhythm nor relative timing among hierarchical levels is accommodated.” (Jones, 2018, p. 263). These are likely to provide an integral aspect of information for hierarchical processing in speech. The current dissertation aims to redress this deficiency and add to the limited literature on what role timing and its variation plays in speech comprehension.

Background

Rhythm in speech and language is so ubiquitous and natural that we only notice it in its absence. Monotonous speech becomes boring speech which is harder to focus on than speech that varies in time, pace, and pitch. This dissertation takes a closer look at these time variations in the context of auditory tone and speech sequences to gain a better understanding of how they contribute to its stimulus prediction and, ultimately, speech comprehension. This dissertation employs both neurophysiological and behavioural tools to explore the interaction of acoustic complexity with temporal variation and to examine how both these things affect language processing. With a better understanding of speech rhythm and time in general, we hope to further our understanding of language comprehension and eventually acquisition. Knowledge of the important rhythmic features in speech can help in both the creation and the study of tools and aids that can improve accessibility and train communication skills. These will have applications involving individuals who struggle with language and communication, as well as those engaging in second language learning.

The idea of timing in speech tends to suggest a concept of pace—how quickly or slowly a speaker is communicating. Indeed, the rate at which various elements of speech are presented and comprehended plays a significant role in communication (Liberman & Whalen, 2000). However, there are many more subtle changes in timing that occur during speech that are separate from its rate. Earlier work that has examined aspects of speech rhythm falls into two categories: work on segmental timing differences with speech stimuli, and work on speech comprehension by altering its timing-related acoustic characteristics. The aim of this dissertation is to bridge the gap between the two areas of study and assess how timing differences influence speech comprehension.

Timing in language

Variation in language is an intrinsic feature of communication that we use to understand real-time speech. Variation can be of many types: phonemic variation, morphological variation, variation in prosody or the speed at which speech is produced. There is a wider issue at hand when looking at the temporal structure of language and all that it entails, however. This includes questions of what keeps speech ordered in time, and how the intrinsic rhythm or phonological patterning of incoming speech allows listeners to remain tethered to the stream of information. Changes to the order of information, such as switching the place of a noun and an adjective, lead to changes in the way the sentence is spoken and impact its rhythm and prosody. These affect the various suprasegmental cues that are critical to speech comprehension (Kotz et al., 2018; Shannon et al., 1995). Changes in these features can stretch over multiple phonemic segments. Known as prosody, these cues can be quantified as changes in the acoustic features of fundamental frequency (perceived as voice pitch), as well as syllabic stress, duration, or tempo. However, this is an insufficient description of prosody as it is limited in accounting for prosodic function, i.e., how lexical items in a sentence relate to each other semantically, syntactically, and pragmatically (Wagner & Watson, 2010). Accounting for both prosodic form and function requires a breadth of topic that is beyond the scope of this work. Nevertheless, a brief discussion on how the encompassing features of prosody and how form and function may coincide follows. The focus of this dissertation will remain on prosodic form, and particularly on how changing temporal cues may affect speech comprehension.

Variations in the phonological units of speech give rise to the prosodic profile of an utterance, and often result in the formation of groups or subgroups within a sentence or phrase comprising of syllables and words (Langus et al., 2012; Lehiste, 1973; Wagner & Watson, 2010). These groups cluster around acoustic cues like changes in pitch, duration, or intensity, and allow inferences to be made about the underlying syntactic structure of the sentence. These prosodic groupings or their boundaries have been suggested by some to exist in a hierarchy that builds up from the smallest prosodic unit (syllable) to the biggest (whole utterance) (e.g., Selkirk, 1986). Prosodic rhythm has been hypothesised to be critical in acquiring language (Curtin, 2010; Gervain, 2018; Langus et al., 2017). Infants recognise the prosody of their native language, and near-term foetuses have been shown to respond to it (Granier-Deferre et al., 2011; Jusczyk et al., 1993; Mehler et al., 1988; Ramus & Mehler, 1999). Differences in prosody are used by babies and adults alike to facilitate word segmentation and recognise speech (Bion et al., 2010; Cohen et al., 2001; Dahan, 2015; Dellwo, 2008; Dilley & McAuley, 2008; Remez et al., 1981; Seidl, 2007; Whalley & Hansen, 2006). Prosody is also stored in verbal memory, acting as an aid to recall language and encode it (Cohen et al., 2001). Prosodic cues help contribute important timing information to the speech envelope, making them a crucial aspect of language processing and comprehension (Rosen, 1992).

Interest in the purpose of these prosodic features, and the intentionality of the speaker have lately attracted increased attention. While the latter is a new area of research (Hellbernd & Sammler, 2016), the study of how changes in prosodic features relate to lexical, semantic, or pragmatic differences has a long history. For example, pause duration can be subconsciously manipulated to signal the end of a phrase or utterance (Choi, 2003; Tseng & Fu, 2005). Pragmatically, shifts in pitch and intonation are used to signal changes in the speaker’s emotions, or even to ask a question (Cole, 2015). Overall, therefore, there is an interplay between the acoustic cues that define prosody (of particular interest in the field of cognitive neuroscience, which demands measurable inputs to make inferences), and the underlying motivations for changes to prosodic feature values. Variation introduced in speech due to changes in pace and rhythm may be one key to understanding language comprehension and processing and possible deficits therein.

Timing in typical language development

Timing variation in the context of this dissertation is defined as differences in duration of intervals *between* syllables (or segments of speech). We get insight into how rhythmic (or prosodic) variations help or harm real-time language comprehension by exploring how variations in timing between the onsets of linguistic elements are perceived and processed. Differences in these cues can further be exploited to facilitate language acquisition. Language impairments have been shown to be strongly associated with rhythmic deficits (e.g., Goswami, 2018; Goswami et al., 2013; Lallier et al., 2013). Therefore, interventions as simple as improving the underlying temporal characteristics of speech-to-text software or introducing rhythm-based exercises can be introduced to improve a sense of timing in language for those who struggle with it. This section will review how rhythm is defined and studied in typical language development before discussing its implications in terms of music and atypical language development.

Timing and rhythm are important aspects of spoken language comprehension, both perceptually (Kotz et al., 2018) and neurophysiologically (Näätänen & Winkler, 1999). According to the rhythm class theory, most languages of the world can be classified into one of three rhythm classes: stress-timed, syllable-timed, or mora-timed (Abercrombie, 1967, as cited by Cummins, 2012; see also Bertinetto & Bertini, 2010; and Grabe & Low, 2002, for a review). Depending on the rhythm class, either stressed units of speech, syllables, or morae are separated by equal intervals of time, making their presentation *isochronous*. Objective measures of rhythm such as the percentage of vowels present and vocalic and consonant duration among others, have been used to categorise languages and validate these rhythmic classes with varying results (Dellwo, 2008; Dellwo & Wagner, 2003; Knight, 2011; Ramus et al., 1999). While some studies have shown successful classification of languages into the rhythm classes using these metrics (e.g., Gibbon & Gut, 2001; Ling et al., 2000), others have faced difficulties in doing the same (e.g., Rathcke & Smith, 2015). In some cases, the classification of languages into rhythm classes has been questioned altogether, with the authors concluding that "…listeners do not have a categorical perceptual sensitivity to rhythm class [… but rather] systematically exploit a range of timing cues to language differences […]" (White et al., 2012, p.14). Variation in vowel duration has been found to be one of the reliable metrics that can be used to quantify rhythm between languages (Knight, 2011). Changes in speech rate and rhythm are often modelled through introducing differences in the duration of consonant or vowel length, in fact, and have reliably been used to study the ability to differentiate between languages (Dellwo, 2008; Dellwo et al., 2015; Dellwo & Wagner, 2003; White et al., 2012). The desire to understand and explain this aspect of language remains constant between studies exploring rhythm and time.

One critical assumption underpinning the rhythm class hypothesis is the role of isochrony in languages. It posits that a repetitive, equally timed pace is the optimal rhythm in language because it allows predictability (and ergo, comprehension) in the listeners. Despite the popularity of the rhythm class hypothesis being used to describe the underlying rhythm of languages, this inherent isochrony assumption has been empirically proven to be false (Cummins, 2012). Speech is anything but periodic, often being described instead as quasi-rhythmic (Aubanel & Schwartz, 2020; Kotz et al., 2018; Poeppel & Assaneo, 2020). Perceptually, naturally timed speech is easier to process in comparison to isochronous speech (Aubanel & Schwartz, 2020). Inter-speaker variation is observed in addition to the intrinsic contrasts that exist between languages. Significant differences in speech rate between speakers were found in a study examining within- and between-speaker speech variability (Dellwo et al., 2015). Perception of these rate differences depended on the native language of the speaker, with "…the percept of 'regularity' and 'irregularity' in syllable- and stress-timed languages respectively […] caused by rate differences between these languages" (Dellwo, 2008, p. 378). It seems, therefore, that it would be most accurate to describe speech as having different rhythms in different contexts. Relatively unpredictable timing is used in conversation, whereas more periodic, rhythmic speech would be favoured with infants, and even more isochronous speech is used in a medium such as rap for artistic effect (Kotz et al., 2018).

Rhythm in speech and music

The discussion so far around rhythm has centred around changes in stress patterns, or differences in phonemic durations specifically in speech and language. The Merriam-Webster dictionary defines rhythm as "a regular, repeated pattern of sounds or movements" (Merriam-Webster, n.d., Essential Meaning 1). This definition is often understood in relation to regularity in music. However, rhythm, music, and speech are not limited to association only as a by-product of variation in the acoustic features of each. A repeated loop of speech and non-speech (environmental sounds) was found to result in the perception of musicality for both, suggesting a natural bias towards detecting rhythm even within tokens that are not typical rhythmic events (Rowland et al., 2019). Additionally, music, speech, and sensorimotor skills have all been found to correlate with each other with rhythm as the underlying unifying element.

Speech and music are often associated through shared elements like stress and meter (Haegens & Golumbic, 2018), and strong correlations have been found linking linguistic and musical abilities (Tillman, 2012). Neural resources between the rhythmic elements of speech and music have been found to be shared (Magne et al., 2016). Furthermore, participants' ability to replicate the rhythm of an incoming auditory sequence by tapping has been associated with how well they perceive the temporal pattern, with better tapping and coordination associated with a stronger sense of rhythm (Elliott et al., 2014; Nozaradan et al., 2016). Another study found that participants were more consistent in tapping along to a sentence with metrical accents when the metrical rhythm aligned with syntactic cues than when it did not (Hilton & Goldwater, 2021). All three of these skill domains—music, speech, and motor movement—rely on an internal sense of regularity, and deficits in one skill often suggest deficits in another. For example, children with developmental language disorders (DLDs) who struggle with language are more likely to struggle with deficits in music, rhythm, or motor skills (Jentschke et al., 2008; Kujala & Leminen, 2017; Law et al., 2014; Molinaro et al., 2016). Similarly, children with poorer rhythm abilities are at greater risk of developing speech or language disorders (Ladányi et al., 2020). Looking at these various aspects holistically is the best way to understand the association and relation of the common denominator—rhythm—and how it might help improve and reduce deficits.

Rhythm and speech perception in dyslexia

Impaired perception of the rhythmic aspects of speech (e.g., stress) have been observed in individuals with dyslexia, and aspects of rhythmic ability have been posited to be predictors for both DLD and dyslexia (Goswami, 2018; Goswami et al., 2013; Lallier et al., 2013, 2017, 2018; Molinaro et al., 2016). Previous work has shown that dyslexic children have poorer beat perception and replication compared to non-dyslexic children (Colling et al., 2017; Huss et al., 2011), and that they are less sensitive to word stress when compared to age-matched controls (Goswami et al., 2013). Intervention training involving rhythm and timing perception may therefore be able to help counter deficits in language that are directly associated with deficits in rhythmic processing (Kraus & White-Schwoch, 2015).

The neural signal itself also contains rhythm elements in terms of regular peaks observable in the signal. *Neural oscillations* are pre-existing oscillatory signals in the brain which can synchronize with (or *entrain* to) incoming sounds, thereby better attending to that stimulus (Bastiaansen et al., 2012; Poeppel & Assaneo, 2020; Jones, 2018. Goswami (2016, 2018) argue that neuronal oscillations are a powerful tool to understand language and language development. Citing evidence from previous studies (e.g., Goswami et al., 2013), Goswami et al. (2016) state that “amplitude envelope ‘rise time’ and measures of phonological awareness [are] unique predictors of prosodic sensitivity” (p. 289). The amplitude envelope rise time refers to how steeply (or otherwise) the oscillatory pattern forms, and by combining this with predictive measures of phonological awareness, rhythm can be used as an adept way to understand speech processing and comprehension. Specifically, considering that developmental dyslexia seems to be linked to a deficit in phonological processing (see Ziegler & Goswami, 2005, for a review), the oscillatory perspective is valuable in allowing us to begin to understand *how* and *what* roles rhythmic elements of language might play in dyslexia. Dyslexic children have shown poorer entrainment to speech compared to typically developing children, suggesting that it is difficult for the dyslexic brain to synchronise with certain rhythms of speech (Goswami et al., 2010; Molinaro et al., 2016). Understanding how neural responses vary for typical compared to non-typical language speakers in relation to rhythmic processing can be critical in identifying the source of developmental language deficits.

Aside from differences in neural oscillations, dyslexic adults also show a different pattern of neural responses during the pre-attentive processing of tones and syllables compared to that shown by non-dyslexic adults. Previous work found that dyslexic adults have deficits in pitch discrimination compared to controls, but not other acoustic features like duration (Kujala et al., 2006). Results to more complex stimuli, like syllables, are mixed. In one study (Sebastian & Yasin, 2008), dyslexic individuals showed diminished pre-attentive neural responses for tones compared to non-dyslexic individuals but had comparable brain responses to syllables. However, other work on syllables has found that dyslexic individuals reliably elicit smaller responses to changes in syllables comprising of stop consonants (e.g., the sounds /p b t d k g/ in Canadian English) (e.g., Schulte-Körne et al., 1998; see Näätänen et al., 2019). While a significant amount of work has been conducted with dyslexia, diminished auditory processing has been found in children with DLDs as well, suggesting that certain neural responses (e.g., the *mismatch negativity*, or MMN) can function as a marker for these language development disorders in certain contexts (Näätänen et al., 2019). This dissertation looks at changes in neural responses to unexpected variations in timing across simple tones, complex waves, and syllables with the aim to understand how underlying auditory processing changes with increasing speechlikeness and how this could help us better understand deficits in language development.

Electroencephalography as a neurophysiological cognitive tool

Electroencephalography, or EEG, is a non-invasive neurophysiological tool that allows a direct look into changes in brain activity on the scale of milliseconds (Kappenman & Luck, 2012; Luck, 2014). This makes it a particularly useful tool to provide insight into the fast, dynamic changes within the neural activity of the brain. EEG is recorded via electrodes that are placed on the scalp. These electrodes pick up additive post-synaptic activity from neurons. Neural data has been studied along two main sets of measures: event-related potentials (ERPs), and changes in synchronised neural oscillations.

Event-related potentials (ERPs)

Event-related potentials or *evoked reaction potentials* (ERP) reflect directed neural activity elicited in response to particular stimulus types. Kappenman & Luck define*"…[an] ERP component as a scalp-recorded voltage change that reflects a specific neural or psychological process*" (2012; p. 3). This definition, while succinct and simple, is hardly the only one that exists, as the relationship between the stimulus being used to evoke the ERP and the underlying process that it represents is not straight-forward or clear-cut (Donchin et al., 1978;

cf., Donchin & Heffley, 1978; Kappenman & Luck, 2012). The observation of a certain deflection in neural activity in response to a postulated neural process does not mean that the process actually occurred, but only that it was hypothesised to. That said, many ERPs reported in the literature across different modalities and representing various neural processes have been studied to a sufficient extent that they have become a reliable correlate of the underlying process they are assumed to represent (Kappenman & Luck, 2012). Factors such as typical timing and amplitude of the response, localised centre of activity on the scalp, and input evoking the response are all used as corroborative convergent evidence when identifying an ERP.

This section reviews ERPs within the domains of attention, working memory, and language, grouped broadly by function and the underlying processes they represent. These three domains are integral to the processing of speech and understanding ERPs in these areas is critical to get a holistic representation of what is happening in the brain during real-time speech comprehension.

ERPs are named for their scalp polarity, with a N or P indicating a negativity or a positivity, and then further distinguished either by their latency (e.g., N100 occurs 100 ms post-stimulus onset) or their ordinal position (e.g., N1 is the first observed negative peak) (Luck, 2014). Other ERPs are sometimes given more opaque names that are indicative of other features related to their elicitation (e.g., *mismatch negativity*; Näätänen, 1992). In the next section, an emphasis is made to discuss ERPs elicited to auditory stimuli as those are directly relevant to this dissertation and its discussion of speech processing. However, it is worth noting that prominent ERPs elicited to cross-modal stimuli, or to stimuli in other domains (e.g., visual) have also been studied extensively (see Luck, 2012; Luck & Kappenman, 2012; Perez & Vogel, 2012; and Wilding & Ranganath, 2012, for reviews).

Attention

In this section, a few prominent ERPs modulated by attention will be discussed. Discussion is limited to components within the auditory modality. Specifically, attention here is defined as “the processes by which the brain selects some sources of inputs for enhanced processing” (Kappenman & Luck, 2012, p. 2). Attention plays some type of role in almost every ERP component. However, for some, it plays a more overt role than others.

N1. The auditory N1 is an attention-modulated sensory response comprising of several subcomponents (Näätänen & Picton, 1987). Due to reliably being evoked at the onset of auditory stimuli it is often used as a biological marker of successful perception in typical populations. The N1 is usually maximal at 100 ms post-stimulus onset and its latency and amplitude are both affected by stimulus features such as intensity, pitch, and timing (see Kappenman & Luck, 2012 for a review).

The N1 has often been observed alongside the P2 response (N1-P2 complex), as well as in tandem with other negative responses such as the MMN (reviewed later).

P2. The auditory P2 is a positive-going response that occurs at a latency of about 180 ms post-stimulus onset (Näätänen & Picton, 1987). Similar to the N1, the P2 response is thought to be modulated by attention and is affected by features of the stimuli being attended to. It often occurs within the N1-P2 complex.

MMN. Another important component that marks detection of unpredicted events in pre-attentive processing is the mismatch negativity, or the MMN. This dissertation uses the MMN to study how timing variations are processed when speechlikeness increases. Therefore, a more detailed overview of this component will be provided in a later section.

P3. The P3 comprises of several different components, the most prominent ones being the P3a (frontally distributed) and the P3b (parietally distributed; see Polich, 2012, for a review). Typically, these responses are elicited 300 ms post-stimulus onset, although some components identified as belonging to the P3 family have been observed with latencies much longer than that. The P3 is related to working memory processing (Donchin, 1981) and has been found to be overtly affected by attention (see Luck & Kappenman, 2012). The P3a has often been observed in tandem with the MMN and is suggested to reflect an attention orienting effect (Escera et al., 2000). The P3b on the other hand is more often discussed in the literature as the ‘default’ P3 response and is often elicited to surprising stimuli in so called oddball paradigms when participants are asked to attend to the stimuli.

Therefore, the P3 responses are thought to represent an engagement of attention. Either inadvertently, as is the case for the P3a, or wilfully, as participants are instructed to do in tasks eliciting the P3b.

Working Memory

Working memory is a limited-capacity system in memory that stores items to be used in a short interim for the performance of cognitive tasks (Baddeley & Hitch, 1974). It is strongly affected by individual differences, and its capacity is a positive predictor of stronger reading and reasoning abilities (see Baddeley, 2010 for a review). Several ERP components have been studied and observed in response to, in particular, visual working memory. Of prominence are the *negative slow wave* (NSW), observed to rehearsal of items in memory, and the *contralateral delay activity*, elicited in response to hemisphere-specific memory tasks and shown to be sensitive to memory load and individual differences (Perez & Vogel, 2012). While not directly pertaining to the work in this dissertation, these responses provide a glimpse of how (visual) working memory has been evaluated in the perceptual domain via neurophysiological methods.

Language

There exist several ERPs that have been observed or elicited specifically to linguistic stimuli. Some have been associated with particular aspects of linguistic processing, such as syntactic analysis. Others have been observed in contexts where a specific linguistic expectancy was violated, either semantically or phonologically. While no specific language ERP was studied in this dissertation (reasons why the MMN was used are discussed in the section titled, "*The mismatch negativity as a marker of cognitive processing*"), these ERPs provide important insight into the time scale of linguistic processes in the brain. In reviewing them here, our hope is that it will help inform the understanding of the results reported in this dissertation. A more detailed review of some language-related ERPs can be found in Swaab et al., 2012.

Phonological Mapping Negativity (PMN). This negative response has been found to peak at somewhat variable times, first reported as between 270 and 300 ms after the onset of a word whose phonological onset does not match expectation (Connolly & Phillips, 1994; Connolly et al., 1992; Connolly et al., 2001). The PMN has been described as reflecting purely phonological processing at a pre-lexical stage.

N400. This ERP was first described as a negative-going marker of semantic mismatch that can be detected as peaking approximately 400 ms post-stimulus onset by the centro-parietal electrodes (Kutas & Hillyard, 1980). It is one of the few language-related components that has been found to be evoked to mismatches regardless of whether the stimuli are visual or auditory. The N400 has been ascribed to lexical process, with a larger N400 associated with greater difficulty in accessing word meaning or a semantic concept (Kutas et al., 2006). An alternative account by Hagoort (2007) suggests that rather than retrieval, the N400 indexes how well the target word semantically settles within the sentential context. A larger N400 is associated with greater work performed in consolidating the target word within its context.

P600. The P600, also known as the Syntactic Positive Shift (SPS) (Hagoort et al., 1993) is a positive response often associated with syntactic reanalysis (Osterhout & Holcomb, 1992). It can be elicited to sentences containing ungrammaticality, or to sentences that require a structural reanalysis to be understood, as is the case with garden path sentences. Discussions in the literature have suggested that the P600 may belong to the P3 family of attention-related responses (e.g., Coulson et al., 1998), or that it may be a marker of predictivity and learning instead of syntactic processing (or reprocessing).

Cortical oscillations

The second type of informative analyses performed on EEG data involves the examination of neural oscillations. Neural signals, recorded via EEG, can be decomposed into the amplitude and frequency of the underlying sine waves that make up the signal (Giraud & Poeppel, 2012). Neural oscillations are informative in the morphology of the response recorded (e.g., high-frequency waves would result in a response with a great number of peaks in a short amount of time), and in the frequency-time decomposition of the waveform being analysed (e.g., power distribution of the different constituent frequencies). These neural oscillations can entrain to incoming speech, meaning that in one way or another, they match with the rhythm of the incoming auditory input and play a role in speech comprehension (Poeppel & Assaneo, 2020). The frequency-time decomposition of neural oscillatory activity shows peak frequency activity over time and gives a sense of entrainment in specific frequency bands. Additionally, measures of phase synchronisation (like phase coherence) can be used to estimate variation in phase angles between neural signals; greater variation implies poorer coherence and synchronisation compared to angles that are better aligned (Cohen, 2014). Other elements, like amplitude rise time can also give a sense of how fast or slow an individual is entraining to incoming stimuli thereby providing a quantification of the degree of entrainment (Goswami et al., 2013).

In speech, alignment of syllabic rhythm with neuronal oscillatory phase aids in the perceptual chunking of incoming speech (Rimmele et al., 2021). Previous work has identified slower frequency ranges in the theta (4-8 Hz) and delta bands (1-3 Hz) to be critical in speech segmentation into syllables and words (Gross et al., 2013; Kaufeld et al., 2020), perception (Doelling et al., 2014; Keitel et al., 2018; Kösem et al., 2018; Peelle & Davis, 2012; Rimmele et al., 2021; Zoefel, 2018), and processing (Giraud & Poeppel, 2012; Hauk et al., 2017; Poeppel & Assaneo, 2020). A role of beta (12-30 Hz) and the fast gamma (25-140 Hz) bands in sentence comprehension (e.g., Lewis & Bastiaansen, 2015), and of theta-gamma coupling in syllable processing (e.g., Hovsepyan et al., 2020) has also been found. In one study (Assaneo & Poeppel, 2018), speech presented at a rate of 4.5 Hz showed the greatest entrainment in the theta band, and synchronisation between auditory and speech-motor cortices was also observed. Neural oscillations entrain to the incoming auditory stream, reflecting peaks in frequency bands that mirror the rate of speech at the level of syllables (theta), words (delta), and even sentences (Ding et al., 2015). More specifically, frequencies ranging from 4-5 Hz have been found to reflect the ‘optimal’ rate for speech, with converging evidence found in both articulatory and phonological domains (see Poeppel & Assaneo, 2020, for a review).

Deficits in neural entrainment to incoming speech have been associated with developmental dyslexia, suggesting that entrainment to speech rhythm may be crucially linked to phonological processing of speech (Goswami, 2018; Lallier et al., 2018). Rhythmic elements in language help foster predictability, improving listener comprehension of incoming speech, and make it easier for different elements of speech to be perceived (e.g., phonemic differences, word boundaries) (Bosker & Kösem, 2017; Peelle & Davis, 2012). Timing of linguistic units as an aspect of rhythm is important in contributing to the predictability and comprehensibility of spoken speech (Gibbon, 2018; Poeppel & Assaneo, 2020).

Neural oscillator models provide a new, neurophysiological method to explore the role of timing and rhythm in language processing. They add to the existing perspectives on rhythm in language by providing new insights into the mechanisms of language processing and comprehension. Examining the underlying neural oscillations in a bid to understand the processing of temporal variations in speech fell outside the scope of this dissertation. However, it is evident that understanding oscillatory behaviour is integral to obtain a more holistic representation of neural activity changes in the brain. The next section provides a brief overview of how neuronal oscillations could be key in the inner workings of the brain.

Dynamic Attending Theory (DAT)

The Dynamic Attending Theory (DAT) (Jones, 2018; Large & Jones, 1999) is a theoretical framework that takes as its starting point the assumption that neural oscillations contribute to sensory (and, therefore, speech) perception. According to this theory, pre-existing sets of internal cortical oscillation frequencies within an individual (*driven rhythms*) synchronise with the temporal rhythmic structure of an external event (*driving rhythm*), leading that individual to *attend* to temporally structured auditory events (Jones, 2018; Large & Jones, 1999). Entrainment, specifically, is defined in this context as “the mechanisms that describe the workings of a synchrony [between the driving and driven rhythms]” (p. 1, Jones, 2018). In simpler words, the DAT posits that when the rhythms on the inside (of the listener/perceiver) match the rhythms of the outside auditory stimuli, perception of the matching events takes place. The strength of the entrainment determines the degree of attending and therefore strength of perception of the incoming sensory stimulus—which, in context of this dissertation, would be auditory and linguistic in nature.

Entrainment to incoming speech is more than simply two rhythms coinciding. Features like oscillatory phase and the underlying frequencies and amplitudes of the driving and driven rhythms are at the forefront of entrainment as it is a mechanistic description of how rhythms synchronise (Giraud & Poeppel, 2012; Jones, 2018; Poeppel & Assaneo, 2020). Timing is an intrinsic focus of DAT, and in language, DAT concentrates on how the rhythms of speech—its prosodic features like pitch, time, and stress—help drive speech perception and comprehension.

Each feature of speech functions at a different level of a hierarchy that builds up at the higher level to include words, and at the lowest level includes individual consonants and vowels (Poeppel & Assaneo, 2020; Ding et al., 2016; Jones, 2018). Each level of this hierarchy forms its own time scale and therefore its own driving rhythms that driven rhythms can be entrained to. Where other linguistic theories consider prosodic markers to be just that—features providing cues about how speech is segmented, DAT considers them to carry temporal information that is crucial to the comprehension of speech (Jones, 2018).

The rhythm class hypothesis, as discussed previously in this chapter, suggests that all the languages of the world inherently belong to one of three rhythm classes. The DAT makes no such claims, and instead aims to unpack *how* the regular rhythms in language—which are present by way of its repetitive phonological, semantic, and syntactic features—help shape its perception and comprehension. The variability inherent in speech is described by DAT as being compensated for by the underlying internal and external rhythms which synchronise with each other. The driven rhythms tend to adjust to the driving rhythm if the latter is strong enough (Jones, 2018).

This thesis explores individual sensitivity to temporal predictability as reflected in immediate recall and ERPs to violations of temporal expectations. Specifically, the MMN brain response will be studied. The next section will review how the mismatch negativity, or the MMN, can be used as a neural marker for cognitive processing.

The mismatch negativity as a marker of cognitive processing

The current dissertation explores how timing variation in language is processed by tracking changes in the EEG-based brain response known as the *mismatch negativity* (MMN). The MMN is a neural response typically used to study pre-attentive cognitive processes and auditory memory trace formation (Näätänen, 1992; 1995). Classically, the MMN has been described as a negative-polarity event-related response that occurs approximately 100 ms to 250 ms after auditory stimulus presentation (e.g., Näätänen, 1992). It has an amplitude ranging from 1-6 μV (e.g., Aaltonen et al., 1987; Ford & Hillyard, 1981) and a fronto-central topographical distribution (Näätänen et al., 2019). MMNs are most often elicited by a rare, deviant stimulus presented within a repetitive stream of frequent, standard stimuli in an experimental paradigm known as the oddball paradigm. Other paradigms have also been suggested to elicit this response (see Näätänen et al., 2019 for an overview). The deviant stimulus differs from a standard auditory stimulus on one or more features, and this difference is critical in eliciting an MMN. The magnitude of the MMN is modulated by the degree of difference perceived between the standard and the deviant; in general, the larger the perceived difference, the larger the MMN amplitude (Näätänen et al., 2019). Due to its sensitivity to differences between sensory inputs within the auditory domain, the MMN has been used as a way to explore the neural sensory perception, or *central sound representation,* of incoming auditory inputs (Näätänen et al., 2001; Näätänen & Winkler, 1999).

Changes in the MMN as a reflection of varying acoustic features helps further the understanding of how those features are auditorily processed. They may be acoustic features pertaining to the stimulus itself (e.g., intensity, pitch, duration) (Ford et al., 1976; Näätänen, 1995; Näätänen, 2000; Näätänen et al., 1978; Näätänen et al., 1989; Näätänen et al., 2007; Näätänen et al., 2019; Nordby et al., 1988a; Nordby et al., 1988b; Novak et al., 1992), the timing of the stimuli, either singularly or in a pattern (Duda-Milloy et al., 2019; Ford & Hillyard, 1981; Lai et al., 2011; Sable et al., 2003), or the presence of an expected stimulus altogether (Yabe et al., 1997).

MMNs have also been successfully evoked in multi-feature paradigms that present multiple deviants differing in contrasting features from the same standard (Pakarinen et al., 2009; Fisher et al., 2011; Shiga et al., 2011). In addition to simple tones, which are the most common type of stimuli used to evoke an MMN, this response has also been elicited to more complex stimuli, including complex tones and linguistic stimuli (Nordby et al., 1994; Shtyrov et al., 2000; Asano et al., 2015; Brückmann, & Garcia, 2020), and to variations in patterns of tones, including deviants varying in the latency of the last tone in a multi-tone sequence (Pato et al., 2002; Sable et al., 2003; Boh et al., 2011; Paavilainen, 2013). Studies have also introduced gaps of varying lengths between two adjoining stimuli (Desjardins et al., 1999; Bertoli et al., 2001; Duda-Milloy et al., 2019), left a stimulus missing altogether (*omission deviant*, Nordby et al., 1988a; Yabe et al., 1997; Yabe et al., 1998; Rüsseler et al., 2001; Bendixen et al., 2009; Horváth et al., 2010; Salisbury, 2012; Ohmae & Tanaka, 2016; Bouwer et al., 2019), or adjusted the stimulus onset asynchrony (SOA), the intervals between onsets of consecutive stimuli, or inter-trial intervals (ITI) between the offsets and onsets of consecutive stimuli (Ford & Hillyard, 1981; Näätänen et al., 1987; Hari et al., 1989; Lai et al., 2011) to elicit an MMN.

The MMN response is thought to be reliant on the formation of a *memory trace*, as evidenced by the fact that multiple standard stimuli need to have been presented before the deviant will elicit the MMN negativity (Cowan et al., 1993; Winkler, Cowan et al., 1996; Winkler, Karmos et al., 1996). This memory trace, alongside the complexity and nature of the stimuli, allows for inferences to be made about higher cognitive processes and for the MMN to function as a marker that reflects how these processes change with the nature of the stimulus or paradigm (Aaltonen et al., 1987; Näätänen, 2000; Näätänen, 2001; Light et al., 2007; see Näätänen et al., 2007, for a review). MMNs evoked to unexpected changes in SOA or ITI can be used to probe memory traces for both short-term and long-term auditory associations formed with simple and complex auditory inputs (Näätänen, 1995; Näätänen et al., 2019). These associations can be formed in conditions where participants are not attending to the stimuli, making the MMN a highly informative tool to study pre-conscious cognitive processes (at one extreme, in coma patients, or in the context of brain injuries; Näätänen, 2000; Näätänen & Escera, 2000; Näätänen et al., 2019). Given its relationship with memory and auditory processing and versatility in the type of stimuli that can be used to evoke it (both linguistic and non-linguistic), the MMN is the ideal tool to study how temporal variation in speech is perceived and processed cognitively.

Memory-based mechanisms of MMN elicitation: the Temporal Window of Integration (TWI)

Memory lies at the core of MMN elicitation, as is evident from the fact that it is elicited to both short-term traces (formed to a particular standard-deviant difference presented within an experiment; Näätänen, 1992, 1995; Näätänen & Picton, 1987) as well as to long-term traces (such as those existing with native language phonemes, Näätänen & Winkler, 1999; Winkler et al., 1999).

Previous work suggests that the MMN memory trace formation for separate stimuli is associated with a sliding temporal window of integration (TWI). Simply put, this memory trace will be formed only when sequential standard and deviant presentations fall within a certain time window (Cowan, 1984; Winkler et al., 1993). Consecutive stimuli in continuous stimulus trains presented at a pace that is faster than the TWI are not perceived completely enough to form an adequate representation of the difference between them, if any, and are perceived as a solitary unit (within the time window) that contains no comparison point. A standard-deviant pair presented at a pace greater than the TWI threshold will, for example, evoke no MMN (Yabe et al., 1997, Yabe et al., 1998). This TWI has been determined to fall between 150 milliseconds (ms) and 200 ms for simple tones (Yabe et al., 1997; Yabe et al., 1998; Shiga et al., 2011), and has been found to be at least 176 ms for speech segments (Asano et al., 2015; Scharinger et al., 2017). The TWI has also been found to be longer in children than young adults (Wang et al., 2005), but not to change thereafter with age (Horváth et al., 2007; see Frisina, 2001, for a review). Musical background also has an effect on the TWI, with musicians being able to elicit omission deviants at greater SOAs compared to non-musicians (Rüsseler et al., 2001; see Näätänen et al., 2007, for a review). In dyslexic populations, MMNs to omission deviants within the TWI are not elicited as reliably compared to non-dyslexic populations (Fisher et al., 2006). As evidence for chunk learning, pairs of stimuli grouped together such that neither appears without the other are processed as a single unit and elicit a single MMN despite varying in acoustic properties (e.g., frequency and intensity) if they are presented within the TWI as a coupled pair (Tervaniemi et al., 1994; Shinozaki et al., 2003, Oceák et al., 2008). Time of presentation of the stimuli is a principal factor of feature-perception, therefore; if a stream of sounds is presented at a rate that falls outside the TWI, the differences in features between the standard and deviant will not be perceived, and no MMN will be elicited.

However, the elicitation of the MMN is not dependent on recognizing the difference in features between individual stimuli alone, but also on recognising differences violating the pattern of the stimuli presented (Bouwer et al., 2019; Bouwer, et al. 2014; Näätänen et al., 2019). This includes different patterns of tones for the standards and the deviants, and includes variations in tone placement (e.g., Sable et al., 2003), or in the presence of the tone altogether (tone omissions, for e.g., Salisbury, 2012). Therefore, the scope of the established memory trace is wider than a single auditory stimulus and can include longer strings of patterns as well, resulting in the elicitation of what is known as an 'abstract feature MMN' (Näätänen et al., 2007).

Theoretical framework for the processing and perception of speech

Perception and processing of incoming speech has been proposed to operate by a system that is entirely separate from the one that processes other non-speech auditory inputs. Theories which ascribe to this understanding of speech perception and production are known to be *domain-specific* and emphasise the importance of speech in the development of the human auditory system (Fowler, 1996; Liberman & Mattingly, 1995; Liberman & Whalen, 2000; Zatorre & Gandour, 2008; cf. Galantucci et al., 2006). In contrast, *cue-specific* models suggest the opposite—that both speech and other auditory inputs are processed along the same neural pathways without there being a particular processing route that caters to speech (see Diehl et al., 2004, for a review). Domain-specificity or domain-generality of speech processing are not predictive models, but rather models that discuss *how* speech is produced, perceived, and processed. The extent to which speech and non-speech share the same mechanisms and machinery is at the heart of the difference between the two, with domain-specificity arguing for speech-specific modularity, and cue-specific views arguing for a general approach that does not discriminate between speech and non-speech at the outset. Recent discussions suggest that it is likely that speech is both (see Zatorre & Gandour, 2008, for a discussion).

In recent years, more and more research has begun to focus on the use of a *predictive coding* (PC) framework to explain perceptual processing. The predictive coding model discusses the role of sensory inputs in terms of perception and processing without specification of the type of input. This makes it compatible with a domain-general point of view. In terms of the PC framework, speech does not function differently than any other auditory input in principle. However, the role of predictive coding in speech perception and comprehension is not as simple as it might seem from the outset.

A distinction needs to be made between *prediction* in speech and language processing compared to its *predictive coding*. Prediction in real-time speech processing has been studied in syntactic, lexical, morphological, and phonological contexts with the perspective that each context helps constrain upcoming possibilities, making one likelier than the other. Therefore, the different contexts each help to calculate the likelihood of different options for the forthcoming speech, with successful predictions aiding in the processing of real-time speech. Sentence contexts and cloze probability (e.g., Staub et al., 2015), neighbourhood density effects (e.g., Magnuson et al., 2007; Metsala, 1997), language-specific consonant clusters and coarticulation (e.g., Salverda et al., 2014), as well as syntactic constraints (e.g., Mattys et al., 2007) are examples of some overt cues that facilitate these predictions.

Predictive coding, on the other hand, was prominently introduced as a method in signal processing to discretise continuous time series data (see Spratling, 2017, for a review). In the perceptual domain it functions as a framework that computes and translates a measure of error when top-down expectations clash with bottom-up sensory inputs. It was originally introduced to explain how visual information is processed (Rao & Ballard, 1999; see also Friston, 2005; 2009; 2010; Gagnepain et al., 2012; Huang & Rao, 2011; Spratling, 2017). Recent work has found strong support for this framework in explaining how speech is attended to, processed, and understood (Denham & Winkler, 2020; Hovsepyan et al., 2020; Lupyan & Clark, 2015; Magnuson et al., 2019; Ylinen et al., 2016; Ylinen et al., 2017). In the literature, prediction in speech is not necessarily examined in the context of the predictive coding framework and may be thought to operate differently from it. However, there is no doubt that prediction in language functions as support for the predictive coding framework, making the latter a strong foundation for understanding speech production, perception, and processing.

The next section provides an overview of the predictive coding framework, with an emphasis on auditory processing and the MMN.

Predictive coding

Predictive coding is a Bayesian hierarchical model that describes the auditory perceptual system as one that creates, updates, and maintains current representations of incoming sensory percepts and uses them to predict and form expectancies regarding incoming future sensations. The framework being hierarchical is key for this type of communication, allowing the perceptual system to gain and store information in the present while simultaneously making predictions about the future. There are distinctions between the types of predictive coding frameworks that have been described in the literature, broadly divided into the original framework, the first adaptation into a neuroscience context, and finally three models of predictive coding of cortical responses (Spratling, 2017).

Linear predictive coding (LPC) was first introduced and extensively used in digital signal processing to reduce noise and extrapolate lost or poor data. Each data point is represented in a time series as a linear sum of the ones preceding it. An extension of the LPC framework was adapted by neuroscience to describe the function of the retina, where "…the predictable component of the signal is removed . . . in order to allow a more efficient transmission." (Spratling, 2017, p. 93). That is, a more efficient and accurate signal for modelling retinal function is obtained by extracting the residual error and transmitting it further up the processing hierarchy to the cortex (Srinivasan et al., 1982). This retinal model aimed to explain specifically how the retina adjusts for intensity and luminance. The adaptation of the LPC framework to retinal function suggests, therefore, that the retina functions by comparing the transmitted predicted value of light intensity on retinal surface is against the real-time intensity value for visual processing (Srinivasan et al., 1982). It is worth noting that this framework does result in the loss of (redundant) information, as only the difference between the predictive and incoming signal is transmitted (Spratling, 2017). Finally, applications to model cortical responses have been proposed, with three main frameworks discussed in the literature: the Rao and Ballard algorithm, PC/BC-DIM, and free energy (Spratling, 2017).

The *Rao and Ballard model* (Rao & Ballard, 1999) reformulates LPC from using time series data to using coefficients that represent sensory inputs weighted by the model's preliminary understanding of what might instigate those sensory inputs. This model presents cortical processing as a hierarchy of different networks (e.g., higher vs. lower-level cortical neurons conveying visual information) that function by processing sensory inputs and tallying them against the predictions made based on previous knowledge of the external environment (Spratling, 2017). This model allows for distinct levels of the hierarchy to interact with each other (in a feedforward/feedback manner) that ensures that the predictions made are consistent with each other. Like the model of retinal processing, the feedforward output signal communicates the residual error, while the feedback signal communicates predictions of what instigated the sensory input.

The second predictive coding framework proposed to account for cortical responses is the *PC/BC-DIM model*, which differs from the Rao and Ballard model in terms of congruity with other models of cortical activation (Biased Competition theories, BC; Spratling, 2008a; 2008b), and how the residual error measuring neuronal activations is calculated (Divisive Input Modulation, DIM; Spratling et al., 2009) (PC/BC-DIM is therefore the Predictive Coding/Biased Competition-Divisive Input Modulation model). Specifically, this model simplifies and re-groups some integral assumptions made by the Rao and Ballard model and reiterates a method that is faster computationally while maintaining the same hierarchical structure.

Finally, the *free energy model* of predictive coding does not model sensory inputs, but the statistical likelihoods of the sensory inputs being perceived from certain probability distributions (the posterior probability density) (Friston, 2009; 2010). While this is the largest difference between this framework and the others discussed, in other regards, the free energy model is similar: it does, like the Rao and Ballard model, still postulate a hierarchy of networks that communicate with each other to make predictions and detect errors.

In this dissertation, reference to the predictive coding framework indicates this hierarchical framework that allows feedforward and feedback predictions based on incoming perceptual (auditory) information. No reference is made to any specific version of the PC framework, as all three iterations of the PC framework agree on the basics.

Perceptual framework of the MMN

In this dissertation, the predictive coding framework is used to evaluate and understand the elicitation of the MMN to temporal cues. This framework is adaptive of the previous theories and models of MMN elicitation that have so far been discussed in the literature. The current section provides an overview of previous MMN elicitation models and what it means for how auditory inputs are processed before discussing the MMN specifically under the predictive coding framework.

One of the earliest theories explaining the MMN generation posits that the response is evoked when a sensory memory trace of past auditory stimuli is compared to incoming stimuli and a difference is perceived (Mäntysalo & Näätänen, 1987; Näätänen, 1988; Näätänen, 1992; Näätänen, 1995; Näätänen & Winkler, 1999; Näätänen et al., 2011). The detection of this difference between the memory trace and the stimulus manifests on scalp electrodes in the form of a negative response, the MMN. The *model-adjustment hypothesis* similarly suggests that the MMN functions as a marker of error when the internal predictions conflict with the external auditory inputs (Winkler et al., 1996; Näätänen & Winkler, 1999). Another explanation for the underlying mechanism driving MMN elicitation lies with the adaptation hypothesis, or *stimulus-specific adaptation (SSA)*, which suggests a decreased (inhibited) neural response to repeated presentations of the same sensory stimuli (May & Tiitinen, 2010). A deviant within a stream of repeated standards ‘breaks’ this chain of adaptation responses when a new stimulus is perceived, therefore leading to the elicitation of an MMN. However, the adaptation hypothesis shows limitations, not being able to account for every context of MMN elicitation. For example, this model is unable to account for the MMN evoked to omission deviants, as the lack of a stimulus presentation should not be enough to disrupt the adaptation responses made to the stream of standards (Fong et al., 2020; Garrido et al., 2009). This framework, therefore, cannot account for the underlying mechanisms of temporal variation in the auditory domain that will be examined in the experiments reported in Chapters 2 and 3 of this dissertation.

MMN and predictive coding

In the context of the predictive coding mechanism in general, the MMN is a response to prediction error. In other words, the MMN is a response elicited when the incoming stimulus (the deviant) differs from the expected pattern of stimuli predicted based on past presentations (the standards). As PC is hierarchical, this prediction error response builds up not just in a feedback manner (comparing every new presentation of the deviant to the previous), but in a feedforward manner as well (by managing expectancy and predicting when the next deviant will be presented). Therefore, this framework predicts that the larger the difference between the deviant and the standard, and the greater the unexpectedness of the presentation, the larger the error response, or MMN, will be (Fong et al., 2020). Although the predictive coding framework itself is not without weaknesses (see Denham & Winkler, 2020, for a discussion), this framework is able to account for some MMN data better than a sensory adaptation model such as the Stimulus-Specific Adaptation (SSA) model (May & Tiitinen, 2010). Indeed, Garrido et al. (2009) suggest that the predictive coding framework is compliant with aspects of both the model-adjustment hypothesis and the SSA, making it more flexible and a better fit as an explanation for MMN elicitation. PC has the advantage of being able to address perception more holistically by incorporating evidence from sources that speak to the psychological aspect of perception as well as the neural aspect. Evidence supporting a predictive coding account for the MMN is plentiful (Baldeweg, 2007; Wacongne et al., 2012; Wacongne et al., 2011), and is most strongly supported by the presence of the MMN to omission deviants (e.g., Yabe et al., 1997). As the type of auditory input is not critical to the PC framework, it can help explain MMN elicitation to both simple tones and more complex stimuli.

The PC framework has also been found to explain how infants learn and recognise words (Ylinen et al., 2017). In this study, EEG was recorded as infants heard disyllabic words that were familiar, or novel, or heard a stream of monosyllables. Mismatch responses (the equivalent to MMNs in infants) were elicited to both novel and familiar words but were smaller for individual syllables compared to words. An effect of familiarity—and therefore, predictability—was found, with more familiar words eliciting a larger mismatch response than novel words, whose lack of familiarity would signal a smaller error response under the predictive coding framework. Other work has encouraged the evaluation of predictive learning (and predictive coding) with neural oscillations, suggesting that by utilising various aspects to evaluate processing, a more complete picture of the underlying mechanisms of language processing might be found (Hovsepyan et al., 2020; Lewis & Bastiaansen, 2015).

