Maturation of Speech Discrimination and Attentional Requirements in Late Childhood

Judith Ann Iannotta

Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Psychology Commons

Recommended Citation
https://academicworks.cuny.edu/gc_etds/979

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
Maturation of Speech Discrimination
and Attentional Requirements in Late Childhood

by

Judith A. Iannotta

A dissertation submitted to the Graduate Faculty in the Program of Speech and Hearing Sciences
in partial fulfillment of the requirements of the degree on Doctor of Philosophy, The Graduate
School and University Center at the City University of New York.

2015
This manuscript has been read and accepted for the Graduate Faculty in Speech-Language-Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Valerie Shafer

Date

Chair of Examining Committee

Klara Marton

Date

Executive Officer

Brett Martin

Hilary Gomes

John Preece

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

Maturation of Speech Discrimination
and Attentional Requirements in Late Childhood

by

Judith A. Iannotta

Advisor: Valerie Shafer

The ability to perceive speech sounds and contrasts continues to be refined throughout the course of development. While emerging models suggest that development is characterized by shifts from an attentionally demanding mode of processing speech sounds to one that occurs relatively automatically, the specific developmental time-course of these changes remains unclear. The present work reports the findings of two experiments that aimed to provide insights into the time-course by which neural processes underlying speech discrimination in children and adolescents becomes automatic. The experiments used event-related potentials (ERP) measures, with a particular focus on mismatch negativity (MMN) - a developmentally-sensitive index of automatic speech discrimination.

The first experiment focused on children ages 6.0-11.9 years old, comparing the MMN responses elicited by English vowel contrasts under differing attentional conditions in an oddball design, to those observed in adults. Specific attentional conditions included: 1) an auditory-attend condition during which listeners silently counted deviants, 2) an auditory-ignore condition in which listeners ignored auditory stimuli while they being required to solve mathematical equations, and 3) an auditory-ignore condition in which listeners ignored auditory stimuli while passively viewing a silent video. Speech perception was hypothesized to be more automatic in children than adults; accordingly, the MMN was predicted to be more sensitive to manipulations of attention in children than in the adults. Consistent with our predictions, attention-related
modulation (i.e., auditory-attend vs. -ignore) of the MMN observed in frontal and central leads was greater in children than adults. Of note, despite obvious differences in the attentional demands of the two ignore conditions (passive, math), the modulation of the MMN produced by the two conditions differed minimally.

The second experiment focused on children ages 10.9-16.9 years old, again comparing their responses to adults. We hypothesized that the maturation of speech discrimination processes would still not be complete in this group and thus expected to find of continued attentional dependencies, in comparison to adults Given the relative lack of differences in modulation produced by the two ignore conditions under experiment 1, we refined our paradigm to include two ignore conditions that differ on a specific cognitive construct – working memory demands. This was accomplished by having participants perform either a 0- or 2-back during the auditory-ignore conditions. Consistent with our prediction, we found greater attention-related modulation of the MMN in the left inferior and anterior pole regions for the child group relative to adults; analyses treating age as a continuous variable were supportive of such distinctions. We also found evidence of continued age-related difference in the late discriminatory negativity (LDN) and P3b.

Together, these two experiments highlight the value of expanding the scope of examination for speech discrimination to consider a broader range of ages than leading models, which tend to emphasize early life. Our findings of continued age-related changes in the MMN, P3b and LDN during adolescence highlight the need to consider the effects of both bottom-up and top-down attentional influences in developmental models of speech perception.
Dedication

This work is dedicated in loving memory to my father, Patrick J. Iannotta who encouraged me to follow my dreams and follow my bliss.
Who shared with me his love of knowledge and showed to me, through his actions, the power of perseverance.
Acknowledgements

To my mentors and guides, thank you for teaching me and helping me on this journey.

To my friends and colleagues, especially Dr. Rebekah Buccheri-Kallas and Emily Zane,
Thank you for your guidance, friendship, and refusal to let me run away!

I want to offer a special show of gratitude to my mother Joan, for her unwavering belief in me.
To my siblings, for cheering me on and being my best friends.
You make my every day beautiful and I am so fortunate to have you in my life.

To my Grace Face and Patrick Henry, for bringing pure joy into my life and heart. Every day.