Predictions generated by the predictive coding framework, therefore, work in tandem with neural entrainment to help identify and comprehend speech. This suggests a significant role of cortical oscillations in speech processing and in managing the error response that is the output of the predictive coding framework.

The current dissertation: motivation

In the current dissertation, temporal cues (lengths of silent pauses) that resulted in rhythm were tracked by way of the neural response known as the mismatch negativity, or the MMN. Manipulating the timing of complex auditory stimuli and monitoring the resulting changes in a neural response reflecting auditory processing would give us insight into how changes in the timing parameters only of auditory stimuli can affect processing.

Timing variation of some form has been manipulated in cognitive or neurocognitive research of language either as differences in the lengths of vowels or consonants (Bertinetto & Bertini, 2010); speech rate (Dellwo, 2008; White et al., 2012; Alexandrou et al., 2018; Assaneo & Poeppel, 2018); as pitch variations in prosody over certain sentence segments (Rosen, 1992; Ramus & Mehler, 1999); or even in terms of duration differences between some speech segments or syllables within a sentence versus others (Bosker & Kösem, 2017; Dahan, 2015; Kotz et al., 2018; Rosen, 1992). Neural correlates of broader aspects of speech have also been recorded. One study, for example, found that the different syntactic units (words, phrases, sentences) were reflected in cortical activity as different temporal hierarchies (Ding et al., 2015). However, to our knowledge, temporal regularity in the linguistic domain has rarely been studied as a structure on its own.

In the current study, the predictive coding (PC) framework was used to explain the results of our hypotheses on how temporal variation in a train of auditory stimuli is perceived, and how the directionality of unexpected variations (temporally early or late) aid or abet auditory processing. The temporal concurrence of feature perception and MMN elicitation suggests that both processes occur at the same processing stage. Study of how sensory perception of the incoming auditory stimulus interacts with the temporal window of integration (TWI) and the underlying predictions will allow us to better understand how auditory temporal patterns are perceived.

Thesis Objectives and Overview

The main question of this dissertation centres around understanding how expectancy-based timing violations affect comprehension of incoming speech. This dissertation employed both electrophysiological and behavioural methods to understand the changing nature of auditory processing as input complexity increases, and to evaluate how that processing interacts with more complex cognitive functions like attention and memory.

The main research question answered in this dissertation concerns the role unanticipated temporal variations play in real-time speech comprehension. This is addressed by exploring the following sub-questions in Chapters 2, 3, and 4:

1. *How do unexpected temporal variations and omissions affect the pre-attentive processing of simple auditory stimuli?*
2. *How does increasing linguistic familiarity (i.e., increasing speechlikeness) alter the pre-attentive processing of unexpected timing variations?*
3. *What is the impact of unexpected timing deviants on verbal short-term memory tested through the perceptual grouping of syllables in a continuous stream?*

There are five chapters in this dissertation. The following summarises the dissertation and provides details on how each chapter addresses one of the research sub-questions.

Chapter 1 supplies a detailed background of the current literature examining time and rhythm in language. Furthermore, this chapter presents and examines the theoretical frameworks at the basis of this research. This chapter concludes with an overview of the dissertation and the examined research questions.

Chapter 2 presents the first two experiments of this dissertation which examine how the brain responds to temporal variations in a train of simple tones. This chapter addresses the research question by setting up a baseline of responses to timing expectancy modulations in auditory stimuli. Additionally, the role of timing variations in a sequence of tones versus omissions within the sequence is explored. The studies outlined in this chapter set up the first of a series of experiments with progressively complex stimuli in the linguistic domain. Results discuss how temporal variation impacts attention and auditory processing for simple tones.

Chapter 3 reports the results of a third experiment that examines the cognitive processing of more acoustically and linguistically complex sounds, and of their interaction with unexpected temporal variations. The impact of unexpected timing variation in relation to pre-attentive processing of speech vs. non-speech sounds is discussed.

Chapter 4 contains the fourth and final experiment of this dissertation. It examines the effect of timing variations within auditory syllable sequences on short-term memory, tested with a recognition task. Implications are discussed regarding the possible effect of unexpected timing variations on perceptual grouping and their impact on verbal short-term memory.

The final Chapter 5 provides a brief discussion and conclusion by summarising how each chapter fits into the wider context of timing in language and briefly discussing the implications of the results. This chapter concludes by discussing some limitations of the work.

References

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*(3), 197–207. https://doi.org/10.1016/0301-0511(87)90002-0

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago, IL: Aldine.

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018). Cortical Tracking of Global and Local Variations of Speech Rhythm during Connected Natural Speech Perception. *Journal of Cognitive Neuroscience, 30*(11), 1704-1719.

Asano, S., Shiga, T., Itagaki, S., & Yabe, H. (2015). Temporal integration of segmented-speech sounds probed with mismatch negativity. *NeuroReport, 26*(17), 1061–1064. https://doi.org/10.1097/WNR.0000000000000468

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, *4*(2), 1–10. https://doi.org/10.1126/sciadv.aao3842

Aubanel, V., & Schwartz, J. L. (2020). The role of isochrony in speech perception in noise. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-76594-1

Bastiaansen, M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0024

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: A predictive coding perspective. *Journal of Psychophysiology*, *21*(3–4), 204–213. https://doi.org/10.1027/0269-8803.21.34.204

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, *29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Bertinetto, P. M., & Bertini, C. (2010). Towards a unified predictive model of Natural Language Rhythm. *Prosodic Universals. Comparative studies in rhythmic modeling and rhythm typology. Rome: Aracne*, 43-78.

Bertoli, S., Heimberg, S., Smurzynski, J., & Probst, R. (2001). Mismatch negativity and psychoacoustic measures of gap detection in normally hearing subjects. *Psychophysiology*, *38*, 334–342. https://doi.org/10.1111/1469-8986.3820334

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2010). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, *54*(1), 123–140. https://doi.org/10.1177/0023830910388018

Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of Complex Auditory Patterns in Musicians and Nonmusicians. *PLoS ONE*, *6*(7), 21458. https://doi.org/10.1371/journal.pone.0021458

Bosker, H. R., & Kösem, A. (2017). An entrained rhythm's frequency, not phase, influences temporal sampling of speech. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2017-August, 2416–2420. https://doi.org/10.21437/Interspeech.2017-73

Bouwer, F. L., Honing, H., & Slagter, H. A. (2020). Beat-based and memory-based temporal expectations in rhythm: similar perceptual effects, different underlying mechanisms. *Journal of Cognitive Neuroscience*, *32*(7), 1221-1241.

Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: An ERP study. *PLoS ONE, 9*(5), e97467. https://doi.org/10.1371/journal.pone.0097467

Brückmann, M., & Garcia, M. V. (2020). Mismatch negativity elicited by verbal and nonverbal stimuli: Comparison with potential N1. *International Archives of Otorhinolaryngology, 24*(2), E80–E85. https://doi.org/10.1055/s-0039-1696701

Cheour, M., Kushnerenko, E., Čeponienė, R., Fellman, V., & Näätänen, R. (2002). Electric brain responses obtained from newborn infants to changes in duration in complex harmonic tones. *Developmental Neuropsychology*, *22*(2), 471–479. https://doi.org/10.1207/S15326942DN2202_3

Choi, J. (2003). Pause length and speech rate as durational cues for prosody markers. *The Journal of the Acoustical Society of America*, *114*(4), 2395–2395. https://doi.org/10.5840/raven20111838

Cohen, H., Douaire, J., & Elsabbagh, M. (2001). The role of prosody in discourse processing. *Brain and Cognition*, *46*(1–2), 73–82.

Cohen, X. M. (2014). Analyzing neural time series data. MIT Press. https://doi.org/10.1007/s13398-014-0173-7.2

Cole, J. (2015). Prosody in context: A review. Language, Cognition and Neuroscience, 30(1-2), 1-31.

Colling, L. J., Noble, H. L., & Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Frontiers in Neuroscience*, *11*, 360.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6(*3), 256–266. https://doi.org/10.1162/jocn.1994.6.3.256

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*(1), 1–18. http://www.ncbi.nlm.nih.gov/pubmed/1643505

Connolly, J. F., Service, E., D'Arcy, R. C., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *NeuroReport*, *12*(2), 237–243. http://www.ncbi.nlm.nih.gov/pubmed/11209927

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21–58. https://doi.org/10.1080/016909698386582

Cowan, N. (1984). On Short and Long Audiology Stores. *Psychonomic Bulletin, 96*(2), 341-370.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 909–921. https://doi.org/10.1037/0278-7393.19.4.909

Cummins, F. (2012). Looking for rhythm in speech, *Empirical Musicology Review, 7*(1-2), 28-35.

Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*, *105*(4), 376–385. https://doi.org/10.1016/j.jecp.2009.12.004

Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(5), 441–452. https://doi.org/10.1002/wcs.1355

Dellwo, V. (2008). Influences of language typical speech rate on the perception of speech rhythm. *The Journal of the Acoustical Society of America*, *123*(5), 3427–3427. https://doi.org/10.1121/1.2934192

Dellwo, V., & Wagner, P. (2003). Relationships between rhythm and speech rate. *International Congress of Phonetic Sciences*, 471–474. https://pub.uni-bielefeld.de/record/1785384#contentnegotiation

Dellwo, V., Leemann, A., & Kolly, M. J. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *The Journal of the Acoustical Society of America*, *137*(3), 1513-1528.

Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: challenges and unresolved questions. *European Journal of Neuroscience*, *51*(5), 1151–1160. https://doi.org/10.1111/ejn.13802

Desjardins, R. N., Trainor, L. J., Hevenor, S. J., & Polak, C. P. (1999). Using mismatch negativity to measure auditory temporal resolution thresholds. *NeuroReport*, *10*(10), 2079–2082. https://doi.org/10.1097/00001756-199907130-00016

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*(1), 149-179.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294–311. https://doi.org/10.1016/j.jml.2008.06.006

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Ding, H., & Ye, D. (2005). Mismatch negativity to different time-frequency distribution complex tones. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, *7 VOLS*, 2083–2086. https://doi.org/10.1109/iembs.2005.1616869

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. https://doi.org/10.1016/j.neuroimage.2013.06.035

Donchin, E., & Heffley, E. (1978). Multivariate analysis of event-related potential data: A tutorial review. *Multidisciplinary Perspectives in Event-Related Brain Potential Research*, 555–572.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP, *Event-Related Brain Potentials in Man*, 349-411. https://doi.org/10.1016/b978-0-12-155150-6.50019-5

Duda-Milloy, V., Tavakoli, P., Campbell, K., Benoit, D. L., & Koravand, A. (2019). A time-efficient multi-deviant paradigm to determine the effects of gap duration on the mismatch negativity. *Hearing Research*, *377*, 34–43. https://doi.org/10.1016/j.heares.2019.03.004

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2014). Moving in time: Bayesian causal inference explains movement coordination to auditory beats. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1786). https://doi.org/10.1098/rspb.2014.0751

Endrass, T., Mohr, B., & Pulvermüller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, *19*(6), 1653–1660. https://doi.org/10.1111/j.1460-9568.2004.03247.x

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro-Otology*, *5*(3–4), 151–166. https://doi.org/10.1159/000013877

Fisher, A. E., Barnes, G. R., Hillebrand, A., Holliday, I. E., Witton, C., & Richards, I. L. (2006). Abnormality of mismatch negativity in response to tone omission in dyslexic adults. *Brain Research*, *1077*(1), 90–98. https://doi.org/10.1016/j.brainres.2005.12.121

Fisher, D. J., Grant, B., Smith, D. M., & Knott, V. J. (2011). Effects of deviant probability on the 'optimal' multi-feature mismatch negativity (MMN) paradigm. *International Journal of Psychophysiology*, *79*(2), 311–315. https://doi.org/10.1016/J.IJPSYCHO.2010.11.006

Fong, C. Y., Law, W. H. C., Uka, T., & Koike, S. (2020). Auditory mismatch negativity under predictive coding framework and its role in psychotic disorders. *Frontiers in Psychiatry*, *11*, 1–14. https://doi.org/10.3389/fpsyt.2020.557932

Ford, J. M., & Hillyard, S. A. (1981). Event-related potentials (ERPs) to interruptions of a steady rhythm. *Psychophysiology*, *18*(3), 322–330. https://doi.org/10.1111/j.1469-8986.1981.tb03043.x

Ford, J. M., Roth, W. T., & Kopell, B. S. (1976). Auditory evoked potentials to unpredictable shifts in pitch. *Psychophysiology*, *13*(1), 32–39. https://doi.org/10.1111/j.1469-8986.1976.tb03333.x

Frisina, R. D. (2001). Subcortical neural coding mechanisms for auditory temporal processing. *Hearing Research*, *158*(1–2), 1–27. https://doi.org/10.1016/S0378-5955(01)00296-9

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology*, *22*(7), 615–621. https://doi.org/10.1016/j.cub.2012.02.015

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361-377.

Gervain. (2018). Gateway to language: The perception of prosody at birth. In *Boundaries crossed, at the interfaces of morphosyntax, phonology, pragmatics, and semantics* (pp. 373–384). Springer.

Gibbon, D. (2018). The future of prosody; It's about time. *Proceedings of the 9th International Conference on Speech Prosody 2018*, 1–9.

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. *Eurospeech 2001 - Scandinavia*, 1–4.

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

Goswami, U. (2016). Educational neuroscience: Neural structure-mapping and the promise of oscillations. *Current Opinion in Behavioral Sciences*, *10*, 89–96. https://doi.org/10.1016/j.cobeha.2016.05.011

Goswami, U. (2018). A neural basis for phonological awareness? An oscillatory temporal-sampling perspective. *Current Directions in Psychological Science*, *27*(1), 56–63. https://doi.org/10.1177/0963721417727520

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. https://doi.org/10.1002/dys.1535

Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing*, *23*(8), 995-1019.

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. https://doi.org/10.1016/j.jml.2013.03.001

Grabe, E., & Low, E. L. (2013). Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology 7, 1982*, 1–16. https://doi.org/10.1515/9783110197105.515

Granier-Deferre, C., Ribeiro, A., Jacquet, A.-Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science, 14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology, 11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, *86*(July 2017), 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Hagoort, P. (2006). The memory, unification, and control (MUC) model of language. *Automaticity and Control in Language Processing, 1*, 1–287. https://doi.org/10.4324/9780203968512

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439–483. https://doi.org/10.1080/01690969308407585

Hari, R., Joutsiniemi, S. L., Hämäläinen, M., & Vilkman, V. (1989). Neuromagnetic responses of human auditory cortex to interruptions in a steady rhythm. *Neuroscience Letters*, *99*(1–2), 164–168. https://doi.org/10.1016/0304-3940(89)90283-8

Hauk, O., Giraud, A.-L., & Clarke, A. (2017). Brain oscillations in language comprehension. *Language, Cognition and Neuroscience, 32*(5), 533–535. https://doi.org/10.1080/23273798.2017.1297842

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86. https://doi.org/10.1016/j.jml.2016.01.001

Hilton, C. B., & Goldwater, M. B. (2021). Linguistic syncopation: Meter-syntax alignment affects sentence comprehension and sensorimotor synchronization. *Cognition*, *217*(August), 104880. https://doi.org/10.1016/j.cognition.2021.104880

Horváth, J., Müller, D., Weise, A., & Schröger, E. (2010). Omission mismatch negativity builds up late. *NeuroReport*, *21*(7), 537–541. https://doi.org/10.1097/WNR.0b013e3283398094

Hovsepyan, S., Olasagasti, I., & Giraud, A. L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature Communications, 11*(1), 1–12. https://doi.org/10.1038/s41467-020-16956-5

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. In *Wiley Interdisciplinary Reviews: Cognitive Science* (Vol. 2, Issue 5, pp. 580–593). John Wiley & Sons, Ltd. https://doi.org/10.1002/wcs.142

Huss, M., Verney, J. P., Fosker, T., Mead, N., & Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex*, *47*(6), 674-689.

Ilvonen, T., Kujala, T., Kozou, H., Kiesiläinen, A., Salonen, O., Alku, P., & Näätänen, R. (2004). The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). *Neuroscience Letters*, *366*, 235–240. https://doi.org/10.1016/j.neulet.2004.05.024

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M., & Alho, K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cognitive Brain Research*, *12*(3), 459–466. https://doi.org/10.1016/S0926-6410(01)00081-7

Jones, M. R. (2018). Time will tell: A theory of dynamic attending. Oxford University Press.

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive N*, *20*(11), 1940–1951.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687.

Kappenman, E. S., & Luck, S. J. (2012). ERP components: The ups and downs of brainwave recordings. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0014

Kaufeld, G., Bosker, H. R., Ten Oever, S., Alday, P. M., Meyer, A. S., & Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*, *40*(49), 9467-9475.

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, *16*(3). https://doi.org/10.1371/journal.pbio.2004473

Knight, R. A. (2011). Assessing the temporal reliability of rhythm metrics. *Journal of the International Phonetic Association*, *41*(3), 271–281. https://doi.org/10.1017/S0025100311000326

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language*, *76*(3), 332–339. https://www.sciencedirect.com/science/article/pii/S0093934X0092426X

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we head. *Current Biology*, *28*(18), 2867-2875.e3. https://doi.org/10.1016/j.cub.2018.07.023

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kujala, T., & Leminen, M. (2017). Low-level neural auditory discrimination dysfunctions in specific language impairment—A review on mismatch negativity findings. *Developmental Cognitive Neuroscience*, *28*(October 2016), 65–75. https://doi.org/10.1016/j.dcn.2017.10.005

Kujala, T., Lovio, R., Lepistö, T., Laasonen, M., & Näätänen, R. (2006). Evaluation of multi-attribute auditory discrimination in dyslexia with the mismatch negativity. *Clinical Neurophysiology, 117*(4), 885–893. https://doi.org/10.1016/J.CLINPH.2006.01.002

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics Electrified II (1994-2005). *Handbook of Psycholinguistics*, 659–724. https://doi.org/10.1016/B978-012369374-7/50018-3

Kraus, N., & White-Schwoch, T. (2015). Unraveling the biology of auditory learning: A cognitive-sensorimotor-reward framework. *Trends in Cognitive Sciences, 19*(11), 642–654. https://doi.org/10.1016/j.tics.2015.08.017.Unraveling

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., & Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(5), 1–32. https://doi.org/10.1002/wcs.1528

Lai, Y., Tian, Y., & Yao, D. (2011). MMN evidence for asymmetry in detection of IOI shortening and lengthening at behavioral indifference tempo. *Brain Research, 1367*, 170–180. https://doi.org/10.1016/J.BRAINRES.2010.10.036

Lallier, M., Lizarazu, M., Molinaro, N., Bourguignon, M., Ríos-López, P., & Carreiras, M. (2018). From auditory rhythm processing to grapheme-to-phoneme conversion: How neural oscillations can shed light on developmental dyslexia. In *Reading and dyslexia* (pp. 147–163). Springer, Cham. https://doi.org/10.1007/978-3-319-90805-2_8

Lallier, M., Molinaro, N., Lizarazu, M., Bourguignon, M., & Carreiras, M. (2017). Amodal atypical neural oscillatory activity in dyslexia: A cross-linguistic perspective. *Clinical Psychological Science, 5*(2), 379–401. https://doi.org/10.1177/2167702616670119

Lallier, M., Thierry, G., & Tainturier, M. J. (2013). On the importance of considering individual profiles when investigating the role of auditory sequential deficits in developmental dyslexia. *Cognition*, *126*(1), 121–127. https://doi.org/10.1016/j.cognition.2012.09.008

Langus, A., Marchetto, E., Bion, R. A. H., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language*, *66*(1), 285–306. https://doi.org/10.1016/j.jml.2011.09.004

Langus, A., Mehler, J., & Nespor, M. (2017). 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), 119–159.

Law, J. M., Vandermosten, M., Ghesquiere, P., & Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00482

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *The Journal of the Acoustical Society of America*, *54*(5), 1228-1234.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155-168. https://doi.org/10.1016/j.cortex.2015.02.014

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in cognitive sciences*, *4*(5), 187-196.

Light, G. A., Swerdlow, N. R., & Braff, D. L. (2007). Preattentive sensory processing as indexed by the MMN and P3a brain responses is associated with cognitive and psychosocial functioning in healthy adults. *Journal of Cognitive Neuroscience*, *19*(10), 1624–1632.

Ling, Grabe, & Nolan. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech*, *43*(4), 377–401.

Luck, S. J. (2012). electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components*. Oxford University Press.

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

Lupyan, G., & Clark, A. (2015). Words and the World. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Magne, C., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language*, *153–154*, 13–19. https://doi.org/10.1016/j.bandl.2016.01.001

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive science*, *31*(1), 133-156.

Magnuson, J. S., Li, M., Luthra, S., You, H., & Steiner, R. (2019). Does predictive processing imply predictive coding in models of spoken word recognition? *Proceedings of the Cognitive Science Society*, 735-740.

Mäntysalo, S., & Näätänen, R. (1987). The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biological Psychology*, *24*, 183–195. https://doi.org/10.1016/0301-0511(87)90001-9

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 960-977. https://doi.org/10.1037/0096-1523.33.4.960

May, P. J. C., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66–122. https://doi.org/10.1111/j.1469-8986.2009.00856.x

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & cognition*, *25*(1), 47-56.

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. https://doi.org/10.1002/hbm.23206

Näätänen, R. (1988). Implications of ERP data for psychological theories of attention. *Biological Psychology*, *26*(1–3), 117–163. https://doi.org/10.1016/0301-0511(88)90017-8

Näätänen, R. (1992). *Attention and brain function*. Lawrence Erlbaum Associates, Inc.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*(1), 6–18.

Näätänen, R. (2000). Mismatch negativity (MMN): Perspectives for application. *International Journal of Psychophysiology*, *37*(1), 3–10. https://doi.org/10.1016/S0167-8760(00)00091-X

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*(1), 1–21. https://doi.org/10.1111/1469-8986.3810001

Näätänen, R., & Escera, C. (2000). Mismatch negativity: clinical and other applications. Audiology and Neurotology, 5(3-4), 105-110.

Näätänen, R., Gaillard, A., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Kujala, T., & Light, G. (2019). Mismatch negativity: a window to the brain. Oxford University Press.

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, *48*(1), 4–22. https://doi.org/10.1111/j.1469-8986.2010.01114.x

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neuroscience Letters*, *107*(1–3), 347–352. https://doi.org/10.1016/0304-3940(89)90844-6

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. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Näätänen, R., Tervaniemi, M., Sussman, E. S., Paavilainen, P., & Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends in Neurosciences*, *24*(5), 283–288. https://doi.org/10.1016/s0030-6657(08)70226-9

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859. https://doi.org/10.1037/0033-2909.125.6.826

Nordby, H., Hammerborg, D., Roth, W. T., & Hugdahl, K. (1994). ERPs for infrequent omissions and inclusions of stimulus elements. *Psychophysiology, 31*(6), 544–552. https://doi.org/10.1111/j.1469-8986.1994.tb02347.x

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988a). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology*, *25*(3), 262–268. https://doi.org/10.1111/j.1469-8986.1988.tb01239.x

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988b). Event-related potentials to time-deviant and pitch-deviant tones. *Psychophysiology*, *25*(3), 249–261. https://doi.org/10.1111/j.1469-8986.1988.tb01238.x

Novak, G., Ritter, W., & Vaughan, H. G. (1992). Mismatch detection and the latency of temporal judgments. *Psychophysiology*, *29*(4), 398–411. https://doi.org/10.1111/j.1469-8986.1992.tb01713.x

Nozaradan, S., Peretz, I., & Keller, P. E. (2016). Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Scientific Reports*, *6*(August 2015), 1–12. https://doi.org/10.1038/srep20612

Oceák, A., Winkler, I., & Sussman, E. (2008). Units of sound representation and temporal integration: A mismatch negativity study. *Neuroscience Letters*, *436*(1), 85–89. https://doi.org/10.1016/j.neulet.2008.02.066

Ohmae, S., & Tanaka, M. (2016). Two different mechanisms for the detection of stimulus omission. *Scientific Reports, 6*(1), 1–9. https://doi.org/10.1038/srep20615

Osterhout, Lee, Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. *International Journal of Psychophysiology*, *88*(2), 109-123.

Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology, 82*(3), 219–226. https://doi.org/10.1016/j.biopsycho.2009.07.008

Pato, M. V., Jones, S. J., Perez, N., & Sprague, L. (2002). Mismatch negativity to single and multiple pitch-deviant tones in regular and pseudo-random complex tone sequences. *Clinical Neurophysiology*, *113*, 519–527.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in psychology*, *3*, 320.

Perez, V. B., & Vogel, E. K. (2012). What ERPs can tell us about working memory. the oxford handbook of event-related potential components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0180

Pettigrew, C., Murdoch, B., Kei, J., Ponton, C., Alku, P., & Chenery, H. (2005). The mismatch negativity (MMN) response to complex tones and spoken words in individuals with aphasia) The mismatch negativity (MMN) response to complex tones and spoken words in individuals with aphasia. *Aphasiology*, *19*(2), 131–163. https://doi.org/10.1080/02687030444000642

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Finnigan, S., Alku, P., Kei, J., Sockalingam, R., & Chenery, H. J. (2004). Automatic auditory processing of English words as indexed by the mismatch negativity, using a multiple deviant paradigm. *Ear and Hearing*, *25*(3), 284–301. https://doi.org/10.1097/01.AUD.0000130800.88987.03

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, *12*(6), 1038–1055. https://doi.org/10.1162/08989290051137567

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 1–13. https://doi.org/10.1038/s41583-020-0304-4

Polich, J. (2012). Neuropsychology of P300. *The Oxford Handbook of Event-Related Potential Components*, 159–188. https://doi.org/10.1093/oxfordhb/9780195374148.013.0089

Prete, D. A., Heikoop, D., McGillivray, J. E., Reilly, J. P., & Trainor, L. J. (2022). The sound of silence: Predictive error responses to unexpected sound omission in adults. *European Journal of Neuroscience, 55*(8), 1972-1985.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R. J., & Näätänen, R. (2001). memory traces for words as revealed by the mismatch negativity. *NeuroImage*, *14*(3), 607–616. https://doi.org/10.1006/NIMG.2001.0864

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*, *79*(1), 49–71. https://doi.org/10.1016/J.PNEUROBIO.2006.04.004

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, *105*(1), 512–521. https://doi.org/10.1121/1.424522

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87. https://doi.org/10.1038/4580

Rathcke, T. V., & Smith, R. H. (2015). Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. *The Journal of the Acoustical Society of America, 137*(5), 2834–2845. https://doi.org/10.1121/1.4919322

Remez, R., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949. https://doi.org/10.1126/science.7233191

rhythm. 2021. In Merriam-Webster.com. Retrieved December 17, 2021, from https://www.merriam-webster.com/dictionary/rhythm

Rimmele, J. M., Poeppel, D., & Ghitza, O. (2021). Acoustically driven cortical delta oscillations underpin perceptual chunking. ENeuro. https://doi.org/10.1101/2020.05.16.099432

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *336*(1278), 367–373. https://doi.org/10.1098/rstb.1992.0070

Rowland, J., Kasdan, A., & Poeppel, D. (2019). There is music in repetition: Looped segments of speech and nonspeech induce the perception of music in a time-dependent manner. *Psychonomic Bulletin & Review, 26*(2), 583–590. https://doi.org/10.3758/s13423-018-1527-5

Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., & Münte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neuroscience Letters, 308*(1), 33–36. https://doi.org/10.1016/S0304-3940(01)01977-2

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): Emitted MMN to violations of an auditory gestalt. *Psychophysiology*, *49*(4), 544–548. https://doi.org/10.1111/j.1469-8986.2011.01336.x

Sanmiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience, 33*(20), 8633-8639.

Sable, J. J., Gratton, G., & Fabiani, M. (2003). Sound presentation rate is represented logarithmically in human cortex. *European Journal of Neuroscience, 17(*11), 2492–2496. https://doi.org/10.1046/j.1460-9568.2003.02690.x

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of memory and language, 71*(1), 145-163.

Scharinger, M., Steinberg, J., & Tavano, A. (2017). Integrating speech in time depends on temporal expectancies and attention. *Cortex*, *93*, 28–40. https://doi.org/10.1016/j.cortex.2017.05.001

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998). Auditory processing and dyslexia: Evidence for a specific speech processing deficit. *NeuroReport*, *9*(2), 337-340.

Sebastian, C., & Yasin, I. (2008). Speech versus tone processing in compensated dyslexia: Discrimination and lateralization with a dichotic mismatch negativity (MMN) paradigm. *International Journal of Psychophysiology*, *70*, 115–126. https://doi.org/10.1016/j.ijpsycho.2008.08.004

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. Journal of Memory and Language, 57(1), 24-48.

Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology, 3*, 371-405.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303-304.

Shiga, T., Yabe, H., Yu, L., Nozaki, M., Itagaki, S., Lan, T.-H., & Niwa, S. (2011). Temporal integration of deviant sound in automatic detection reflected by mismatch negativity. *NeuroReport*, *22*(7), 337–341. https://doi.org/10.1097/WNR.0b013e3283462db6

Shinozaki, N., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Matsuoka, T., & Kaneko, S. (2003). Spectrotemporal window of integration of auditory information in the human brain. *Cognitive Brain Research*, *17*(3), 563–571. https://doi.org/10.1016/S0926-6410(03)00170-8

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, *12*(6), 657–663. https://doi.org/10.1006/NIMG.2000.0646

Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *NeuroReport*, *13*(4), 521–525. https://doi.org/10.1097/00001756-200203250-00033

Sorokin, A., Alku, P., & Kujala, T. (2010). Change and novelty detection in speech and non-speech sound streams. *Brain Research*, *1327*, 77–90. https://doi.org/10.1016/j.brainres.2010.02.052

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97. https://doi.org/10.1016/j.bandc.2015.11.003

Spratling, M. W. (2008a). Predictive coding as a model of biased competition in visual attention. *Vision Research*, *48*(12), 1391–1408. https://doi.org/10.1016/j.visres.2008.03.009

Spratling, M. W. (2008b). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, *2*(OCT), 1–8. https://doi.org/10.3389/neuro.10.004.2008

Spratling, M. W., De Meyer, K., & Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational intelligence and neuroscience, 2009.*

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205), 427–459

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language*, *82*, 1-17.

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0197

Tervaniemi, M., Ilvonen, T., Sinkkonen, J., Kujala, A., Alho, K., Huotilainen, M., & Näätänen, R. (2000). Harmonic partials facilitate pitch discrimination in humans: electrophysiological and behavioral evidence. *Neuroscience Letters*, *279*, 29–32.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, *161*(1), 1–10. https://doi.org/10.1007/s00221-004-2044-5

Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport*, *5*(7), 844-846.

Tervaniemi, M., Schröger, E., & Näätänen, R. (1997). Pre-attentive processing of spectrally complex sounds with asynchronous onsets: An event-related potential study with human subjects. *Neuroscience Letters*, *227*(3), 197–200. https://doi.org/10.1016/S0304-3940(97)00346-7

Tillmann, B. (2012). Music and language perception: Expectations, structural integration, and cognitive sequencing. *Topics in Cognitive Science*, *4*(4), 568–584. https://doi.org/10.1111/j.1756-8765.2012.01209.x

Tseng, C.-Y., & Fu, B.-L. (2005). Duration, intensity and pause predictions in relation to prosody organization. *Ninth European Conference on Speech Communication and Technology*, 1405–1408.

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Wacongne, C., Labyt, E., Van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(51), 20754–20759. https://doi.org/10.1073/pnas.1117807108

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes*, *25*(7-9), 905-945.

Wang, J., Friedman, D., Ritter, W., & Bersick, M. (2005). ERP correlates of involuntary attention capture by prosodic salience in speech. *Psychophysiology*, *42*(1), 43–55. https://doi.org/10.1111/j.1469-8986.2005.00260.x

Whalley, K., & Hansen, J. (2006). The role of prosodic sensitivity in children’s reading development. *Journal of Research in Reading*, *29*(3), 288–303. https://doi.org/10.1111/j.1467-9817.2006.00309.x

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wilding, E. L., & Ranganath, C. (2012). Electrophysiological Correlates of Episodic Memory Processes. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0187

Winkler, I., Cowan, N., Csépe, V., Czigler, I., & Näätänen, R. (1996). Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. *Journal of Cognitive Neuroscience, 8(*5), 403–415. https://doi.org/10.1162/jocn.1996.8.5.403

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*(1-2), 239-252.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, a, Czigler, I., Csépe, V., Ilmoniemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology*, *36*(5), 638–642.

Winkler, I., Reinikainen, K., & Näätänen, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Perception & Psychophysics*, *53*(4), 443–449. https://doi.org/10.3758/BF03206788

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport, 8*(8), 1971–1974. https://doi.org/10.1097/00001756-199705260-00035

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilomoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*(5), S0048577298000183. https://doi.org/10.1017/S0048577298000183

Ylinen, S., Bosseler, A., Junttila, K., & Huotilainen, M. (2017). Predictive coding accelerates word recognition and learning in the early stages of language development. *Developmental Science, 20*(6). https://doi.org/10.1111/desc.12472

Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P. (2016). Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain and Language*, *162*, 72–80. https://doi.org/10.1016/j.bandl.2016.08.007

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 363, Issue 1493, pp. 1087–1104). Royal Society. https://doi.org/10.1098/rstb.2007.2161

Zevin, J. D., Datta, H., Maurer, U., Rosania, K. A., & McCandliss, B. D. (2010). Native language experience influences the topography of the mismatch negativity to speech. *Frontiers in Human Neuroscience*, *4*, 212. https://doi.org/10.3389/fnhum.2010.00212

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.

Zoefel, B. (2018). Speech entrainment: Rhythmic predictions carried by neural oscillations. *Current Biology, 28*(18), R1102-R1104.

CHAPTER TWO

Introduction

As is evident from the background presented in the first chapter, the question of timing in language and speech is a vast one. Not only are there numerous meanings to the way timing can be defined, there is also a lack of coherence to the way it is dealt with in literature. Out of so many different aspects of timing in speech, are time cues that play on expectation important? And if so, how, and within which modalities? It is possible that cues for timing may be relevant in specific contexts, but not in others. It is not a stretch, then, to consider that if expectancy-related time cues react only to simple sounds, then they have no role in speech whatsoever. The aim of this chapter is to examine how anticipation affects the perception and processing of time cue variations in a series of isochronous simple tones. This question will further be developed in the next chapter (*Chapter three*) to continue exploring how the evoked neural responses to expectancy-based modulations of time cues change for decidedly non-speech stimuli compared to decidedly speech-related tokens.

Temporal variation and auditory processing

The current chapter explores how timing variation in language is processed by tracking changes in the EEG-based brain response known as the *mismatch negativity* (MMN). As was reviewed in the previous chapter, the MMN has been extensively used as a tool to study pre-attentive processing. In a standard oddball paradigm, a MMN is elicited to a "deviant" stimulus presented within a stream of identical stimuli. Often, the deviant stimulus differs on one or more acoustic features, eliciting a negative-going response about 150-200 ms after its presentation in the auditory stream. Rarely does the deviant differ on other features, like its timing of presentation. The following section describes the existing work on timing variations in isochronous sequences, and how the MMN changes in response.

Limited work has been conducted that makes use of deviants that differ in time of presentation. Deviants that differ from the standard, isochronously-presented stimuli in their time of presentation alone, with all other acoustic features the same, are referred to in this dissertation as *timing deviants*. *Early timing deviants* refer to those that occur earlier than expected (that is, at a stimulus onset asynchrony (SOA) that is shorter than that of the standards). *Late timing deviants* refer to those that occur later than expected (that is, at an SOA that is longer than that of the standards). The experiments reported in this chapter seek to understand how varying SOAs as a model for temporal variation predicts processing of these differences. Previous work has used SOA and inter-trial interval (ITI) differences as a measure of feature perception to track the threshold of MMN elicitation and make predictions about the limits of the temporal window of integration (TWI). These timing differences stimuli or trains of stimuli (trials) have also been used as a measure of beat perception and musical processing. However, to the best of our knowledge, no work has looked at temporal variation specifically in the context of speech, and how the cognitive processing of these timing differences changes from simple auditory inputs to more complex ones, if at all.

Omission deviants

MMNs have been elicited to missing or omitted deviants in several experiments. One of the most commonly employed experimental designs to elicit this type of MMN is the classic oddball paradigm, which uses a missing (or omitted) stimulus as the deviant. Responses are recorded to the onset of the omission, and a distinct negativity (the MMN) is found in response to the omissions when the stimuli have been presented at an SOA that falls within the temporal window of integration (TWI), often under 200 ms (Yabe et al., 1997; 1998). Salisbury (2012) found that omission deviants could also be elicited at longer SOAs if they were presented within perceptual groups formed using gestalt principles. However, omission MMNs have been shown to be much more reliable when presented amongst a steady train of tones, rather than within a pattern of tones (Horváth et al., 2010). Within patterns that made use of musical beats, omissions at strong metrical positions elicited a bigger MMN compared to omissions at weak metrical positions (Bouwer et al., 2019). Musical background of the participants also plays a role in MMN elicitation, with musicians showing a MMN elicited at a longer SOA (bigger TWI) compared to non-musicians (Rüsseler et al., 2001). Difficulties with language, like dyslexia, have been found to affect the omission MMN, with one study reporting no MMNs elicited to omissions for dyslexic adults when compared to neurotypical adults when an SOA of 175 ms was used (Fisher et al., 2006). The predictability of stimulus omission also affects responses. In one study, participants heard tones presented isochronously, but paired in frequency such that every other tone was a repetition of the first (Bendixen et al., 2009). Either the first tone (unexpected) or second tone (predictable) was omitted from the pairs a fixed number of times and presented to participants alongside a control condition where unpaired tones were omitted randomly. Predictability was driven from the assumption that the second tone would be perceived as the repetition of the first, and so an omission for the second tone would be more expected than that for the first. Results showed that the highly predictable tone-pair condition evoked an omission response that was like the control tone omissions, but not to the unexpected tone-pair condition, which showed a differential (more negative) response. Bendixen et al. (2009) concluded that this result supports the hypothesis that the auditory system makes predictions when processing incoming auditory stimuli. A study that controlled stimulus predictability and omission based on participant response showed comparable results, with a larger negativity for a predicted omission versus an unpredicted sound (Sanmiguel et al., 2013). Omissions of word-final syllables presented in both a word and sentential context, too, showed a larger negativity (omission MMN) when the speech segment was more predictable compared to when it was not (Bendixen et al., 2014).

Finally, recent work has shown that omission deviants can be elicited beyond the TWI when the unanticipated omissions are compared to anticipated omissions (Prete et al., 2022). This study examined MMNs to stimuli presented at an SOA of 500 ms (well beyond the TWI) that contained either expected silences (translated as longer SOAs between stimuli) or unexpected silences (translated as occasionally longer SOAs). Steady-state EEG was recorded to complete silence as well, and results found a small but reliable MMN response elicited to the unexpected omissions, despite the longer SOA. This study finds evidence for the predictive coding framework as a mechanism for auditory processing in the brain.

Omission deviants, thus, seem to be a highly enlightening source of information about perceptual processing. Variations in omission-related MMN provide insights into these cognitive processes. They can also help separate the cognitive profiles of different special populations from each other. Introducing SOA differences to a train of standard stimuli delivered with constant SOAs will inevitably lead to ‘gaps’ where the expected presentation of the stimuli should have occurred but did not because it was presented either too late, or too early. While timing deviations perceptually account for the change in SOA and no response is recorded to this gap (Sable et al., 2003)—ostensibly because the larger pattern in the stimuli overshadows the resulting ‘omission’—understanding the omission deviant is nevertheless an important step to understanding how this timing variation is processed in general.

Unanticipated temporal variation

For simple tones presented earlier than expected, either in the context of a pattern of tones or in isolation, a negative response has been observed that matches the latency of a standard MMN (Ford & Hillyard, 1981; Hari et al., 1989; Näätänen et al., 1993; Alain et al., 1999; Lai et al., 2011). Specifically, it has been shown that changes in the inter-stimulus interval (ISI) elicit a negative response that is proportional to the magnitude of the difference between the deviant and the standard, so that a larger temporal difference results in an MMN of greater amplitude (Alain et al., 1999; Kisley et al., 2004). However, in other studies, this MMN-like response has been either minimal and has not reached significance, or undetectable, and apparently modulated by whether the stimulus was attended to, or not (Nordby et al., 1988b, 1988a). Timing deviants presented earlier than the standards have also been reported to affect the standard N1-P2 event-related response complex, with at least one study reporting a bigger complex for the early deviant compared to the standard (Ford & Hillyard, 1981). Typically, the N1-P2 complex is an exogenous response elicited to the onset of auditory stimuli, often modulated by changes in participant attention or alertness (Lightfoot, 2016). Ford and Hillyard (1981) suggest that the presence of the larger N1-P2 complex to stimuli with shorter ISI might be a response different from the typical N1-P2, and more like a mismatch negativity (MMN), despite the fact that the response they report is morphologically similar to an N1-P2 response. Tavakoli et al. (2021) report similarly a complex with the N1-P2 and MMN responses known as the *deviant-related negativity* (DRN), elicited to auditory stimuli with gaps (see also Duda-Milloy et al., 2019).

Work involving the use of late timing deviants, that is, timing deviants where the ISI or SOA is increased from the standards, is limited. Previous work using tones presented in a pattern found that varying the timing of the last tone in the sequence elicited an MMN that was proportional in latency to the timing difference from the standard, and increased in amplitude as the difference in SOA increased (in either direction) (Sable et al., 2003). In other words, deviants that were presented with a shorter SOA (‘early’) showed an earlier latency compared to deviants that were presented with a longer SOA (‘late’), which showed a later latency. In terms of amplitude, the bigger the difference in SOA between the deviant and the standard, the larger the MMN elicited (with the largest amplitude observed for the early deviant compared to the late). Another study examined changes in the P3 component in response to variations in the pacing of musical beats for sequences of tones where the last tone in the pattern was presented earlier or later than the standard interval (Jongsma et al., 2007). The P3 (or P300) is an attention-modulated ERP, often elicited by similar paradigms to the MMN but requiring the participants to engage with the stimuli in terms of providing a response related to an acoustic feature of the stimuli (see Polich, 2020 for a review). This response comprises mainly of the P3a and P3b subcomponents: the P3a is often elicited to rare deviants in stimuli trials and is commonly observed alongside the MMN with a similar fronto-central scalp distribution and a latency of about 250-350 ms; the P3b response is elicited when participants are attending and responding to stimuli, and has a more posterior distribution and later latency compared to theP3a (~300-600 ms) (Comerchero & Polich, 1999; Jongsma et al., 2007). Jongsma et al. (2007) found that early timing deviants led to a decrease in the amplitude of early-latency P3 subcomponents (P3a), but an amplitude increase in the late P3 subcomponents (P3b). Conversely, late timing deviants led to an increase in the P3a-like responses observed. These results suggest that the early and late timing deviants are not necessarily processed the same way. Jongsma et al. (2007) interpret these results in terms of an ‘internal oscillator’ model of attention. An unexpectedly early deviant arrives when the attention oscillator is still reaching its target maxima and the late P3 subcomponent functions as a “surprise” response similar to the P3b (Donchin, 1981). The late deviant, on the other hand, is presented after expectation has built for it, and the early P3 subcomponent (i.e., the P3a) functions as an “anticipation” response (Jongsma et al., 2007, p. 226).

Finally, it is worth noting in these discussions of the timing deviants that it is difficult to say whether the temporal differences that drive expectations are a result of the timing difference between stimulus offset-to-onset (ISI), or stimulus onset-to-onset (SOA). Both these measures are affected when either is manipulated. The fact remains, however, that a change in timing between subsequent stimuli leads to differential responses, and therefore, the most important factor in manipulating timing is consistency. In the experiments reported in this chapter, therefore, SOA was manipulated to create timing differences and drive expectancy.

In the current experiment, simple tones with different pacing were presented to participants. The aim of this experiment was to explore how variation in timing within a highly controlled environment effects the MMN, and to create a baseline of responses that can then be used for further work on linguistic timing variation. While previous work has explored the effect of specific timing variations (or beat acceleration/deceleration, see Jongsma et al., 2007) on the MMN, to our knowledge little work has been performed on timing pattern recognition and the MMN in a wider scope. That is to say, most timing variation or pattern-based stimulus experiments used stimuli that were finite sequences grouped together temporally. In our experiment, we make use of timing deviants that exist within the stream of isochronous standards, and we also present two different types of timing deviants in one block, modeling temporal variation that exists naturally (in terms of early/late elements in variations of any type of sound, including speech).

Hypotheses

In this experiment, we hypothesised that unexpected variation in presented stimuli will elicit the MMN response, and this MMN will differ based on the type of variation. Two types of variations were examined: temporal variation as a result of a missing stimulus (omission deviant), and temporal variation due to changing SOAs. The omission deviant task was a follow-up replication experiment to the Yabe et al. (1997) study and was conducted to assess whether the temporal window of integration (TWI) identified in that study would replicate in our laboratory setting. This included the use of more modern equipment and software, as well as changes to the methodologies and analyses (including but not limited to: more electrodes, higher online sampling rate, wider online bandpass filtering, and more conservative measures of artifact rejection). In line with the original experiment, we hypothesised that SOAs shorter than 150 ms would elicit an MMN to a missing (omission) deviant, but not for SOAs longer than 150 ms.

For timing deviants, that is, deviants presented earlier or later than the standards (varying SOAs), we expected the MMN response to be modulated accordingly. Previous work using tones presented in a pattern found early timing deviants to elicit a larger-amplitude MMN compared to a late timing deviant (Sable et al., 2013). This experiment presented these timing variations within the context of a sequence of tones, whereas the current chapter reports an experiment where timing manipulations are made within a continuous sequence of tones. However, we hypothesised similar results: the MMN would be earlier and bigger in amplitude compared to a typical MMN for early deviants, and for the late deviants, we expect the MMN to be later in latency, and larger in amplitude compared to a typical MMN (Sable et al., 2003). Only tone-locked MMN responses were analysed in our experiment, as previous work found no responses to the ‘expected’ (but empty) tone position in conditions of temporal variation, (Sable et al., 2003).

Participants completed two experiments consisting of two and three blocks, respectively. The first experiment was a replication of the study conducted by Yabe et al. in 1997. The motivation to replicate this experiment was to observe the MMN to an ‘absent’ stimulus in the event that these analyses would be conducted for the second experiment that made use of timing deviants. Yabe et al. (1997) hypothesised that tones presented within the temporal window of integration would elicit an MMN when a tone was missing, but not when the tones were presented at an SOA outside the temporal window of integration. By looking at responses across a range of SOAs, the threshold of the temporal window of integration could be observed by noting the last SOA that elicited the MMN. For the current replication, two SOAs were chosen, one within the temporal window of integration identified to elicit an MMN (SOA = 125 ms), and one outside the temporal window of integration that would *not* elicit an MMN (SOA = 250 ms).

Experiment 1

Methods

Participants

30 Canadian University students (mean age = 20.2, S.D. = 2.1; 21 female; 2 participants did not report age, and 2 did not report sex) with reported normal hearing and normal or correct-to-normal vision were recruited to participate in this experiment. Participants were restricted in age (30 years or under), as perception thresholds go down with age (Alain et al., 2004). Based on the Edinburgh Handedness Inventory (Oldfield, 1971), 22 participants were right-handed, 5 were left-handed, and 3 were ambidextrous. Data from 6 participants were removed due to excessive noise or technical issues, and partial data were used from 7 participants. All participants provided written consent to participate in this experiment in line with the ethical standards of the Declaration of Helsinki, and were either compensated with money or course credit, or volunteered to

participate in this experiment. This study was cleared under the Hamilton Integrated Research Ethics Board (HiREB) in Hamilton, Ontario, Canada.

Experimental procedure

An auditory oddball paradigm was used. For the first experiment, one type of standard and one type of deviant were presented per block. Four blocks were presented in total with two different SOAs (125 ms and 250 ms), counterbalanced across presentations. A total of 810 standards (90%) and 90 deviants (10%) were presented in each block, as per the original paper.

Stimuli

The standard was a 1000 Hz tone generated in Praat (Boersma & Weenink, 2019) (50 ms duration, 5 ms linear rise/fall time) presented at a sound pressure level (SPL) of 80 dB (time-weighting: slow; frequency rating: C; model: RadioShack SPL meter; intensity ratings were obtained by attaching the SPL meter directly to the earphones, so the conditions were identical to those at the listener's ear canal/tympanic membrane). In the first experiment, the deviant was a 'missing' stimulus (omission) presented as a gap between two standards, such that the time between one standard and the next was twice the length of the regular SOA (SOA*2) (see **Figure 1** for an illustrated example). The stimuli were presented to participants via ER-1 Insert earphones (Etymotic Research, Inc., Elk Grove Village, IL, www.etymotic.com) using Presentation® software (Version 18.0 Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The order of the stimuli was pseudo-randomised according to the following parameters: 7 standards presented at the start of a block; minimum 2 standards in a row; and, maximum 4 standards in a row.

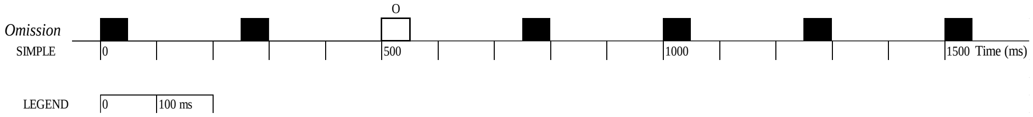

Figure 1. A visual representation of stimuli timing across the omission condition. The left edge of the block represents the onset of the stimulus. All stimuli were 50 ms in length. Standard pure tones are represented in black. The white blocks represent the expected presentation of the tone. For the Omission deviants, this represents a missing stimulus presented in a sequence of steady tones (in this example, at a fixed SOA of 250 ms).

Testing procedure

After providing informed consent at the beginning of the testing session, participants were asked to fill out a short demographic survey that asked for information such as age, years of education, languages spoken and order of usage frequency, any visual or hearing impairments, and any relevant medical history (for example, history of concussion, etc.). Participants were also asked to fill out the Edinburgh Handedness Inventory (Oldfield, 1971), but this information did not affect eligibility.

Once the forms were filled out, participants took part in the EEG experiment. They were seated in a comfortable chair, about 90 cm from a computer monitor and asked to watch a silent film, and to not pay attention to the sounds presented to them. No response was required from the participants for the duration of the experiment. The first experiment with the omission deviant ran for approximately 12 minutes. Participants were given brief breaks between blocks, with the option of taking a longer break at any point during the experiment.

Electroencephalography (EEG) recordings

EEG was recorded using a BioSemi ActiveTwo system with 64 Ag/AgCl electrodes (International 10-20 system) digitally sampled at 512 Hz and bandpass filtered at 0.01-100 Hz. Five Ag/AgCl external electrodes were placed on the participant’s nose, left and right mastoids, and above and beside their left eye. The electrodes placed above and beside the left eye were used to record the EOG (electrooculogram) with the same bandpass and sampling rate. Online EEG acquisition was referenced to the DRL (driven right-leg) and re-referenced offline to the average of the left and right mastoids.

EEG data analyses

EEG data were pre-processed and cleaned using BrainVision Analyzer (v2.1.2.327, BrainVision Analyzer, Brain Products GmbH, Gilching, Germany). Data were re-filtered to 0.1-30 Hz (24 dB/oct), and any noisy channels were interpolated from the surrounding channels and replaced (up to a maximum of four channels) (Duda-Milloy et al., 2019). Data were visually inspected for artifacts (e.g., due to movement, or electrode noise), and sections with artifacts greater than 100 µV were removed. Blocks where > 20% of deviant epochs were removed were not used for analyses, so the deviant grand average for each participant contained at least 75/90 segments (range = 77-90, S.D. = 3.5). Ocular Independent Component Analysis (ICA) was performed to remove vertical and horizontal eye movements. The data were then segmented into epochs of -100 ms to 295 ms (for the omission deviant) and averaged. Difference waveforms were produced by subtracting the averaged waveform of the standard condition from the deviant condition. Automated peak detection (Barr et al., 1978) was conducted to find the maximum amplitude pertaining to the MMN within a window of 100 ms to 200 ms. The window of the MMN for the omission deviant was kept consistent with the parameters used by Yabe et al. (1997). For further analyses, the 64 electrodes were divided into groups of 4-6 electrodes that formed pre-defined regions of interest (ROIs). These ROIs were broadly identified using 3 sagittal planes (right, middle, and left) and 3 coronal planes (frontal, parietal, and occipital) along the scalp for a total of nine ROIs: Right Frontal (RF), Middle

Frontal (MF), Left Frontal (LF), Right Central (RC), Middle Central (MC), Left Central (LC), Right Parietal (RP), Middle Parietal (MP), and Left Parietal (LP). For the purposes of our analyses, only four ROIs (RF, MF, LF, MC) were included as the MMN is a fronto-central component (Näätänen et al., 2019).

One sample *t*-tests were conducted to compare the mean peak amplitude and latency values of the observed responses to baseline. A 2 x 4 Repeated Measures ANOVA for the MMN response was conducted evaluating the effect of SOA (Short, Long) and ROI (RF, MF, LF, MC) on the mean peak amplitude and latency. Four ROIs were chosen to maximise power, centering around the fronto-central regions where the MMN response is maximal. Post-hoc paired *t*-tests were used to explore any significant main effects. Greenhouse-Geisser corrections for sphericity were applied, where needed; uncorrected degrees of freedom, and corrected *p*-values and generalised eta-squared (η^2) values are reported. *p*-values were interpreted as significant at an alpha (α) level of 0.05, unless otherwise noted. All parametric statistics were conducted using the R Studio (v. 1.4.1106, R Studio Team, 2021) environment for R 4.1.0 (R Core Team, 2021). The following packages were used to reshape data, calculate statistics, and generate figures: rstatix (0.7.0 Kassambara, 2021); dplyr (v. 1.0.6, Wickham, François, Henry, & Müller, 2021), ggplot2 (Wickham, 2016), and ggpubr (v. 0.4.0, Kassambara, 2020).

Results

This experiment replicated findings by Yabe et al. (1997) by evaluating the elicitation of the MMN within and outside the TWI. As per the original paper, an MMN to omissions was elicited only when each presented stimulus falls within a specific window of time. This temporal window of integration was found by Yabe et al. (1997) to be about 150 ms for simple tones. Therefore, in this reported experiment, participants were presented with simple tones at two SOAs: a short SOA that fell within the TWI, 125 ms, that was hypothesised to elicit an MMN, and a long SOA that fell outside the TWI, 250 ms, that was hypothesised to not elicit an MMN. The two different SOAs were presented in different blocks, and only the timing differed between them. All other parameters regarding mode of presentation and stimulus were kept the same.

Difference waves were calculated by subtracting the averaged response to the standard from the averaged response to the deviant for each participant in each block. Mean peak amplitude and mean peak latency values for the MMN were extracted from this subtraction by identifying the negative-most value within a pre-defined time window of 100 ms – 200 ms (MMN, Näätänen et al., 2019). One sample *t*-tests were conducted to evaluate whether the mean peak amplitude of this difference wave was significantly different from 0 μV (baseline) or not.

One-sample *t*-tests showed a significant difference from baseline for the Short SOA, $t(22) = -4.53$, $p < 0.001$, as well as for the Long SOA, $t(22) = -7.78$, $p < 0.001$. Visually, the two responses differ, with the Short SOA eliciting a response in the designated time window of 100 ms – 200 ms that is more negative compared to that observed for the Long SOA (**Table 1**). A larger negativity can be observed in **Figure 2** (left) around 100 ms for the Short SOA; a smaller negativity is observed for the Long SOA (right) within the same time window.

A 2 x 4 repeated measures ANOVA was conducted to evaluate the effect of rate of presentation (Short/Long SOA) and ROI (LF, MF, RF, MC) on the mean peak amplitude for the MMN response. A significant main effect of ROI was observed, $F(3, 66) = 10.018$, $p = 0.000016$, $\eta^2 = 0.028$. The simple main effect of Rate of Presentation (Short SOA vs. Long SOA) was not significant, $F(1, 22) = 0.031$, $p = 0.862$, $\eta^2 = 0.0004$, nor was the interaction effect between Rate of Presentation and ROI significant, $F(3, 66) = 0.110$, $p = 0.954$, $\eta^2 = 0.0002$). MMN latency was not analysed as there were no original hypotheses pertaining to it (see Yabe et al., 1997).

A reversal of amplitude, from negative to positive, was observed for the shorter SOA when comparing the difference wave in the frontal regions (LF, MF, RF) to the waveform at the averaged mastoids. This amplitude reversal is one of the features used to identify an MMN (Paavilainen et al., 1991). This amplitude reversal is missing for the Long SOA block (Figure 3).

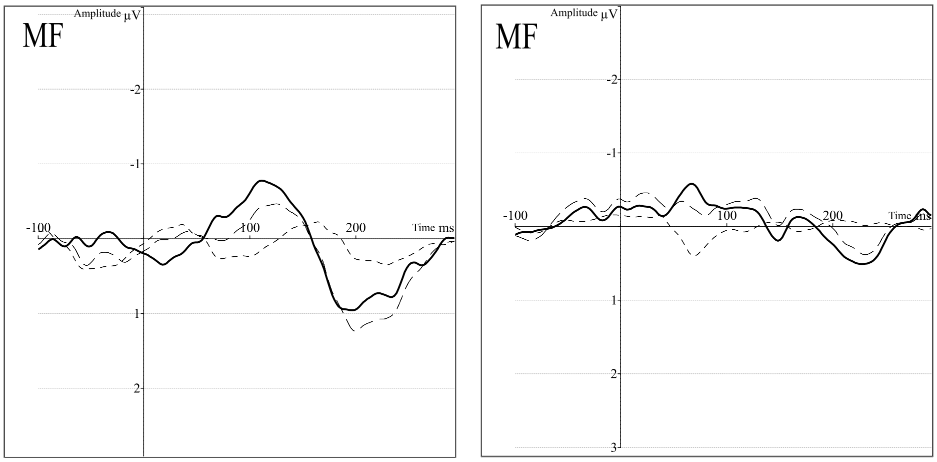

Figure 2. Figure showing difference waves created by subtracting the standard from the omission deviant in the Short SOA condition (left) and long SOA condition (right). The responses are from the Middle-Frontal ROI and have been filtered at 1-25 Hz. The solid line shows the difference wave; the long dashed line shows the deviant response and the short dashed line shows the standard response.

Table 1. Mean amplitude and latency values for the omission deviant across the two levels of rate of presentation (Short SOA, and Long SOA) for the MMN response. Values reported were recorded from the difference waves at the Middle-Frontal ROI.

Peak mean amplitude (µV) and latency (ms) by rate of presenation (ROI: MF)

		MMN	
Rate of presentation	*n*	*Mean amplitude (µV)*	*Mean latency (ms)*
Short (SOA = 125 ms)	23	-1.57 ± 1.66	123 ± 23.6
Long (SOA = 250 ms)	23	-1.59 ± 0.98	148 ± 33.4

mean ± standard deviation (s.d.)

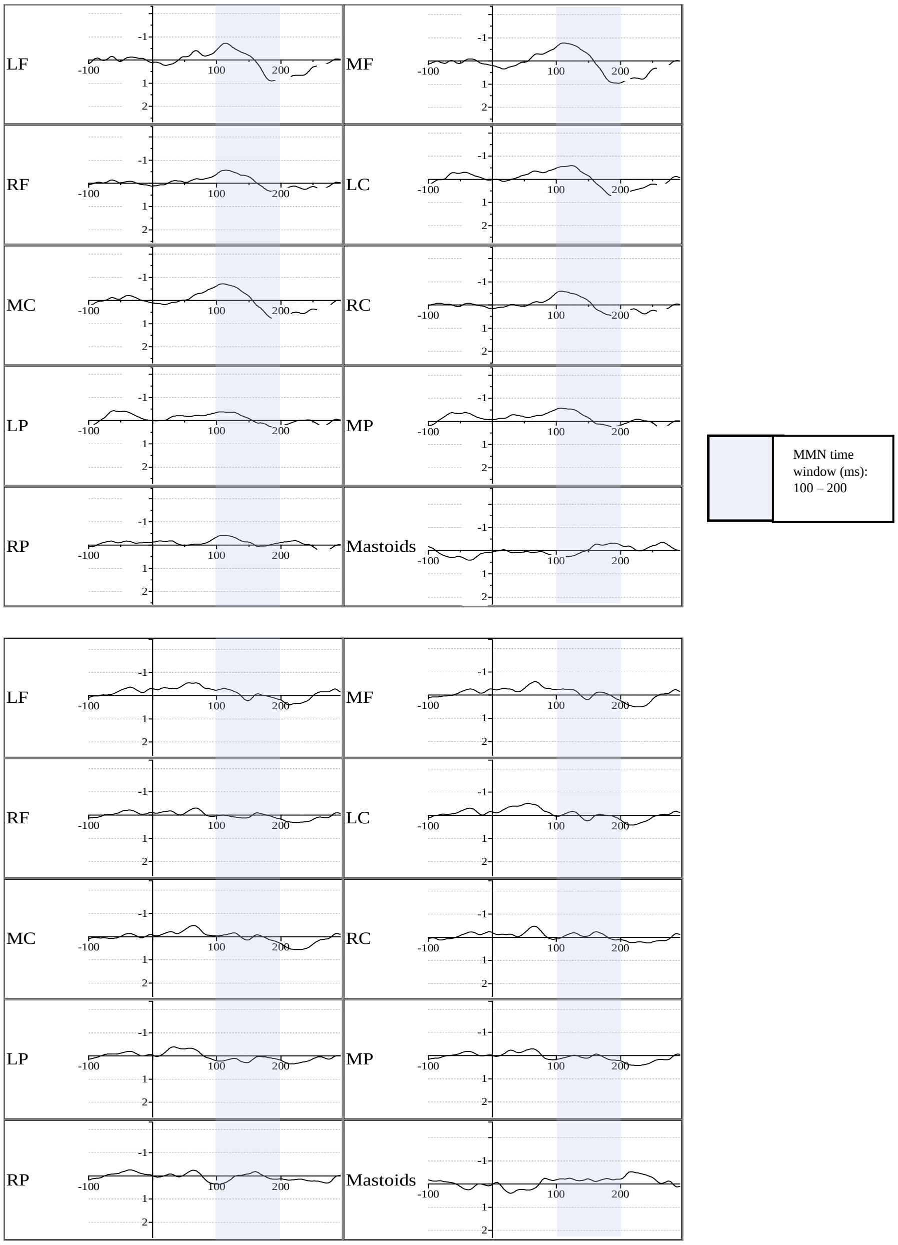

Figure 3. Difference waves created by subtracting the standard from the omission deviant in the Short SOA condition (left) and the Long SOA condition (right). Responses are shown across all nine ROI, starting from frontal electrodes (LF, MF, RF) to posterior, parietal ones (LP, MP, RP). Responses from the average of the mastoids are shown as well. Waveforms presented have been filtered at 1-25 Hz, with negative amplitude presented upwards by convention.

Topographical distribution

Responses to omissions for both SOAs show a negativity in the front-central regions, 100 ms – 135 ms post-stimulus onset, and then a subsequent positivity (Figure 4). For the short SOA, this positivity begins around 175 ms and continues until the end of the defined window (200 ms). For the longer SOA, the positivity is apparent in the same window but, like the negativity, its amplitude is relatively smaller. The location of the positivity for the long SOA is also posterior to that seen for the short SOA.

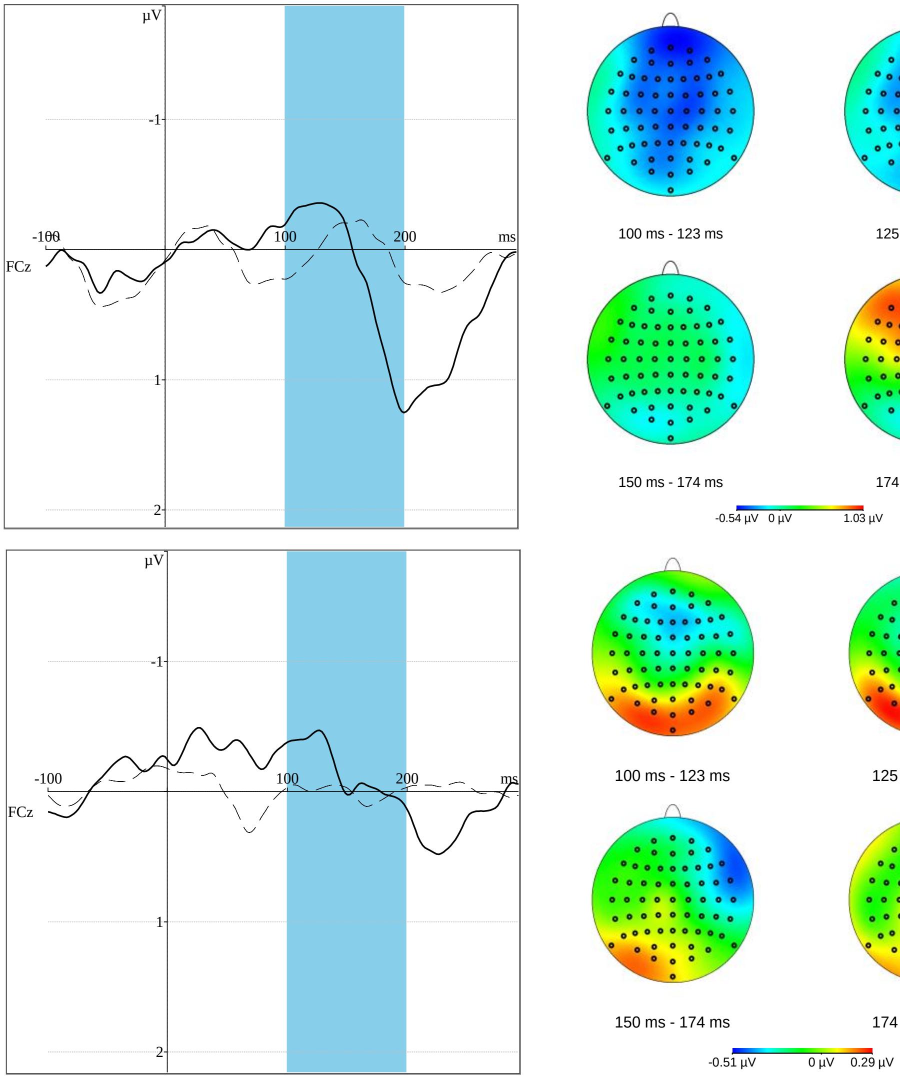

Figure 4. Scalp distribution maps for SOA = 125 ms (top row) and SOA = 250 ms (bottom row) at the FCz electrode are illustrated here. The figure shows the top-view representation of a head, with equidistant electrode placement on the scalp. Solid line shows responses to the omission, dotted shows responses to the standard stimuli. Changes in amplitude across the scalp are shown in four maps for the 100 ms to 250 ms window. The corresponding ERP is shown on the left. The data were re-filtered between 1-25 Hz for visual representation.

Summary of Results

Overall, an MMN of varying amplitude, but significantly different from baseline, was elicited for both the Long SOA and Short SOA. The MMN response appeared to be more robust for the Short SOA condition than the Long SOA condition, although this difference was not significant in the omnibus analysis of MMN amplitude data.

Discussion

Experiment 1 was conducted in order to replicate the study by Yabe et al. (1997) and evaluate how the omission MMN changes with a change in rate of presentation (i.e., SOA). The original experiment found a TWI of about 150 ms within which an MMN is elicited. Any rate of presentation outside that would not elicit an omission MMN. Our results found that both SOAs elicited an MMN, although visually, the response was more robust for the shorter SOA compared to the longer SOA.

The waveforms observed in our experiment for both SOAs are visually similar to those reported by Yabe et al. (1997). The original study found distinct peaks for SOAs below 150 ms, and no peaks or smaller differences between standard and deviants for SOAs 150 ms and above. In our experiment, an SOA below that threshold (125 ms) and above that threshold (250 ms) was used, and visually, a similar pattern was observed. A larger negativity was observed for shorter SOAs compared to longer SOAs. However, unlike the original paper, we found a statistically significant MMN for both SOAs above and below the TWI, suggesting that the TWI might be longer than 170 ms, as initially reported by Yabe et al., (1997; 1998).

Previous work has observed longer TWIs for musicians compared to non-musicians (Rüsseler et al., 2001), as well as in children compared to young adults (Wang et al., 2005). It is possible that participants had a musical background that contributed to the elicitation of MMNs at the higher SOA. However, this information was not collected, making it difficult to draw any conclusions about musicality as a factor contributing to the elicitation of the MMN at the longer SOA. All participants tested were adults, removing the question of age as a factor entirely. To our knowledge, there is one other study only that has demonstrated the elicitation of omission MMNs at an SOA beyond the TWI cited by Yabe et al. (1997; 1998). Prete et al. (2022) found omission MMNs elicited to tones presented at an SOA of 500 ms. The critical difference between this study and previous ones was the methodology used to extract the MMN difference wave. Typically, the difference wave is calculated by subtracting activity elicited to the presented tone from the activity elicited to the missing (omitted) tone. It is this difference wave that has been used as the hallmark of a typical MMN in this experiment, replicating the methodology used by Yabe et al. (1997) and, to our knowledge, other experiments that studied the omission MMN. However, Prete et al. (2022) argue that this comparison is invalid as it compares sound-related neural activity to silence-related neural activity, making it natural that an effect should be found. In their study, the authors compared neural activity for expected silences (that is, longer SOAs between tones) to activity towards unexpected silences (occasionally longer SOAs between tones) and found a reliable negativity elicited to the unexpected silences when compared to the expected silences. This, the authors argue, is evidence for predictive coding in auditory processing, and demonstrates a fallacy in the way omission MMNs are typically calculated—therefore suggesting that omission MMNs are observable at longer SOAs.

In our experiment, responses to the omissions are compared to the presented standards (as is typically done) in order to calculate the MMN difference wave. Critically, previous work has *not* observed reliable MMNs to SOAs above 200 ms, regardless of the methodology used to evaluate the response, meaning that our results are still novel. It is difficult to directly compare our results with that of Prete et al. (2022). Considering the greater control and rigor with which this experiment was performed, however, the MMN observed at an SOA of 250 ms may provide evidence for predictive coding in auditory processing, as Prete et al. (2022) argue. Therefore, the negativity observed to the omission deviant is an ‘error response’ generated to the expectation of a tone where none was presented. Expectation here is driven by the sequence of standard tones presented. For the short SOA (125 ms) condition especially, our results corroborate the MMN visually with a polarity reversal at the mastoids. For the long SOA (250 ms), the response is smaller and harder to decipher. Additionally, the fact that we observed robust MMNs to omission deviants at both SOAs suggests that the typical difference wave methodology is valid (unlike what is argued by Prete et al, 2022). Specifically, if comparing the presence of a sound to its absence was enough to show activity where (arguably) there was none, then the original Yabe et al., (1997) study should have reported significant MMN responses for all omission deviants, regardless of the SOA. This was not the case. Therefore, we suggest that a deviant-minus-standard subtraction is perfectly valid as an MMN calculation. The comparison presented by Prete et al. (2022) contrasts expected activity with unexpected activity; while corroborative of the predictive coding framework, it is difficult to say whether it was the silence that was expected within the sequence, or the delayed tone.

Experiment 2 involves the elicitation of an MMN to temporal deviants that are presented earlier or later than expected. For the late deviant in particular, the delay in presentation can be interpreted as an ‘omission’ deviant (up until the actual presentation of the sound), suggesting a motivation for analysing the time window of the *expected* presentation. However, the aim of Experiment 2 is to study how the evoked responses differ for a sound that was presented late.

Therefore, responses are only analysed at the onset of the sounds, and not between them.

Experiment 2

The second experiment tracked the MMN to unexpected variations in pace in a train of pure tones. Although previous work has looked at temporal variation, the scope has been limited to observing changes in the MMN and making inferences about stimulus perception. In our current experiment, the aim of this temporal variation is to understand how auditory stimuli are processed when unexpected interruptions of a steady rhythm are presented within an otherwise steady rate of stimulus presentation. Here the deviant stimulus is identical to the standard stimulus in all respects except the timing with which it is presented. The MMN in this context is used as a measure of tracking this processing, and we hypothesise that changes in the amplitude of the MMN will give us insight into how well the change in timing of unexpectedly early or unexpectedly late stimuli is perceived.

Methods

In the second experiment, the participants and technical details of the recording and stimuli were the same as in Experiment 1. Data from 22 participants was used for analyses. Parts of the testing protocol that differed from that of Experiment 1 are explained below.

Experimental procedure and Stimuli

An auditory oddball paradigm was used again, with two types of deviants. The first type of deviant differed from the standard in frequency only (1200 Hz), and the second type differed from the standard in the SOA it was presented at. One standard and two deviants were presented per block (see Figure 5). Participants were presented with three blocks; each block presented the same standard and a frequency deviant that differed in pitch. The blocks differed only in the timing deviant that was presented. The first block contained a timing deviant that was earlier than expected (*Early deviant*, SOA = 250 ms), the second block contained a timing deviant that was later than expected (*Late deviant*, SOA = 750 ms), and the last block contained both the early and the late deviant in the same block (*Mixed block*). Presentation order of the Early/Late deviant blocks was counterbalanced across participants, with both being presented first and last alternately. The ratio of standards to deviants was kept fixed across all blocks at 80:20. For the blocks with a single timing deviant, 800 standards were presented with 100 deviants of each type, for a total of 1000 tokens. For the mixed block, 1200 standards were presented with 100 deviants of each type, for a total of 1500 tokens.

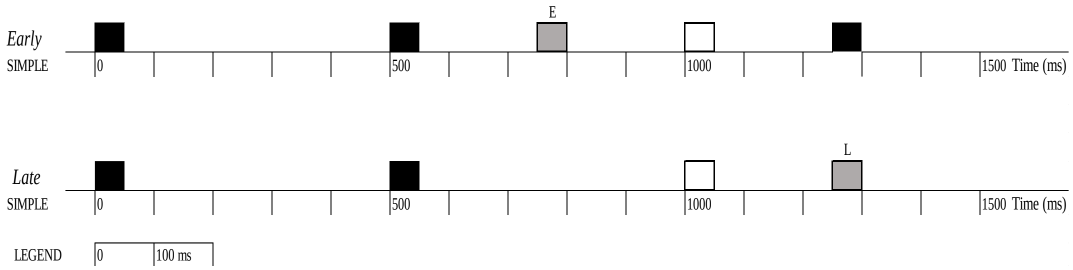

Figure 5. A visual representation of stimuli timing across the Early and Late timing. The left edge of the block represents the onset of the stimulus. All stimuli were 50 ms in length. Standard pure tones are represented in black, whereas deviants (early and late) are represented by grey. As the Early and Late deviants are temporal deviants, only the onset of the stimuli is different (i.e., different SOA). The early deviant is presented 250 ms earlier than expected (SOA = 250 ms), and the late deviant is presented 250 ms later than expected (SOA = 750 ms). The white blocks represent the expected presentation of the tone. For the Early and Late deviants, this is at the standard SOA of 500 ms.

Testing procedure

As Experiment 2 was run during the same session as Experiment 1, the testing procedures (including consent) are the same as those for Experiment 1. The second experiment with the timing deviants was longer, and ran for a total of 30 minutes, making the full session duration with the two experiments to be about 45 minutes.

EEG data analyses

EEG data were pre-processed and cleaned using the same pipeline used for Experiment 1. The deviant grand average for each participant contained at least 80/100 deviants (range: 85-100, S.D. = 3.0). The only difference was in the length of the epochs, which were changed to -100 ms to 500 ms before averaging. Automated peak detection (Barr et al., 1978) was conducted to find the maximum amplitude pertaining to the MMN within a window of 100 ms to 300 ms.

Results

The second experiment with the timing deviants sought to establish a baseline of responses by investigating how response varied to stimuli that were identical in all regards except for the SOA they were presented at. In this experiment, deviants were presented either earlier or later than expected, and we hypothesised that an MMN would be elicited to each timing deviant. A Frequency deviant was also added to each block as a deviant differing in pitch reliably elicits an MMN, with the motivation of providing a reference for what the MMN looks like for each participant. A third block was presented to participants with both timing deviants to evaluate whether the two types of timing deviants would elicit different responses when presented together compared to when they are presented in separate blocks. Therefore, participants heard the timing deviants twice during the experimental session, and the frequency deviants three times (once with each timing deviant, and then once in the combined block).

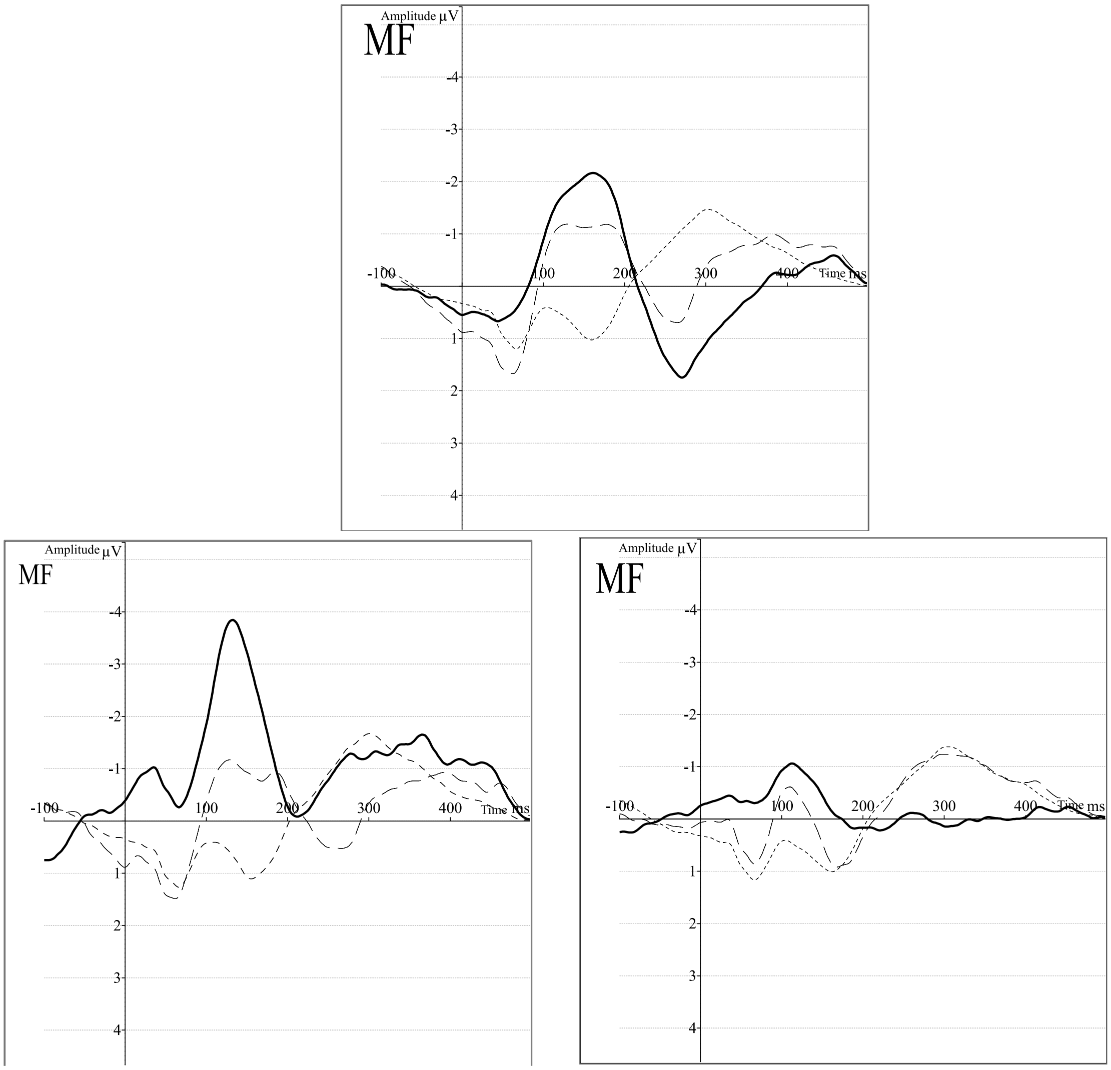

Figure 6. Difference waves created by subtracting the mean response amplitude to the standard from the mean response amplitude to the deviant. The responses are from the Middle-Frontal ROI and have been filtered at 1-25 Hz. The top row shows the responses to the Frequency deviant, and the bottom row shows the response to the Early timing deviant (left) and the Late timing deviant (right). Within the figures, the solid line shows the difference wave; the long dashed line shows the deviant response and the short dashed line shows the standard response.

Pairwise comparisons using *t*-tests were conducted for mean peak amplitude that compared between deviants of the same type across the different blocks to see if the context made any difference. No significant differences *p* > 0.05) were found between deviants of the same type presented in different blocks (see **APPENDIX A, Table A** for detailed comparisons). Therefore, difference wave data was consolidated by averaging timing deviants across two blocks, and the Frequency deviant across three. The analysis and figures in this section are based on these consolidated numbers

Difference waves were calculated for the responses to each deviant as described for Experiment 1. Mean peak amplitude and mean peak latency values for the MMN response were obtained by identifying the negative-most value within a pre-defined time window of 100 ms – 300 ms for the MMN (Näätänen et al., 2019).

All of the deviants (timing deviants and Frequency deviant) presented in each block elicited an MMN. One-sample *t*-tests were conducted to evaluate whether the mean peak amplitude of this difference wave was significantly different from 0 μV (baseline). The result of these comparisons corroborated the visual presence of the MMN, and were all significant ($p < 0.05$) (Table 2). Figure 6 shows the difference waves for each Deviant Type for the ROI MF.

Table 2. One-sample t-test comparing peak mean amplitudes for the three Deviant types against the baseline, 0 μV, for the Middle-Frontal ROI.

One-sample *t*-test comparing peak mean amplitude against baseline (0 µV) by Deviant type (ROI: MF)

Deviant type	*Mean amplitude (μV)*	*t*	*df*	*Conf. low*	*Conf. high*	*p*	*Significance*
Frequency	-2.94 ± 0.89	-10.9	21	-3.50	-2.38	3.87E-10	**
Early	-4.29 ± 2.39	-9.13	21	-5.27	-3.32	9.30E-09	**
Late	-1.83 ± 1.38	-11.2	21	-2.17	-1.49	2.74E-10	**

Note: mean ± standard deviation (s.d.)

- Indicates no significance p > 0.05

* Indicates significance p < 0.05

** Indicates significance p < 0.01

The effect of temporal variation on MMN amplitude

A 3 x 4 repeated measures ANOVA was conducted to evaluate the effect of Deviant Type (Frequency deviant, Early timing deviant, Late timing deviant) and Region of Interest (ROI: LF, MF, RF, MC) on the mean MMN amplitude. An alpha level of 0.05 was used, and a Greenhouse-Geisser correction applied where necessary. Uncorrected values were reported for all except the *p* values. Follow-up pairwise comparisons were conducted for each main effect using *t*-tests, and *p*-values were corrected using the Bonferroni multiple testing correction method where needed.

The main effects of Deviant Type ($F(2, 42) = 19.84$, $p = 0.0000009$, $\eta^2 = 0.295$) and ROI ($F(3, 63) = 65.54$, $p < 0.0001$, $\eta^2 = 0.109$) were both significant. Additionally, a significant interaction effect between Deviant Type and Condition was found for mean MMN amplitude ($F(6, 126) = 10.61$, $p < 0.0001$, $\eta^2 = 0.018$. Follow-up pairwise comparisons between pairs of all deviants for each ROI found that the three deviants differed significantly from each other ($p < 0.0042$ across all comparisons, Bonferroni correction applied) for all ROI except one (see Figure 7). Early timing deviant and Frequency deviant showed no significant differences for the ROIs LF and RF only. No significant effects of MMN latency were found, $p > 0.05$.

The effect of timing variation on standardised MMN amplitudes

Figure 8 shows the representative scalp distribution maps at the FCz electrode, 100 ms to 300 ms post-stimulus onset. Scalp maps of the amplitude distribution across electrodes are shown to the Early timing deviant (top row) and the Late timing deviant (bottom row). For all maps, a distinct negativity is seen from 100 ms to 135 ms.

For the Early deviant, an N100 can be seen at 100 ms post-stimulus onset for both the standard and deviant. An adjoining negativity, the MMN, follows immediately, continuing up until about 175 ms. This N100-MMN complex is also known as the Deviant-Related Negativity, or DRN. In the current experiment, any negativity after ~ 100 ms will be referred to as the MMN. A positivity in the range of 215 ms to 250 ms is also observed for the Early timing deviant. The topographical map (right) shows a distinct negativity for the fronto-central electrodes between 100 ms and 150 ms. A fronto-central positivity, the P3a, is observed between 200 and 250 ms.

For the Late deviant, a negativity similar in morphology to that elicited by the Early deviant but smaller in size is observed at 100 ms. This negativity is also followed by a positivity in the range of 135 ms to 214 ms approximately, possibly the P200 response that often occurs in tandem with the N100. The topography shows a strong fronto-central positivity in the time range of 150 ms to 199 ms, and a fronto-central negativity at 250 ms to 299 ms, although it is not the MMN and is elicited to both the standard and the deviant equally.

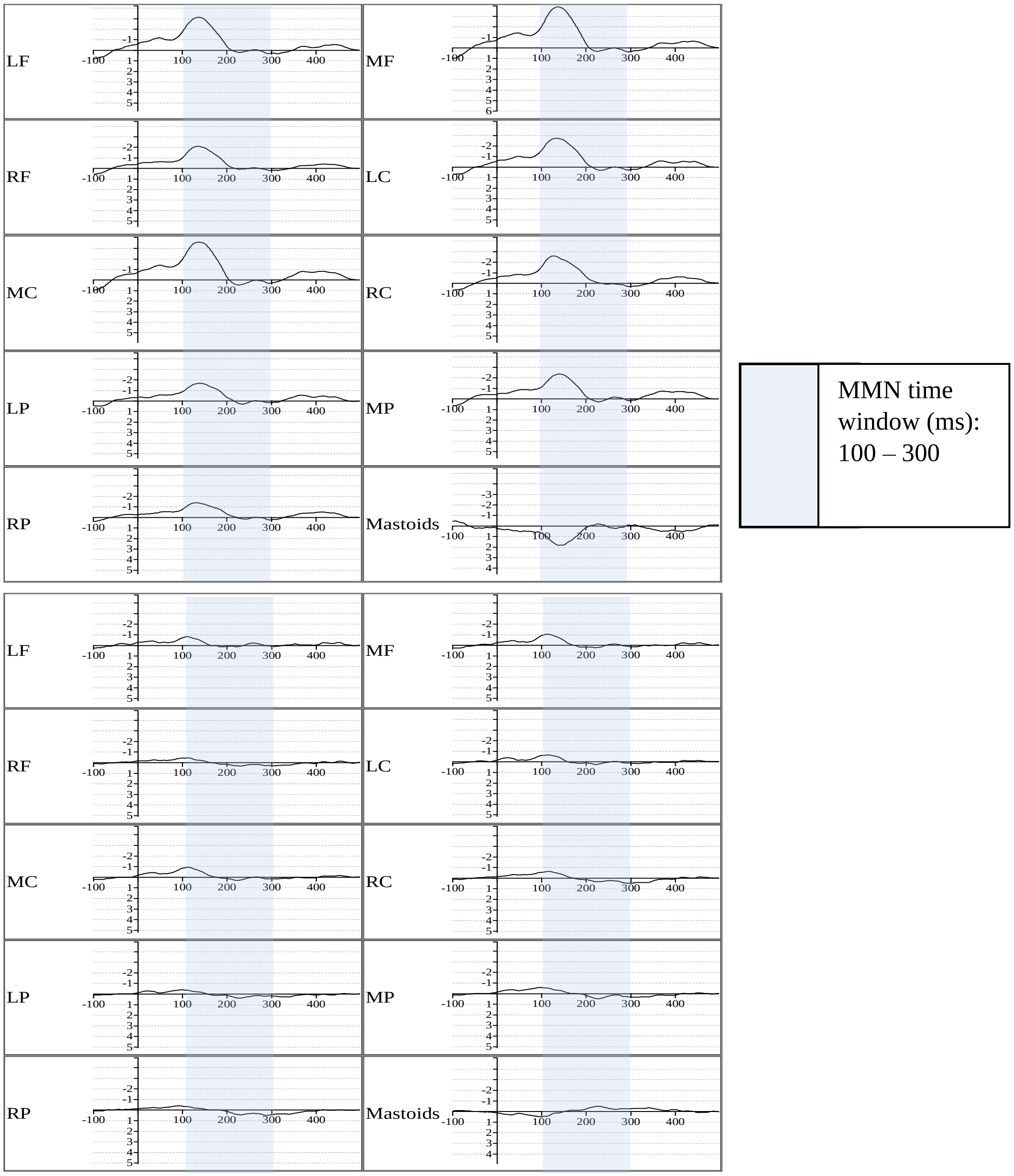

Figure 7. Difference waves created by subtracting the standard from the Early timing deviant (left) and Late timing deviant (right). Responses are shown across all nine ROI, starting from frontal electrodes (LF, MF, RF) to posterior, parietal ones (LP, MP, RP). Responses from the average of the mastoids are shown as well. Waveforms presented have been filtered at 1-25 Hz, with negative amplitude presented upwards by convention.

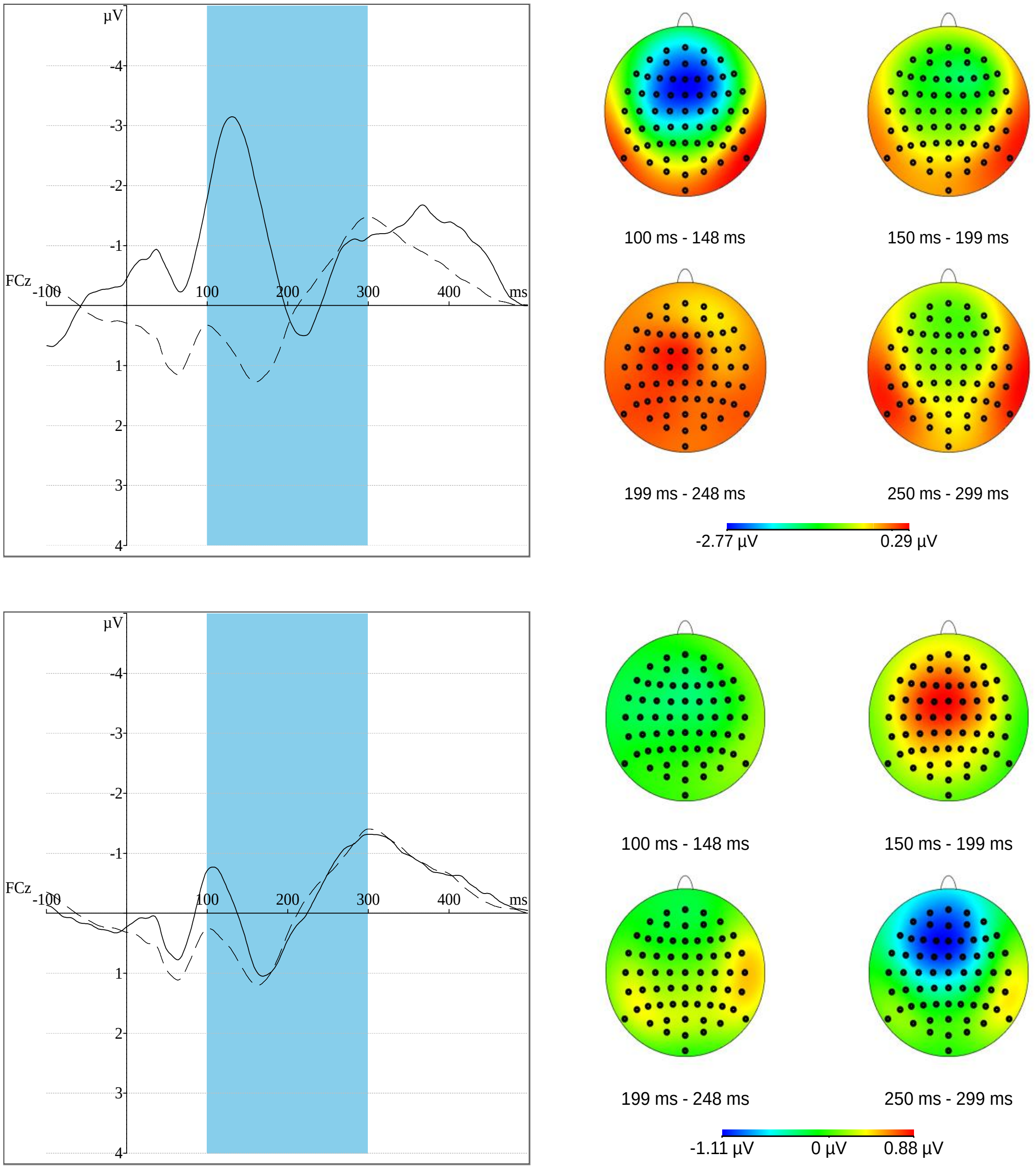

Figure 8. Scalp distribution maps for the Early timing deviant and Late timing deviant are presented at the FCz electrode. The figure shows the top-view representation of a head, with equidistant electrode placement on the scalp. The solid line shows responses to the deviant, dashed line shows responses to the standard stimuli. Changes in amplitude across the scalp are shown in four maps for the 100 ms to 300 ms window. The corresponding ERP is shown on the left. The data were re-filtered between 1-25 Hz for visual representation.