And to my MM
Thank you for making all of my dreams come true.
ILYWAMHAAF
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>i.</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii.</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv.</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v.</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>vi.</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi.</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi.</td>
</tr>
<tr>
<td>List of Supplementary Figures</td>
<td>viii.</td>
</tr>
<tr>
<td>Chapter 1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Learning &amp; Attention in Classic Models of Speech Perception Development</td>
<td>1</td>
</tr>
<tr>
<td>1.3 A Primer on Automaticity</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Towards An Automaticity-based Perspective of Speech Perception</td>
<td>8</td>
</tr>
<tr>
<td>1.5 Probing the Automaticity of Speech Perception with EEG</td>
<td>11</td>
</tr>
<tr>
<td>1.6 Mapping the Development of Automatic Speech Perception</td>
<td>13</td>
</tr>
<tr>
<td>1.7 Remaining Obstacles</td>
<td>17</td>
</tr>
<tr>
<td>1.8 Summary</td>
<td>19</td>
</tr>
<tr>
<td>1.9 The Present Study</td>
<td>19</td>
</tr>
<tr>
<td>Chapter 2. Experiment 1</td>
<td>20</td>
</tr>
<tr>
<td>2.1 Methods</td>
<td>21</td>
</tr>
<tr>
<td>2.1.1. Participants</td>
<td>22</td>
</tr>
<tr>
<td>2.1.2. Stimulus Materials</td>
<td>22</td>
</tr>
<tr>
<td>2.1.3. Paradigms and Procedures</td>
<td>23</td>
</tr>
<tr>
<td>2.1.4. Electrophysiological Recordings</td>
<td>23</td>
</tr>
<tr>
<td>2.1.5. Experimental Design</td>
<td>24</td>
</tr>
<tr>
<td>2.1.6. Data Analysis</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Results</td>
<td>25</td>
</tr>
<tr>
<td>2.3 Discussion</td>
<td>34</td>
</tr>
<tr>
<td>2.4 Limitations of the First Experiment and Considerations</td>
<td>37</td>
</tr>
</tbody>
</table>
List of Appendices
Appendix 1. Primary Analysis with Age Treated as a Dimensional Variable __________ 81

List of Tables
Table 1. Waveforms to the standard stimuli in the attend condition at Cz ____________ 51
Table 2. Repeated measure ANOVAs for the 100-200ms, 200-300ms, and 300-400ms time windows ________________________________ 105

List of Figures
Figure 1a. Children Grand Means to standards and deviants averaged across all conditions at C3, Cz, C4, F3, Fz, and F4. ____________________________ 29
Figure 1b. Adult Grand Means to standards and deviants averaged across all conditions at C3, Cz, C4, F3, Fz, and F4. ____________________________ 30

Figure 2a. Children’s subtraction waves (deviant minus standard) at C3, Cz & C4 for attend, passive-ignore and math-ignore conditions. ____________________________ 31
Figure 2b. Children’s subtraction waves (deviant minus standard) at F3, Fz & F4 for the attend, passive-ignore and math-ignore conditions. ____________________________ 32

Figure 2c. Adult subtraction waves (deviant minus standard) at C3, Cz & C4 for the attend, passive-ignore and math-ignore conditions. ____________________________ 33
Figure 2d. Adult subtraction waves (deviant minus standard) at F3, Fz & F4 for the attend, passive-ignore and math-ignore conditions. ____________________________ 34

Figure 3. First component from the PCA for 100-400msec time window. __________ 47

Figure 4. Second component from the PCA for 100-400msec time window. __________ 48

Figure 5. First component from the PCA for 450-600msec time window. __________ 49

Figure 6. Second component from the PCA for 450-600 msec time window. __________ 50

Figure 7a. Waveforms of Stadnard Stimui in the Attend Condition. _______________ 53

Figure 7b. Plot of GFP, Mastoids and Fz with selected time-windows (100 and 450-600ms __ 54
Figure 8. Correlation between GFP and age under each of the attentional conditions (attend, 0-back, 2-back). 55

Figure 9. Child vs. adult responses for attend condition. 56

Figure 10. Child vs. adult responses for 0-back condition 57

Figure 11. Child vs. adult responses for 2-back condition. 58

Figure 12. Condition by age-group interaction for mean amplitude in the 100-400msec time window at LI 59

Figure 13. Child vs. adult responses for attend condition at RP 61

Figure 14. Mean amplitude for the left inferior lead for the attend and 2-back conditions in the 100-400 time window. 62

viii. List of Supplementary Figures

Figure 15. Children’s MMN waveforms at C3, C4 and Cz in the Attend, 0-Back and 2-Back conditions 72