Summary of results

Overall, an MMN of varying amplitude, but significantly different from baseline, was elicited for both the Frequency deviant and the Early and Late timing deviants. The largest MMN response was elicited to the Early timing deviant, and the smallest to the Late timing deviant, suggesting an ease of processing for the former.

Discussion

The aim of Experiment 2 was to explore how unexpected temporal variation in isochronous sequences is auditorily processed. Participants were presented with sequences of simple tones at the same rate except for when a timing deviant was presented unexpectedly early or unexpectedly late. Each block also included a Frequency (pitch) deviant. Results showed that all deviants elicited an MMN, but the MMN to the Early timing deviant was the largest in size, while that to the Late timing deviant was smallest.

The amplitudes of the MMNs elicited to each deviant varied, with the early deviant eliciting the biggest MMN, and the late deviant the smallest. Previous work exploring MMN responses to changes in inter-onset interval found a similar result, with a bigger and earlier MMN to shorter intervals than longer ones (Lai et al., 2011). This was suggested to reflect the difference between the deviant and the standard, with an increased difference leading to a proportional increase in amplitude (Kisley et al., 2004; Sams et al., 1985; Tervaniemi, 1999). In our experiment, however, the Early and Late timing deviants did not elicit responses that were similar to each other in size, as would be predicted. Rather, the Early timing deviant evoked a much larger response, suggesting that it was easier to auditorily process compared to the Late timing deviant. The differences between the Early timing deviant and the standards may have been more easily encoded pre-attentively, making it much more salient and discriminable compared to the Frequency deviant and the Late timing deviant. The difference in size for the response to the Early timing deviant compared to the Late also suggests the co-incidence of the N1 response with the former but not the latter, leading to a much more robust deviant-related negativity (DRN) observed to one timing deviant type (Early) over the other (Late) (Duda-Milloy et al., 2019; Tavakoli et al., 2021).

In terms of TWI, all stimuli were presented at an SOA that was greater than the TWI found for simple tones in previous literature (Yabe et al., 1997; 1998). In the context of the Early and the Late timing deviants, both were presented at an SOA that was 250 ms from the standard SOA. The Early timing deviant, therefore, was presented no earlier than 250 ms after the preceding standard. In Experiment 1, this SOA was found to still elicit MMNs. However, in the case of the Early timing deviant, this would have no impact on the results for two reasons: one, all standards were presented at an SOA greater than TWI, so there would be no integration of successive stimuli; and two, the unexpectedly early token means there is no point during the stimuli stream when there is no tone where there should be one. Therefore, no omission MMNs would be generated in either case. The TWI is relevant for trace formation in the niche context of the generation of the omission deviant, which requires a steady stimulus presentation at a rate on or below the TWI.

General Discussion

The purpose of the current experiment was to replicate previous findings on the elicitation of an MMN to non-regular timing of tones, and to create a baseline with which to compare future work on language variation. We also examined the elicitation of the MMN to an omission deviant using modern equipment, a greater number of electrodes, and conservative parameters of analyses and data rejection to create a more tightly controlled environment.

The first experiment with the omission deviant replicated the study by Yabe et al. (1997), corroborating a temporal window of integration (TWI) for SOAs that are shorter than 150 ms, but is exceeded for SOAs longer than that threshold. Results did not support our hypothesis; an MMN to omissions presented in a stimulus train was elicited regardless of whether the SOA was above or below 150 ms. The current study found these results in an experiment with additional rigor. A portion of the data was re-sampled offline at both 200 Hz (original parameter) and 512 Hz, and visually inspected. No difference was observed, and all data thereafter was re-sampled at 512 Hz (lab standard). Therefore, the MMNs observed to both SOA are reliable, and suggest that the TWI may be wider than originally thought.

The second experiment with the timing deviants was designed in order to look at how neural responses change to stimuli presented earlier or later than expected that are identical in all other regards. We hypothesised that the early and late timing deviants would elicit responses that were larger in amplitude relative to the regular, standard time of presentation. Our results supported this hypothesis, and MMNs were elicited to both timing deviants. The size of the responses varied between the two, with the early timing deviant eliciting a much larger MMN compared to the late deviant. The difference in response suggests that the early and the late timing deviant do not elicit the same type of response. The early timing deviant elicits a MMN that occurs alongside an N1 response, leading to the robust negativity observed (Tavakoli et al., 2021). This is not the case for the late timing deviant. While a significant negativity is elicited in the pre-defined MMN time window when the Late timing deviant is presented, this negativity is visually small and could pertain to just N1 activity.

N1-P2 complex. Another aspect of interest is the change in the N1-P2 complex. This complex is a coupled pre-attentive response involving a negativity at ~100 ms, and a positivity at ~200 ms post-stimulus onset, evoked in response to the presentation of any auditory stimuli. Ford and Hillyard (1981) found the N1-P2 complex to be larger in response to unexpectedly early stimuli compared to standard stimuli. This is similar to the results observed in the current experiment. It was seen for the late deviant, too, although the increase in amplitude was relatively smaller. Interestingly, although Ford & Hillyard (1981) suggest it as a possibility, it is worth noting that their results do not show any prolonged negativity that might be the MMN, whereas our current results show a clear, prolonged negativity alongside the N1-P2 complex, suggesting an MMN (or DRN) response.

Pre-attentive auditory processing of temporal variation

Perceptual grouping

Presenting both timing deviants within one block was an exploratory condition included in the experiment in order to better recreate temporal variation in the auditory domain. By having both shorter *and* longer intervals in the same block, a better model of temporal variation could be replicated within the paradigm. However, the results suggested that there was no difference in expectation or pattern perception depending on whether these short and long deviants were blocked separately (deviants presented individually) or not (deviants presented together). The MMN responses elicited to the timing deviants presented in the same block did not statistically differ to the ones elicited to the timing deviants presented independently (in separate blocks). No perceptual grouping effect was observed. Given the large SOA for the late deviant, it was possible that the stimuli would be grouped together from one late deviant to the other, such that participants were perceiving patterns of tones where none existed. However, presenting the stimuli pseudo-randomly helped circumvent these potential grouping effects, as evidenced by the similar results between blocks.

A predictive coding account of temporal variation processing

Based on previous work, we hypothesised that the timing deviants would both elicit similar responses that would differ in latency alone. This was not what we observed in our results; the early deviant did elicit a large amplitude MMN as predicted. However, the response did not differ in latency from a typical MMN, and neither did the MMN elicited to the late deviant. The MMN to the late deviant was also considerably smaller than that elicited to the early deviant. Potentially, this difference in timing deviant amplitudes could be due to a perceived lack of reliability of the late deviant.

The stimulus-specific adaptation (SSA) hypothesis based on the idea of the MMN resulting from the refractoriness of specific neural populations involved in representing the standard stimulus is not able to fully account for this data. The timing variation in stimulus presentation led to perceptually empty spots in the stream of tones due to the late timing deviant, but not so for the early timing deviant. The standards and the deviants in this experiment were identical except for the time of presentation. Under the SSA hypothesis, therefore, it could be predicted that either the temporal variation would not elicit a dishabituated MMN response, or, that it would, but the response would vary for the timing deviant type. In the first case, if the SSA does not code temporal features, our results provide evidence against this hypothesis as an explanation for our results. However, May & Tiitinen (2010) argue that complex temporal patterns can be represented using this framework. In that case, we would expect late timing deviants to be more likely to trigger new neural group activity compared to early timing deviants, because the latter still fall within the SOA of the standards and might be responded to by the same neurons. In this case, we would observe a larger MMN to the late timing deviants than to the early; our results do not support this, either, as a larger MMN was observed to the early timing deviant compared to the late. Arguably, it is also possible that both the timing deviants are treated as novel stimuli. However, in that case, similar responses would be expected to both. However, the early timing deviant elicits a response several magnitudes larger than that of the late timing deviant. Therefore, the SSA hypothesis fails to adequately explain our results, and we turn to the predictive coding framework.

An integral aspect of the predictive coding framework is its ability to make predictions about the stimuli; it is the dissonance between the prediction and the sensory input that results in the elicitation of an MMN (or lack thereof) (e.g., Friston, 2005; 2009; Garrido et al., 2009; see also *Chapter 1: Introduction*). In the context of our results, therefore, given that the late deviant elicited a smaller MMN in amplitude compared to the early deviant, we can conclude that the late deviant elicited a smaller prediction error compared to the early deviant. In other words, the late deviant was better predicted than the early deviant. There are several possible explanations for this, the first being that perhaps late temporal variations are just more expected in general in the context of the auditory domain, compared to early temporal variations. Assuming that the predictive coding framework also takes into account an individual's personal sensory experience across a lifetime, humans may be more used to hearing pauses and delays in auditory inputs (for example, hesitations in the context of speech, or processing delays when audio is buffering) compared to hearing an auditory input that 'jumps the queue', so to speak, and arrives earlier than expected. In that context, the early deviant would then result in a larger prediction error than the late deviant, as the 'delays' would be more expected (and therefore predicted) than the opposite. A better estimation of this will be explored in *Chapter 3*, where temporal variation in the context of increasing speechlikeness (stimuli with increasing linguistic complexity) will be examined.

Alternatively, pre-attentive processing of the timing deviants may be dependent on the SOA of the stimuli. Although MMNs have been elicited with stimuli using standard ISIs up to 2000 ms (Mäntysalo, & Näätänen, 1987), it is possible that the SOA of the late deviant (750 ms) fell outside the processing window for some participants, therefore resulting in a smaller MMN amplitude. In this context, the late deviant could function as a marker for individual differences in future work.

Limitations

Although effort was made to control variables in stimulus construction, experiment presentation, and participant recruitment, several allowances were made that could be better controlled in future. For example, a more detailed language questionnaire and musical background questionnaire could have been administered. This would have allowed us to correlate the responses observed with participant history and make inferences about the influence of language and musical background on the observed MMN. In turn, we could then have better analysed individual differences between the participants.

For future experiments, better control of the participant demographics would help in recording better quality data. The above-mentioned uncontrolled variables may have added variation to our data.

Conclusion

The purpose of the current study was to increase our understanding of responses elicited to temporal variation in the delivery of simple tones as an auditory input. The biggest differences in MMN morphology were to stimuli presented earlier than expected, suggesting a potential role for the early deviant as a marker to model individual differences in perceiving temporal patterns in the auditory domain. Next, in order to gain a more holistic overview and understanding of how temporal variation in language is processed, the responses elicited to these non-linguistic auditory stimuli will be compared to stimuli of varying linguistic and auditory complexity. We expect to observe differences in how the timing deviants are perceived and processed when increasing linguistic information is added to the cognitive load and hope to understand better how temporal variation in language is therefore processed.

References

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, *24*(3), 197–207. https://doi.org/10.1016/0301-0511(87)90002-0

Alain, C., Cortese, F., & Picton, T. W. (1999). Event-related brain activity associated with auditory pattern processing. *NeuroReport*, *10*(11), 2429–2434. https://doi.org/10.1097/00001756-199908020-00038

Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2004). Aging: a switch from automatic to controlled processing of sounds? *Psychology and aging, 19*(1), 125.

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018). Cortical Tracking of Global and Local Variations of Speech Rhythm during Connected Natural Speech Perception. *Journal of Cognitive Neuroscience, 30*(11), 1704-1719.

Asano, S., Shiga, T., Itagaki, S., & Yabe, H. (2015). Temporal integration of segmented-speech sounds probed with mismatch negativity. *NeuroReport*, *26*(17), 1061–1064. https://doi.org/10.1097/WNR.0000000000000468

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, *4*(2), 1–10. https://doi.org/10.1126/sciadv.aao3842

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: a predictive coding perspective. *Journal of Psychophysiology, 21*(3-4), 204-213.

Barr, R. E., Ackmann, J. J., & Sonnenfeld, J. (1978). Peak-detection algorithm for EEG analysis. *International Journal of Bio-Medical Computing, 9*(6), 465-476.

Bendixen, A., Scharinger, M., Strauß, A., & Obleser, J. (2014). Prediction in the service of comprehension: Modulated early brain responses to omitted speech segments. *Cortex, 53*, 9–26. https://doi.org/10.1016/j.cortex.2014.01.001

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience, 29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Bertinetto, P. M., & Bertini, C. (2010). Towards a unified predictive model of Natural Language Rhythm. Prosodic Universals. Comparative studies in rhythmic modeling and rhythm typology. Rome: Aracne, 43-78.

Bertoli, S., Heimberg, S., Smurzynski, J., & Probst, R. (2001). Mismatch negativity and psychoacoustic measures of gap detection in normally hearing subjects. *Psychophysiology, 38*(2), 334–342. https://doi.org/10.1111/1469-8986.3820334

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2010). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech*, *54*(1), 123–140. https://doi.org/10.1177/0023830910388018

Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PLoS ONE*, *6*(7), 21458. https://doi.org/10.1371/journal.pone.0021458

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program].: Vol. Version 6.

Bosker, H. R., & Kösem, A. (2017). An entrained rhythm's frequency, not phase, influences temporal sampling of speech. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2017-August, 2416–2420. https://doi.org/10.21437/Interspeech.2017-73

Bouwer, F. L., Honing, H., & Slagter, H. A. (2019). Beat-based and memory-based temporal expectations in rhythm: Similar perceptual effects, different underlying mechanisms. *BioRxiv*, 613398. https://doi.org/10.1101/613398

Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: An ERP study. *PLoS ONE*, *9*(5), e97467. https://doi.org/10.1371/journal.pone.0097467

Brückmann, M., & Garcia, M. V. (2020). Mismatch negativity elicited by verbal and nonverbal stimuli: Comparison with potential N1. *International Archives of Otorhinolaryngology*, *24*(2), E80–E85. https://doi.org/10.1055/s-0039-1696701

Cohen, H., Douaire, J., & Elsabbagh, M. (2001). The role of prosody in discourse processing. *Brain and Cognition*, *46*(1–2), 73–82.

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical neurophysiology*, *110*(1), 24-30.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP*). Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 909.

Curtin, S. (2010). Young infants encode lexical stress in newly encountered words. *Journal of Experimental Child Psychology*, *105*(4), 376–385. https://doi.org/10.1016/j.jecp.2009.12.004

Dahan, D. (2015). Prosody and language comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(5), 441–452. https://doi.org/10.1002/wcs.1355

Dellwo, V. (2008). Influences of language typical speech rate on the perception of speech rhythm. *The Journal of the Acoustical Society of America*, *123*(5), 3427–3427. https://doi.org/10.1121/1.2934192

Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: challenges and unresolved questions. *European Journal of Neuroscience, 51*(5), 1151–1160. https://doi.org/10.1111/ejn.13802

Desjardins, R. N., Trainor, L. J., Hevenor, S. J., & Polak, C. P. (1999). Using mismatch negativity to measure auditory temporal resolution thresholds. *NeuroReport, 10*(10), 2079-2082.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. Journal of Memory and Language, 59(3), 294–311. https://doi.org/10.1016/j.jml.2008.06.006

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Donchin, E. (1981). Surprise!...Surprise? *Psychophysiology, 18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Duda-Milloy, V., Tavakoli, P., Campbell, K., Benoit, D. L., & Koravand, A. (2019). A time-efficient multi-deviant paradigm to determine the effects of gap duration on the mismatch negativity. *Hearing Research, 377*, 34–43. https://doi.org/10.1016/j.heares.2019.03.004

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*(5), 590-604.

Fisher, D. J., Grant, B., Smith, D. M., & Knott, V. J. (2011). Effects of deviant probability on the 'optimal' multi-feature mismatch negativity (MMN) paradigm. *International Journal of Psychophysiology*, *79*(2), 311–315. https://doi.org/10.1016/J.IJPSYCHO.2010.11.006

Fong, C. Y., Law, W. H. C., Uka, T., & Koike, S. (2020). Auditory mismatch negativity under predictive coding framework and its role in psychotic disorders. *Frontiers in Psychiatry*, *11*, 1–14. https://doi.org/10.3389/fpsyt.2020.557932

Ford, J. M., & Hillyard, S. A. (1981). Event-Related Potentials (ERPs) to Interruptions of a Steady Rhythm. *Psychophysiology*, *18*(3), 322–330. https://doi.org/10.1111/j.1469-8986.1981.tb03043.x

Ford, J. M., Roth, W. T., & Kopell, B. S. (1976). Auditory evoked potentials to unpredictable shifts in pitch. *Psychophysiology, 13*(1), 32–39. https://doi.org/10.1111/j.1469-8986.1976.tb03333.x

Frisina, D. R., Frisina Jr, R. D., Snell, K. B., Burkard, R., Walton, J. P., & Ison, J. R. (2001). Auditory temporal processing during aging. In *Functional neurobiology of aging* (pp. 565-579). Academic Press.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology*, *120*(3), 453-463.

Gervain. (2018). Gateway to language: The perception of prosody at birth. In *Boundaries crossed, at the interfaces of morphosyntax, phonology, pragmatics and semantics* (pp. 373–384). Springer.

Goswami, U. (2018). A neural basis for phonological awareness? an oscillatory temporal-sampling perspective. *Current Directions in Psychological Science*, *27*(1), 56–63. https://doi.org/10.1177/0963721417727520

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. https://doi.org/10.1016/j.jml.2013.03.001

Granier-Deferre, C., Ribeiro, A., Jacquet, A.-Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Hari, R., Joutsiniemi, S. L., Hämäläinen, M., & Vilkman, V. (1989). Neuromagnetic responses of human auditory cortex to interruptions in a steady rhythm. *Neuroscience Letters*, *99*(1–2), 164–168. https://doi.org/10.1016/0304-3940(89)90283-8

Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology*, *87*(2), 165–172. https://doi.org/10.1016/J.IJPSYCHO.2012.12.005

Honbolygó, F., Kóbor, A., & Csépe, V. (2019). Cognitive components of foreign word stress processing difficulty in speakers of a native language with non-contrastive stress. *International Journal of Bilingualism*, *23*(2), 366–380. https://doi.org/10.1177/1367006917728393

Honbolygó, F., Kolozsvári, O., & Csépe, V. (2017). Processing of word stress related acoustic information: A multi-feature MMN study. *International Journal of Psychophysiology*, *118*, 9–17. https://doi.org/10.1016/J.IJPSYCHO.2017.05.009

Horváth, J., Czigler, I., Winkler, I., & Teder-Sälejärvi, W. A. (2007). The temporal window of integration in elderly and young adults. *Neurobiology of aging*, *28*(6), 964-975.

Horváth, J., Müller, D., Weise, A., & Schröger, E. (2010). Omission mismatch negativity builds up late. *NeuroReport*, *21*(7), 537–541. https://doi.org/10.1097/WNR.0b013e3283398094

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20*(11), 1940–1951.

Jongsma, M. L. A., Meeuwissen, E., Vos, P. G., & Maes, R. (2007). Rhythm perception: Speeding up or slowing down affects different subcomponents of the ERP P3 complex. *Biological Psychology, 75*(3), 219–228. https://doi.org/10.1016/J.BIOPSYCHO.2007.02.003

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. Child Development, 64(3), 675–687.

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr

Kisley, M. A., Davalos, D. B., Layton, H. S., Pratt, D., Ellis, J. K., & Seger, C. A. (2004). Small changes in temporal deviance modulate mismatch negativity amplitude in humans. *Neuroscience Letters, 358*(3), 197–200. https://doi.org/10.1016/j.neulet.2004.01.042

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The evolution of rhythm processing. *Trends in Cognitive Sciences, 22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kraus, N., & White-Schwoch, T. (2015). Unraveling the biology of auditory learning: A cognitive-sensorimotor-reward framework. *Trends in Cognitive Sciences, 19*(11), 642–654. https://doi.org/10.1016/j.tics.2015.08.017.Unraveling

Kujala, T., & Leminen, M. (2017). Low-level neural auditory discrimination dysfunctions in specific language impairment—A review on mismatch negativity findings. *Developmental Cognitive Neuroscience, 28*, 65–75. https://doi.org/10.1016/j.dcn.2017.10.005

Lai, Y., Tian, Y., & Yao, D. (2011). MMN evidence for asymmetry in detection of IOI shortening and lengthening at behavioral indifference tempo. *Brain Research, 1367*, 170–180. https://doi.org/10.1016/J.BRAINRES.2010.10.036

Lallier, M., Lizarazu, M., Molinaro, N., Bourguignon, M., Ríos-López, P., & Carreiras, M. (2018). From auditory rhythm processing to grapheme-to-phoneme conversion: How neural oscillations can shed light on developmental dyslexia. In Reading and dyslexia (pp. 147–163). Springer, Cham. https://doi.org/10.1007/978-3-319-90805-2_8

Lallier, M., Molinaro, N., Lizarazu, M., Bourguignon, M., & Carreiras, M. (2017). Amodal atypical neural oscillatory activity in dyslexia: A cross-linguistic perspective. *Clinical Psychological Science, 5*(2), 379–401. https://doi.org/10.1177/2167702616670119

Lallier, M., Thierry, G., & Tainturier, M. J. (2013). On the importance of considering individual profiles when investigating the role of auditory sequential deficits in developmental dyslexia. *Cognition*, *126*(1), 121–127. https://doi.org/10.1016/j.cognition.2012.09.008

Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews, 81*, 158-166.

Law, J. M., Vandermosten, M., Ghesquiere, P., & Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00482

Light, G. A., Swerdlow, N. R., & Braff, D. L. (2007). Preattentive sensory processing as indexed by the MMN and P3a brain responses is associated with cognitive and psychosocial functioning in healthy adults. *Journal of Cognitive Neuroscience*, *19*(10), 1624–1632.

Lightfoot G. (2016). Summary of the N1-P2 cortical auditory evoked potential to estimate the auditory threshold in adults. *Seminars in hearing*, *37*(1), 1–8. https://doi.org/10.1055/s-0035-1570334

Mäntysalo, S., & Näätänen, R. (1987). The duration of a neuronal trace of an auditory stimulus as indicated by event-related potentials. *Biological Psychology*, *24*, 183–195. https://doi.org/10.1016/0301-0511(87)90001-9

May, P. J., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66-122.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*(2), 143–178. https://doi.org/10.1016/0010-0277(88)90035-2

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. https://doi.org/10.1002/hbm.23206

Näätänen, R. (1988). Implications of ERP data for psychological theories of attention. *Biological Psychology*, *26*(1–3), 117–163. https://doi.org/10.1016/0301-0511(88)90017-8

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*(2), 201-233.

Näätänen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear and Hearing*, *16*(1), 6–18. http://www.ncbi.nlm.nih.gov/pubmed/7774770

Näätänen. (2000). Mismatch negativity (MMN): Perspectives for application. *International Journal of Psychophysiology*, *37*(1), 3–10. https://doi.org/10.1016/S0167-8760(00)00091-X

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *3*8(1), 1–21. https://doi.org/10.1111/1469-8986.3810001

Näätänen, R., & Escera, C. (2000). Mismatch negativity: clinical and other applications. *Audiology and Neurotology*, *5*(3-4), 105-110.

Näätänen, R., Gaillard, A. W. K. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329. https://doi.org/10.1016/0001-6918(78)90006-9

Näätänen, R., Kujala, T., & Light, G. (2019). Mismatch negativity: a window to the brain. Oxford University Press.

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. *Psychophysiology*, *48*(1), 4–22. https://doi.org/10.1111/j.1469-8986.2010.01114.x

Näätänen, R., Jiang, D., Lavikainen, J., Reinikainen, K., & Paavilainen, P. (1993). Event-related potentials reveal a memory trace for temporal features. *NeuroReport*, *5*(3), 310–312.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1987). Inter-stimulus interval and the mismatch negativity. In *Evoked potentials III* (pp. 392-397). Butterworths.

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neuroscience Letters*, *107*(1–3), 347–352. https://doi.org/10.1016/0304-3940(89)90844-6

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. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826–859. https://doi.org/10.1037/0033-2909.125.6.826

Nordby, H., Hammerborg, D., Roth, W. T., & Hugdahl, K. (1994). ERPs for infrequent omissions and inclusions of stimulus elements. *Psychophysiology*, *31*(6), 544–552. https://doi.org/10.1111/j.1469-8986.1994.tb02347.x

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988a). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology*, *25*(3), 262–268. https://doi.org/10.1111/j.1469-8986.1988.tb01239.x

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988b). Event-related potentials to time-deviant and pitch-deviant tones. *Psychophysiology*, *25*(3), 249–261. https://doi.org/10.1111/j.1469-8986.1988.tb01238.x

Novak, G., Ritter, W., & Vaughan, H. G. (1992). Mismatch detection and the latency of temporal judgments. *Psychophysiology, 29*(4), 398–411. https://doi.org/10.1111/j.1469-8986.1992.tb01713.x

Oceák, A., Winkler, I., & Sussman, E. (2008). Units of sound representation and temporal integration: A mismatch negativity study. *Neuroscience Letters, 436*(1), 85–89. https://doi.org/10.1016/j.neulet.2008.02.066

Ohmae, S., & Tanaka, M. (2016). Two different mechanisms for the detection of stimulus omission. *Scientific Reports, 6*(1), 1–9. https://doi.org/10.1038/srep20615

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology, 88*, 109-123. https://doi.org/10.1016/j.ijpsycho.2013.03.015

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1991). Right hemisphere dominance of different mismatch negativities. *Electroencephalography and Clinical Neurophysiology, 78*(6), 466–479. https://doi.org/10.1016/0013-4694(91)90064-B

Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology, 82*(3), 219-226.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience, 21*(6), 322-334.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J. (2012). Neuropsychology of P300. The Oxford Handbook of Event-Related Potential Components, 159–188. https://doi.org/10.1093/oxfordhb/9780195374148.013.0089

Polich, J. (2020). 50+ years of P300: Where are we now? *Psychophysiology, 57*(7). https://doi.org/10.1111/psyp.13616

Prete, D. A., Heikoop, D., McGillivray, J. E., Reilly, J. P., & Trainor, L. J. (2022). The sound of silence: Predictive error responses to unexpected sound omission in adults. *European Journal of Neuroscience, 55*(8), 1972-1985.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America, 105*(1), 512–521. https://doi.org/10.1121/1.424522

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Remez, R., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949. https://doi.org/10.1126/science.7233191

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London. Series B, *Biological Sciences*, *336*(1278), 367–373. https://doi.org/10.1098/rstb.1992.0070

RStudio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., & Münte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neuroscience Letters*, *308*(1), 33–36. https://doi.org/10.1016/S0304-3940(01)01977-2

Sable, J. J., Gratton, G., & Fabiani, M. (2003). Sound presentation rate is represented logarithmically in human cortex. *European Journal of Neuroscience*, *17*, 2492–2496. https://doi.org/10.1046/j.1460-9568.2003.02690.x

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): Emitted MMN to violations of an auditory gestalt. *Psychophysiology*, *49*(4), 544–548. https://doi.org/10.1111/j.1469-8986.2011.01336.x

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *62*(6), 437-448.

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience*, *33*(20), 8633-8639.

Scharinger, M., Steinberg, J., & Tavano, A. (2017). Integrating speech in time depends on temporal expectancies and attention. *Cortex*, *93*, 28–40. https://doi.org/10.1016/j.cortex.2017.05.001

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303-304.

Shiga, T., Yabe, H., Yu, L., Nozaki, M., Itagaki, S., Lan, T. H., & Niwa, S. I. (2011). Temporal integration of deviant sound in automatic detection reflected by mismatch negativity. *NeuroReport*, *22*(7), 337-341.

Shinozaki, N., Yabe, H., Sato, Y., Hiruma, T., Sutoh, T., Matsuoka, T., & Kaneko, S. (2003). Spectrotemporal window of integration of auditory information in the human brain. *Cognitive Brain Research*, *17*(3), 563–571. https://doi.org/10.1016/S0926-6410(03)00170-8

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of Speech and of Complex Nonspeech Sounds of Different Temporal Structure in the Left and Right Cerebral Hemispheres. *NeuroImage*, *12*(6), 657–663. https://doi.org/10.1006/NIMG.2000.0646

Spratling, M. W. (2008a). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience, 2*(OCT), 1–8. https://doi.org/10.3389/neuro.10.004.2008

Spratling, M. W. (2008b). Predictive coding as a model of biased competition in visual attention. *Vision Research, 48*(12), 1391–1408. https://doi.org/10.1016/j.visres.2008.03.009

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition, 112*, 92–97. https://doi.org/10.1016/j.bandc.2015.11.003

Spratling, M. W., De Meyer, K., & Kompass, R. (2009). Unsupervised learning of overlapping image components using divisive input modulation. *Computational intelligence and neuroscience*, 2009.

Tavakoli, P., Duda, V., Boafo, A., & Campbell, K. (2021). The effects of sleep on objective measures of gap detection using a time-efficient multi-deviant paradigm. *Brain and Cognition, 152*, 105772.

Tervaniemi, M. (1999). Pre-Attentive Processing of Musical Information in the Human Brain. *Journal of New Music Research, 28*(3), 237–245. https://doi.org/10.1076/jnmr.28.3.237.3109

Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *NeuroReport, 5*, 844-846.

Tillmann, B. (2012). Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing. *Topics in Cognitive Science, 4*(4), 568–584. https://doi.org/10.1111/j.1756-8765.2012.01209.x

Pato, M. V., Jones, S. J., Perez, N., & Sprague, L. (2002). Mismatch negativity to single and multiple pitch-deviant tones in regular and pseudo-random complex tone sequences. *Clinical Neurophysiology, 113*, 519–527.

Wang, W., Datta, H., & Sussman, E. (2005). The development of the length of the temporal window of integration for rapidly presented auditory information as indexed by MMN. *Clinical Neurophysiology, 116(*7), 1695–1706. https://doi.org/10.1016/j.clinph.2005.03.008

Wang, J., Friedman, D., Ritter, W., & Bersick, M. (2005). ERP correlates of involuntary attention capture by prosodic salience in speech. *Psychophysiology*, *42*(1), 43–55. https://doi.org/10.1111/j.1469-8986.2005.00260.x

Whalley, K., & Hansen, J. (2006). The role of prosodic sensitivity in children’s reading development. *Journal of Research in Reading, 29*(3), 288–303. https://doi.org/10.1111/j.1467-9817.2006.00309.x

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Wickham, Hadley, François, Romain, Henry, Lionel, & Müller, Kirill (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.6. https://CRAN.R-project.org/package=dplyr

Winkler, I., Cowan, N., Csépe, V., Czigler, I., & Näätänen, R. (1996a). Interactions between transient and long-term auditory memory as reflected by the mismatch negativity. *Journal of Cognitive Neuroscience*, *8*(5), 403-415.

Winkler, I., Karmos, G., & Näätänen, R. (1996b). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*(1-2), 239-252.

Winkler, I., Reinikainen, K., & Näätänen, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Perception & Psychophysics*, *53*(4), 443-449.

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, *8*(8), 1971–1974. https://doi.org/10.1097/00001756-199705260-00035

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilomoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*(5), S0048577298000183. https://doi.org/10.1017/S0048577298000183

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: early and late mismatch negativity to tone patterns. *NeuroReport*, *16*(18), 2015-2019.

CHAPTER THREE

Introduction

The previous chapter found that expectancy cues for timing variation in an isochronous stream of tones are differentially sensitive to anticipated tones that occur earlier or later than expected. In this chapter, more linguistically complex stimuli were used to understand how this sensitivity evolves as a function of the complexity of the auditory input. This chapter ends with a comparison of how the response to this timing sensitivity changes as the ‘speechlikeness’—inherent similarity to speech— of the stimuli increases.

In much of the previous work (see for example, Ding & Ye, 2005), complex sound waves have been used as a ‘middle step’ between simple tones and language. The rationale has been that the acoustic complexity of these sounds without any lexical or semantic association makes them an ideal candidate to study how complex auditory inputs are processed, and what it might mean for speech perception and language processing. Previous literature compares both the processing of simple tones and complex waves, and complex waves and language stimuli, allowing an exploration of how increasing speechlikeness and linguistic complexity affect processing.

One study examining the elicitation of long-latency auditory evoked potentials (LLAEPs) using stimuli of all three types (simple tone, complex tone, and a synthesised vowel) in children found that neural activity recorded to each stimulus elicited responses that were largest in magnitude for complex waves, and smallest for simple tones (Čeponiené et al., 2001). This work primarily investigated the evoked response complex of P100-N250-N450. However, the results corroborated previous work looking at stimulus complexity in adults using other ERPs like the N100, where enhanced responses were observed for more acoustically and linguistically complex stimuli compared to simple tones (Tiitinen et al., 1999; Woods & Elmasian, 1986).

This pattern was subverted when linguistic complexity was introduced as a comparison, with components like the N1-P2 complex showing a larger response to non-linguistically complex stimuli than to syllables in adults (Čeponiené et al., 2005). However, for responses that occur later in latency, like the N200 or the N400, syllables elicited a larger response, suggesting a priority in processing speech sounds within that time window post-stimulus.

The study reported in this chapter uses the mismatch negativity (MMN) to explore the processing of timing differences in stimuli that increase in speechlikeness. Based on previous work, the MMN typically occurs earlier than other ERPs. This latency is a critical advantage when evaluating speech as most psycholinguistic models posit early levels of speech processing to begin well before 400 ms (for a detailed review, see Pulvermüller, & Shtyrov, 2006). The advantage of an early latency combined with the fact that the MMN has been described to be a pre-attentive response makes it a particularly useful tool for exploring speech processing.

Complex wave processing

Whereas simple tones are constructed with one base frequency (or pitch), complex waves comprise of a base fundamental frequency (f0), and a number of harmonics that are multiples of the f0. Depending on the manner of construction, the complex wave may comprise of just the f0 and one additional harmonic, or multiple additional harmonics. Previous work has found that complex waves elicit MMNs similarly to simple tones, when presented in a sequence containing a deviant that differs in pitch from some of the harmonics comprising of the complex waves, or in duration (Ding & Ye, 2005; Tervaniemi et al., 1997; Tervaniemi et al., 2000). Complex waves elicit MMNs in infants as well (Cheour et al., 2002), and are processed differentially in aphasics, with diminished responses elicited compared to controls (Pettigrew et al., 2005). Compared to the MMNs elicited by simple tones, those elicited by complex waves have been found to occur earlier in latency and to be bigger in amplitude (Tervaniemi et al., 2000). This difference may reflect a difference in the way complex waves are processed. Tervaniemi et al. (2000) suggest that this MMN response pattern may reflect the facilitative nature of processing complex sounds. Complex sounds are frequently encountered and processed in daily life which may make them easier to process compared to simple tones, which are almost never heard.

Linguistic processing

The MMN response has been used to explore language processing in several limited ways, falling largely into work of two types: phonemic perception, and speech vs. non-speech perception.

Phonemic perception work has utilised the presence of the MMN as a means for understanding perceptual boundaries within and between individuals, with the response being elicited only once the listener is able to discriminate between the different phonemes presented (Aaltonen et al., 1987; Näätänen et al., 2019; Zevin et al., 2010; cf. Phillips et al., 2000). It has also been used to evaluate the impact of phoneme training or participant native language on MMN elicitation (Saloranta et al., 2020; Shestakova et al., 2003; Tremblay et al., 2001). Results of these studies help us understand the perceptual thresholds between different phonemes, and how the learning curve for non-native speakers varies. As well, with further research, inferences can then be made on how more complex aspects of speech, beyond the phonetics, may be processed.

Comparisons of speech processing vs. non-speech processing have found variable MMN responses elicited to speech, depending on context. For example, an MMN was elicited to word deviants within a stream of non-speech sounds, but not the other way around (Pettigrew et al., 2005). Comparisons of pre-attentive speech vs. non-speech sounds also found a larger response elicited to speech when compared to acoustically matched non-speech stimuli (Jaramillo et al., 2001; Kuuluvainen et al., 2014; Sorokin et al., 2010). Studies comparing words to pseudo-words found enhanced MMNs to syllables that completed a word compared to those that completed a pseudoword (Endrass et al., 2004; Pulvermüller et al., 2000; Shtyrov, & Pulvermüller, 2002). Although both deviants were syllables, the results of this study demonstrate the importance of lexical and semantic meaning in auditory processing. In an experiment with deviants that differed in speech-relevant features such as vowel duration, pitch, and intensity in addition to vowel and consonant variations, researchers found that MMNs were larger for speech compared to non-speech sounds for frequency and vowel duration deviants in particular (Sorokin et al., 2010). Considering the participants in this study were Finnish, whose language makes use of vowel length contrastively, the results of this study emphasise the importance of acoustic features used in daily speech perception, and suggest that a participant’s native language has an impact on how they perceive and process speech vs. non-speech sounds.

Korpilahti et al. (2001) examined linguistic processing in children, presenting complex waves, pseudowords, and naturally spoken words. They found MMNs elicited to each stimulus category. Two types of MMNs were observed, categorised based on their latency: early MMNs (150-200 ms), and late MMNs (400-450 ms). Differences in MMN amplitude were found depending on stimulus type, with biggest (late) MMNs observed for words, then complex waves, then pseudowords. The early MMN was bigger for complex waves than other types of stimuli. The authors suggest that the early MMN reflects purely acoustic processing, and the late MMN reflects lexical processing. The presence of a larger, later-latency response to the linguistic stimuli corroborates previous research with LLAEPs, where patterns of neural responses were observed that were similarly larger for syllables than non-linguistic stimuli within similar (later) latency time windows (Čeponiené et al., 2005). It is worth noting that this late MMN, also observed with simple tones, has been suggested to be an indication of long-term memory transfer and may be more accurately defined as the *late discriminative negativity* (LDN) instead (Peter et al., 2012; Zachau et al., 2005). In this case, the bigger late MMNs observed to words may suggest long-term memory access generally rather than lexical access specifically (Korpilahti et al., 2001).

An exception to enhanced speech MMNs was found by Wunderlich and Cone-Wesson (2001), who observed little to no MMNs elicited to CV syllables or words despite (or perhaps, because of) strictly controlling acoustic measures within their stimuli. A similar result was found by Pettigrew et al. (2005), who semi-synthesised CV syllables and presented them in standard-deviant pairs of words and non-words along with tone stimuli that differed on features like frequency, duration, and intensity. These elicited clear MMNs in the majority of participants while speech stimuli did not, replicating the previous work by Wunderlich and Cone-Wesson (2001).

What is evident is that speech and non-speech sounds are processed differently from each other. In experiments exploring the effect of speechlikeness on MMN elicitation in clinical populations such as dyslexic children and adults (e.g., Schulte-Körne et al., 1998; Sebastian & Yasin, 2008; see Hämäläinen et al., 2013, for a review) or persons with aphasia (e.g., Ilvonen et al., 2014; Pettigrew et al., 2005), diminished responses to speech stimuli were observed in comparison to non-speech stimuli. Results from these clinical groups suggest that speech and non-speech sounds are processed via mechanisms that differ from one another, and that deficits in language processing reflect acutely in the responses elicited to stimuli that are linguistic in nature.

Temporal variation processing with increasing speechlikeness as a model for speech prosody

The literature that has thus far been reported has evaluated the elicitation of the MMN in terms of linguistic processing by looking at either manipulations of acoustic features, or through the presence/absence of language (e.g., Korpilahti et al., 2001; Pettigrew et al., 2005; Pulvermüller et al., 2001). These studies fail to capture in detail the role of prosody and stress—and, ultimately, temporal variation—in language processing. To our knowledge, work that focuses on the neural processing of timing variability in speech is limited. Research questions in the literature addressing speech variability have centred around the processing of stress placement in mono- or di-syllabic pseudo-word segments to model prosody (Honbolygó & Csépe, 2013; Honbolygó et al., 2004; Honbolygó et al., 2017). In these studies, stress placement was realised by manipulating various acoustic features of one of the syllables in sequence (e.g., intensity, f0, etc.). Results showed MMN elicitation to depend on whether natural stress was present or not, as well as on whether the heard stress pattern matched the expected stress pattern (i.e., whether the stress pattern was 'legal' in the given language). Other studies have examined how variation in vocalic and intervocalic gaps—that is, lengths of consonants and vowels—is processed and used to discriminate languages (Dellwo, 2008; Dellwo & Wagner, 2003; Ramus et al., 1999; cf. White et al., 2012). Other work has studied speech rate and syllable duration as prosodic markers in speech (Choi, 2003; Grabe & Low, 2002; Tseng & Fu, 2005). However, so far relatively little work has been conducted in understanding how the temporal features of prosody and rhythm of the incoming speech may facilitate language comprehension. No studies to date have extensively examined processing of temporal variation in language, particularly when aspects of speech variability are maintained. In the current experiment, for example, multiple different syllables within a stream are presented as a baseline rather than a repetitive syllable or syllable pair (as has often been used in previous MMN work). Research on the neural processing of prosodic cues has revealed an early role in L1 acquisition (Cole, 2015; Gervain, 2018; Langus et al., 2017), making this an important area of research.

The current chapter aims to identify the temporal factors that affect language processing and understand how these features can help in additional language (Ln) acquisition. This information can then be exploited to facilitate language learning and comprehension and improve speech production in special populations. In Experiment 3 reported here, we studied MMN responses to track how changes in the expected timing of complex non-linguistic versus linguistic stimuli are processed cognitively. We recorded ERP responses to stimuli in a predominantly isochronous sequence. The deviants of interest occurred either too early or too late with respect to the standard stimulus onset asynchrony (SOA). In order to elicit a more ‘speech-like’ stream, multiple tokens were chosen for both the standards and the deviants. Choosing to introduce such variation in the stimulus stream for both standards and deviants is not commonly done. However, when discussing the presentation of a multi-deviant paradigm, Pettigrew, Murdoch, Ponton et al. (2004) suggest that increased variation within stimulus presentation “…[is] more ecologically valid as a paradigm investigating responses to speech stimuli." (p. 286). Considering the aim of our study is to evaluate reactions to timing differences in speech, we presume that introducing the same variation for both standards *and* deviants will not interfere with the processing of the temporal patterns (and may even facilitate it).

Observation of MMNs elicited to patterns of stimuli, rather than to deviants from acoustic repetition, is not new. These responses are often referred to collectively as abstract-feature MMNs; see Näätänen et al., 2007; Näätänen et al., 2001; Saarinen et al., 1992, for reviews). The advantage of an abstract MMN paradigm is that it allows for the processing of the patterns between the stimuli, rather than focusing on the acoustic features of the stimuli themselves (Peter et al., 2012). This better reflects how speech is naturally processed and allows us to examine the underlying mechanisms more precisely. Our study, while motivated by the need to mimic natural speech variation, functions on a similar premise. To our knowledge, however, this is the first study focused on patterns in the context of timing stimuli while maintaining some acoustic variability between both standards and deviants.

Hypotheses

The aim of this experiment was to study how temporal variation in language is processed, and how it changes with the increasing linguistic content of the stimuli. The main question addresses the effect of temporal variation within a steady stream of stimuli that are acoustically more complex than simple tones. We expect an MMN to be elicited to both the temporal deviants that differ depending on the type of timing deviant (Early or Late). Based on the results observed for the MMN elicited to the timing deviants with simple tone stimuli (Experiment 2, Chapter 2), we expect any difference in the observed response to lie primarily in the amplitude domain. It is difficult to predict whether the Early or Late deviant will differ from one another in terms of size, and if so, in what direction. However, based on the fact that the Early timing deviant elicited a larger-sized MMN in the experiment with simple tones, we predict a similar pattern of results. Therefore, we expect that the Early timing deviant will elicit a larger MMN response than the Late timing deviant. Such a response pattern would indicate a greater saliency of the Early timing deviant, with the expectancy formation being violated more strongly for a response that is unexpectedly early than one that is unexpectedly late (even though the two timing deviants differ from the standard by exactly the same latency). We expect an MMN to be elicited to the non-timing, Frequency deviant as well, although we do not make any predictions comparing the timing deviant MMNs to the Frequency deviant MMN. The motivation for including the Frequency deviant in the experiment was to establish an individual baseline for each participant to which the relative size of the response to all deviants (for that participant) could be compared.

In addition to the variation observed between deviants, this experiment aimed to explore the effect of acoustic and linguistic complexity (speechlikeness) on the MMN responses observed, particularly for the timing deviants of interest. This part of our experiment was largely exploratory. We did expect an MMN to be elicited to both complex wave non-linguistic deviant stimuli and to syllable deviants. The stimuli used in this experiment were constructed such that both sets of stimuli were artificially synthesised, but vocal tract characteristics and non-harmonic structure was applied to only one (syllables). Therefore, given the nature of the difference between them, it was possible that there would be no difference in the elicited MMNs for the two stimulus types. Such an outcome would suggest that complex waves and artificially synthesised syllables are processed similarly, with no particular attention attributed to speech stimuli when presented one by one. Alternatively, the MMN elicited may differ in size between the two stimulus types. A larger MMN amplitude elicited by the complex waves might suggest better pattern-forming and expectancy-building for the non-linguistic, purely acoustic, stimuli compared to the artificial speech syllables. A bigger MMN for the syllables could, however, suggest that an intrinsic familiarity leads to priority in the processing of linguistic (speech) stimuli over that for other auditory stimuli, leading to a stronger violation response. We had no expectations regarding latency differences between the two stimulus types, as that MMN feature has not been found to be critical in previous literature examining speech and non-speech processing (e.g., Korpilahti et al., 2001).

In some cases, the MMN response might be accompanied by a fronto-central positivity known as the P3a, signalling that the stimulus triggered attention during its processing (Escera et al., 2000; Polich, 2007). It is worth noting that previous literature has found a strong trend for bigger MMN responses elicited to speech stimuli compared to non-speech stimuli (Korpilahti et al., 2001; Sorokin et al., 2010). However, this seemingly linear relationship of stimulus complexity to response magnitude is not so straightforward. It predicts that simple tones would elicit responses that would be comparatively the smallest in size. This has not been the case. Previous work has reported mixed results, with complex waves eliciting bigger MMNs compared to simple tones (Tervaniemi et al, 2000), but syllable stimuli eliciting bigger MMNs than those elicited by complex waves, yet smaller than those elicited by simple tones (Pettigrew, Murdoch, Kei et al., 2004; Sorokin et al., 2010). If all these results are reliable, the linear pattern of complexity to MMN size is disrupted, and we cannot truly hypothesise that a larger response to the speech stimuli will be elicited.

Experiment 3

Methods

Participants

Twenty Canadian university students (mean age = 21.8; 15 female) with no reported visual/auditory problems were recruited to participate in this experiment. Based on the Edinburgh Handedness Inventory (Oldfield, 1971), 14 participants were right-handed, 4 were left-handed, and 2 were ambidextrous. Full data sets from 2 participants and partial data sets from 3 participants were rejected due to excessive artifacts and noise in the data. Subsequent analyses were conducted on data from 18 participants, with partial data used for some. All participants provided written consent to participate in this experiment in line with the ethical standards of the Declaration of Helsinki, and were either compensated with money or course credit, or volunteered to participate. The study was cleared by the Hamilton Integrated Research Ethics Board (HiREB) in Hamilton, Ontario, Canada.

Experimental procedure

Each session consisted of two variants of the experiment. These differed in the nature of the stimuli only (complex waves, or syllables). An auditory oddball paradigm was used with one frequent standard stimulus type and three different deviant types. Each deviant type was presented in a separate block for a total of three blocks in each experiment variant. The presentation order of the blocks was counterbalanced between participants, as was the order of experiment variant presentation. A total of 900 standards (90%) and 100 deviants (10%) were presented in each block with a standard SOA of 650 ms. Pitch deviants were also presented at the same SOA, but the SOA for timing deviants differed, with early deviants presented at 450 ms, and late deviants at 850 ms. Complex waves and syllable tokens were presented at the same SOA; it was not deemed necessary to change the SOA for the speech stimuli, and indeed, previous work with similar

SOAs have shown that responses to speech stimuli do not vary between SOAs in the 600 ms range, or SOAs which are longer (Pettigrew, Murdoch, Kei et al., 2004).

Each deviant type was presented in its own block, for a total of three blocks presented within each experiment variant: the frequency (pitch) deviant was presented at the same SOA as the standards in its own block; a timing deviant that was earlier than expected (*Early deviant,* SOA = 450 ms) was presented in another block; and a timing deviant that was later than expected (*Late deviant*, SOA = 850 ms) was presented in a third block. Similar to the standards, the timing deviant could be any one of three tokens, distinct from the standards only in the time of presentation. **Figure 9** shows a visual representation of the timing deviants for both complex waves and syllables.

To mimic the variation present in naturally spoken speech, in the blocks with syllable timing deviants, three different tokens were presented equiprobably for both standards and deviants (300 x 3 types of tokens for a total of 900 standards, ~33 x 3 for a total of 100 deviants). The standards consisted of a stream of these three tokens, pseudo-randomised according to the following parameters: 15 standards presented at the start of a block; minimum 2 standards in a row and, maximum 4 standards in a row. The same token was restricted from repeating more than three times in a row.

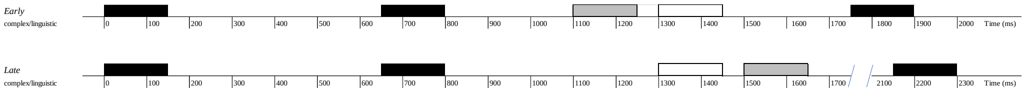

Figure 9. A visual representation of stimulus timing across the Early and Late blocks, for both experiment variants (with Complex waves, and Syllables). The left edge of the block represents the onset of the stimulus at time 0 ms. All stimuli presented were 150 ms in length. Standards are represented in black, whereas the timing deviants are presented in grey. As the Early and Late deviants are temporal deviants, only the time of the onset of the stimulus is different from that of the standard stimulus (i.e., different SOA). The Early deviant was presented 200 ms earlier than expected (SOA = 450 ms), and the Late deviant was presented 200 ms later than expected (SOA = 850 ms). The white blocks represent the expected time of presentation of the stimulus—that is, where the standard stimulus would have been, if it had not been temporally shifted. For the Early and Late deviants, this is at the standard SOA of 650 ms.

Stimuli

Syllables

Three different syllables all beginning with the plosive phoneme /t/ were presented as the standard to emulate natural speech variation. These three syllables were also used as the timing deviants with equal probability, such that the changing stream of syllables differed only in the timing of presentation when the deviant occurred. See **Figure 10** for a visual representation of the stimulus stream.

Three syllables, /ti/, /tu/, /tə/ (f0 = 100 Hz, 150 ms duration, 70 dB SPL_C) were presented as standards, and one syllable with a different phoneme onset, /ka/ (f0 = 120 Hz), was presented as the frequency/pitch deviant. The voiceless plosives /t / and /k/ were chosen to form syllables because these plosives allow precise trigger placement in ERP experiments and are amongst the most frequently occurring consonants in the world (Maddieson & Disner, 1984). This allowed us to test a participant pool that was not limited to native Canadian English speakers. The vowels /i /, /u/, /ə/, and /a/ were similarly chosen to include the most mutually distinctive sounds that also occur most frequently amongst the world’s languages (Maddieson & Disner, 1984). All syllables were constructed using the state-of-the-art articulatory synthesizer VocalTractLab (Birkholz et al., 2019), and constructed to be 150 ms long (5 ms rise/fall times). The first 50 ms contained the voiceless plosive consonant information, i.e., the plosive ‘burst’ (sudden spike in acoustic energy) and frication/aspiration, and the following 95 ms consisted of the synthesised vowel. All resulting syllables were loudness-normalised using SoundForgePro (v.3.0).

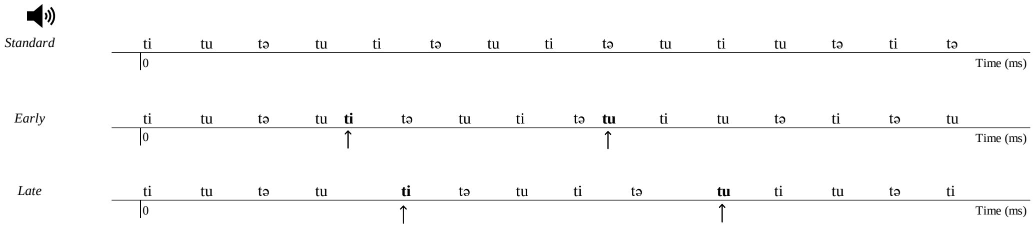
Standard
ti tu tə tu ti tə tu ti tə tu ti tu tə ti tə
0
Time (ms)
Early
ti tu tə tu ti tə tu ti tə tu ti tu tə ti tə tu
0
Time (ms)
Late
ti tu tə tu ti tə tu ti tə tu ti tu tə ti
0
Time (ms)

Figure 10. A Visual representation of the syllables auditory stream presented to participants in Experiment 3. This figure represents an example auditory stream that a participant may have heard during the experiment variant with syllables. Participants heard a stream of three different syllables beginning with /t/ that varied in pace only (that is, the SOA). The stream was pseudo-randomised for each block, although the same stream is represented here for clarity. The timing deviant was one of three possible syllables, /ti/, /tu/, or /tə/, which also comprised the standards. In the figure, the timing deviants are bolded and marked with an arrow. The Early timing deviant was presented earlier than expected, at an SOA of 450 ms, and the Late timing deviant was presented later than expected, at an SOA of 850 ms. The standards were all presented at 650 ms. The distances between subsequent stimulus presentations reflect relative pacing of the syllables only and are not an accurate temporal representation of stimulus pace.

Complex waves

The stimuli for the complex wave experiment variant were developed in tandem with those for the syllable variant. One of the aims of Experiment 3 was to observe how linguistic complexity factored into the perception of pace. Therefore, stimuli were constructed such that those used in one experiment variant were both acoustically complex (in contrast to pure sine tones) and were intensity matched to syllables in the other experiment variant but had no linguistic content.

The four syllable stimuli were used to shape the complex waveforms presented in the second variant of the experiment. The standard complex wave stimulus was a complex acoustic signal constructed by taking an initial frequency (f0) component of 100 Hz, and then adding additional sine wave harmonics that were multiples of the f0 and thus increased in 100 Hz steps up until 8000 Hz, with a spectral slope of – 6 dB/octave. This, thus, effectively generated a complex waveform with a f0 of 100 Hz, and a large number of harmonics gradually decreasing in intensity. This signal was generated using Praat (Boersma & Weenink, 2019). The deviant waveform differed in the base frequency (120 Hz instead of 100 Hz), resulting in differing harmonics; all other features were identical to the standards.

To keep the linguistic and non-linguistic sets of stimuli as acoustically similar as possible, the lengths of the stimuli were kept the same. The generated complex waveforms were processed to be 150 ms in length, with a 5 ms rise and fall time. Intensity envelopes of the three different speech syllables used as the standards, /ti tu tə/, and one for the deviant, /ka/, were used to transfer the amplitude-over-time information onto the generated complex waveform. This resulted in four different waveforms with differing amplitude envelopes, each with a specific syllabic speech amplitude distribution but no linguistic content (see **Figure 11** for a visual comparison between the syllable /ka/ and its complex wave counterpart).

The stimuli were presented to participants via ER-1 Insert earphones (Etymotic Research, Inc., Elk Grove Village, IL, www.etymotic.com) using Presentation® software (Version 18.0 Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The volume was adjusted (i.e., lowered) at the participant's request, if needed.

Testing procedure

After providing informed consent at the beginning of the testing session, participants were asked to fill out a short demographic survey that asked for information such as age, years of education, languages spoken and order of fluency, any visual or hearing impairments, and any relevant medical history (for example, history of concussion, etc.). Participants also filled out the Edinburgh Handedness Inventory (Oldfield, 1971) for an objective assessment of their handedness.

Once the forms had been filled out, participants took part in the EEG experiment. They were seated in a comfortable chair, about 90 cm from a computer monitor, and asked to watch a silent film while ignoring the sounds presented to them. Participants did not have to make any responses during this experiment. The experiment duration was approximately 30 minutes, and participants were given brief breaks between blocks with the option of taking a longer break at any point during the experiment.

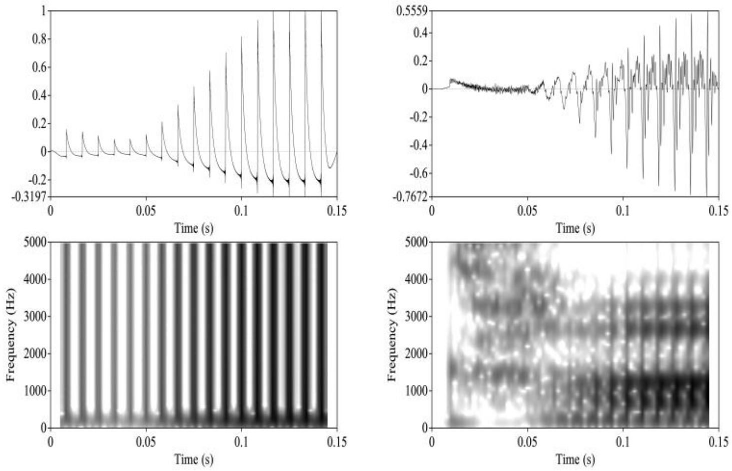

Figure 11. Oscillogram and spectrogram for the generated complex wave (left) containing amplitude-over-time information from the syllable /ka/ (right).

The testing procedure was adjusted during the COVID-19 pandemic to introduce increased health and safety protocols for participants and researchers. Most of the changes implemented affected lab preparation before and after the session, and steps taken to minimise the time participants spent on-site. Participant consent and demographic information was obtained online several days before their testing session. Participants also additionally provided online consent to being at the testing site after understanding the risks of the COVID-19 virus, and were required to share the results of the Ontario COVID-19 self-assessment tool (https://covid-19.ontario.ca/self-assessment/) with the researchers a day before and an hour before the testing session (as per the lab COVID-19 testing protocol). Only participants who were identified as not having any symptoms and not being part of an at-risk group were tested. Other adjustments to the protocol included checking participant temperature and noting their time of entry for contact tracing purposes, as well as ensuring their adherence to the university indoor mask policy. Researchers also recorded their time of entry and exit, temperature, and the results of the Ontario COVID-19 self-assessment tool before each testing session. Researchers were responsible for cleaning the lab and every contact-surface before and after each testing session and wore personal protective equipment (PPE) during the testing session (mask, gloves, face shield). A total of thirteen participants were tested using the COVID-19 protocol.

Electroencephalography (EEG) recordings

EEG was recorded using a BioSemi ActiveTwo system with 64 Ag/AgCl electrodes (placement as per the international 10-20 system; Jasper, 1958) digitally sampled at 512 Hz and bandpass filtered at 0.01-100 Hz. Five Ag/AgCl external electrodes were placed on the participant's nose, left and right mastoids, and above and beside their left eye. The electrodes placed above and beside the left eye were used to record the EOG (electrooculogram) with the same bandpass and sampling rate. Online EEG acquisition was referenced to the DRL (driven right-leg) and re-referenced offline to the average of the left and right mastoids.

EEG data analyses

EEG data were pre-processed and cleaned using BrainVision Analyzer (v2.1.2.327, BrainVision Analyzer, Brain Products GmbH, Gilching, Germany). Data were re-filtered to 0.1-30 Hz (24 dB/oct), and any noisy channels were interpolated from the surrounding channels and replaced (up till a maximum of four channels) (Duda-Milloy et al., 2019). Data were visually inspected for artifacts (e.g., due to movement, or electrode noise), and sections with artifacts greater than 100 µV were removed. The deviant grand average for each participant contained 83 deviants on average (range: 25-99, mean = 83.21, S.D. = 14.3). Participant blocks with fewer than 50% of deviant retention (less than 50/100 after data pre-processing) were evaluated individually on data quality through comparisons with other blocks recorded during that participant session. If the deviant retention was high for other blocks (> 70%), it was assumed that the quality of data obtained from that participant was reliable with a low chance of noise distortion or other artifacts. Therefore, the block with fewer deviants could be retained for further analysis.

The grand average of each standard for each participant contained a maximum of 700 segments; standards presented directly before and after the deviant were not included. Triggers were inserted at the onset of each stimulus to allow the averaging of responses recorded to the stimulus beginnings. Triggers for the timing deviants were inserted to the onset of the presentation of that stimulus. Ocular Independent Component Analysis (ICA) was performed to remove vertical and horizontal eye movements. The data were then segmented into epochs of -100 ms to 600 ms. Difference waveforms were produced by subtracting the averaged waveform of the standard condition from the deviant condition. Automated peak detection (Barr et al., 1978) was conducted to find the maximum amplitude pertaining to the MMN within a window of 100 ms to 300 ms post-stimulus onset, and for the P3a within a window of 250 ms – 500 ms (Polich, 2007). For further analyses, the 64 electrodes were divided into groups of 4-6 electrodes that formed pre-defined regions of interest (ROIs). These ROIs were broadly identified using 3 sagittal planes (right, middle, and left) and 3 coronal planes (frontal, parietal, and occipital) along the scalp for a total of nine ROIs: Right Frontal (RF), Middle

Frontal (MF), Left Frontal (LF), Right Central (RC), Middle Central (MC), Left Central (LC), Right Parietal (RP), Middle Parietal (MP), and Left Parietal (LP). As the MMN and P3a are both fronto-central components, four ROIs corresponding to that area (RF, MF, LF, and MC) were chosen for our analyses to preserve power (Escera et al., 2000; Näätänen et al., 2019).

One-sample *t*-tests were conducted to compare the mean peak amplitude values of the responses to the baseline. To evaluate differences based on the speechlikeness of stimuli and ROI, a 2 x 4 Repeated Measures ANOVA on the mean peak amplitude and latency for both the MMN and P3a responses was conducted, with post-hoc paired *t*-tests. Similar parametric analyses (Repeated Measures ANOVA) were performed in order to explore the role of timing deviant type (2) and ROI (4) on mean peak amplitude and latency of the MMN/P3a responses. Greenhouse-Geisser corrections for sphericity were applied where needed; uncorrected degrees of freedom, and corrected *p*-values and generalised eta-squared (η^2) values are reported. The *p*-values were interpreted at an α level of 0.05, unless otherwise noted. All parametric statistics were conducted using the R Studio (v. 1.4.1106, R Studio Team, 2021) environment for R 4.1.0 (R Core Team, 2021). The following packages were used to reshape data, calculate statistics, and generate figures: dplyr (v.1.0.6, Wickham et al., 2021), ggplot2 (Wickham, 2016), ggpubr (v.0.4.0, Kassambara, 2020), rstatix (v.0.7.0, Kassambara, 2021) and tayloRswift (v.0.1.0, Stephenson, 2021). Although the different stimulus types (Complex waves and Syllables) are referred to here as being presented within different experiment variants, results will be reported as one experiment alone. The variable ‘speechlikeness’ functions as a distinguishing factor between the two experiment variants.

Results

Experiment 3 explored responses elicited to temporal deviants among acoustically complex stimuli. In the two experiment variants, stimuli of two types were used, respectively: syllables, and ‘complex waves’ that contained multiple sine waves layered over each other, and with a syllable-shaped acoustic envelope applied. Three types of deviants were presented in each experiment variant (Early timing deviant, Late timing deviant, Frequency deviant). These three kinds of deviants were presented in separate blocks. The only difference between the experiment variants containing complex waves with minimal linguistic content vs. syllables was the type of stimuli. All other experimental parameters were kept the same.

MMN responses are typically reported using difference waves. For each participant, difference wave responses were calculated to each deviant for each stimulus type. The averaged response to the standards was subtracted from the averaged response to the deviants for each participant, in each block. Mean peak amplitude and mean peak latency values for two ERP components, the MMN and the P3a, were extracted from this subtraction by identifying the most negative value within a pre-defined time window of 150 ms – 300 ms for MMN (Näätänen et al., 2019), and the most positive value within a later pre-defined time window of 250 ms – 500 ms for P3a (Polich, 2007). The latter ERP is an attention-orienting response that is often elicited in tandem with the MMN and follows similar topographical distributions (both are maximal over the fronto-central regions; Escera et al., 2000).

All deviants (Early timing, Late timing, and Frequency) elicited an MMN and P3a for both Complex Waves and Syllables. One-sample *t*-tests were conducted to evaluate whether the mean peak amplitudes of these difference waves were significantly different from 0 (baseline). The MMN and P3a peaks were found to be significantly different from baseline, $p < 0.05$ (see **Table 3** for detailed comparisons). **Figure 12** shows the difference waveforms for each Deviant Type by level of Speechlikeness for the ROI MF.

Table 3. One-sample t-tests comparing the mean peak MMN amplitude to baseline (0 µV) across the three Deviant Types (Early timing, Late timing, and Frequency), at the ROI: MF.

One-sample t-test comparing mean peak MMN and P3a amplitude from baseline (0 µV) by Deviant type (ROI: MF)

	Speechlikeness	*Deviant type*	*Amplitude (µV)*	*t*	*df*	*Conf. low*	*Conf. high*	*p*	*Significance*
MMN	COMPLEX WAVES	Early	-1.81	-4.16	17	-2.73	-0.89	0.000649	**
		Late	-2.90	-11.3	17	-3.45	2.36	2.61E-09	**
		Frequency	-3.90	-8.83	17	-4.84	-2.97	9.32E-08	**
	SYLLABLES	Early	-2.14	-4.91	16	-3.06	-1.21	0.000158	**
		Late	-1.89	-7.32	16	-2.44	-1.35	0.000002	**
		Frequency	-2.30	4.80	16	-3.32	-1.28	0.000197	**
P3a	COMPLEX WAVES	Early	2.41	5.50	17	1.49	3.33	0.00004	**
		Late	1.37	3.27	17	0.49	2.26	0.0045	**
		Frequency	2.44	4.23	17	1.22	3.65	0.00056	**
	SYLLABLES	Early	3.08	9.59	16	2.40	3.76	4.92E-08	**
		Late	1.81	3.61	16	0.75	2.88	0.00236	**
		Frequency	2.59	6.99	16	1.81	3.38	0.000003	**

Note:

- Indicates no significance p > 0.05

* Indicates significance p < 0.05

** Indicates significance p < 0.01

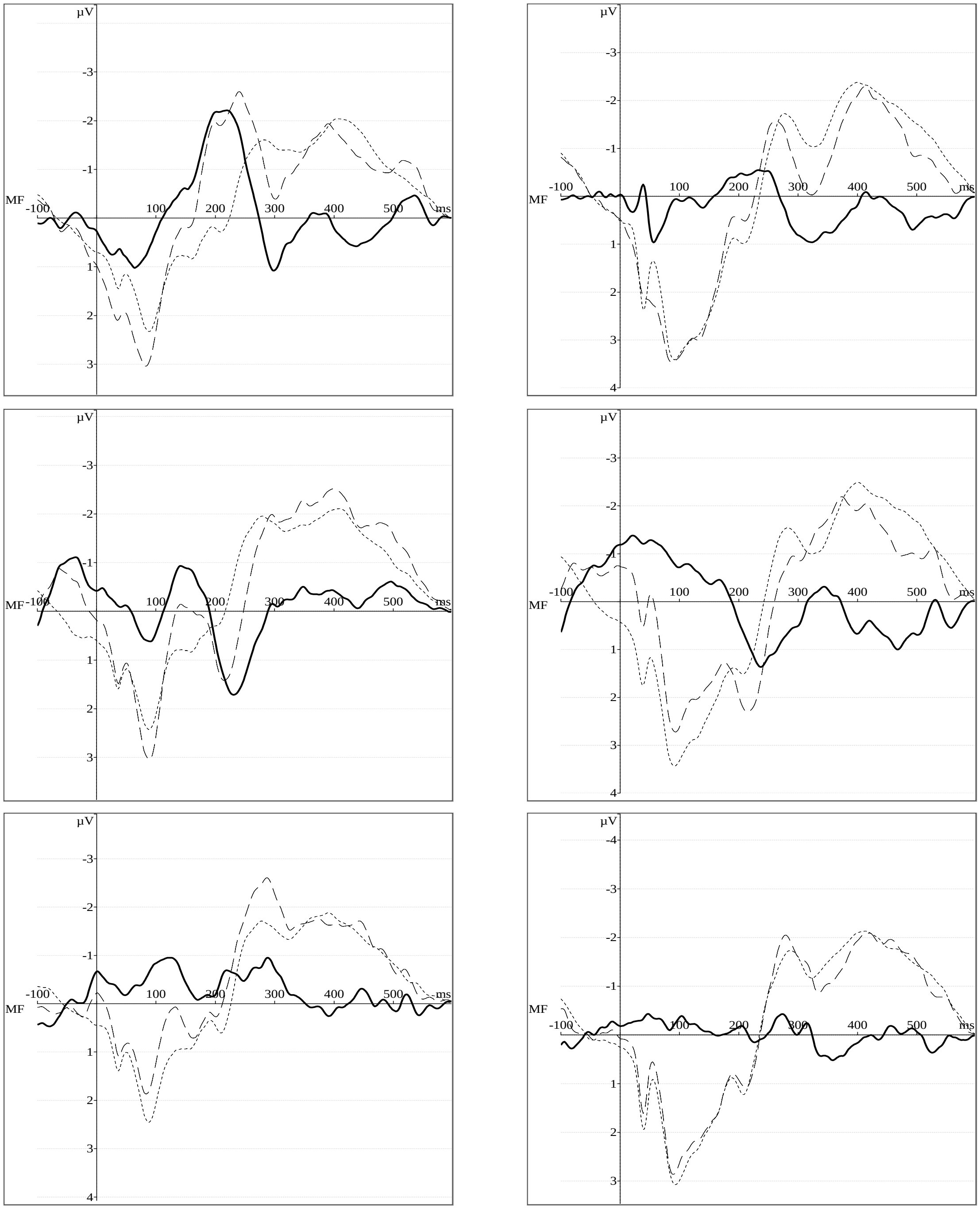

Figure 12. Difference waves created by subtracting the mean response amplitude to the standard from the mean response amplitude to the deviant. The responses are from the Middle-Frontal ROI and have been filtered at 1-25 Hz. The top row shows the responses to the Frequency deviant, the middle row to the Early timing deviant, and the last row to the Late timing deviant. The left column shows the response to the Complex Waves, and the right column shows the response to the Syllables. Within the figures, the solid line shows the difference wave; the long dashed line shows the deviant response and the short dashed line shows the standard response.

The effect of temporal variation and speechlikeness on MMN and P3a amplitude

A summary of the MMN and P3a mean peak amplitude and latency values recorded across the three deviants, for both Complex waves and Syllables, is reported in **Table 4**.

A 2 x 3 x 4 repeated measures ANOVAs was conducted for the mean peak amplitude to evaluate the effect of Speechlikeness (Complex waves/Syllables), Deviant Type (Frequency, Early timing, Late timing), and ROI (RF, MF, LF, MC) on the amplitude of the MMN and P3a responses respectively.

MMN

There were significant main effects of Speechlikeness ($F(1, 15) = 5.46$, $p = 0.034$, $\eta^2 = 0.006$), Deviant Type ($F(2, 30) = 6.42$, $p = 0.005$, $\eta^2 = 0.089$), and ROI ($F(3, 45) = 32.09$, $p < 0.001$, $\eta^2 = 0.071$) on MMN amplitude. There were significant interaction effects of Speechlikeness and Deviant Type ($F(2, 30) = 5.36$, $p = 0.010$, $\eta^2 = 0.082$) and of Speechlikeness and ROI ($F(3, 45) = 10.06$, $p < 0.00001$, $\eta^2 = 0.008$). The effect of Deviant Type at the two levels of Speechlikeness was evaluated separately using one-way ANOVAs. A significant main effect was found in the Complex waves analysis only ($F(2, 34) = 8.93$, $p = 0.0008$, $\eta^2 = 0.23$). The more speechlike syllable stimuli, therefore, did not show significant differences between Deviant Types, but the Complex wave stimuli did. Complex wave stimuli elicited significantly larger responses averaged across all Deviant Types (- 2.49 μV) compared to Syllable stimuli (-1.84 μV) (paired $t(15) = -2.34$, $p = 0.034$). This could be why differences in elicited responses between Deviant Types were more reliable for Complex wave stimuli than Syllable stimuli.

Follow-up paired comparisons showed a significant difference between Early timing and Frequency deviants ($t(17) = 3.95$, $p = 0.001$; $\alpha = 0.017$, Bonferroni corrections applied) only, with a larger-amplitude MMN being elicited by the Frequency deviant (-3.27 μV) compared to the Early deviant (-1.56 μV). The difference between the two timing deviants was not significant ($t(17) = -2.11$, $p = 0.05$), nor did the Late timing deviant (-2.40 μV) and Frequency deviant show a significant difference ($t(17) = -2.29$, $p = 0.035$). This pattern of results suggests that the frequency (pitch) variations were more salient and, therefore, more easily processed acoustically in comparison to the temporal variations within the same time window, although the results reached significance only in the comparisons between the Frequency and the Early timing deviants.

The interaction between Speechlikeness and ROI was investigated in another pair of post hoc analyses. The effect of ROI was evaluated for Complex waves and Syllables separately. The effect was significant for both Complex waves ($F(3, 51) = 33.42$, $p < 0.00001$, $\eta^2 = 0.179$) and Syllables ($F(3, 51) = 17.98$, $p < 0.00001$, $\eta^2 = 0.116$). **Figure 13** shows the MMN difference waves across ROIs. Visually, differences can be observed between each, with mean peak amplitude values greater for the fronto-central regions than the parietal, more posterior ROIs. This is as expected; MMN and P3a responses are usually larger in magnitude in the fronto-central regions (ROI: LF, MF, RF, MC) than the more posterior ones (e.g., LP, MP, RP). The responses shown in Figure 13 demonstrate this pattern.

There were no specific hypotheses regarding MMN peak latencies for different Deviant types. However, a pattern of earlier MMN latencies was observed for the Early timing deviant compared to the other two deviants. A paired *t*-test evaluating mean latency between the different deviant types averaged over the two levels of Speechlikeness showed a significant difference between the Early timing deviant (175 ms) and the Frequency deviant (210 ms) only, $t(17) = -2.66$, $p = 0.016$ ($\alpha = 0.017$, Bonferroni correction applied). Comparisons of latency differences were not significant between the Late timing deviant (198 ms) and the Frequency deviant ($t(17) = 0.93$, $p = 0.366$), nor between the two timing deviants ($t(17) = -1.48$, $p = 0.157$). This significantly shorter Early deviant latency reflects a response to the onset of the unexpectedly early presentation of the timing deviant, where the triggers were placed to create epochs, and not to the expected point of presentation. Therefore, the earlier latency may reflect the stimulus being processed at an earlier time point than the expected temporal pace of the standards. Similar differences therefore were not observed for the Frequency deviant, where the stimuli were presented at the same rate throughout, nor for the Late timing deviant, which was presented with a delay and would therefore not have any processing buffers in terms of time when compared to the standard pace of presentation.

P3a

2 x 3 x 4 omnibus analyses including Speechlikeness, Deviant Type, and ROI found a significant main effect of ROI on P3a amplitude, ($F(3, 45) = 18.863$, $p < 0.001$, $\eta^2 = 0.022$). The main effects of Speechlikeness ($F(1, 15) = 0.504$, $p = 0.489$, $\eta^2 = 0.005$) and Deviant Type ($F(2, 30) = 2.716$, $p = 0.082$, $\eta^2 = 0.005$) failed to reach significance. However, there was a significant interaction between Deviant Type and ROI, $F(6, 90) = 3.24$, $p = 0.006$, $\eta^2 = 0.006$. The simple main effect of ROI for the three Deviant Types was evaluated using one-way ANOVAs. Significant effects of ROI were found for the Early timing deviant ($F(3, 51) = 25.55$, $p < 0.00001$, $\eta^2 = 0.097$) and the Frequency deviant ($F(3, 51) = 11.45$, $p < 0.0001$, $\eta^2 = 0.08$), but not for the Late timing deviant ($F(3, 51) = 1.86$, $p = 0.17$, $\eta^2 = 0.007$). These results suggest differences in the distribution of P3a amplitude by ROI for all deviants except the Late timing deviant. Figure 13 displays the differences in P3a amplitude across ROI. Overall, there was no

significant effect of speechlikeness on P3a amplitude, suggesting that this attention-orienting response is not modulated by linguistic familiarity. Paired *t*-tests comparing P3a amplitude between Deviant Types showed no significant differences, $p > 0.016$ ($\alpha = 0.016$, Bonferroni correction applied).

P3a peak latencies were not analysed as they had no relevance to the research question. However, observationally, Early timing deviants showed a similar trend as MMN latencies. Earlier P3a responses were detected overall for the Early timing deviants (338 ms) compared to the P3a latencies for the Late timing deviant (389 ms) and the Frequency deviant (378 ms).

Table 4. Mean difference wave amplitude and latency values for each deviant type (Frequency, Early timing and Late timing deviant) across the two levels of speechlikeness (Complex waves, and Syllables) for both the MMN and P3a responses (recorded at the Middle-Frontal ROI).

Peak mean amplitude (µV) and peak mean latency (ms) for MMN and P3a responses by Deviant Type and Speechlikeness

			MMN		**P3a**	
Speechlikeness	*Deviant type*	*n*	*Mean amplitude (µV)*	*Mean latency (ms)*	*Mean amplitude (µV)*	*Mean latency (ms)*
Complex waves	Early	72	-1.56 ± 1.53	182 ± 65.6	2.15 ± 1.64	326 ± 86.2
	Late	72	-2.40 ± 1.15	194 ± 74.0	1.45 ± 1.73	407 ± 75.2
	Frequency	72	-3.27 ± 1.67	211 ± 39.9	2.18 ± 2.04	382 ± 79.6
Syllables	Early	68	-1.85 ± 1.47	161 ± 58.7	2.68 ± 1.19	349 ± 85.0
	Late	68	-1.65 ± 0.99	201 ± 76.2	1.69 ± 1.90	371 ± 61.0
	Frequency	68	-2.14 ± 1.71	213 ± 52.6	2.20 ± 1.40	373 ± 76.7

mean ± standard deviation (s.d.)

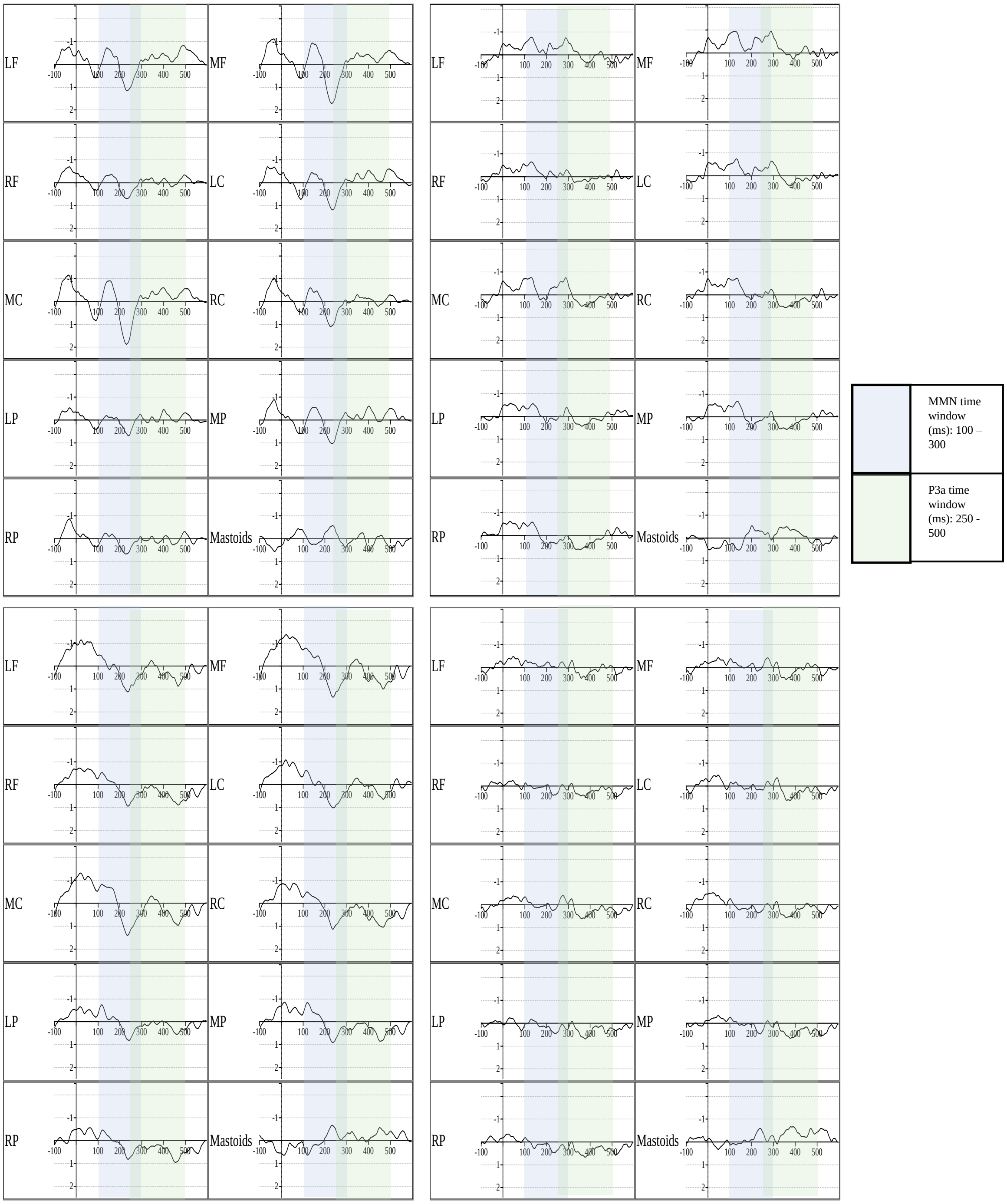

Figure 13. Difference waves to timing deviants created by subtracting the standard from the Early timing deviant (left) and Late timing deviant (right). Responses are shown across all nine ROI, starting from frontal electrodes (LF, MF, RF) to posterior, parietal ones (LP, MP, RP). Responses from the average of the mastoids are shown as well. Waveforms presented have been filtered at 1-25 Hz, with negative amplitude presented upwards by convention. The top panels show the response to the Complex Waves, and the bottom panels show the response to the Syllables.

Topographical distribution of MMN and P3a to timing deviants

The MMN and P3a peaks for the Timing Deviant Type by Speechlikeness are shown in **Figure 14**. This section will report on the topographical distribution of MMN and P3a activity for the timing deviants separately for Complex waves and Syllables.

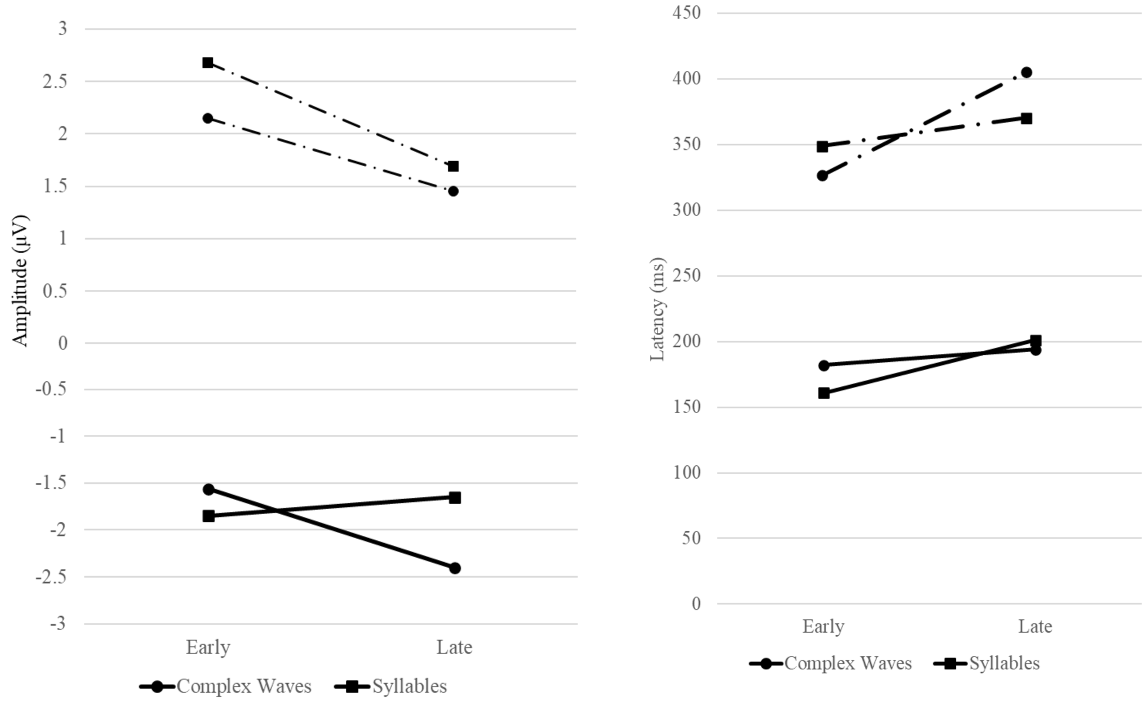

Figure 14. MMN and P3a mean amplitudes (µV) and mean latency (ms) values for the Early and Late deviants as a function of Speechlikeness (ROI: MF). The solid line shows MMN values, and the dashed line shows P3a values.

Complex Waves

The scalp distribution of Timing Deviant effects is presented in Figure 15, displaying the responses to the standard stimuli and timing deviants at the frontal electrode FCz. For the Early timing deviant (upper panel), the perceptual N100 component can be seen peaking at approximately 150 ms post-stimulus onset for both the standard and the deviant stimuli. The waveform is slightly more negative for the deviants than it is for the standards, suggesting the presence of the MMN response in combination with the N100 (*deviant-related negativity*; Duda-Milloy et al., 2019; Tavakoli et al., 2021). Beginning at around 200 ms, the deviant shows a more positive response compared to the standard up until approximately 300 ms post-stimulus onset. While this response could be a P200 response (often evoked in combination with the N100 to form the N1-P2 complex), the prolonged latency of the response means that this could also be an early P3a response. This positivity is distinctly reflected in the topographical map as well within the 199 – 248 ms time window, along the central electrodes. Finally, while outside our pre-determined MMN time window, the deviant does show a prolonged, if slight, negativity compared to the standard during the last half of the epoch (300 ms – 550 ms).

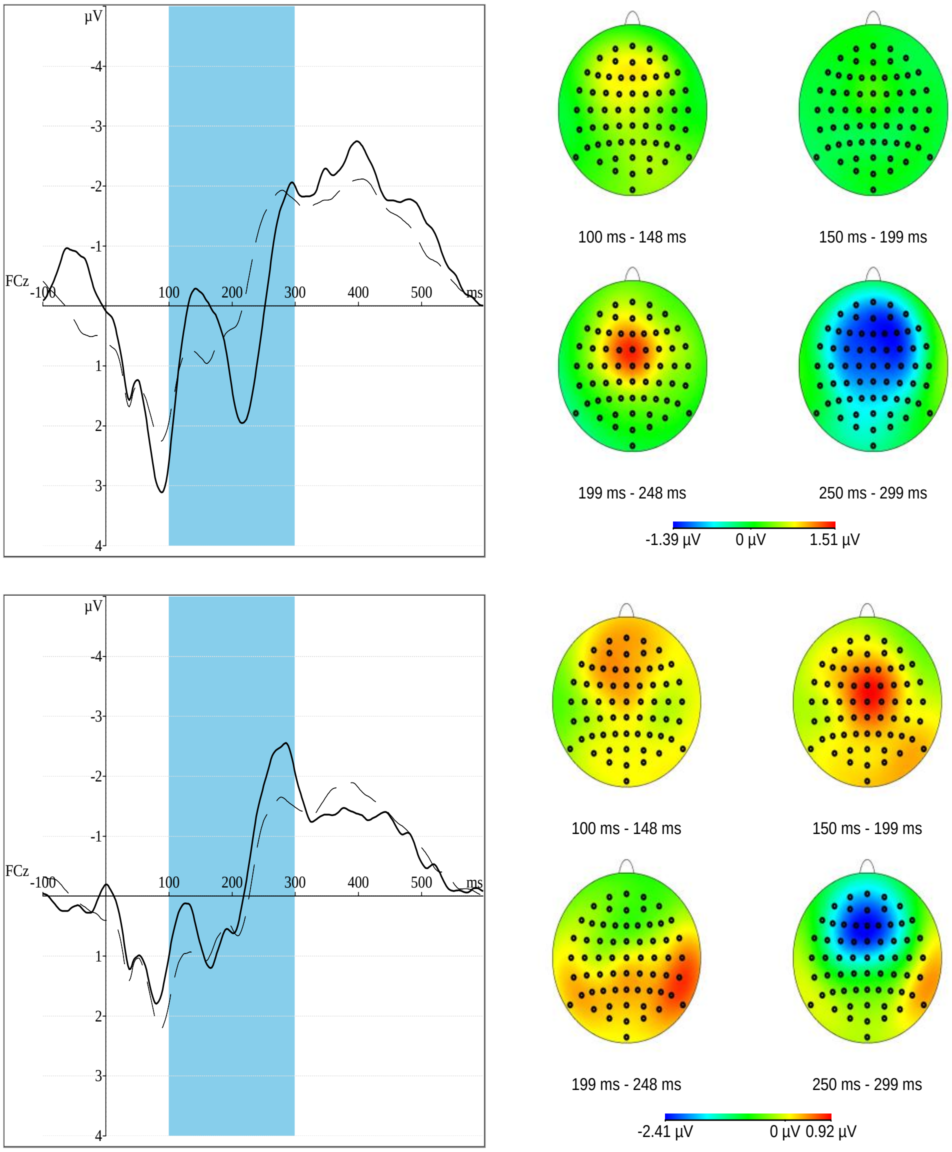

Figure 15. Waveform graphs in the left panel and scalp distribution maps in the right panel for the Early and Late timing deviants at the FCz electrode for the responses to the Complex Wave stimuli, Solid lines show responses to the deviants, dashed lines to the standard stimuli. The upper panels show responses to the Early Deviant and the lower panels to the Late Deviant. Changes in amplitude across the scalp are shown in four maps corresponding to 50-ms time steps inside the 100-ms to 300-ms window. The spheres are top-view representations of a head, with equidistant electrode placement on the scalp. All figures are filtered between 1-25 Hz.

For the Late timing deviant, (lower panel, Figure 15), a sharp N40-P50 complex is visible at the beginning to the onset of the acoustic stimuli. The N100 peaks earlier at ~120 ms compared to the same response for the Early timing deviant, which peaked at ~150 ms. There is a prolonged negative response for the deviant compared to that for the standard. It begins at around 50 ms and continues up until 350 ms. The topographical maps reflect a slight negativity, or general positivity for the response to the deviant in the same time window, however. A slight positive response between 350 ms and 450 ms is observed also.

Syllables

Figure 16 shows representative scalp distribution maps for ERPs at the FCz electrode within a time window of 100 ms to 300 ms post-stimulus onset. The neural response to the standard stimuli and the timing deviants in the left panels, and the topographical maps in the right panels. The upper panels of the figure shows the response elicited to the Early timing deviant. It is more negative in the first half of the epoch (before 250 ms), and more positive later.

The negativity extends from the pre-stimulus baseline up until about 180 ms post-stimulus, suggesting the presence of some artifact in addition to the response being recorded[i]. Peaks around 50 ms are again observed, suggesting the presence of an N40/P50 complex[1]. The negativity present before 200 ms is briefly seen from 350 ms to 400 ms as well, perhaps suggesting the presence of a late MMN. For most of the pre-defined window for the MMN (100 ms to 300 ms), however, the Early timing deviant evoked a response that is more positive compared to the standard. This is reflected in the topographical map between 199 – 248 ms, located around the central electrodes. A lingering negativity is also observed on the topographical map from 250 – 299 ms, although not visible clearly in the waveform as the magnitude of the negative response is small.

The response to the Late deviant, observable in the bottom panel of the figure, shows a negative-going peak at 50 ms, and then two peaks at around 270 ms and 400 ms. The response to the deviant is only more negative from 250 ms to 300 ms; a positivity is observed between 320 ms and 400 ms. Otherwise, the difference between the Late deviant and standard responses is minimal. The

[1] This response was more visually salient for syllables but is present in data for both stimulus types. As the response is consistent across all types of deviants (timing and non-timing), and reliably occurs within 50 ms of the stimuli onset, it is possibly an artifact present at the onset of the stimulus sound file, like a ‘click’. Nordby et al. (1988a, 1988b) report a similar deflection to the one observed in our data, although they do not explain its presence beyond called it an ambiguous response.

topographical maps for the Late deviant reflect a similar pattern to the ones for the other two deviants in the syllable condition, with a positivity early in the time window, and a negativity in the last 50 ms (250 – 299 ms).

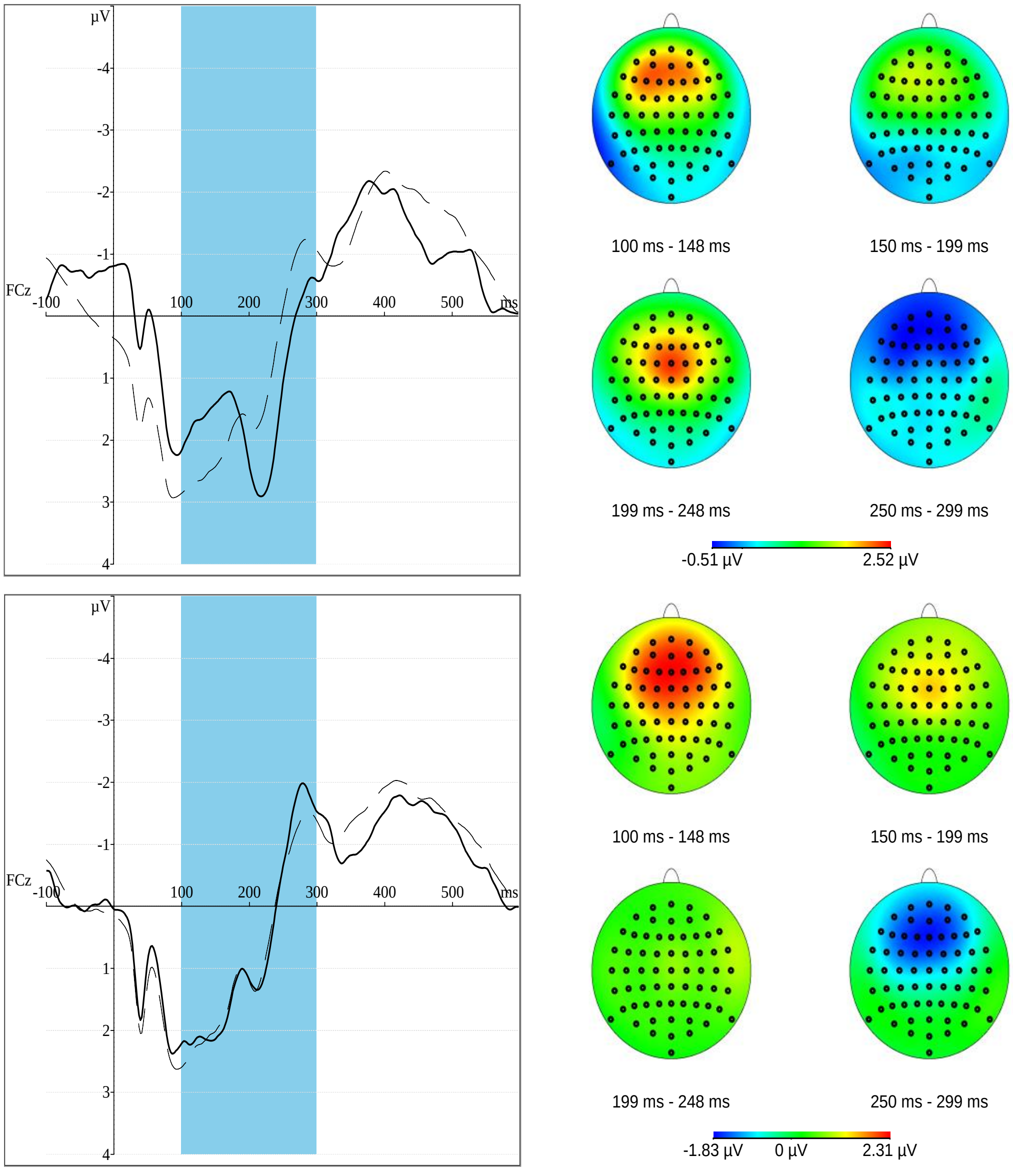

Figure 16. ERP responses and scalp distribution maps for the responses to Early and Late timing deviants for the Syllable stimuli at the FCz electrode. Solid line shows responses to the deviants, dashed line shows responses to the standard stimuli. The upper panels show responses to the Early timing deviant; the lower panels to the Late timing deviant The figure shows the top-view representation of a head, with equidistant electrode placement on the scalp. Changes in amplitude across the scalp are shown in four 50-ms time-step maps for the 100 ms to 300 ms window.. All figures are filtered between 1-25 Hz.

Summary of results

Overall, an MMN of varying amplitude, but significantly different from baseline (0 μV), was elicited for all deviant types, as expected. Robust positivities, the P3a response, were also observed for all deviant types. The MMN response appeared to be more robust for the Complex wave stimuli than it was for the Syllable stimuli, and this difference was significant in the omnibus analysis of amplitude data for the Early timing and Frequency deviants. Conversely, a more dominant positivity was observed amongst the responses to the Syllables. However, analyses showed a significant difference for the P3a amplitude only. This was observable for the Frequency deviant and the Early timing deviant but not the Late timing deviant. MMN difference wave amplitude was significantly different between the Early and the Frequency timing deviants for Complex waves, suggesting that the two are differentially processed for non-linguistic stimuli.

Discussion

The purpose of the current experiment was to explore how temporal variation affects the processing of acoustically complex unfamiliar and linguistically familiar stimuli to better understand processing of time variation in speech. Neural responses reflected in ERPs were explored. Acoustically complex stimuli, in this context, were stimuli consisting of fundamental frequency and harmonics ("complex waves"), or with stimuli that had the same structure (f0 and a number of harmonics) but additionally require a higher level of cognitive processing (linguistic elements, for example). In the experiment reported in this chapter, participants were presented with synthesised syllables and complex waves tokens that varied unexpectedly in their timing. We predicted an MMN to be elicited to all deviant types and hypothesised the early timing deviant to elicit a larger response than the late timing deviant. While our examination of the effect of Speechlikeness on MMN amplitude was exploratory, we expected there to be some effect on the elicited brain responses based on how similar to speech the stimuli were.

An MMN was observed for all deviant types in this experiment, regardless of the nature of the stimuli. The magnitude of the elicited response differed by both Deviant Type and the familiarity (Speechlikeness) of the stimuli. Complex waves, overall, elicited larger MMNs on average than did syllables. Post-hoc analyses revealed significant differences between Deviant Types depending on level of Speechlikeness, with Complex waves showing a significant difference between the Early timing and Frequency deviants. No differences were observed between Deviant Types for Syllables.

A P3a effect suggesting attentional capture was also observed for all the deviants presented in this experiment, for both Complex Waves and Syllables.

The predictive coding model posits that the elicitation of the MMN is dependent on the wider context of the stimulus sequence, as well as the direct differences between the standard and deviant stimuli (Fong et al., 2020). We expected the timing deviants to elicit an MMN the same way as the acoustically different Frequency deviant or any other deviant varying in acoustic features would. Within the predictive coding framework, timing as a feature should be processed the same way as any other acoustic feature. While there may be predictive differences in the way different types of temporal variations are processed, previous work has suggested that temporal variation proportionally affects MMN amplitude, with a larger difference being reflected as a bigger response (Alain et al., 1999; Kisley et al., 2004). Sable et al. (2003) found that this proportional amplitude change occurred regardless of the direction of the timing variation. The authors presented trains of tones where the last tone varied in time of presentation to different degrees with the aim of studying changes in the MMN as a reflection of changes in auditory processing of temporal information. This is different from the current experiment, where participants were presented with temporal variation of acoustically complex stimuli in separate blocks where only one type of temporal deviant was presented at a time.

If there are familiarity effects, a predictive coding framework in the context of speech would lead us to expect the resulting ‘error response’, or MMN, to be modulated by the prior expectations of timing variability in speech. Considering natural pauses and delays in real-time communication, it can be posited that Late timing deviants would be more expected than Early timing deviants, and so would elicit a smaller MMN comparatively. Although our results failed to show any significant differences between the Early and the Late timing deviants, visually the Late timing deviant is smaller. Further work with larger sample sizes could help provide stronger evidence for how Late timing deviants are processed compared to Early timing deviants.

Finally, an alternative explanation in the form of the competing stimulus-specific adaptation (SSA) hypothesis of MMN generation can be considered (see May & Tiitinen, 2010, for an extensive overview). This hypothesis predicts an inhibitory effect of repeated stimulus presentation on neural firing rates, with the MMN being a novelty response elicited to a new, rare deviant presentation within a stream of old, repetitive standards. May & Tiitinen (2010) argue that this model can consolidate temporal rate of presentation via neural clusters that respond strongly to different rates of presentation. This would suggest the same degree of response elicited to both the Early and the Late timing deviant presented in our experiment. While the Early and the Late timing deviant amplitudes did not differ significantly, visually there is a difference between them that trends in the direction of the Early timing deviant eliciting a larger MMN than the Late timing deviant. Indeed, for simple tones, the Early timing deviant was found to elicit a significantly larger MMN than the Late (*Chapter 2*). While it appears that the results of our experiment are not adequately explained by the SSA model, more work investigating the differences between the two timing deviants is necessary to form any conclusions. However, in terms of speechlikeness in particular, our results do not reconcile with the predictions of the SSA model, which again predicts an equal response elicited to the complex acoustic features that make up the auditory percept in our experiment. A significant effect of Speechlikeness was observed in our results, wherein a larger MMN was elicited to the Complex waves than to the Syllables. This model, therefore, may not fully explain our results as compared to the predictive coding framework, which is able to accommodate timing and acoustic complexity differences.

The effect of temporal variation on acoustic processing

Temporal variation effects on the MMN

The main research question we aimed to address with this experiment was how temporal variation is processed depending on acoustic and linguistic complexity. Unexpectedly Early and unexpectedly Late deviants that were otherwise identical to the standards were presented to participants. Overall, visually larger MMNs were observed to the Early timing deviants compared to those elicited by the Late timing deviants. This response pattern is as we hypothesised, suggesting that the Early timing deviant is acoustically processed more easily than the Late. Previous literature has studied varying temporal patterns in streams of simple tones and found the elicited MMN amplitudes to be proportional to the absolute timing difference between the standard and the deviant (Alain et al., 1999; Kisley et al., 2004; Sable et al., 2003). In our experiment, both the Early and Late deviants differed from the standard to the same degree (i.e., by 200 ms). Based on these results, we would expect the MMN, if elicited, to be of the same amplitude for both deviant types. This is not what we observed; our results found that the Early deviant reliably elicited an MMN that was more negative than that observed to the Late deviant. This difference is also visually apparent. Although the conducted analyses failed to reach significance for the comparison, the Early and Late timing deviants observationally differ from one another in terms of the mean peak amplitudes recorded for both. This casts some doubt on the suggestion that the timing difference between standard and deviant may proportionally affect MMN amplitude. Indeed, this pattern of results was observed in *Chapter 2* as well, with simple tones eliciting a larger response to an Early timing deviant compared to a Late deviant. However, the effect in the present experiment was not statistically significant, in large due to individual variation within the data. Further research on individual differences in processing timing deviants will help shed light on how the type of temporal variation is processed relative to each other (Early vs. Late).

The larger amplitude MMN to the Early deviant might be more accurately described as the DRN (deviant-related negativity; e.g., Tavakoli et al., 2021).

However, there are no latency differences observed for the responses elicited to both timing deviants. We can assume, therefore, that the N100 response affects the observed response for both timing deviants equally. The larger response elicited to the Early timing deviant should not be disproportionally driven by the N100 response compared to the observed response to the Late timing deviant. Therefore, we can consider the difference in size of the observed MMN to be related to the nature of the timing deviant. Although conventionally, N100-MMN complexes are referred to as the DRN, here the discussion will refer to the observed/elicited responses as the MMN.

According to the predictive coding model, the error response that is the MMN is elicited when predictions by an internal model do not match input (Ylinen et al., 2017). The better the predictions are met, however, the smaller the error response and therefore the MMN. In this context, therefore, the non-significant response to the Late timing deviant suggests that there was very little difference between the anticipatory prediction and the actual presentation of the late timing deviant. In contrast, the Early timing deviant elicited a larger error response. Based on the observed results, therefore, we can posit that the internal model was able to accommodate the Late timing deviant better than the Early timing deviant. Considering that the internal model is based on information from both the input stream and from prior experience with the stimuli (or anything similar to them), it is possible that the Late timing deviant was better accommodated within the predictions as it was the most ‘acceptable’ and widely experienced form of variation in speech.

Alternatively, it is possible that the internal model relied entirely on the context created by the repeating stream of consecutive standard stimuli. Considering the isochronous presentation of the standards and therefore the anticipated pace of the incoming stimuli, the Early timing deviant would have been a surprise event that fell within the scope of this window. The Late timing deviant, on the other hand, would have fallen outside; therefore, perhaps, there would be no error response generated to the Late timing deviant because there was no comparison being made between it and the internal model. This explanation would suggest that there would be no MMN elicited to the Late timing deviant. However, this is not the case; significant MMN responses were observed to the Late timing deviant for both the Complex wave and Syllables stimuli, suggesting that there is some predictive mechanism underlying this elicitation. Work specifically exploring timing variation in acoustically complex stimuli is limited. While temporal variation and pattern violations have been studied with simple tones (see *Chapter 2*), the literature has overwhelmingly focused on the Early timing deviant as a marker of timing expectation violation, and limited work on the Late timing deviant exists. One study found a similar pattern of results to ours, with an increased inter-stimulus onset (IOI; time between onsets of tones, similar to SOA but modality-specific) leading to smaller amplitude response in the MMN compared to a decreased IOI (Lai et al., 2011). However, the size of the elicited response was smaller compared to that observed in our study. Additional evidence for neural differences in the processing of early versus late timing deviants comes from a P3 study exploring temporal variation in beat presentation, whereby the authors suggest that the way the unexpectedly early vs. late deviants are processed differs (Jongsma et al., 2007). Therefore, while it is possible that the Late timing deviant elicited a smaller response because it fell outside the internal predictive model, it seems unlikely that this is the case as a reliable MMN was elicited to it. It is more likely that the Late timing deviant is processed differently from the Early timing deviant. Again, either because it is better accommodated by the internal predictive model that is 'attuned' to the delays. Or perhaps because the delayed deviant creates anticipation which naturally leads to a smaller error response because the deviant is now expected.

When the frequency deviants were excluded from the analysis, the MMN amplitudes elicited by the two timing deviants did not show any differences between the Complex wave and the Syllable stimuli. However, visually, the two timing deviants appeared to elicit larger MMN responses to the Complex wave stimuli than they did to the Syllables. There is only a limited amount of previous work that varied the speechlikeness of stimuli when exploring responses to timing variations. In general, increased speechlikeness has been associated with larger responses (see e.g., Korpilahti et al., 2001), and following the observed pattern of larger responses elicited to Early timing deviants (in this chapter and previously, e.g., Lai et al., 2011), we would expect Speechlikeness to interact with timing deviant type such that the Early timing deviants elicited a larger MMN than the Late timing deviants, and that this response was larger overall to Syllables compared to Complex waves. However, while our results find that the Early timing deviant did, indeed, elicit a larger response compared to the Late timing deviant, there is no significant effect of Speechlikeness between the two, suggesting that the two timing deviants are not processed differently with increasing acoustic complexity in any robust way. The next section discusses the effect of speechlikeness on acoustic processing in more detail.

Temporal variation effects on the P3a

Visually, the Early timing deviant elicited a larger-amplitude P3a than the Late timing deviant for Syllables, suggesting that the former engages attention to a greater degree than the latter in speechlike stimulus streams. This difference was not significant so it is difficult to conclude with any certainty whether the Late timing deviant was processed differently from the Early timing deviant. However, if this result was robust, it could provide further evidence for the Late timing deviant being deemed more 'acceptable' within the context of the internal model, particularly in the context of the Syllable stimuli. Future work exploring temporal processing within a more controlled, homogenous participant group would help provide definite conclusions.

It is worth noting that previous work examining timing differences in sequences of simple tones has found that Early and Late timing deviants show differential P3 activity in terms of the P3 subcomponents observed to each (Jongsma et al., 2007). Early timing deviants led to larger late-P3 subcomponents (P3b responses) compared to earlier P3 subcomponents, whereas the opposite was true for Late timing deviants, where larger P3a responses were observed. Jongsma et al. (2007) explain these results with reference to an oscillator model of attention, suggesting that the difference in the patterns observed is indicative of the role attention plays in a given sequence: the early timing deviant elicits a ‘surprise’ response, whereas the late timing deviant elicits an ‘anticipation’ response. In the current experiment, results were analysed in reference to one time window for the P3a only, so a comparison was not made to see if any P3b responses were elicited to the early timing deviants. However, it is worth noting that the P3b response is generally elicited when attention is actively engaged and participants are geared to respond to the stimuli they are attending to (Polich, 2007). Given the nature of the task in the reported experiment where participants were instructed not to attend to the auditory stimuli, we do not expect any P3b responses to be elicited.

The effect of speechlikeness on acoustic processing

The effect of Speechlikeness on MMN amplitude

Another aspect of our research question addressed the role of increasing acoustic and linguistic complexity on the neural responses observed to our timing and frequency deviants. While we did not make specific predictions about how the responses would differ between levels of speechlikeness, if at all, previous literature at large has reported bigger MMN responses elicited to speech stimuli when compared to non-speech (and non-simple tone) stimuli (e.g., Sorokin et al., 2010). Our results did not replicate this trend. Instead they corroborated the results of Wunderlich & Cone-Wesson (2001) and Pettigrew et al. (2005). Both of these studies found small to no MMNs elicited to carefully controlled speech stimuli.

In our experiment, overall, Complex wave stimuli elicited larger MMN responses than syllable stimuli. This was particularly evident in the responses observed to the Frequency deviant, with a large negative response elicited for the Complex waves, and a comparatively smaller negativity observed for the Syllables. Responses to the timing deviants preserved this pattern.

Considering the overall pattern of MMN responses observed, there are three possible explanations for the fact that the syllables elicited responses that were smaller in amplitude compared to that of the complex waves. The first is that this may be due to the use of synthesised syllables in our experiment. Work by Wunderlich & Cone-Wesson (2001) and Pettigrew et al., (2005) has suggested that MMNs to synthesised or semi-synthesised CV tokens are hard to observe. However, this finding contradicts the majority of literature that has observed MMNs to CV syllables (e.g., Honbolygó et al., 2017) and more complex linguistic stimuli such as words (e.g., Korpilahti et al., 2001). The issue, perhaps, may lie with synthesised nature of the syllable itself, which has been discussed by Pettigrew et al. (2005) as a possible factor in contributing to the smaller responses. Although the technology used to synthesise the syllables in our experiment resulted in clearly discernible CV-tokens, it is possible that the controlled nature of their construction led to the lack of some natural speech variation that is helpful in eliciting MMNs to speech tokens. However, if the Syllables were not perceived as sufficiently speechlike, then the responses elicited to them should have been similar to those observed to Complex waves. This was not the case; the Complex wave and Syllable MMNs patterned differently enough for there to be a significant interaction between Speechlikeness and both Deviant Type and ROI. Additionally, we would not expect the nature of the tokens to have an impact on the responses in terms of the timing deviants, as our study design did not rely on a change in any acoustic or phonetic property of the syllables but specifically rather on the timing of the CV syllables. Yet, the timing deviants showed a diminished MMN as well, with similar response patterns for both Complex waves and Syllables. Therefore, although the synthesised nature of the stimuli may possibly explain the difference in responses, and has been cited as a reason by Pettigrew et al., (2005) in their discussion, we do not expect this to explain our results completely. Interestingly, Pettigrew et al., (2005) also suggest a habituation effect due to the repetitive nature of a single syllable standard and deviant pair being presented in one testing session leading to smaller MMN responses, and motivating a subsequent study that used a multi-feature paradigm to circumvent this issue. In our experiment, the temporal variation and natural speech variation already led to the inclusion of a more diverse stream of syllables, reducing the possibility of any habituation effects on the results.

A second possible explanation for the reduced speech responses relates to the experimental design. In order to model natural speech variation, three different syllables were presented to form the stream of standards. This variability among standard tokens was maintained for the timing deviants, whereas a fourth syllable with a different onset and vowel was introduced as the frequency deviant. Previous work has robustly found that MMNs can be, and are, elicited to abstract-feature stimuli, with standards being identified as sharing a common acoustic feature (e.g., Peter et al., 2012; see Näätänen et al., 2007, for a review). However, it is possible that in our experiment, the inherent variation within the standards resulted in weak predictability for the deviant due to the shallow contrast in the ratio of standards to deviants. The MMN is a response that relies on the difference between the frequent standard and the rare deviant. "…[W]hen a deviant stimulus deviates from the standard in two or several attributes, then the MMN amplitude shows additivity" (Näätänen et al., 2007, p. 2558). As the stimuli in this experiment were comprised of three different syllables presented at varying SOAs, the proximity to modelling variation in speech could have resulted in a less noticeable difference between the standards and the deviants. In terms of the predictive coding framework, this means that it is possible that the amount of variation present in the pseudo-random stream of syllables would have accommodated a greater number of predictions for what the deviant could be, meaning that any deviant would elicit a small error response because it would not, by most accounts, be an error. The difference in features between the standards and deviants is the critical piece of information that causes the elicitation of an MMN (either as an error response or a memory trace disruption); if it is weighted as less important because of its links to language, it might explain the shallower amplitudes observed to syllables compared to complex waves. Arguably, multi-feature paradigms should therefore also show similar results, with increasing number of deviants resulting in smaller elicited MMNs. However, this has reliably not been found to be the case, and clearer MMNs have been recorded to a paradigm with five deviants, for example, compared to three (Fisher et al., 2011). The critical difference between multi-feature paradigms of this nature and our experiment lies in the ratio of standards to deviants. When an increasing number of deviants are introduced in a paradigm with a fixed number of standards, the overall probability of encountering one of the deviants goes down, meaning the ratio of standards to deviants goes up. This has been reliably found to elicit larger MMNs, as MMN elicitation relies on the deviant being presented less frequently than the standard. However, in our experiment, variation was introduced for the standards as well as the timing deviants. The presentation of three different syllables for the standards (as well as deviants) could have led to the perception of a shallower ratio of standards to deviants, resulting in the elicitation of a smaller MMN. This variation was introduced in order to encourage an association to speech during stimuli presentation. Therefore, under the predictive coding framework, prior experience with hearing varying patterns of syllables could weight more heavily due to a language familiarity effect, leading to a smaller 'error' response generation. This would not be true for a similar sequence within a multi-feature paradigm, which would predominantly feature one syllable (standard) and rarely feature others (deviants).

As a third possibility, the nature of the stimuli themselves may explain the difference in the responses observed. Participants heard acoustically novel Complex waves and familiar Syllables within the experiment. It is possible that the syllables, in particular, were processed linguistically, rather than auditorily, meaning that any deviant present in the stream was accounted for as natural language variation, and therefore elicited only a small error response. As the predictive coding framework assumes an internal set of priors that affect the predictions made, it is reasonable to assume that the variation in language—both phonemic and temporal—would allow a more flexible criterion for elucidating what token counts as being different from the stream of standards enough to elicit an error response. Therefore, shallower responses would have been generated for the syllables than for the complex waves because the criteria of acceptability would have been wider. Additionally, the syllable stimuli were synthesised using vocal tract characteristics that introduced some co-articulatory effects in the onset consonant as the vowel was changed. This effect on the onset consonant may have impacted the perception of the sound, leading to a weaker abstraction and possibly processing of the standard syllables as if they were deviants (like the syllable frequency deviant presented).

It is also possible that enhanced MMNs are observed to speech when one or more linguistic aspects are engaged beyond phonology. For example, many studies that reported enhanced speech MMNs (e.g., Shtyrov, & Pulvermüller, 2002; Sorokin et al., 2010) made use of words and pseudowords in their stimuli. Engaging semantic processing may result in mismatches (and ergo, expectancy violations) that output a larger 'error' response and, consequently, generate a bigger MMN compared to non-linguistic stimuli without these associations. The syllables used in our experiment contained no semantic information whatsoever. However, other work with similar results as those found here did make use of words and non-words that could engage semantic processing, making this explanation unlikely (e.g., Wunderlich & Cone-Wesson, 2001).

This discussion has so far assumed that enhanced MMNs elicited to speech is the more reliable finding. However, our opposite results provide support to the possibility that speech stimuli do *not* elicit enhanced MMN responses. Indeed, considering a predictive framework as an explanation, it makes sense that speech stimuli would be less likely to trigger 'error' responses and elicit MMNs compared to non-speech stimuli.

The effect of Speechlikeness on MMN latency

There were no MMN latency effects observed for any of the deviant types across the two levels of Speechlikeness in the pre-defined time window of 100 ms to 300 ms. However, a smaller negativity was observed between 400 ms and 500 ms, visually apparent for Complex wave and Syllable responses, and elicited to all of the deviants to some degree. Previous work has reported the presence of a 'late MMN', also interchangeably known as the *late discriminative negativity (LDN)*. This response has been characterised as a marker of lexical processing (Korpilahti et al., 2001), although some studies have referred to it as a reflection of pattern learning (Zachau et al., 2005), or, perhaps, as a marker for attention reorientation (Escera et al., 2000). Therefore, in our experiment, we would expect an LDN to be most evident for responses to the syllables, and particularly so to the timing deviants as they rely the most on the tacit rule-learning that the LDN has been hypothesised to indicate. This was not found, however. Although this response was not pre-defined or analysed, the presence of an LDN-like negativity for both levels of Speechlikeness (and for all deviants, including the non-timing frequency deviants) suggests that in our experiment, at least, this negativity might be indicating a reorientation of attention. This is further supported by the fact that the P3a (discussed in the next section) was elicited for all conditions across the three deviant types and two levels of Speechlikeness. As the P3a marks an engagement of attention at some point during the stimulus presentation, it makes sense for a reorientation of attention response (the LDN) to follow.

Speechlikeness effects on P3a amplitude

The P3a is often observed alongside the MMN as an ‘attention-orienting’, or attention-capturing response, suggesting that participants begin attending to the stimuli at some point during the course of the experiment. Overall, however, no statistically significant effect of Speechlikeness on P3a elicitation could be detected, suggesting that attention-engagement may not depend on the linguistic familiarity of the stimuli presented. No significant differences in P3a amplitude between Deviant Types were observed either.

The MMN functions as an error response within the predictive coding framework, elicited to the stimuli when they differ from expectations. The magnitude of the MMN corresponds directly to the degree of error. Ylinen et al. (2016) suggest that a P3a response following an MMN response might be indicative of increased resources allocated to processing the unexpected deviant, with attention being consciously diverted to it. All deviants in our experiment elicited a P3a response that was significantly different from baseline (0 μV), suggesting perhaps the allocation of extra resources in processing them. No differences were observed between Deviant Types so no conclusions can be drawn about the degree to which each deviant might be more attentionally demanding than another. However, considering the complex acoustic nature of the Complex wave and Syllable stimuli, the elicited P3a response for all Deviant Types may be reflecting the increased auditory processing demands of a non-simple stimulus.

Attention, temporal variation, and speechlikeness

The MMN has been described as a pre-attentive response, and reflects the auditory processes active without a conscious task. This chapter aimed to address the question of how temporal variation in increasingly-linguistically complex stimuli are processed with the purpose of examining how linguistic familiarity interacts with pre-attentive auditory processing of unexpected temporal variations.

There are two main findings in the pattern of results observed. The first is that, overall, unexpectedly early timing deviants resulted in a larger MMN response compared to late timing deviants, regardless of linguistic familiarity. This suggests a conclusion that the internal auditory framework assigns a larger error response associated to a stimulus encountered at a time point before it is expected compared to one that results in prolonged anticipation before it is available for processing. Thus, the seeming acceptability of the late timing deviant (the smaller response) could be due to the very fact that it is anticipated—and therefore, predictable—within the context of the overall auditory stream. Another explanation depends on experience in extra-experimental contexts that the late deviant is not considered an error as pauses and delays are common.

The second conclusion is that increased speechlikeness elicits smaller responses compared to Complex Wave stimuli. A general trend in previous literature, as discussed in the Introduction, has found enhanced MMNs elicited to speech stimuli (e.g., Korpilahti et al., 2001). Our results seem to suggest the contrary, aligning instead with those reported by Wunderlich & Cone-Wesson (2001) and Pettigrew et al. (2005). Overall, we found that increased speechlikeness elicits smaller MMNs, with larger responses observed to Complex Waves than to Syllables.

N1-P2

While present, the N1-P2 responses elicited to the Complex waves and Syllables are visually less defined than expected. This is due to the almost concurrent elicitation of the MMN, resulting in a N100-MMN complex. Previous literature has sometimes referred to this as *deviant-related negativity*, or DRN (e.g., Duda-Milloy et al., 2019). In our experiment, this composite MMN + N100 response is more apparent for the Syllable condition than it is for the Complex wave condition, where the MMN is distinct. The MMN is also more distinct for the frequency deviant than it is for the timing deviants in both experiments. This result is unexpected given previous research, which has found that the N1 and MMN are temporally closer together for non-verbal stimuli, compared to verbal stimuli (Brückmann & Garcia, 2020). Although analyses on the latency difference between the N1 and MMN were beyond the scope of the current project, a visual inspection shows a result that is almost the opposite, with greater N1-MMN overlap for the more verbal stimuli than the non-verbal tones. This difference could, in part, be related to the temporal differences of the timing deviants themselves.

One other possible reason for this difference might be the length of the stimuli. Tervaniemi (1999) suggests that the more complex the stimuli, the longer the stimuli need to be to elicit an MMN, i.e., to track the differences between standards and deviants enough for there to be a response when they do not match. Tervaniemi (1999) suggest the stimuli to be at least 500 ms in length for more acoustically complex stimuli. The stimuli used in our experiments were 150 ms in length for the complex waves and syllables. Perhaps the acoustic and linguistic complexity of the stimuli interacted with the length, leading to insufficient inputs to generate the expected output (i.e., the MMN). This does not explain the difference in the responses observed by timing deviant type, however. While both timing deviants elicited an MMN, the response to the early timing deviant was larger in amplitude compared to the late. If, indeed, longer stimuli are required for more complex stimuli, then one might posit that a longer SOA would be more beneficial as well, as it would give sufficient time for information to be encoded and processed to generate the MMN. This was not found to be the case for our stimuli, where the early deviant, presented at an SOA 200 ms shorter than that of the standard, elicited a larger MMN compared to the late timing deviant, which was presented at an SOA that was 200 ms longer than the standard SOA.

Limitations

Due to the COVID-19 pandemic and testing time restrictions imposed as a result, the sample size for this experiment was reduced from the original goal of 30, to 20. With artifact rejection and loss of data, the resulting sample size was reduced further to 18. Although an MMN response has reliably been evoked with sample sizes smaller than this (e.g., n = 10 tested by Winkler & Näätänen, 1992), a greater number of participants would have added power to our study and provided scope for further analyses that explored individual differences between participants. Additionally, data recorded during the pandemic were observationally found to be noisier than those recorded pre-pandemic, possibly due to the new COVID-19 testing protocols which necessitated masks and plastic covers on the testing chair. Participants who did not remove their masks during the session may have been in more discomfort, and the plastic covers on the testing chair, used for easy disinfecting, may have increased that discomfort. Additionally, it is possible that participants were nervous about the close contact of equipment during the pandemic, although maximum care was taken to disinfect it. That said, the individual responses recorded across participants did not differ between those obtained pre-pandemic. The difference lay mostly in the volume of data that had to be rejected during pre-processing.

The other main limitation for this study was the lack of consistency between the experimental designs for the simple tones (Chapter 2, and complex tones and syllables (current Chapter 3), which made it difficult to perform any direct comparative analyses. The reason for the changes in experimental design was to 1. improve the stimulus stream so it modelled natural language variation more accurately, and 2. adjust the SOA so that there was adequate time to process the more complex stimuli. Additionally, participants who heard the two types of complex stimuli were tested within one testing session, whereas the participants who heard the simple tone stimuli were tested in another. This limited our ability to draw direct conclusions from differences in the responses recorded to simple tones, and those recorded to complex waves and syllables. One long session for all three types of stimuli was not feasible; however, a single testing session would have provided us with the opportunity to more closely follow the growth and decline of the MMN and P3a within an individual in response to temporal deviants, as stimuli increased in speechlikeness (from simple tones, to syllables).

Finally, although the participants tested were predominantly English dominant speakers, future work testing Canadian English monolinguals or Canadian English dominant speakers would better eliminate any effect of bilingualism and foreign language knowledge that may have influenced the results in our current study. The aim of constructing stimuli to test from a wider language background was to facilitate data collection considering the participant pool available (which was largely non-monolingual). However, stricter language questionnaires would help parse out effects of language background on the responses observed (if any).

Conclusion

The purpose of the current study was to explore how temporal variation in stimulus trains of varying speechlikeness is processed. MMN responses to non-linguistic auditory stimuli (simple sine and complex waves) and linguistic auditory stimuli (syllables) were compared to examine if, and how, the MMN changes when resemblance to linguistic elements is increased. Results did not find any robust differences between the two timing deviants. However, there was a significant effect of speechlikeness, with deviants in the sequence of Complex waves (non-speech) eliciting larger responses than deviants among the Syllables (modelling speech). These results suggest that the tempo interruptions are more costly in terms of processing for unfamiliar complex stimuli compared to acoustically matched stimuli that resemble familiar syllables.

This chapter considers the ramifications of unexpected timing variations on pre-attentive attention, drawing conclusions about how external patterns of time align with the internal predicted representations of it. In the next chapter, the effect of timing on verbal short-term memory is considered as participants hear sequence of paired syllables with temporal variations and make selections at the end as to which target pair was present in the auditory sequence. The aim is to consider how unexpected temporal variations impact other aspects of cognitive processing such as memory, and what that might mean for how speech is perceived.

References

Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2004). Aging: a switch from automatic to controlled processing of sounds? *Psychology and Aging*, *19*(1), 125.

Alonso-Búa, B., Díaz, F., & Ferraces, M. J. (2006). The contribution of AERPs (MMN and LDN) to studying temporal vs. linguistic processing deficits in children with reading difficulties. *International Journal of Psychophysiology, 59*(2), 159-167.

Barr, R. E., Ackmann, J. J., & Sonnenfeld, J. (1978). Peak-detection algorithm for EEG analysis. *International Journal of Bio-Medical Computing, 9(*6), 465-476.

Birkholz, P., Drechsel, S., & Stone, S. (2019). Perceptual optimization of an enhanced geometric vocal fold model for articulatory speech synthesis. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2019-Septe*, 3765–3769. https://doi.org/10.21437/Interspeech.2019-2410

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program].: Vol. Version 6.

Boh, B., Herholz, S. C., Lappe, C., & Pantev, C. (2011). Processing of complex auditory patterns in musicians and nonmusicians. *PLoS ONE*, *6*(7), 21458. https://doi.org/10.1371/journal.pone.0021458

Brückmann, M., & Garcia, M. V. (2020). Mismatch negativity elicited by verbal and nonverbal stimuli: Comparison with potential N1. *International Archives of Otorhinolaryngology*, *24*(2), E80–E85. https://doi.org/10.1055/s-0039-1696701

Čeponienė, R., Alku, P., Westerfield, M., Torki, M., & Townsend, J. (2005). ERPs differentiate syllable and nonphonetic sound processing in children and adults. *Psychophysiology, 42*(4), 391-406.

Čeponienė, R., Service, E., Kurjenluoma, S., Cheour, M., & Näätänen, R. (1999). Children's performance on pseudoword repetition depends on auditory trace quality: evidence from event-related potentials. *Developmental Psychology*, *35*(3), 709-720.

Cheour, M., Kushnerenko, E., Čeponienė, R., Fellman, V., & Näätänen, R. (2002). Electric brain responses obtained from newborn infants to changes in duration in complex harmonic tones. *Developmental Neuropsychology*, *22*(2), 471–479. https://doi.org/10.1207/S15326942DN2202_3

Choi, J. (2003). Pause length and speech rate as durational cues for prosody markers. *The Journal of the Acoustical Society of America, 114*(4), 2395–2395. https://doi.org/10.5840/raven20111838

Cole, J. (2015). Prosody in context: A review. Language, *Cognition and Neuroscience, 30*(1-2), 1-31.

Dellwo, V. (2008). Influences of language typical speech rate on the perception of speech rhythm. *The Journal of the Acoustical Society of America*, *123*(5), 3427–3427. https://doi.org/10.1121/1.2934192

Dellwo, V., & Wagner, P. (2003). Relationships between rhythm and speech rate. *International Congress of Phonetic Sciences*, 471–474.

Duda-Milloy, V., Tavakoli, P., Campbell, K., Benoit, D. L., & Koravand, A. (2019). A time-efficient multi-deviant paradigm to determine the effects of gap duration on the mismatch negativity. *Hearing Research*, *377*, 34–43. https://doi.org/10.1016/j.heares.2019.03.004

Endrass, T., Mohr, B., & Pulvermüller, F. (2004). Enhanced mismatch negativity brain response after binaural word presentation. *European Journal of Neuroscience*, *19*(6), 1653-1660.

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro-Otology*, *5*(3–4), 151–166. https://doi.org/10.1159/000013877

Fong, C. Y., Law, W. H. C., Uka, T., & Koike, S. (2020). Auditory Mismatch Negativity Under Predictive Coding Framework and Its Role in Psychotic Disorders. *Frontiers in Psychiatry*, *11*, 1–14. https://doi.org/10.3389/fpsyt.2020.557932

Gervain, J. (2018). Gateway to language: The perception of prosody at birth. In *Boundaries crossed, at the interfaces of morphosyntax, phonology, pragmatics and semantics* (pp. 373-384). Springer, Cham.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*(1982), 515-546.

Hämäläinen, J. A., Salminen, H. K., & Leppänen, P. H. (2013). Basic auditory processing deficits in dyslexia: systematic review of the behavioral and event-related potential/field evidence. *Journal of Learning Disabilities*, *46*(5), 413-427.

Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology*, *87*(2), 165–172. https://doi.org/10.1016/j.ijpsycho.2012.12.005

Honbolygó, F., Csépe, V., & Ragó, A. (2004). Suprasegmental speech cues are automatically processed by the human brain: A mismatch negativity study. *Neuroscience Letters*, *363*(1), 84–88. https://doi.org/10.1016/j.neulet.2004.03.057

Honbolygó, F., Kolozsvári, O., & Csépe, V. (2017). Processing of word stress related acoustic information: A multi-feature MMN study. *International Journal of Psychophysiology*, *118*, 9–17. https://doi.org/10.1016/J.IJPSYCHO.2017.05.009

Ilvonen, T., Kujala, T., Kozou, H., Kiesiläinen, A., Salonen, O., Alku, P., & Näätänen, R. (2004). The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). *Neuroscience Letters*, *366*, 235–240. https://doi.org/10.1016/j.neulet.2004.05.024

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M., & Alho, K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cognitive Brain Research*, *12*(3), 459–466. https://doi.org/10.1016/S0926-6410(01)00081-7

Jasper, H.H. (1958) The Ten-Twenty Electrode System of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371-375.

Jongsma, M. L. A., Meeuwissen, E., Vos, P. G., & Maes, R. (2007). Rhythm perception: Speeding up or slowing down affects different subcomponents of the ERP P3 complex. *Biological Psychology*, *75*(3), 219–228. https://doi.org/10.1016/J.BIOPSYCHO.2007.02.003

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0.

Kisley, M. A., Davalos, D. B., Layton, H. S., Pratt, D., Ellis, J. K., & Seger, C. A. (2004). Small changes in temporal deviance modulate mismatch negativity amplitude in humans. *Neuroscience Letters*, *358*(3), 197–200. https://doi.org/10.1016/j.neulet.2004.01.042

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. H. (2001). Early and Late Mismatch Negativity Elicited by Words and Speech-Like Stimuli in Children. *Brain and Language*, *76*(3), 332–339. https://doi.org/10.1006/brln.2000.2426

Kuuluvainen, S., Nevalainen, P., Sorokin, A., Mittag, M., Partanen, E., Putkinen, V., Seppänen, M., Kähkönen, S., & Kujala, T. (2014). The neural basis of sublexical speech and corresponding nonspeech processing: A combined EEG-MEG study. *Brain and Language, 130*, 19–32. https://doi.org/10.1016/j.bandl.2014.01.008

Lai, Y., Tian, Y., & Yao, D. (2011). MMN evidence for asymmetry in detection of IOI shortening and lengthening at behavioral indifference tempo. *Brain Research*, *1367*, 170-180.

Langus, A., Mehler, J., & Nespor, M. (2017). Rhythm in language acquisition. *Neuroscience & Biobehavioral Reviews, 81*, 158-166.

Maddieson, I., & Disner, S. F. (1984). Patterns of sounds. Cambridge University Press.

May, P. J., & Tiitinen, H. (2010). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*, *47*(1), 66-122.

Näätänen, R., Kujala, T., & Light, G. (2019). Mismatch negativity: a window to the brain. Oxford University Press.

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. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, *24*(5), 283-288.

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988a). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology*, *25*(3), 262–268. https://doi.org/10.1111/j.1469-8986.1988.tb01239.x

Nordby, H., Roth, W. T., & Pfefferbaum, A. (1988b). Event-related potentials to time-deviant and pitch-deviant tones. *Psychophysiology*, *25*(3), 249–261. https://doi.org/10.1111/j.1469-8986.1988.tb01238.x

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Peter, V., McArthur, G., & Thompson, W. F. (2012). Discrimination of stress in speech and music: a mismatch negativity (MMN) study. *Psychophysiology*, *49*(12), 1590-1600.

Pettigrew, C. M., Murdoch, B. E., Kei, J., Ponton, C. W., Alku, P., & Chenery, H. J. (2005). The mismatch negativity (MMN) response to complex tones' and spoken words in individuals with aphasia. *Aphasiology*, *19*(2), 131–163. https://doi.org/10.1080/02687030444000642

Pettigrew, C. M., Murdoch, B. E., Chenery, H. J., & Kei, J. (2004). The relationship between the mismatch negativity (MMN) and psycholinguistic models of spoken word processing. *Aphasiology*, *18*(1), 3–28. https://doi.org/10.1080/02687030344000463

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Finnigan, S., Alku, P., Kei, J., Sockalingam, R., & Chenery, H. J. (2004). Automatic Auditory Processing of English Words as Indexed by the Mismatch Negativity, Using a Multiple Deviant Paradigm. *Ear and Hearing*, *25*(3), 284–301. https://doi.org/10.1097/01.AUD.0000130800.88987.03

Pettigrew, C. M., Murdoch, B. M., Kei, J., Chenery, H. J., Sockalingam, R., Ponton, C. W., Finnigan, S., & Alku, P. (2004). Processing of English words with fine acoustic contrasts and simple tones: A mismatch negativity study. *Journal of the American Academy of Audiology*, *15*(1), 47–66. https://doi.org/10.3766/jaaa.15.1.6

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., & Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, *12*(6), 1038–1055. https://doi.org/10.1162/08989290051137567

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0089

Polich, J. (2020). 50+ years of P300: Where are we now? *Psychophysiology*, *57*(7). https://doi.org/10.1111/psyp.13616

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R. J., & Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage*, *14*(3), 607–616. https://doi.org/10.1006/NIMG.2001.0864

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in neurobiology*, *79*(1), 49-71.

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

R Studio Team (2021). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, *105*(1), 512–521. https://doi.org/10.1121/1.424522

Rosburg, T., Trautner, P., Korzyukov, O. A., Boutros, N. N., Schaller, C., Elger, C. E., & Kurthen, M. (2004). Short-term habituation of the intracranially recorded auditory evoked potentials P50 and N100. *Neuroscience letters*, *372*(3), 245-249.

Saarinen, J., Paavilainen, P., Schöger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport*, *3*(12), 1149-1151.

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

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (1998). Auditory processing and dyslexia: evidence for a specific speech processing deficit. *NeuroReport*, *9*(2), 337-340.

Sebastian, C., & Yasin, I. (2008). Speech versus tone processing in compensated dyslexia: Discrimination and lateralization with a dichotic mismatch negativity (MMN) paradigm. *International Journal of Psychophysiology*, *70*, 115–126. https://doi.org/10.1016/j.ijpsycho.2008.08.004

Shestakova, A., Huotilainen, M., Čeponiene, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology, 114*(8), 1507–1512. https://doi.org/10.1016/S1388-2457(03)00134-2

Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *NeuroReport*, *13*(4), 521–525. https://doi.org/10.1097/00001756-200203250-00033

Sorokin, A., Alku, P., & Kujala, T. (2010). Change and novelty detection in speech and non-speech sound streams. *Brain research*, *1327*, 77-90.

Stephenson, A. (2021), tayloRswift: Colour Palettes Generated by Taylor Swift Albums. R package version 0.1.0. https://github.com/asteves/tayloRswift

Tavakoli, P., Duda, V., Boafo, A., & Campbell, K. (2021). The effects of sleep on objective measures of gap detection using a time-efficient multi-deviant paradigm. *Brain and Cognition*, *152*, 105772.

Tervaniemi, M. (1999). Pre-Attentive processing of musical information in the human brain. *Journal of New Music Research*, *28*(3), 237–245. https://doi.org/10.1076/jnmr.28.3.237.3109

Tervaniemi, M., Ilvonen, T., Sinkkonen, J., Kujala, A., Alho, K., Huotilainen, M., & Näätänen, R. (2000). Harmonic partials facilitate pitch discrimination in humans: electrophysiological and behavioral evidence. *Neuroscience Letters*, *279*, 29–32.

Tervaniemi, M., Schröger, E., & Näätänen, R. (1997). Pre-attentive processing of spectrally complex sounds with asynchronous onsets: An event-related potential study with human subjects. *Neuroscience Letters*, *227*(3), 197–200. https://doi.org/10.1016/S0304-3940(97)00346-7

Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J., & Näätänen, R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Cognitive Brain Research, 8*(3), 355-363.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, A. B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79–90. https://doi.org/10.1097/00003446-200104000-00001

Tseng, C. Y., & Fu, B. L. (2005). Duration, intensity and pause predictions in relation to prosody organization. In *Ninth European Conference on Speech Communication and Technology*.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York

Wickham, Hadley, François, Romain, Henry, Lionel, & Müller, Kirill (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.6. https://CRAN.R-project.org/package=dplyr

Winkler, I., & Näätänen, R. (1992). Event-related potentials in auditory backward recognition masking: A new way to study the neurophysiological basis of sensory memory in humans. *Neuroscience Letters*, *140*(2), 239–242. https://doi.org/10.1016/0304-3940(92)90111-J

Woods, D. L., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *65*(6), 447-459.

Wunderlich, J. L., & Cone-Wesson, B. K. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *The Journal of the Acoustical Society of America*, *109*(4), 1526–1537. https://doi.org/10.1121/1.1349184

Ylinen, S., Bosseler, A., Junttila, K., & Huotilainen, M. (2017). Predictive coding accelerates word recognition and learning in the early stages of language development. *Developmental Science*, *20*(6). https://doi.org/10.1111/desc.12472

Ylinen, S., Huuskonen, M., Mikkola, K., Saure, E., Sinkkonen, T., & Paavilainen, P. (2016). Predictive coding of phonological rules in auditory cortex: A mismatch negativity study. *Brain and Language*, *162*, 72–80. https://doi.org/10.1016/j.bandl.2016.08.007

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: Early and late mismatch negativity to tone patterns. *NeuroReport*, *16*(18), 2015–2019. https://doi.org/10.1097/00001756-200512190-00009

Zevin, J. D., Datta, H., Maurer, U., Rosania, K. A., & McCandliss, B. D. (2010). Native Language Experience Influences the Topography of the Mismatch Negativity to Speech. *Frontiers in Human Neuroscience*, *4*, 212. https://doi.org/10.3389/fnhum.2010.00212

CHAPTER FOUR

Introduction

The expectation of prosodic units in languages to be temporally equidistant is the basis of most of the work conducted on language rhythm processing. Isochrony implies a sense of predictability, and evenly paced incoming speech is therefore processed more easily—and faster—compared to an auditory input that is not predictable. However, while speech may be perceived to be relatively regular, real speech fails to meet the assumption of isochrony.

Timing itself is a nebulous concept in the wide array of literature that has tackled it. Depending on context, its definition may change (cf. Gibbon, 2018). In some areas of literature, variations in language timing refer to the obvious changes in tempo, or speech rate. In other areas, the timing variations refer to changes in the duration of individual segments or syllables that make up a speech utterance. Further, combining both aspects, there exists work where variations in timing in language refer to changes in timing between phonemes and syllables; a case that contains both changes in tempo and syllable durations but does not consistently belong to either. This is the focus of the current chapter, which aims to understand how unexpected timing perturbations in an otherwise predictable sequence of speech affects its perception. The next section provides a summary of timing in speech comprehension.

Timing in language

Timing variation as changes in speech rate (tempo)

One of the most intuitive interpretations of timing in language concerns the rate at which it is spoken. Speech rate can be quantified in terms of syllable rate or phonemic segment rate. However, in cases where the two conflict—for example, constant syllable rate with variable segment rate due to syllable complexity—listeners' perceived tempo aligns with the syllable rate (Plug & Smith, 2018). Syllables in speech, therefore, seem to have an implicit processing preference and importance compared to other sub-lexical units. Speech rate manipulations altering the rate of presentation of syllables could function well as a proxy for natural speech. The experiment reported in this chapter makes use of this by presenting participants with syllable sequences with timing variations.

Dilly and McAuley (2008) studied whether changes in speech rate modulate the perception of words or syllables. Distal (non-local) prosody was manipulated by changing duration and pitch cues earlier in the sentence to study effects on the perception of the final words in the sentence. Each sentence heard by participants ended with three syllables that formed two ambiguous words (e.g., *notebook worm* vs. *note bookworm*). The version that was heard depended on the distal prosody, with one prosodic profile favouring a monosyllabic interpretation of the sentence-final word (*worm*), and another favouring a disyllabic interpretation (*bookworm*). Over a series of experiments, the researchers showed that distal prosody influences perception of speech by affecting the perceived rate of speech, even when the sentences used identical words differing in prosodic elements only. Similar results were found in another study where speech rate manipulations resulted in participants hearing function words in sentences where there were none (Dilley & Pitt, 2010). Speech rate, therefore, plays a key role in word segmentation and comprehension. These aspects of language processing are intrinsically tied to the timing of speech. An important role for speech rate is found in memory as well, with rates of recall and rehearsal both associated with verbal memory span (Cowan, 1994; Cowan et al., 1998).

Timing variation as changes in phoneme/syllable duration

The role of segmental (phonemic) length has been examined in terms of speech comprehension and language rhythm class. So-called s*yllable-timed* and *stress-timed* languages were found to be distinguishable from each other based on various measures of vocalic and consonantal interval length (Dellwo & Wagner, 2003; Ramus et al., 1999). Similarly, variations in the duration of consonantal and vowel units were taken as a measure of varying speech rate and found to affect cross-language identification (Dellwo, 2008). Vocalic length, specifically, was found to vary the most between stressed and unstressed syllables in naturally spoken American English utterances (Greenberg et al., 2003). In addition to segmental duration, the duration of whole syllables was also evaluated as an important feature of speech rhythm and tempo. Language acquisition studies suggest that children produce more vowels than adults, but more slowly and with less variability compared to adults (Ordin & Polyanskaya, 2015; Polyanskaya & Ordin, 2015). The authors suggest that stress-timed speech, therefore, develops later in age compared to syllable-timed speech, which functions as a default (see also Ordin & Polyanskaya, 2014, 2015). In one study, both the timing and length of syllables was manipulated such that they were equal. Adjustments were made such that the temporal distances between syllables were of equal length (syllable-timed), or temporal distances between syllable *stresses* were of equal length (stress-timed). Sentences approaching syllable-timing were more intelligible to both speakers of stress- and syllable-timed languages compared to sentences that were isochronously presented (Aubanel & Schwartz, 2020). According to the authors, their results "…[indicate] that the temporal scale associated with the syllable is key—rather than its linguistic functional value." (Aubanel & Schwartz, 2020, p. 7). Indeed, a body of previous work has explored the role of syllable duration in speech perception and segmentation. Studies have found that adults prefer syllable duration as a marker for segmentation (Bion et al., 2010), and that these durational cues allow better categorisation across language rhythm classes and even individual languages (White et al., 2012). Syllable boundaries have also been shown to better help contextualise phonological processing compared to word boundaries (Hooper, 1972). Thus, the findings suggest that the syllable as a unit and syllable duration as a feature are both important in speech perception and segmentation.

Timing variation or the lack thereof (isochrony)

A sustained rate of speech (or specific type of rhythm) that is often used to present the stimuli in timing studies of speech perception is not naturally observed in speech. Rather, dynamic changes in durations between segments—observed as brief increases in speech rate, or as delays or pauses—are far more common. Temporal variations in speech deserve more research attention. The study reported in this chapter aims to address this limitation in the literature. However, previous work has often examined speech under the assumption of isochrony, and made use of these timing variations as an acoustic feature, or as a basis for language discrimination only.

The rhythm class hypothesis, as previously discussed in the Introduction chapter, posits that all the languages of the world fall into one of three rhythm classes: stress-timed, syllable-timed, or mora-timed. Most of the emphasis in work on rhythm and language has been put on the differences between stress- and syllable-timed languages. However, White et al. (2012) suggest that rhythm classes do not inherently exist as a categorical percept. Rather, languages contain a range of temporal cues that are used by listeners to differentiate between them and categorise them into broad rhythm classes. The isochrony assumption at the core of the rhythm class hypothesis has empirically been disproven over the course of the years, with evidence showing the variability in language (Cummins, 2012; see Grabe & Low, 2008, for a discussion). For instance, Gibbon and Gut (2001) attempted to classify Ibibio (syllable-timed) and British as well as dialects of Nigerian English (stress-timed) into their respective rhythm classes using a modified version of what they deemed to be an objective measure of rhythm, the Pairwise Variability Index (PVI). The PVI is an index that attempts to provide an objective measure for gauging the amount of variability in an utterance. This index is computed by taking measurements of vocalic and intervocalic durations and entering them into a pre-determined formula (Low & Grabe, 1995). This formula also allows for rate-normalisation. The PVI has been used extensively as a measure of speech rhythm and vocalic intervals (He & Dellwo, 2016; Gibbon & Gut, 2001; Grabe & Low, 2008; Ordin & Polyanskaya, 2014). Gibbon and Gut (2001) modified the PVI to a value between 0 and 100, scaling the index for more intuitive interpretation. While the results showed differences between the two languages in terms of syllable durations, suggesting corroboration of the rhythm class hypothesis, no statistical analyses were conducted to validate the results. Therefore, while the study provides some interesting comparisons in terms of syllable durations of spoken words between languages and dialects belonging to different rhythm classes (stress vs. syllable-timed), it failed to provide concrete evidence for the rhythm class hypothesis. In fact, if we can take one conclusion from this paper, it might be that it is difficult to find *any* objective measure of rhythm which can be used to compare languages. That said, the theoretical framework of the rhythm class hypothesis allows for the descriptive grouping of the languages of the world, and while that may not be empirically rigorous, it provides a starting point for the investigation into rhythm in language.

A considerable amount of time and effort over the years has gone into identifying which prosodic cues are important for marking rhythm and time in language, as well as into determining how these cues can be used in language perception, processing, and discrimination (see White et al., 2012). Literature tends to agree on the importance of vocalic and consonantal durations as a measure of rhythm in language (Dellwo, 2008; Dellwo & Wagner, 2003; He & Dellwo, 2016; Gibbon & Gut, 2001; Greenberg et al., 2003; Plug & Smith, 2008; Ramus et al., 1999). As important as these cues are, they do not directly address how timing variation *between* segments (between consecutive syllables or vowel-consonant clusters) affects language processing.

Language processing using syllable sequences

The research on the mechanisms of perceptual chunking and language comprehension has heavily relied on sequences of syllables. This has been the methodology of choice to explore questions of comprehensibility, speech segmentation, and word perception. The studied syllable sequences have often been non-semantic (‘nonsense’) in nature (Daikoku et al., 2017; Rassili & Ordin, 2020), or formed the basis of an artificial language containing some inherent constructed predictability (Isbilen et al, 2020; Ordin & Nespor, 2013, 2016). Sometimes, these sequences of syllables were constructed such that they would form words in the participants’ native language (e.g., Ding et al., 2015), or in a foreign language unfamiliar to the participants (Rimmele et al., 2019). Within these studies, acoustic features (like pitch or durational information) were adjusted. In other studies, the predictability or learnability of the ‘words’ in the sequence was manipulated (through adjusting frequency of occurrence). By observing how well participants were able to identify recurring parts of the sequences presented to them (accuracy) and how fast (reaction time), researchers found evidence for perceptual boundaries of syllables and words (Isbilen et al., 2020; Ordin et al., 2020), preferences for isochrony in language (e.g. Rassili & Ordin, 2020) as well as preferences for native language rhythm over others (Ordin & Nespor, 2013, 2016; Rimmele et al., 2019).

The length of syllables has been found to be a reliable cue for speech segmentation (Ordin & Nespor, 2013; Ordin & Nespor, 2016; Ordin & Polyanskaya, 2014; Ordin & Polyanskaya, 2015; Polyanskaya & Ordin, 2015; Ordin & Nespor, 2016; Ordin et al., 2017). Using variability in vocalic and consonantal length as an indicator of the rhythmic class of a language, this research has explored the roles of native language on segmentation and distinguishing between stress- and syllable-timed rhythmic structures. Auditory sequences of syllables have also been widely used in the fields of statistical learning and neurophysiological research to understand the subconscious constraints of language processing.

Statistical learning paradigms have been most popular in language acquisition studies with infants but have also been employed with adults to understand language learning and individual differences (Erickson & Thiessen, 2015; Siegelman et al., 2017). By presenting carefully controlled pseudo-words in an artificial language, these paradigms allow researchers to study linguistic acquisition phenomena using 'words' consisting of syllable strings that contain no semantic or syntactic information (Aslin et al., 1998; Saffran, 2001, 2003; Saffran et al., 1996; Saffran, et al., 1999). In a typical statistical learning paradigm, there is first a 'learning phase', where participants listen to a stream of syllables at a steady rate. *Transitional probabilities (TP)*—the probability of one given syllable following another syllable present in the sequence—are typically manipulated in this auditory stream to make certain combinations more likely than others. The variation in transitional probabilities perceptually gives rise to word boundaries in the listeners' mind. Following the learning phase, participants undergo a 'testing phase', where they are asked to choose between two syllable strings to indicate the one they heard within the auditory stream (*two-alternative forced-choice task, 2AFC*). The probed target "words" are based on the manipulated transitional probabilities, and often, participants are asked to choose between pairs of words that are both possible in the given stream, with one being more frequent than the other. The idea is that the higher the transitional probabilities within certain groups of syllables, the higher the likelihood that they will be remembered as "words".

During the testing phase, participants typically perform above chance in picking out the target words predicted by the perceptual grouping mechanisms hypothesised by the researchers. Results of these experiments give insights into how participants implicitly group syllables into words and create perceptual boundaries. The disadvantages of a task such as the 2AFC are that participants rarely score above 60% accuracy, and that there is a high likelihood of participants guessing (chance-level results; that is, a 1 in 2 chance of getting the answer right, or 50% probability of getting the correct answer based on a guess), and that the data obtained obfuscate details about individual differences or learning patterns (Isbilen et al., 2020). Variations in the implementation of the 2AFC have been developed to circumvent these issues (see Isbilen et al., 2020, for a detailed discussion). The current experimental design uses a task that resembles aspects of statistical learning. However, the present task is a short-term memory (STM) rather than a statistical learning task. In the retrieval phase, a 2AFC recognition task is used. Participants hear a target 'word', two consecutive syllables present in the sequence they just heard, along with a non-target foil pair of syllables that were not present in the sequence.

During the stimuli presentation part of the trial, syllables are presented in a sequence isochronously, with the idea that no other cue than serial order will be provided to participants. However, an inherent, internally generated, rhythm, is created when random syllables are being repeated for an extended period of time in various permutations (cf., the tick-tock rhythm of a clock). This was observed during data collection for Experiment 3 (*Chapter 3)*, where a participant commented on the rhythmic nature of the syllable stream. While there is ample evidence that participants in statistical learning experiments group syllables according to the transitional probabilities built into the sequences they hear, little work has examined the role of timing deviants for STM of such sequences of syllables. As natural speech is so variable in rhythm, it is likely that the timing of these syllables in sequences (or perturbations introduced throughout the stream) could interact with the serial order cues and influence memory for embedded ordered syllable pairs.

In the current chapter, the role of timing variations between syllables is discussed in verbal memory processing. The next section briefly summarises the previous literature on timing in short-term verbal memory.

Timing in verbal short-term memory

Working memory (WM) is a three-part conceptualisation of a memory system that holds information for shorter periods of times, allowing for it to be retrieved and processed immediately (Baddeley, 1986; Baddeley, 1992a; Baddeley, 2010; Baddeley & Hitch, 1974). It comprises of the *central executive*, which maintains most of the information, the *visuospatial sketch pad*, that maintains visual imagery, and the *phonological loop*, often associated with rehearsal and repetition (Baddeley & Hitch, 1974). An addendum to this model was added in the form of the *episodic buffer*, which allows “temporary storage” of items, as well as cross-modal integration (Baddeley, 2000; 2010).

Verbal short-term memory is an aspect of WM that deals with the repetition and rehearsal of verbal items for short durations. While the terms ‘short-term memory’ and ‘working memory' are often used in literature distinctly, both capture similar mechanisms, and WM is more realistically seen as an attentional mechanism that uses short-term memory to function (see Cowan, 2008, for a discussion).

Articulatory rehearsal by way of the phonological loop is one of the widest used explanations for the maintenance of recall items in memory. This model (initially proposed by Baddeley & Hitch, 1974) proposes covert rehearsal of items as a mechanism for maintaining them within the memory trace for recall. As rehearsal stops, the items decay until they are forgotten, unless rehearsal begins again before that point. A critical assumption of this model is the separation between memory and attention, with both mechanisms occupying separate parts of the whole that makes up the Working Memory model. An alternative explanation for verbal item recall, on the other hand, posits a concurrent processing and retrieval mechanism by which memory decay is stalled by attention switching between the various processes (Barrouillet & Camos, 2004). The *Time-Based Resource-Sharing (TBRS)* model has been found to be a great descriptor of memory recall processes over other explanations such as the Towse & Hitch (1995) model of decay, which suggests decay that is proportional to the time between storage and retrieval, or the Interference model by Saito & Miyake (2004) that suggests decay to be a function of cognitive load and, therefore, limited attentional resources (Camos et al., 2009; Hudjetz & Oberauer, 2007; Saito & Miyake, 2004; Towse & Hitch, 1995).

Previous literature has found a long history of covert speech rate correlations to verbal memory spans that determine that faster articulations lead to bigger spans (Cowan, 1992; Cowan et al., 1992; Cowan et al., 1994; Cowan et al., 1998). Predominantly evaluated in children, these studies have provided evidence for the presence of the articulatory loop as first postulated in the Baddeley and Hitch (1974) model of Working Memory. Previous work has found facilitatory effects of pauses (longer time durations) between items during recall (Cowan, 1994; Cowan et al., 1998). However, limited work exists otherwise that examines the role of unexpected temporal variations in speech and their impact on recall. The current chapter aims to redress this gap and expand the current area of research on the effect of unpredictable temporal variations on real-time speech and verbal short term memory.

Timing and order in verbal short-term memory

There have been many models presented in order to account for the encoding of items in verbal STM. Although a detailed overview of these models is outside of the scope of this chapter, it is worth talking about one type of model specifically: context-based competitive queuing (CQ) models. CQ models in general posit a two-stage selection process for the item being recalled, whereby all items in memory are competing to be selected for recall. Context-based CQ models specifically suggest the presence of a context signal that changes with the items being encoded (see Hitch et al., 2022, for a review).

There are three types of context-based CQ models: *event-based*, *timebased*, or a hybrid. Event-based models such as the C-SOB (C Serial-Order-in-a-Box) posit that encoding of items is dependent on their features contingent upon their order of presentation (Lewandowsky & Farrell, 2008). On the other hand, time-based models such as OSCAR (Oscillator-Based Associative Recall model), or Burgess & Hitch (1999) suggest an intrinsic role of the time of presentation of the items (in general and relative to each other) in the way they are recalled.

Evidence has suggested that order and time information is stored separately in STM (Farrell & McLaughlin, 2007). More recently, Gorin (2020) found support for event-based models of recall over time-based models through the evaluation of verbal recall for sequences that were presented at regular or irregular rhythms through a series of experiments. No effect was found of the regularity (or lack thereof) on recall, suggesting that items were encoded based on their serial order and not their temporal proximity. More strongly, this supports claims by Lewandowsky & Farrell (2008) for example, who "…suggest that time-based forgetting is no longer viable as an explanatory construct." (p. 42).

The best fit, however, is a model that encompasses features of both event-based and time-based models of recall. The Bottom-Up Multiscale Population (BUMP) oscillator model by Hartley et al. (2016) is an example of a hybrid model that does just that.

BUMP oscillator model

In the brain, *oscillations* refer to rhythmic changes in neural activity that follow a rise-and-fall pattern akin to that of a sine wave. Neural oscillations can be used as clocks to encode the order and timing characteristics of an incoming auditory signal. In fact, individual 'oscillators' (i.e., neurons) can be said to respond to different aspects of incoming sensory percepts, thereby giving the neural system a way to encode specific features as well as broad patterns that are embedded within the incoming signal. The BUMP oscillator model describes a facet of this processing, namely limited to memory.

This model was posited to account for irregular or non-temporally grouped data (Hartley et al., 2016). Within this model, oscillators paired together resonate to a specific intrinsic rhythm and are off-set in phase by 90°. Therefore, these neurons would oscillate and *entrain* to (i.e., synchronise with) different levels of the incoming signal, thereby capturing different aspects of it. Hartley et al. (2016) showed that top-down effects did not modulate temporal grouping effects in recall, thereby giving evidence for the bottom-up nature of this model and its ability in being able to explain both regular and irregular temporal grouping effects. The authors also ran simulations that corroborated this result and suggested bottom-up processes were integral to temporal grouping.

The BUMP model has been influential in describing methods of memory encoding. Although it has been discussed in the context of speech perception as well, its primary supportive evidence has come from serial recall methodologies. Neural oscillations have long been used to describe language perception and processing, however, most influentially through the *Dynamic Attending Theory* (DAT). The premise of DAT is similar to the BUMP model—pre-existing neural oscillations synchronise with the rhythm of incoming percepts, with better entrainment associated with stronger attending (and better perception) of the incoming event (Hitch et al., 2022).

Therefore, taking the elements of the BUMP oscillator model and DAT together, it is evident that rhythm and neural oscillations have a role in perception and memory. The current chapter explores how participants retain stimulus order in an auditory sequence of syllables that contained timing expectancy violations, whereby a stimulus was presented either earlier or later than the expected isochronous timing of presentation. Analyses were conducted based on the accuracy scores in target pair recognition, as well as the time it took participants to select a response (reaction time, RT).

The current chapter explores how variation in the rhythm of spoken syllables is perceived, and how that compares to isochronously presented tokens when participants' retention of stimulus order in a perceived sequence was tested. Sequences of syllables were auditorily presented to participants at a fixed rate. Participants' retention of stimulus order in a perceived sequence was tested at the end of each sequence. They were asked to identify which ordered pair of syllables of two options had been present in the sequence. A portion of the syllable sequences contained a timing expectancy violation, where a stimulus was presented either earlier or later than the expected isochronous timing of presentation. Analyses were conducted based on the accuracy scores in target pair recognition, as well as the time it took participants to make a selection (reaction time, RT).

Hypotheses

The current chapter aims to understand the role of unexpected timing deviants in temporally grouping syllables in a continuous stream for immediate memory. In other words, we want to understand whether one type of timing disruption is 'better' than the other in terms of interrupting the processing of speech. In this experiment, we hypothesise that target pair identification will be negatively impacted by the presence of a timing deviant within the sequence. The timing deviants used in this experiment do not, when present, disrupt the probed syllable pair in the experiment. Rather, they are presented somewhere within the whole syllable sequence. The presence of the timing deviant would impact the grouping of syllables in the sequence at points *other* than the position of the target syllable pair. Therefore, we predict lower accuracy (poorer recall) of the target syllable pair in sequences that contain timing deviants compared to sequences with no timing deviants.

We also hypothesised that the type of timing deviant would affect recall as well. The Early timing deviant introduces a shorter time gap between syllables compared to the late timing deviant which introduces a longer, more noticeable time gap. In the late timing deviant condition therefore, we expect recall of the target syllable pair to be interrupted by an orienting effect that leads to increased attention at the position of the deviant. The deviant could, therefore, lead to a stronger syllable grouping effect for parts of the syllable sequence other than the target pair, which would lead to disruption of target recall. In the Early timing deviant, we do not expect the smaller time gap to lead to a similar attention effect that would be disruptive to the recall of the target pair. We expect the Early timing deviant conditions to result in better recall for the target pair than the Late timing deviant conditions. Therefore, we hypothesise that participants will be faster and more accurate in identifying targets when a stimulus in the auditory sequence has been presented earlier than expected (Early deviant) compared to being presented later than expected (late deviant). Sequences with no timing deviants would show highest accuracy and fastest reaction times as there would be no intervening timing deviant to interrupt the memory trace formed of the syllable sequence.

Across all conditions, we also expect to see a serial position effect of target position, with targets presented later in the sequence showing higher accuracy and faster identification times than those that were presented earlier in the sequence (Baddeley & Hitch, 1993).

Methods

Participants

126 Canadian university students (mean age = 19.6, s.d. = 3.31; 106 female, 2 N/A) with no reported visual/auditory problems were recruited for the study. The experiment was hosted online in two parts, with an initial survey that linked to an external website with the online task portion of the experiment. Out of 126 students who signed up online for this experiment, 95 completed the experimental portion to some degree (that is, completed at least 1 trial or more). After removing duplicates and data with fewer than 50% responses (135/270 trials), data from 80 participants was used for further analyses. All participants provided online consent to participate in this experiment in line with the ethical standards of the Declaration of Helsinki and were compensated with course credit. This study was cleared by the McMaster Research Ethics Board (MREB) in Hamilton, Ontario, Canada.

Stimuli

Syllables included in this experiment comprised of consonant-vowel (CV) pairs with the consonants /p t k m/ and the vowels /i u a/. Each consonant was paired with each vowel, resulting in a total of 12 different syllables. A free online text-to-speech service (fromtexttospeech.com) was used to generate the 12 syllables. The following settings were used: United States (US) English (language); George (voice); and slow (speed). Syllables were generated in triplets (for e.g., /ti ta tu/) for convenience, and then spliced and processed using Praat (Boersma & Weenink, 2019). All syllables were intensity-normalised to 75 dB and adjusted to approximately the same length (300 ms).

Pairs of syllables were created by joining together the individual-syllable files with a gap of 500 ms of silence between them (standard ISI) through custom Python code (Python Software Foundation, https.//www.python.org/), via the Anaconda Python distribution platform (Anaconda Software, www.anaconda.com). Trials with no timing deviants were presented at an isochronous rate of 500 ms (2 Hz). In conditions containing an Early or late deviant, the ISIs for one of the syllables was either shortened to 350 ms (Early; 2.86 Hz) or lengthened to 650 ms (Late; 1.54 Hz), respectively. The shortened or lengthened ISI did not interrupt any syllable pairs, and any timing manipulations always occurred between syllable pairs, not within them (see **Figure 17** for a visual representation of one trial). Trials were created using the same method to concatenate individual files according to a pre-determined pattern of 20 syllables. An original set of 340 trials was created such that each syllable appeared an approximate even number of times across them. A total of 270 trials were used in the final iteration of the experiment to account for time constraints and ensure good quality of data. All syllables in each trial sequence were arranged to ensure that: 1. no syllable was repeated twice in a row; 2. no foil (non-target) pairs of syllables were present in the sequence and 3. no target pairs of syllables were repeated twice in the trial sequence.

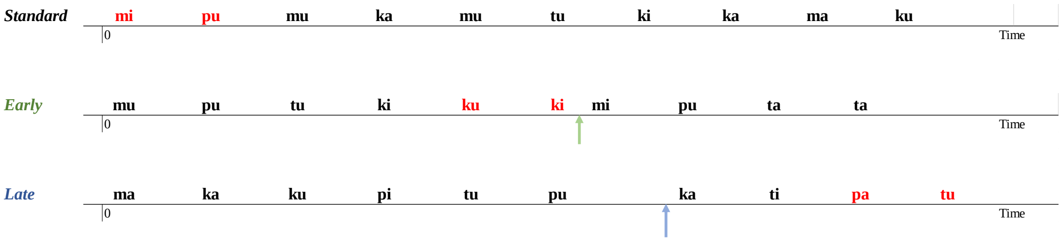

Figure 17. The figure below shows a visual representation of three trials presented in the experiment, with an example of one of each type: Standard (no timing deviant; first row), Early (middle row), and Late (last row). The syllables in red form the syllable pairs that were probed at the end of the trial, along with a foil (distractor) pair. The arrows represent the position of the timing deviant in the form of an ISI difference, with the green arrow representing an Early timing deviant (ISI = 350 ms), and the blue arrow representing a late timing deviant (ISI = 650 ms).

The experiment was presented entirely online to comply with COVID-19 restrictions. Participants listened to the stimuli and provided responses using their own computers. They were asked to confirm that they were using good quality headphones at the beginning of the experiment. Out of 80 total participants whose data were used for analyses, 74 reported using headphones, 3 reported not using headphones, and 3 did not respond. Responses from participants who did not use headphones were NOT excluded. The aim of asking participants to use headphones was to ensure that most of them did so, and as 74/80 (92.5%) complied, that goal was assumed to be met.

Design and experimental procedure

This experiment used a within-subjected design to investigate the effect of unexpected timing variations on participant accuracy and reaction time (RT). An online behavioural two-alternative forced choice task was used to explore memory for ordered syllable pairs in a sequence of syllables when no deviants were introduced, or when timing deviants (Early or late) were introduced into the auditory sequence. Participants heard trials consisting of 20 syllables presented at a standard, fixed, interstimulus interval (ISI) of 500 ms. These ISIs were chosen based on results from a pilot experiment as the most informative[2]. Target pair presentation was controlled within trials such that an equal number of pairs appeared in the first quarter, second quarter, third quarter and last quarter of the syllable stream. The position of timing deviants (Early or late) was also systematically varied, in this case across three positions within the syllable sequences, discounting the first two pairs and the last two pairs of syllables. The position of the deviant was not dependent on the target probe sequences, and both were varied independently from each other. At the end of each auditory sequence, participants were presented with two pairs of auditory ordered syllables and asked to use a mouse or trackpad to make a choice response to indicate which one of the pairs they had heard. Among the probes, target pairs appeared as the first pair (left click for correct) half the time, and as the second pair (right click for correct) the remaining time. The order of presentation of the stimuli was reversed for every other participant. A total of 270 trials were presented to participants. They were divided into 30 blocks of 9 trials each. Each block contained an equal mix of standard, Early, and late trials. All participants began with a practice block.

Testing procedure

This experiment was advertised on the Student Research Participation (SONA) systems at McMaster University across two departments (Linguistics & Languages SONA, and Psychology, Neuroscience, and Behaviour (PNB) SONA). As the experiment was hosted entirely online, participants signed up for it

[2] This experiment was originally run as a pilot with a longer sequence of 28 syllables and at different ISIs ranging from 350 ms to 600 ms in a similar 2AFC task. 500 ms was determined to be the optimal ISI to use as the results for each participant differentiated the most (that is, were most informative) at this rate of presentation. The trends showed that either participants performed better at this ISI compared to the other ISIs, or worse. This suggests that presenting the syllable sequence at the standard ISI of 500 ms gave us the best insight into individual variation in processing and accurately identifying target pairs.

independently and could participate at any time. Upon signing up, participants were linked to a LimeSurvey (online survey platform) questionnaire where they could provide informed consent by selecting one of two options. The survey would terminate if participants chose not to continue. If participants consented to the experiment, they were asked to fill out some brief demographic questions about their auditory/visual health, language background, and the computer they were using (operating system, etc.). After the survey was completed, they were provided with the link to an external online experiment hosted on pavlovia.org, where they would complete the experimental portion of the session.

Once participants click on the link provided by the researcher, experiments hosted on pavlovia.org are run locally from an internet browser. Participants were asked to enter a personally identifying key (not linked to their name) at the start of the experimental session, that would link their dataset to the demographic information they had entered on LimeSurvey. Participants used either Windows or Mac Operating System (OS) machines to run the experiment, with most of them being run through Safari, Chrome, or Firefox browsers. Participants started the experiment with a practice block of three trials and were asked to make a mouse-click response after each. The experiment ran for approximately 60 minutes. Participants were prompted to take a short break (if they so required it) at the end of each block of nine trials. They could take as long a break as they wanted at any point between the blocks.

Data analysis

Data were pre-processed using Microsoft Excel and the R Studio (v. 1.4.1106, R Studio Team, 2021) environment for R 4.1.0 (R Core Team, 2021). Duplicate data sets (ascertained as data sets with the same unique keys and operating systems, or data that was saved twice) and data sets with fewer than 50% trial responses (less than 135/270) were discarded. Outliers were removed using a script that employed Tukey's method of identifying outliers; mean values for each condition for each participant that lay beyond 1.5 times the interquartile range were flagged and replaced with *null* (Dhana, 2016). These values were dropped from further analyses.

A 3 x 4 repeated measures ANOVA was conducted in order to evaluate the effect of timing deviant presence (Standard trials (no timing deviant), Early timing deviant trials, Late timing deviant) on participant accuracy and reaction time. Further pairwise comparisons were conducted to evaluate any significant main effects observed.

The following packages were used to reshape data, calculate statistics, and generate figures: dplyr (v. 1.0.6, Wickham, François, Henry, & Müller, 2021), ggplot2 (v. 3.3.3., Wickham, 2016), ggpubr (v. 0.4.0, Kassambara, 2020), lme4

(v. 1.1-27, Bates, Maechler, Bolker, & Walker, 2015), rstatix (0.7.0 Kassambara, 2021), and tayloRswift (0.1.0, Stephenson, 2021).

Results

Participant accuracy (correctly choosing the target pair) and participant RT in seconds (time taken to respond) were both recorded and analysed across the three different ISI conditions: Standard (no timing deviant), Early (one syllable presented earlier than expected), and Late (one syllable presented later than expected).

Participant accuracy

At the end of each trial, participants selected which pair of syllables had been present within the sequence of syllables they had heard for that trial. A correct response was scored as a 1, and an incorrect response was scored as a 0. As each trial probed both a target and a foil pair of syllables, a mean accuracy of 0.5 represented chance-level responses. **Table 5** shows mean accuracy for participants in each of the three conditions.

We hypothesised that the trials containing the timing deviants (Early and Late) would differ from the Standard (no timing deviant) trials as the added timing interruption would disrupt recall for the target syllable pair by orienting attention to parts of the sequence that were not being probed. We expected participants to perform better (with higher accuracy) on standard trials than on timing deviant trials. We also hypothesised that participants would perform better on the Early timing deviant trials than on the Late timing deviant trials, as the increased gap in ISI for the Late timing deviant would lead to a greater recall disruption than the smaller time gap present within the Early timing deviant trials. Overall, we found that the task proved harder than the pilot data had indicated. On average, participants were most accurate in the Standard condition that did not contain any timing deviants with a mean accuracy of 0.54 (s.d. = 0.068). Accuracy was lower for trials with a timing deviant (either Early or Late) compared to the standard, non-timing deviant trials. Overall, the Late timing deviant trials showed the lowest mean accuracy score amongst participants **Table 5**).

Table 5. Summary of mean accuracy scores averaged across participants for each Trial Type.

Mean accuracy by Trial Type (Condition)			
Trial Type	*n*	*Mean*	*s.d.*
Standard	80	0.540	0.068
Early	80	0.529	0.073
Late	80	0.513	0.070

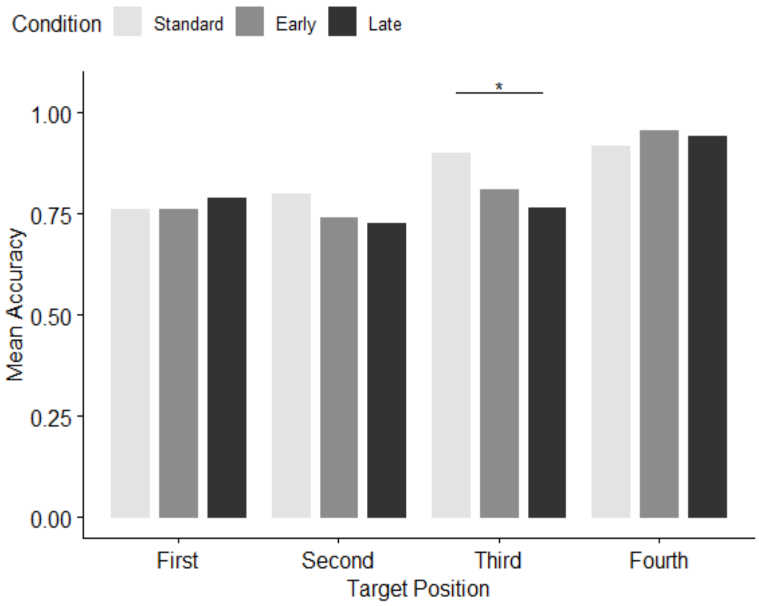

Figure 18. Mean accuracy scores between Trial Type (Standard, Early, Late), and Target Position.

The effect of temporal variation and target position on mean accuracy

A 3 x 4 Repeated Measures ANOVA was conducted to evaluate the effect of Trial Type (Standard, Early, Late) and Target Position (First, Second, Third, or Fourth quarter) on mean accuracy. Greenhouse-Geisser corrections were applied where necessary; uncorrected degrees of freedom, corrected *p*-values and corrected η^2 values are reported. There was a statistically significant main effect of Trial Type ($F(2,158) = 5.354$, $p = 0.006$, $\eta^2 = 0.009$), and Target Position ($F(3, 237) = 29.851$, $p < 0.00001$, $\eta^2 = 0.092$) on accuracy. The two-way interaction effect between Trial Type and Target Position was not significant ($F(6, 474) = 0.757$, $p = 0.604$, $\eta^2 = 0.004$). Follow-up pairwise comparisons were conducted using a paired *t*-test to evaluate the hypotheses concerning Trial Type and Target Position. Bonferroni corrections were applied and an adjusted alpha value of 0.017 was used (**Figure 19**).

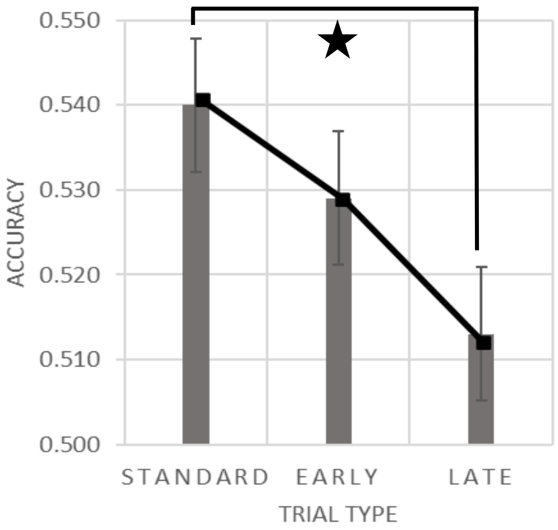

Figure 19. Mean accuracy scores averaged across participants for each Trial Type (Condition: Early timing deviant, Late timing deviant, or Standard (no timing deviant)). A significant difference between the Late and Standard Trial Types was observed (marked by a star).

The effect of Trial Type was significant between the Late Trial Type and Standard Trial Type (*t*(79) = 3.12, *p* = 0.003), with participants performing better on the Standard trials (accuracy = 0.540) than on Late trials (accuracy = 0.513).The difference between the Early Trial Type (accuracy = 0.529) and the Standard Trial Type was not significant (*p* = 0.214), and neither was the difference between the Early and Late Trial Types (*p* = 0.076), although the comparison trended in that direction (see **Table 6** for details on all comparisons). This suggests that participants were able to perform better on trials with no timing deviants, as we predicted, and that the presence of the Late timing deviant in particular affected performance negatively.

The mean accuracy scores increased with Target Position, with the lowest accuracy for targets presented at the beginning of the auditory sequence, and the highest accuracy scores for targets presented at the end of the auditory sequence (**Table 7**). Mean accuracy differed significantly when paired comparisons were made between all Target Positions (except for the first and second, first and third, and second and third positions), *p* < 0.008 (Bonferroni correction applied). This suggests that participants were significantly better able to remember target pairs when they were presented in the last half of the sequence (Third and Fourth positions) compared to when they were presented in the first half of the sequence (First and Second positions) (see **APPENDIX B, Table B** for detailed comparisons).

Table 6. The results of pairwise comparisons using a paired t-test, evaluating the difference in mean accuracy scores between the different Trial Types.

Pairwise comparisons of mean accuracy by Trial Type					
Comparison	*n*	*t*	*df*	*p*	*Significance*
Standard/Early	80	1.25	79	0.214	-
Standard/Late	80	3.12	79	0.003	*
Early/Late	80	1.80	79	0.076	-

Note: Alpha adjusted (Bonferroni correction applied)

- Indicates no significance p > 0.017

* Indicates significance p < 0.017

Table 7. Mean accuracy scores averaged across participants for each Target Position.

Mean accuracy by Target Position			
Target Position	*n*	*Mean*	*s.d.*
First	80	0.499	0.069
Second	80	0.495	0.062
Third	80	0.522	0.083
Fourth	80	0.594	0.114

Logistic regression

In consideration of the fact that the dependent variable was a binary correct/incorrect response, a logistic regression was run in addition to the ANOVA. As individual trials were included in this analysis, trial-level control variables such as Trial Number were able to be included. This analysis was used to validate the results of the ANOVA for participant accuracy scores. A maximum likelihood generalised linear mixed model (GLMM) was fitted using Laplace approximation (**Table 8**). A logistic regression was conducted for the accuracy data to evaluate how the dependent variable varied as a function of the fixed effects of Trial Type (Condition), Target Position, and Trial Number. Individual participants were added as a random group intercept. This binomial model was run using the lme4 package (v. 1.1-27, Bates et al., 2015), with the bobyqa optimiser.

Results of the Wald test based on the mixed logistic regression indicated a statistically significant difference for the intercept (χ^2(1, N = 80) = 51.84, $p <$ 0.0001). All factors evaluated were found to be significantly involved in the model fit: Trial Number (χ^2(1, N = 80) = 30.01, $p <$ 0.0001), Target Position (χ^2(3, N = 80)= 126.61, $p <$ 0.0001), and Trial Type (χ^2(2, N = 80) = 10.47, $p <$ 0.01).

The intercept of this model represents the estimated baseline log-odds of the accuracy being 1 for any given trial when the Trial Number is 1, Target Position is the First, and Trial Type is Standard (reference values). The observed intercept estimate ($ß_0$ = -0.25; see **Table 8**) suggests that the likelihood of participants correctly identifying a target syllable pair when the predictors are at reference values is 0.44 ($\frac{e^{ß_0}}{1+ e^{ß_0}}$ = 0.44). Therefore, participants are likelier to identify target syllable pairs incorrectly at the start of the experiment and when the syllable pair is presented in the beginning of the sequence. For Trial Number, participants were likelier to identify target syllable pairs correctly with a unit increase in Trial Number (assuming all other predictors are held constant). Although the observed difference here is small ($e^{ß_1} = 1.0011$), it has a significant impact to the model fit.

Participants had a lower log-odds of accurately identifying the target syllable pair as the Target Position increased (Second, Third, Fourth) when compared to the log-odds of accuracy for the First Target Position (see Table 4). In terms of odds, participants were 1.13, 1.14, and 1.03 times likelier to identify the target syllable pairs correctly for the Second, Third, and Fourth Target Positions respectively compared to the First. This corroborates the results observed with the ANOVA reported in the previous section; participants are more accurate as the target is presented later on in the sequence.

Finally, the Trial Type was found to have an effect on the log odds of a participant identifying the target syllable pair correctly. The presence of an Early Deviant was associated with a significant drop in the likelihood of accurately identifying target syllable pairs of 0.053 (5.3%). While not significant, the presence of a Late Deviant shows a similar trend, with a drop in the likelihood of an accurate response by 0.008 (0.8%). It is interesting to note that this trend of patterns is different from that reported in the previous section for the results of the ANOVA, where the presence of the Late Deviant resulted in a significantly poorer performance compared to the Standard Trials, but the Early Deviant did not. Considering this model reports the increase in likelihood of an accurate target identification with a change in Trial Type, but with all other predictors held at baseline, this result can be interpreted as the likelihood of an accurate judgement at the beginning of the experiment, when the target is presented early on in the sequence. In that regard, the Early Deviant shows a facilitatory effect of recall compared to the Late Deviant.

Table 8. Generalised Linear Mixed Model evaluating participant accuracy for the fixed effects of Trial Number, Target Position, and Trial Type.

Fixed effects - Generalised Linear Mixed Model fit by Maximum Likelihood (Laplace Approximation)					
Predictor (n)	β_n	*S.E.* β_n	*z-value*	*p*	*Significance*
(Intercept)	-0.2496	0.0346	-7.21	< 0.001	***
Trial Number	0.0011	0.0001	5.48	< 0.001	***
Target Position - Second	0.1207	0.0254	4.74	< 0.001	***
Target Position - Third	0.1331	0.0258	5.17	< 0.001	***
Target Position - Fourth	0.0303	0.0258	1.18	0.2400	-
Trial Type - Early	-0.0546	0.0210	-2.59	0.0095	**
Trial Type - Late	-0.0079	0.0210	-0.37	0.7080	-

Note: Intercept is n = 0

- Indicates no significance p > 0.05

* Indicates significance p < 0.05

** Indicates significance p < 0.01

*** Indicates significance p < 0.001

Participant Reaction Time (RT)

At the end of each auditory sequence, participants selected which pair of syllables had been present within the sequence of syllables they had heard in that trial. The time they took to select their response was recorded. Participant reaction time was analysed to explore how reaction time differences compare to patterns of participant accuracy observed across trials. In the data summarised below, mean reaction time (RT) was calculated for responses by each participant separately for each condition.

Response RT for only correct responses was analysed to evaluate whether there was an effect of the different conditions when participants correctly identified the target. We expected any effects in the data to be stronger when only correct responses were analysed under the assumption that participants had correctly identified the target syllable pair when they made a response and were not guessing. However, no difference was found between different conditions, and so subsequent data analyses were conducted with all responses included.

On average, participants were faster to respond in the Standard condition (2.31 s, s.d. = 0.595) that did not contain any timing deviants compared to the timing deviants. RT was longer for both the Early and Late timing deviants, with faster reaction times for the Early timing deviant condition (2.34 s, s.d. = 0.572) compared to the Late timing deviant condition (2.35 s, s.d. = 0.602). Overall, participants were slowest to respond for the Late timing deviant conditions compared to the other two (**Figure 20**). However, paired *t*-tests evaluating the difference between the three timing deviants did not find any significant differences, *p* > 0.016 (Bonferroni corrections applied). In evaluating participant accuracy, significantly higher accuracy was observed for the Standard trials than for the Late timing deviant trials. This accuracy difference is not reflected in RT, with the Late deviant trials averaging a response time that is 80 ms longer than that of the Standard trials. Therefore, the better accuracy observed for the Standard trials does not stem from the participants taking longer to complete those trials.

The (comparatively) longer RT for Late deviant trials could also suggest a lag due to a processing delay or distraction event attended to due to the presence of the Late timing deviant. The Standard trials contain no timing deviant and so participants respond faster for those. The RT for the Early timing deviant is more comparable to that observed for the Standard trials than the Late deviant trials, suggesting that as we hypothesised, the Early timing deviant may be forming a less distracting event than the Late timing Deviant.

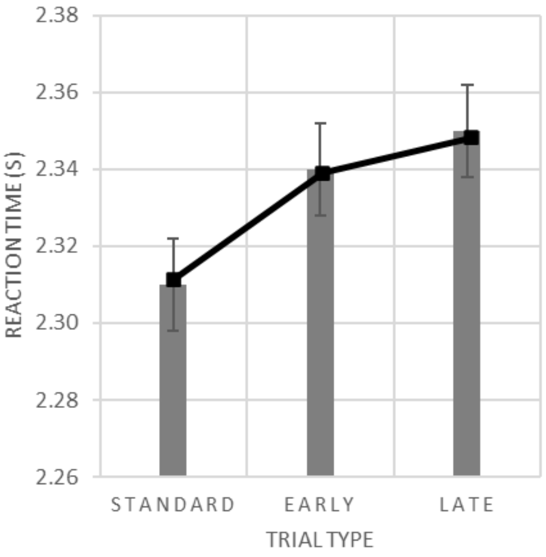

Figure 20. Mean RT (in seconds) averaged across participants for each Trial Type (Condition: Early timing deviant, Late timing deviant, or Standard (no timing deviant). A significant difference between the Late and Standard Trial Types was observed.

A 3 x 4 two-way Repeated Measures ANOVA was conducted to evaluate the effect of Trial Type and Target Position on mean RT. Greenhouse-Geisser corrections were applied where necessary. There was a statistically significant main effect of Target Position in the sequence ($F(3, 237) = 5.662$, $p = 0.000922$, $\eta^2 = 0.004$). However, the main effect of Trial Type failed to reach significance ($F(2, 158) = 1.39$, $p = 0.252$, $\eta^2 = 0.0006$), as did the interaction effect between Trial Type and Target Position, $F(6, 474) = 1.14$, $p = 0.339$, $\eta^2 = 0.002$. Follow-up pairwise comparisons using paired *t*-tests were conducted to evaluate the main effect of Target Position, and significant differences were found between the Second and Fourth positions only, with the RT being faster (shorter) for targets presented at the end of the sequence (RT = 2.28 s) compared to when it was presented in the Second position (2.40 s) (*p* < 0.008, Bonferroni correction applied). The difference between the other positions were not significant, *p* > 0.008 (**Figure 21**; see also **APPENDIX A, Table C**).

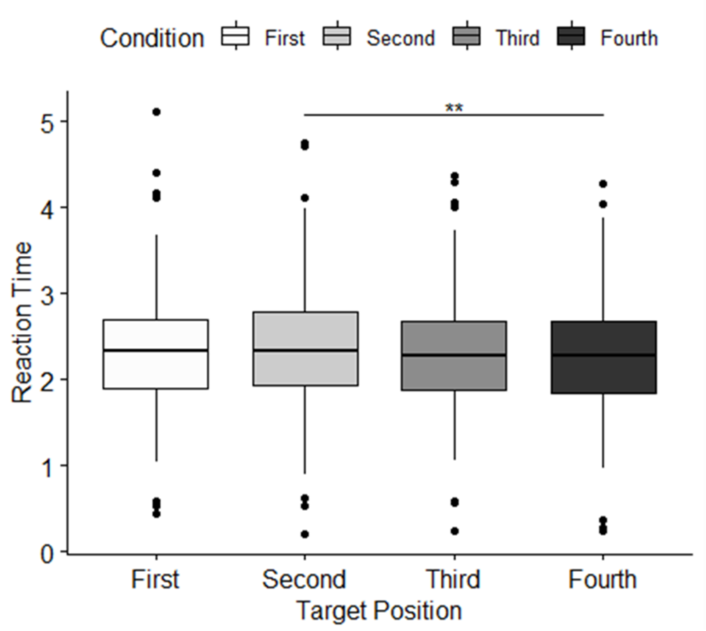

Figure 21. The figure shows mean RT averaged across participants for each Target Position.

Summary

Results showed a significant difference in mean accuracy between the Late deviant condition and the Standard (no timing deviant) condition, supporting our hypothesis that the Late deviant trials would show greater interference in target pair recognition compared to the other two types of trials. Mean accuracy also differed based on Target Position, with targets presented later in the sequence being easier to identify (higher accuracy) compared to those that were presented earlier in the sequence. Although no significant differences were observed between Standard trials and Early deviant trials, nor Early deviant trials and Late deviant trials, a trend was observed in the data that followed hypotheses, with Standard trials resulting in the highest mean accuracy scores, and Early deviant trials resulting in mean accuracy scores higher than that of the Late deviant trials, which showed the lowest mean accuracy scores.

No significant differences were found between conditions for mean RT, although data trends demonstrated a marginally faster RT for standard trials than for the timing trials. Based on Target Position, participants were significantly faster for targets presented towards the end of the sequence than for those presented earlier. These results suggest that time-accuracy trade-off did not play a role in this experiment. The fastest RTs were seen for probes matching pairs in the later positions, which also had higher recognition rates.

Discussion

The purpose of the current experiment was to explore how temporal unpredictability affects target order recognition in a sequence of auditorily-presented syllables. Specifically, the aim was to evaluate whether Early/Late variations were facilitative in grouping syllables or not. We hypothesised that Early/Late deviants introduced within an otherwise isochronous stream of auditory syllables would make it harder to recognise ordered target syllable pairs embedded in the sequence compared to no timing variation. However, comparing Early versus Late variations, we hypothesised that Late timing variations would be more disruptive of grouping and therefore target pair identification compared to Early timing variations. Our results support this hypothesis, with Late deviants resulting in significantly poorer mean accuracy than Standards, and performance for Standards being the best.

The effect of temporal variation on verbal short-term memory accuracy

Overall, a significant difference between accuracy for the Late timing deviant and the Standard condition was observed, with participants performing more poorly in identifying targets when a Late timing deviant was present in the sequence. Although there were no significant differences in accuracy between the Early and Late timing conditions, or the Early timing condition and the Standard, the mean accuracy for each condition did trend in the hypothesised direction.

The Late deviant, therefore, seemed to be more disruptive to short-term memory and implicit recognition of the syllable sequences compared to the Early deviant and the no deviant (standard) trials. The fact that unexpected temporal variation—even as little as a single disruption within a given sequence—can negatively affect verbal short-term memory suggests the importance of timing cues in cognitive processing.

An explanation for this detrimental effect may lie in the possibility of increased attention being diverted at the location of the Late timing deviant compared to the location of the target, thereby resulting in poorer target identification for the trial. Recent work by Mizark & Oberauer (2021), however, suggests that gaps in presentation of items being encoded into STM presents a proactive advantage for recall, with the ‘free time’ present during item presentation facilitating participant performance. The authors suggest that this provides evidence for an “…encoding resource that depletes with each item being encoded and recovers with time.” (p. 1334). While the authors of this study used a serial recall task as opposed to our recognition task, we expect that this ‘free time’ advantage should still be reflected in the context of the Late timing deviant as both task types make use of verbal STM. However, our results do not support this, suggesting that this renewing encoding resource model might not be the most accurate explanation for the results observed in the Mizark & Oberauer (2021) study.

Our results observed are best explained by the BUMP oscillator model (Hartley et al., 2016). This model falls under the wider umbrella of context-based competitive queuing models that suggest STM recall takes place via a two-step process involving target activation and subsequent competition and selection (see Hitch et al., 2022 for an overview). The next section discusses the results reported in this chapter under the BUMP model.

Implications of a delayed timing expectancy violation on language processing and verbal short-term memory

According to the BUMP oscillator model, predictability does not affect temporal grouping of items in a sequence (Hartley et al., 2016). Although the authors discuss the model in terms of temporal grouping of items in serial recall, here the model will be used in reference to temporal grouping and recognition of target items. Hartley et al. (2016) found that grouping effects in verbal items were not affected by their predictability; both predictable and unpredictable items showed an effect of temporal grouping. This, the authors argue, is evidence for the bottom-up nature of the mechanism that drives encoding. In the experiment reported in this chapter, participants heard sequences of syllables and were then asked to identify a target pair at the end of each sequence. Although no temporal groups were present in the syllable sequences, one-third of the trials contained an Early timing deviant, and a third contained a Late timing deviant, both of which could perceptually have elicited grouping effects within the perceived stimuli.

According to the BUMP oscillator model (and, indeed, general literature that discusses neural oscillations and their entrainment), oscillators would more strongly entrain to the overall isochronous rhythm of the sequence than they would to the irregular and unexpected timing deviants. Although, by definition of this model and other similar models (like the Dynamic Attending Theory, DAT), even irregular rhythms are entrained to. However, irregular rhythms or unexpectedly early or late item presentations entrain existing rhythms to a lesser extent than regular, repeating, predictable rhythms. Therefore, these models predict that a delay in some of the syllable sequences should have no consequential effect on verbal STM, or speech perception.

However, this is not the case. Our results show that there is a significant effect of an unexpected pause or late deviant on target syllable pair recognition when compared to no timing deviants. Although the isochronous presentation does result in the highest accuracy, as would be predicted under any neural oscillatory framework, there is also a converse effect of a late timing deviant. Interestingly, no similar effect is observed for an early timing deviant.

One explanation is that a pre-existing internal oscillator exists that entrains strongly to a delayed timing deviant (here presented at a rate of 1.54 Hz) but does not entrain as strongly to an early timing deviant (2.86 Hz). As a comparison, the isochronous rate of presentation was 2 Hz. Previous literature has shown that syllables rates of 2 – 8 Hz facilitate speech comprehension, and that intelligibility has been shown to improve for entrainment in the theta band (4 Hz to 8 Hz; Doelling et al., 2014; Poeppel & Assaneo, 2020). In our experiment, the rate of presentation for the early timing deviant and the standard falls within the range of 2-8 Hz, while that of the Late timing deviant does not. It is possible it is the anomaly of this that conversely leads to a stronger entrainment. However, that would not work because speech entrainment would not optimally fall between 2 – 8 Hz if entrainment got STRONGER outside that range; entrainment is possibly strongest to speech in the range of 2 – 8 Hz.

The effect of Target Position on mean accuracy and RT

In our experiment, we found that there was no interaction effect between Trial Type (condition) and Target Position. However, accuracy scores trended to be higher as the target syllable pair were presented late in the sequence, compared to earlier on. This trend of improving accuracy reflects the *serial position effect* (see Baddeley & Hitch, 1993, for a discussion), wherein items presented at the beginning and end of a sequence are recalled better than those presented in the middle. Particularly, the better recall of items at the end of the sequence (*recency effect*) is the pattern that we observed in our data. Although participants were not asked to recall items in our experiment, they did have to recognise them to select the correct target. The pattern of accuracy scores observed demonstrated an increase in accuracy (i.e., in recognition of targets) as the targets were presented at the end of sequences. This suggests the retention of these syllable sequences in short-term memory, making targets presented at the end easier to retrieve than those presented at the beginning.

Previous work comparing irregularly timed syllables to regularly timed (isochronous) syllables in a similar di-syllabic word identification task found that reaction times were lower for the former compared to the latter. However, this was only when targets were presented later in the sequence (Rassili & Ordin, 2020). In our experiment, too, there was a significant result of Target Position, with targets presented later in the sequence showing a higher accuracy of identification compared to targets presented earlier in the sequence. However, no interaction effect between Trial Type and Target Position was observed for accuracy scores or RT.

Summary

In terms of language processing, there seem to be two conclusions, therefore. First is that isochronous presentation and predictability support recognition of target syllable pairs the most. That is, there is a preference for a constant rate of presentation of syllables that allows better recognition of embedded targets compared to syllables presented with some forms of temporal variation. This supports the ideas of the rhythm class theory (see *Chapter 1: Introduction*). Even though speech is quasi-rhythmic in nature, this experiment fails to capture the cadence of natural speech, making it unlikely for anything but isochronous presentation to yield an advantage in identifying target syllable pairs. The focus of this experiment was on evaluating how timing deviants would be processed compared to isochronous speech. Future work extrapolating these results to more natural speech with a semantic component would provide further insight into how the processing of temporal variation evolves as factor of the complexity of speech. This experiment is one of the first, to the best of our knowledge, that evaluates timing variation between syllables and seeks to explore how variation in that pace would affect target identification.

The second conclusion is that delays in speech are less facilitative to syllable grouping, meaning that it is harder to follow speech with a lot of pauses and disruptions (although this is dependent on factors like the syntactic structure of the utterance; Reich, 1980). Early timing disruptions seem to lie in the middle of the spectrum of disruption and facilitation. Therefore, while there is a pattern in the data that show a lower accuracy for timing deviants in general, including the Early deviant, in some ways, the Early timing deviant seem to be more predictable (or more facilitative *of* predictability) compared to the Late timing deviant. Therefore, it is easier to identify target syllables when the unexpected timing disruption falls within the expected window of presentation (unexpected ISI < expected ISI; Early deviant) than when it falls outside of it (unexpected ISI > expected ISI; Late deviant). This modulation of expectancy in predicting and facilitating syllable groupings should be investigated further in future work.

Limitations

Considering that this is one of the first few studies to consider the temporal dynamics of intervals between syllables, there were improvements that could have been made to the design and were slated to be implemented in follow-up studies. Accuracy rates observed in this experiment were overall lower than expected[3], even though the 2AFC task was adjusted to have a clear target in each

[3] In typical target-identification experiments evaluating temporal expectations, evaluating participant accuracy and RT serves as a proxy for examining how short-term memory processes are affected by the different conditions (e.g., Ball et al., 2019). Typically, higher accuracy scores (any value beyond 0.5, but the higher

trial. An obvious reason for this is the difficulty of the task, either due to the length of the sequences or the length of the experiment itself. If true, this would mean that even participants who were engaging with the task attentively were, at some point, forced to guess which target syllable pair was present in the sequence when they were asked to make the choice at the end of the trial. A staircase method of presenting the task that would have presented participants with longer and more trials the better they performed may have alleviated these concerns and given us a better gauge of individual differences between participants as well. Additionally, this task was run online, which may have contributed to poorer accuracy scores as participants would have completed the experiment under less-rigorous standards than if it had been run in-person in a research facility. Future work with fewer and shorter trials, run in-person, might help improve observed participant accuracy scores. Additionally, differences between conditions could be elevated by presenting the timing deviants more than once per sequence. This may have better highlighted the impact, if any, of the timing deviants on target recall.

Finally, the transitional probabilities of syllables were not controlled for in the sequences. Although each syllable was controlled such that it appeared approximately the same number of times across the length of the experiment, probabilities of subsequent syllables occurring next to each other, as well as probabilities of the sequences that made up the target pair, were not calculated. It was also difficult, due to the controlled sequences structured, to control for the frequency of occurrence of targets chosen. This meant that some target pairs appeared more often within the sequence compared to others. However, care was taken to ensure that most target pairs appeared similar number of times, and that there was not a huge disparity in the range (of how many times each target pair was presented in the experiment).

the better) suggests that the target was easily identified and there were fewer interfering cognitive loads. A lower accuracy score (closer to 0.5, which would be chance given two choices) would suggest that there was impedance in identifying the target. In our experiment, the timing conditions, and specifically, the Early and Late timing variations, were imposed on the auditory sequences to see if, and how, they would affect recall. Participant accuracy was low overall, with values below 60%. A typical range for 2AFC experiments involving statistical learning is around 60% (Isbilen et al., 2020). However, the lower range of values observed (with some participants showing below-chance accuracy) means that much of the results were less meaningful in parsing out differences between conditions than they could have been had another type of task been used. The use of a target-distractor pair of syllables for the 2AFC did help curb some of the issues that are often associated with this task (see section "*Language processing using syllable sequences*" for a discussion).

Conclusion

The purpose of the current study was to examine how memory of auditory sequence recognition was impacted when timing variations were introduced within syllable sequences. Mean accuracy and reaction time of participant responses were recorded, and we hypothesised that participants would be faster and more accurate in identifying targets for no deviant (standard) trials than for timing deviant trials. Further, within the timing deviant trials, accuracy would be higher (and RT lower) for Early timing deviant trials than Late. Results showed a significant difference in mean accuracy for Late timing deviants compared to the standard trials, with accuracy being significantly lower in the Late timing deviant trials. There were no significant effects of RT. Although the data showed a trend in the hypothesised direction, with overall accuracy being higher for Early timing deviants than Late, but lower than that of the standard trials, these differences were not significant. The results of this study suggest that the Late timing deviant could help us answer questions about real-time speech processing. This study is the first of its kind, to our knowledge, to investigate timing differences in a stream of syllables, modelling real-time speech perception.

References

Anaconda Software Distribution. Computer software. Vers. 2-2.4.0. Anaconda, Nov. 2016. Web. <https://anaconda.com>.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Aubanel, V., & Schwartz, J. L. (2020). The role of isochrony in speech perception in noise. *Scientific Reports, 10*(1), 1–12. https://doi.org/10.1038/s41598-020-76594-1

Baddeley, A. (1992a). Working memory. *Science, 255*(5044), 556-559.

Baddeley, A. (1992b). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience, 4*(3), 281-288.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning And Motivation, 8*, 47-89, Academic Press.

Baddeley, A. D., & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition, 21*(2), 146–155. https://doi.org/10.3758/BF03202726

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.

Ball, F., Groth, R. M., Agostino, C. S., Porcu, E., & Noesselt, T. (2020). Explicitly versus implicitly driven temporal expectations: No evidence for altered perceptual processing due to top-down modulations. *Attention, Perception, and Psychophysics, 82*(4), 1793–1807. https://doi.org/10.3758/s13414-019-01879-1

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: Genera*l, *133*(1), 83-100.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bion, R. A. H., Benavides-Varela, S., & Nespor, M. (2010). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54*(1), 123–140. https://doi.org/10.1177/0023830910388018

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program].: Vol. Version 6.

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.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, *31*(5), 668-684.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*(1), 1-17.

Cowan, N., Keller, T. A., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language*, *33*(2), 234-250.

Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General*, *127*(2), 141.

Daikoku, T., Yatomi, Y., & Yumoto, M. (2017). Statistical learning of an auditory sequence and reorganization of acquired knowledge: A time course of word segmentation and ordering. *Neuropsychologia*, *95*, 1–10. https://doi.org/10.1016/j.neuropsychologia.2016.12.006

Dellwo, V. (2008). Influences of language typical speech rate on the perception of speech rhythm. *The Journal of the Acoustical Society of America*, *123*(5), 3427–3427. https://doi.org/10.1121/1.2934192

Dellwo, V., & Wagner, P. (2003). Relationships between rhythm and speech rate. *International Congress of Phonetic Sciences*, 471–474.

Dhana, K. (2016, April 30). Identify, describe, plot, and remove the outliers from the dataset. *R bloggers*. https://www.r-bloggers.com/2016/04/identify-describe-plot-and-remove-the-outliers-from-the-dataset/

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294–311. https://doi.org/10.1016/j.jml.2008.06.006

Dilley, L. C., & Pitt, M. A. (2010). Altering Context Speech Rate Can Cause Words to Appear or Disappear. *Psychological Science*, *21*(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

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.

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

Farrell, S., & McLaughlin, K. (2007). Short-term recognition memory for serial order and timing. *Memory & Cognition*, *35*(7), 1724-1734.

Gibbon, D. (2018). The Future of Prosody: It's about Time. In *Proceedings of the 9th International Conference on Speech Prosody* 2018 (pp. 1-9).

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. *Eurospeech* 2001 - Scandinavia, 1–4.

Gorin, S. (2020). The influence of rhythm on short-term memory for serial order. *Quarterly Journal of Experimental Psychology*, *73*(12), 2071-2092.

Grabe, E., & Low, E. L. (2008). Durational variability in speech and the rhythm class hypothesis. In *Laboratory phonology* 7 (pp. 515-546). De Gruyter Mouton.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech - A syllable-centric perspective. *Journal of Phonetics*, *31*, 465–485. https://doi.org/10.1016/j.wocn.2003.09.005

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.

He, L., & Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech Language and the Law*. https://doi.org/10.1558/ijsll.v23i2.30345

Hitch, G. J., Hurlstone, M. J., & Hartley, T. (2022). Computational Models of Working Memory for Language. In *Schwieter, J. W., & Zhisheng, W. The Cambridge Handbook of Working Memory and Language*. Cambridge University Press.

Hooper, J. B. (1972). The Syllable in Phonological Theory. *Language*, *48*(3), 525–540.

Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition*, *35*(7), 1675-1684.

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*(7). https://doi.org/10.1111/cogs.12848

Jones, M. R. (2018). Time will tell: A theory of dynamic attending. Oxford University Press.

Kassambara, A. (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr

Kassambara, A. (2021). rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.0. https://CRAN.R-project.org/package=rstatix

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

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *Psychology of Learning and Motivation*, *49*, 1-48.

Low, E. L., & Grabe, E. (1995). Prosodic patterns in Singapore English. In *Proceedings of the International Congress of Phonetic Sciences*, Stockholm (Vol. 3, pp. 636-639).

Ordin, M., & Nespor, M. (2013). Transition probabilities and different levels of prominence in segmentation transition probabilities and different levels. *Language Learning, 63*(4), 800–834. https://doi.org/10.1111/lang.12024

Ordin, M., & Nespor, M. (2016). Native language influence in the segmentation of a novel language. *Language Learning and Development*, *12*(4), 461–481. https://doi.org/10.1080/15475441.2016.1154858

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, *42*(February), 244–257. https://doi.org/10.1016/j.system.2013.12.004

Ordin, M., & Polyanskaya, L. (2015). Acquisition of English speech rhythm by monolingual children. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition*, *45*, 863–876. https://doi.org/10.3758/s13421-017-0700-9

Ordin, M., Polyanskaya, L., Soto, D., & Molinaro, N. (2020). Electrophysiology of statistical learning: Exploring the online learning process and offline learning product. *European Journal of Neuroscience, 51*(9), 2008–2022. https://doi.org/10.1111/ejn.14657

Plug, L., & Smith, R. (2018). Segments, syllables, and speech tempo perception. *Proceedings of the International Conference on Speech Prosody*, 2018-June(June), 279–283. https://doi.org/10.21437/SpeechProsody.2018-57

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, *21(*6), 322-334.

Polyanskaya, L., & Ordin, M. (2015). Acquisition of speech rhythm in first language. *The Journal of the Acoustical Society of America, 138*(3), EL199-EL204.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Rassili, O., & Ordin, M. (2020). The effect of regular rhythm on the perception of linguistic and non-linguistic auditory input. *European Journal of Neuroscience*, September, 1–8. https://doi.org/10.1111/ejn.15029

Reich, S. S. (1980). Significance of pauses for speech perception. *Journal of Psycholinguistic Research, 9*(4), 379-389.

Rimmele, J. M., Sun, Y., Michalareas, G., Ghitza, O., & Poeppel, D. (2019). Dynamics of functional networks for syllable and word-level processing. BioRxiv, 584375.

Saffran, J. R. (2001). The Use of Predictive Dependencies in Language Learning. *Journal of Memory and Language*, *44*, 493–515. https://doi.org/10.1006/jmla.2000.2759

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current directions in psychological science*, *12*(4), 110-114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language*, *50*(4), 425-443.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711). https://doi.org/10.1098/rstb.2016.0059

Stephenson, A. (2021). tayloRswift: Color Palettes generated by Taylor Swift Albums. R package version 0.1.0

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *The Quarterly Journal of Experimental Psychology*, *48*(1), 108-124.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*(4), 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.6. https://CRAN.R-project.org/package=dplyr

CHAPTER FIVE: DISCUSSION AND CONCLUSION

The aim of this dissertation was to explore the effect of unanticipated changes in the temporal structure of incoming speech on its comprehension and processing. Both neurolinguistic and behavioural measures were employed to evaluate how temporal variation was processed at different levels of complexity and linguistic familiarity, from simple tones, to complex waves, to syllables. *Chapters 2 and 3* evaluated the pre-attentive processing of unexpected timing variations for increasingly speech-like stimuli, while *Chapter 4* looked at the effect of temporal variation on verbal short-term memory (vSTM).

In order, the chapters answered the main research question regarding the role of unanticipated temporal variations on real-time speech comprehension by addressing the following specific questions: 1. How do unexpected temporal variations and omissions affect the pre-attentive processing of simple auditory stimuli? 2. How does increasing linguistic complexity (i.e., increasing speechlikeness) alter the pre-attentive processing of unexpected timing variations? 3. What is the impact of unexpected timing deviants on verbal short-term memory tested through the perceptual grouping of syllables in a continuous stream?

The current chapter aims to provide a summary of the dissertation findings and briefly discuss the implications of these results within a more narrow context (relative to the research objectives), as well as a wider context (relative to the existing literature). Further, this chapter will provide a summary of the research contributions to existing literature, and will conclude with a discussion on research limitations and suggested future research directions. The desire to understand more about the role timing plays in speech comprehension lies at the core of this dissertation. Timing is a fundamental aspect of prosody and language but has often been studied only in the context of other acoustic features such as vocalic duration, or tempo. This dissertation focuses on timing expectations in non-linguistic and linguistic sequences. In a series of four experiments, it better examined the effect of unanticipated timing variability on the cognitive processes of attention and memory. The next section summarises the results with respect to the research questions posited at the beginning of the dissertation, and how they add to our overall understanding of the cognitive processing of the temporal aspects of speech.

Research Questions

How do unexpected temporal variations and omissions affect the pre-attentive processing of simple auditory stimuli?

The auditory processing of simple tones showed differential patterns of pre-attentive processing depending on the temporal deviant type. Early timing deviants elicited a bigger response compared to Late deviants. This suggests that unanticipated early variations may violate expectancy patterns to a larger degree because they are more easily processed compared to stimuli presented unexpectedly late. Additionally, the results of an omission MMN experiment suggest that the TWI for simple tones may be wider than previously thought. This suggests that the memory trace for the pre-attentive sound representation of the incoming simple tones may last longer than previously thought.

How does increasing linguistic complexity (i.e., increasing speechlikeness) alter the pre-attentive processing of unexpected timing variations?

There was a significant effect of increasing speechlikeness on pre-attentive processing, although it did not interact with type of temporal variation. Overall, the results suggest that speech is processed differentially from non-speech, even, and especially, when both are artificially synthesised using the same acoustic/articulatory methods. In terms of timing variation, the Late timing deviant was found to elicit a smaller MMN than the Early timing deviant. Within the predictive coding framework, this points to a greater ‘acceptability’ of the Late deviant within the context. Alternatively, it might be a result of reduced perception of a token that falls outside the standard rate of presentation. The latter explanation is less likely, however, as attention-orienting responses like the P3a were observed to both timing deviant types. Rather than unexpected lateness affecting perception, lesser importance could have been assigned to the Late timing deviant when updating the internal model because it fell outside the modelled rate of presentation, resulting in the reduced responses regardless of speechlikeness.

What is the impact of unexpected timing deviants on verbal short-term memory tested through the perceptual grouping of syllables in a continuous stream?

Chapter 4 examined how timing deviants effected verbal short-term memory for syllable order in sequences. The order of two syllables embedded in sequences that either included early or late stimuli elsewhere in the sequence, or were isochronous, was probed. Results revealed a detrimental effect of late timing deviant presence compared to sequences with no time-shifted syllables. This suggests that an unexpected delay somewhere in the sequence makes it difficult to later remember the order of stimuli elsewhere in the sequence. Although not reaching significance, a trend was also observed for trials with early timing variations to have better accuracy than trials with late stimuli, suggesting that even within timing deviants, one is more ‘preferable’ than the other. The results here suggest that isochrony may be best for vSTM, but, that in terms of timing variation, delays impair recognition of order detail more than stimuli presented unexpectedly early.

Ramifications for speech perception

Compared to earlier than expected temporal variations, delays or later than expected auditory inputs were found to be more detrimental to both pre-attentive processing and vSTM. Considering that natural speech is disfluent, what does this mean for real-time speech processing? Particularly, might it add to difficulties for people with atypical language development, or those struggling to learn a new language? The next section focuses on the processing of stimuli arriving late and discusses how the current data measure up against previous literature.

Implications and Contributions to Knowledge

Prosody generally refers to changes in suprasegmental cues of speech, such as fundamental frequency (perceived as voice pitch), loudness, duration, or tempo, among other parameters. These cues often change or co-occur in tandem, signalling meaningful changes in an utterance. For example, stress differences are often perceived to associated changes in pitch, intensity, and vocalic/consonantal length (Honbolygó & Csépe, 2013; Honbolygó et al., 2004; Honbolygó et al., 2017). Changes in vocalic duration or duration of intervocalic gaps have been studied as measures of durational variation as well as tempo (e.g., Dellwo, 2008; Dellwo & Wagner, 2003). Word segmentation and perception have been found to be affected by variations in speech rate (Dilley & McCauley, 2008; Dilly & Pitt, 2010). Studies of prosody have often relied on features like consonantal or vocalic duration as a measure of temporal variability in an utterance (e.g., Gibbon & Gut, 2001; He & Dellwo, 2016; Low & Grabe, 1995; Ordin & Polyanskaya, 2014). The aim of many of these studies has been to evaluate the rhythmic class of a language and provide an objective measure of linguistic rhythmicity (Choi, 2003; Grabe & Low, 2002; Tseng & Fu, 2005). Studies that have specifically examined how non-isochronous speech is processed have looked at varying rhythms to find preferential processing for isochrony versus irregular speech (Aubanel & Schwartz, 2020; Dellwo et al., 2015). However, as discussed in *Chapter One: Introduction*, the majority of earlier work has remained split between examining segmental duration time differences in speech, or examining the effect of altering timing-related acoustic features (such as tempo) on speech perception. This dissertation bridges the gap between the two areas by examining temporal changes at a more abstract level to understand how they affect attention and memory in auditory contexts.

The role of timing in auditory processing has been more robustly studied for simple tones, such as pure sine waves, as compared to complex waves and syllables. As such, the research reported here provides novel insight into the pre-attentive processing of unanticipated temporal variation in the context of increasing acoustic complexity. The main focus of this dissertation has been on the timing of stimulus presentation, with careful manipulations made for rare deviant stimuli to be presented earlier or later than the expected rate of presentation. Both early and late presentations differed from the standard rate of presentation to the same amount of time in all experiments, thereby leaving the only difference between early and late stimuli to be the direction of the time difference compared to standard times of presentation. One of the questions we aimed to address was whether or not the timing deviants differed in the way they were processed, and how increasing speechlikeness would interact with it. This would then lend us insight into how timing as a prosodic feature was processed, and ergo how it could contribute to language acquisition.

Results from *Experiments 2, 3,* and *4* show that delays are processed differently compared to unanticipated early presentations, suggesting that processing of later than expected tokens is detrimental to attention and memory processes. Previous work on how timing deviants are pre-attentively processed is limited. The general consensus remains that MMNs elicited to timing deviants should be proportionally bigger in amplitude as the difference in SOA between the timing deviant and the standard rate of presentation increases, regardless of whether the timing deviant is early or late (Alain et al., 1999; Kisley et al., 2004; Sable et al., 2003). While both early and late timing deviants showed this proportional increase in MMN as SOA increased, Sable et al. (2003) also found that the early timing deviant elicited a larger response compared to the late timing deviant. The results reported in this dissertation corroborate this pattern for simple tones, and further describe the same pattern for complex waves and syllables, something which has to our knowledge not previously been reported in the literature.

These observed differences between the late and early timing deviants further support claims that these unexpected timing variations are processed differently, as posited by Jongsma et al. (2007). However, there are two important distinctions here. The first is that the Jongsma et al. (2007) study evaluated changes in the P3a and P3b responses in the context of early and late timing deviants. The second is that these changes were observed to simple tones in order to study beat perception. Direct comparisons, therefore, are difficult to make between that study and our results. However, this dissertation finds support for the idea that early and late timing deviants are processed differently, as evidenced by the differential MMN responses observed to each.

As is evident from the literature reviewed so far, the majority of work on timing differences and the MMN has investigated simple tones. Work on timing with complex waves and syllables is highly limited, and typically has evaluated measures other than the direct rate of presentation of stimuli. For example, studies have evaluated the perception of highly regular speech (isochronous speech) compared to irregular, or non-isochronous speech, and found preferences for the former (see Aubanel & Schwartz, 2020). The results for the vSTM task reported in *Chapter 4* corroborates this result, with higher accuracy for syllable recognition found for isochronous sequences compared to those including a late timing deviant.

To our knowledge, the results reported in this dissertation are amongst the first to provide an objective evaluation of how temporal variation affects attention and memory in speech. Timing is an intrinsic element of rhythm in both speech and music, which, along with motor skills, have been found to be highly correlated with each other. Children who struggle with non-typical language development have also been found to often struggle with rhythm in music and motor tasks (Jentschke et al., 2008; Kujala & Leminen, 2017; Law et al., 2014; Molinaro et al., 2016). Therefore, understanding rhythm is particularly important, and if unexpectedly long pauses have a detrimental effect on attention and memory, care can be taken to devise resources that incorporate isochrony instead.

Other contributions

TWI

The results of *Experiment 1* (*Chapter 2)* suggested that the Temporal Window of Integration (TWI), previously suggested to be about 170 ms for simple tones, may be wider than reported by previous research (Yabe et al., 1997; 1998). Our results showed MMNs elicited to omission deviants at an SOA of 250 ms, suggesting the window to be at least 250 ms for simple tones. This would have unique ramifications for our understanding of the extent of the memory trace formation for the MMN. A wider TWI suggests, for example, that more time can pass between subsequent tones but still allow the elicitation of an MMN should one be omitted from sequence (thus forming an omission deviant).

Speechlikeness

Experiment 3 (Chapter 3) directly compared the elicitation of the MMN to temporal deviants in sequences of complex waves and syllables. Although no effect of increasing speechlikeness was found to interact with how temporal variation was processed, a significant main effect of speechlikeness on MMN elicitation *was* observed, with complex waves eliciting larger MMNs compared to syllables. This suggested that the speech stimuli were differently processed compared to the complex wave stimuli. This result has important implications for future research, and particularly so for MMN research, on two counts.

Firstly, a significant difference was observed between the response elicited to complex waves and syllables. It is worth reiterating here that both stimulus types were completely artificially synthesised (using identical acoustic/articulatory methods) and highly controlled for, with the aim to make the only difference between them the presence of linguistic attributes for the syllable condition. As discussed in *Chapter 3*, the explanation for the difference in how complex waves and syllables were processed seems to lie in the fact that the syllables were processed as speech because they sounded like familiar syllables. This aspect of linguistic familiarity was absent for the complex (non-linguistic) waves. The elicitation of an MMN to both conditions means that the use of carefully synthesised speech could be a good candidate for EEG experiments allowing different acoustic parameters to be carefully and tightly controlled, something that may not be possible for natural speech stimuli without losing the special quality that make them *speech*. Parameters such as inter- and intra-speaker variation, voice quality differences and acoustic variability cannot be controlled for in natural speech the same way they can be for synthesised tokens.

Secondly, the observed responses to the syllables were reduced compared to those to the complex waves. Previous work has suggested that synthesised syllables elicit reduced or no MMNs (Pettigrew et al., 2005; Wunderlich & Cone-Wesson, 2001). *Chapter 3* discusses these differences and examines possible reasons in detail. However, the critical takeaway here is that our results provide a good comparison point for future research that plans to use synthesised syllables. The response recorded to increasing speechlikeness in *Chapter 3* helps establish a baseline of how synthesised stimuli are processed, thereby providing a stronger foundation for future research on how acoustically and linguistically complex tokens are auditorily processed.

Limitations and Future directions

This section examines a few general limitations of the research reported in this dissertation and provides context for the interpretation of the results.

Firstly, the experiments reported in this dissertation were advertised to students at McMaster University and surrounding areas (Hamilton, ON, Canada). This meant that all recruited participants (in-person or online) were exposed predominantly to Canadian English (and any regional variations associated with that). As prosody and timing both strongly correlate with native language, participant language background is important in understanding how timing differences were processed. Relatedly, language background was not controlled for in the results reported here, and participants that spoke or had exposure to languages in addition to Canadian English were tested. Although none of the stimuli was language-specific and, indeed, care was taken to construct stimuli so they would not adhere to any particular prosodic or phonological linguistic profile, it would be naïve to say that language background would have had no effect on how the more linguistically complex stimuli (syllables) were processed. Future work could address this limitation by exploring how MMN responses to timing variations change based on varying language and language family background. A between-subjects design exploring MMN size to Early and Late timing deviants for participants of varying language backgrounds would help provide further insight into temporal variability processing. Groups to be compared could include Canadian English speakers with minimal exposure to

French, balanced bilingual Canadian English and French speakers and bilingual Canadian English and Other speakers, ostensibly participants who speak or have been exposed to any language or languages as part of their heritage. These comparisons would provide ample insight into: 1. Whether increased exposure to a second language (French) changes how pre-attentive responses to temporal variability pattern, and, and 2. whether bilingualism background (one additional language with Canadian English, or multiple additional languages with Canadian English) makes any difference to the perception of timing variability. This will help inform the results of the current dissertation as well as provide awareness into the degree to which language background affects the processing of unexpected temporal variability, if at all. Related to participant demographics, musical knowledge background (i.e. if participants play and practice and instrument and/or have formal musical education) in addition to language background can also provide increased insight into how individuals process timing. Exploring individual differences between participants will provide greater insight into the extent to which timing influences processing. This will pave the way for better understanding deficits in language acquisition and development with regard to processing rhythm and timing.

This would be valuable to explore further in specialised populations (e.g., dyslexic individuals) to better understand their perception of timing uncertainties. Disruptions in fluent speech can be more difficult to process for certain populations over others. For example, when there are delays in a real-time speech conversation, it is possible that someone struggling with a language disorder would find it harder to entrain to and keep track of the conversation compared to someone who is neurotypical. Although the literature on dyslexia and rhythm is robust (e.g., Goswami, 2018; Goswami et al., 2013), little of it specifically addresses unexpected temporal variations in speech processing. Examining specialised clinical populations with realistic (synthesised) speech tokens would be the next best step to understand the underlying neural mechanisms of speech perception, how typically or otherwise they work, and how we can provide tools to minimise the impact of any atypical processing.

Another limitation of the results reported in this dissertation was the inability to directly compare the results for *Chapters 2* and *3*. The experiments were designed with the aim of being able to directly follow a change in MMN elicitation for stimuli that progressed in acoustic complexity and speechlikeness from simple tones to complex waves to syllables. However, in order to preserve the quality of the data obtained within each testing session, it was impractical to present all three stimulus types and have sufficient tokens to observe the MMN. Furthermore, aspects of the methodology were adjusted to make it easier to see the observed MMN for the more acoustically complex stimuli (*Chapter 3*). Future research could redress this by exploring changes between just the Early and Late timing deviants for the three stimulus types. By removing certain redundant conditions and the presentation of the Frequency deviant and therefore reducing the total number of blocks presented to the participants, a more manageable testing session can be run that would allow a direct comparison of how the MMN elicitation changes with increasing speechlikeness for each individual.

Finally, the use of behavioural tools for *Experiment 4* (*Chapter 4*) made it difficult to compare aspects of temporal processing to the results reported in *Chapter 2* and *Chapter 3*. Developing the same paradigm for use with EEG can help address this and make the results more comparable.

Conclusion

This dissertation provides novel insight into how unanticipated temporal variation affects pre-attentive processing of auditory stimuli and verbal short-term memory. While previous literature has looked at isochronous speech at large, or examined how patterns of tones are processed, studies have not looked at the role of unexpected timing variations on attention and memory within the context of acoustically complex stimuli. The results reported here show that care should be taken during experimental design when choosing timing of stimulus presentation. Additionally, researchers should recognise that temporal variability with delays will introduce greater cognitive processing load compared to tokens that fall within the average rate of presentation. This dissertation helps address the gap in our understanding about how timing variation in language is processed, and further raises the question of the extent to which the late timing deviant affects cognitive processing. Timing is an intrinsic part of speech, and understanding more about the fundamentals of how it affects cognition will, with further work, allow for a better understanding of prosody and speech at large.

References

Alain, C., Cortese, F., & Picton, T. W. (1999). Event-related brain activity associated with auditory pattern processing. *NeuroReport*, *10*(11), 2429–2434. https://doi.org/10.1097/00001756-199908020-00038

Aubanel, V., & Schwartz, J. L. (2020). The role of isochrony in speech perception in noise. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-76594-1

Choi, J. (2003). Pause length and speech rate as durational cues for prosody markers. The *Journal of the Acoustical Society of America*, *114*(4), 2395–2395. https://doi.org/10.5840/raven20111838

Dellwo, V. (2008). Influences of language typical speech rate on the perception of speech rhythm. *The Journal of the Acoustical Society of America*, *123*(5), 3427–3427. https://doi.org/10.1121/1.2934192

Dellwo, V., Leemann, A., & Kolly, M. J. (2015). Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *The Journal of the Acoustical Society of America*, *137*(3), 1513-1528.

Dellwo, V., & Wagner, P. (2003). Relationships between rhythm and speech rate. *International Congress of Phonetic Sciences*, 471–474. https://pub.uni-bielefeld.de/record/1785384#contentnegotiation

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294–311. https://doi.org/10.1016/j.jml.2008.06.006

Dilley, L. C., & Pitt, M. A. (2010). Altering Context Speech Rate Can Cause Words to Appear or Disappear. *Psychological Science*, *21*(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Gibbon, D., & Gut, U. (2001). Measuring speech rhythm. Eurospeech 2001 - Scandinavia, 1–4.

Goswami, U. (2018). A Neural Basis for Phonological Awareness? An Oscillatory Temporal-Sampling Perspective. *Current Directions in Psychological Science*, *27*(1), 56–63. https://doi.org/10.1177/0963721417727520

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.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*(1982), 515-546.

He, L., & Dellwo, V. (2016). The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech Language and the Law*, *23*(2). https://doi.org/10.1558/ijsll.v23i2.30345

Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology*, *87(*2), 165–172. https://doi.org/10.1016/J.IJPSYCHO.2012.12.005

Honbolygó, F., Csépe, V., & Ragó, A. (2004). Suprasegmental speech cues are automatically processed by the human brain: A mismatch negativity study. *Neuroscience Letters*, *363*(1), 84–88. https://doi.org/10.1016/j.neulet.2004.03.057

Honbolygó, F., Kolozsvári, O., & Csépe, V. (2017). Processing of word stress related acoustic information: A multi-feature MMN study. *International Journal of Psychophysiology*, *118*, 9–17.

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience*, *20*(11), 1940–1951.

Jongsma, M. L. A., Meeuwissen, E., Vos, P. G., & Maes, R. (2007). Rhythm perception: Speeding up or slowing down affects different subcomponents of the ERP P3 complex. *Biological Psychology*, *75*(3), 219–228. https://doi.org/10.1016/J.BIOPSYCHO.2007.02.003

Kisley, M. A., Davalos, D. B., Layton, H. S., Pratt, D., Ellis, J. K., & Seger, C. A. (2004). Small changes in temporal deviance modulate mismatch negativity amplitude in humans. *Neuroscience Letters*, *358*(3), 197–200. https://doi.org/10.1016/j.neulet.2004.01.042

Kujala, T., & Leminen, M. (2017). Low-level neural auditory discrimination dysfunctions in specific language impairment—A review on mismatch negativity findings. *Developmental Cognitive Neuroscience*, *28*(October 2016), 65–75. https://doi.org/10.1016/j.dcn.2017.10.005

Law, J. M., Vandermosten, M., Ghesquiere, P., & Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00482

Low, E. L., & Grabe, E. (1995). Prosodic patterns in Singapore English. In Proceedings of the *International Congress of Phonetic Sciences*, Stockholm (Vol. 3, pp. 636-639).

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. https://doi.org/10.1002/hbm.23206

Ordin, M., & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System*, *42*(February), 244–257. https://doi.org/10.1016/j.system.2013.12.004

Pettigrew, C. M., Murdoch, B. E., Kei, J., Ponton, C. W., Alku, P., & Chenery, H. J. (2005). The mismatch negativity (MMN) response to complex tones' and spoken words in individuals with aphasia. *Aphasiology*, *19*(2), 131–163. https://doi.org/10.1080/02687030444000642

Sable, J. J., Gratton, G., & Fabiani, M. (2003). Sound presentation rate is represented logarithmically in human cortex. *European Journal of Neuroscience*, *17*(11), 2492–2496. https://doi.org/10.1046/j.1460-9568.2003.02690.x

Tseng, C.-Y., & Fu, B.-L. (2005). Duration, intensity and pause predictions in relation to prosody organization. *Ninth European Conference on Speech Communication and Technology*, 1405–1408.

Wunderlich, J. L., & Cone-Wesson, B. K. (2001). Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. *The Journal of the Acoustical Society of America*, *109*(4), 1526–1537. https://doi.org/10.1121/1.1349184

Yabe, H., Tervaniemi, M., Reinikainen, K., & Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *NeuroReport*, *8*(8), 1971–1974. https://doi.org/10.1097/00001756-199705260-00035

Yabe, H., Tervaniemi, M., Sinkkonen, J., Huotilainen, M., Ilomoniemi, R. J., & Näätänen, R. (1998). Temporal window of integration of auditory information in the human brain. *Psychophysiology*, *35*(5), S0048577298000183. https://doi.org/10.1017/S0048577298000183

APPENDIX A

Table A. The table below shows the results of pairwise comparisons using a paired t-test, evaluating the difference in mean amplitude for each deviant depending on the block it was presented in during Experiment 2. Each block was devised to present a timing deviant and a frequency deviant. In the Mixed block, both timing deviants and the frequency deviant were presented.

Paired *t*-test comparing peak mean amplitude between blocks

Deviant type	*t*	*df*	*Conf. low*	*Conf. high*	*p*	*Significance*
Early vs. Early_Mixed	-2.05	17	-1.68	0.023	0.06	-
Late vs. Late_Mixed	0.015	17	-0.80	0.81	0.99	-
Freq_Early vs. Freq_Late	0.38	17	-0.61	0.88	0.71	-
Freq_Early vs. Freq_Mixed	1.43	17	-2.1	1.-9	0.17	-
Freq_Late vs. Freq_Mixed	1.21	17	-0.23	0.83	0.24	-

Note: mean ± standard deviation (s.d.)

- Indicates no significance p > 0.05

* Indicates significance p < 0.05

APPENDIX B

Table B. The table below shows the results of pairwise comparisons using a paired t-test, evaluating the difference in mean accuracy scores between the different Target Positions.

Pairwise comparisons of mean accuracy by Target Position

Comparison	*n*	*t*	*df*	*p*	*Significance*
First/Second	80	0.372	79	0.071	-
First/Third	80	-2.07	79	0.041	-
First/Fourth	80	-6.85	79	1.43E-09	**
Second/Third	80	-2.57	79	0.0012	-
Second/Fourth	80	-8.05	79	6.90E-12	**
Third/Fourth	80	-5.31	79	9.70E-07	**

Note: Alpha adjusted (Bonferroni correction applied)

- Indicates no significance p > 0.008

* Indicates significance p < 0.008

** Indicates significance p < 0.0001

Table C. The table below shows the results of pairwise comparisons using a paired t-test, evaluating the difference in mean RT between the different Target Positions.

Pairwise comparisons of mean accuracy by Target Position

Comparison	*n*	*t*	*df*	*p*	*Significance*
First/Second	80	-1.99	79	0.0500	-
First/Third	80	0.92	79	0.3620	-
First/Fourth	80	2.15	79	0.0350	-
Second/Third	80	2.67	79	0.0090	-
Second/Fourth	80	3.74	79	0.0003	*
Third/Fourth	80	1.18	79	0.2440	-

Note: Alpha adjusted (Bonferroni correction applied)

- Indicates no significance p > 0.008

* Indicates significance p < 0.008

[i] *Artifacts within the data*

The responses to the Early timing deviant for both Complex waves and Syllables show widespread negativity during the pre-stimulus baseline, carrying over into the first 100 ms of the post-stimulus time window. This type of ‘unstable’ baseline is not observed for either the Frequency deviant, or the Late timing deviant, suggesting, perhaps, that it is an artifact that is unique to the Early timing deviant itself. As the Early timing deviant is presented earlier than expected within a stream of standards, the presentation of this deviant would be 450 ms after the onset of the last standard. Or, considering that each syllable is 150 ms in length, only 300 ms after the end of the last standard. The pre-stimulus baseline for the Early timing deviant, therefore, is possibly reflecting activity elicited to the standard reliably enough that the grand average across all participants does not eliminate it.