Figure 16. Children’s MMN waveforms at F3, F4 and Fz in the Attend, 0-Back and 2-Back conditions 73

Figure 17. Adult’s MMN waveforms at C3, C4 and Cz in the Attend, 0-Back and 2-Back conditions 74

Figure 18. Adult’s MMN waveforms at F3, F4 and Fz in the Attend, 0-Back and 2-Back conditions 75

Figure 19. Children’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend condition 76

Figure 20. Adult’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend condition 76

Figure 21. Children’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend condition 77

Figure 22. Adult’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend condition 77

Figure 23. Children’s MMN at Fz, Left Mastoid and Right Mastoid in the 2-Back condition 78

Figure 24. Adult’s MMN at Fz, Left Mastoid and Right Mastoid in the 2-Back condition 78

Figure 25. Grand Means Standards and Deviants in the Attend condition at C3, C4, Cz, F3, F4 and Fz 79
Figure 26. Grand Means Standards and Deviants in the 0-Back condition at C3, C4, Cz, F3, F4 and Fz

Figure 27. Grand Means Standards and Deviants in the 2-Back condition at C3, C4, Cz, F3, F4 and Fz

Figure 28. Standard vs. Deviant Responses for Attend condition with standard error at C3, C4 and Cz

Figure 29. Standard vs. Deviant Responses for Attend condition with standard error at F3, F4 and Fz

Figure 30. Standard vs. Deviant Responses for 0-Back condition with standard error at C3, C4 and Cz

Figure 31. Standard vs. Deviant Responses for 0-Back condition with standard error at F3, F4 and Fz

Figure 32. Standard vs. Deviant Responses for 2-Back condition with standard error at C3, C4 and Cz

Figure 33. Standard vs. Deviant Responses for 2-Back condition with standard error at F3, F4 and Fz

Figure 34. Difference waves in the attend condition by age at F3, C3, Fz, Cz, F4 and C4

Figure 35. Grand Means difference waves in 3 conditions at F3, C3, Fz, Cz, F4 and C4

Figure 36. Difference waves in three attentional conditions at Fz and Cz

Figure 37. Difference waves in three attentional conditions at F3 and C3

Figure 38. Difference waves in three attentional conditions at F4 and C4

Figure 39. Grand Means difference waves at F3, C3, Fz, Cz, F4 and C4 with Mastoids

Figure 40. Condition by Time interaction at C3 in the 100-200 msec time window.

Figure 41. Condition by Time interaction at C4 and Cz in the 300-400 msec time window

Figure 42. Children and Adult MMN waveforms in Attend condition at C3, C4 and Cz

Figure 43. Children and Adult MMN waveforms in Attend condition at F3, F4 and Fz

Figure 44. Children and Adult MMN waveforms in 0-Back condition at C3, C4 and Cz
Figure 45. Children and Adult MMN waveforms in 0-Back condition at F3, F4 and Fz 99
Figure 46. Children and Adult MMN waveforms in 2-Back condition at C3, C4 and Cz 101
Figure 47. Children and Adult MMN waveforms in 2-Back condition at F3, F4 and Fz 101
Figure 48. Condition by Time interaction shown at F3 and C3 in the 100-200 msec window 102
Figure 49. Time by Group interaction shown at Fz in the 200-300 msec time window 103
Figure 50. Condition by Time interaction at C3 and C4 in the 200-300 msec time window 103
Figure 51. Condition by Time by Age-group at F3 in the 200-300 msec time window 104
Figure 52. Condition by Time by Age-group at Fz, Cz, C3, C4. In the 300-400ms window 105
Chapter 1. Introduction

1.1 Statement of the Problem

Despite decades of work, the field of speech science remains without a consensus model that explains speech perception and its development throughout childhood and adolescence. While the specific conceptualizations and details of the many models proposed over the years can differ notably, there are two key questions that most attempt to address. First, how learning shapes innate (e.g., genetically determined) auditory perception of speech, and secondly, whether speech perception at the various developmental stages depends on attention as opposed to being subconscious or automatic processing. In the following sections, we provide an overview of key insights from classic models of speech perception development; then we describe an emerging model, which draws upon well-established theories of automaticity and is amenable to behavioral and electrophysiological examination via manipulations of stimulus complexity and task demands. We review experimental support for this model, as well as gaps in the extant literature and next steps towards a more robust characterization of speech perception development.

1.2 Learning & Attention in Classic Models of Speech Perception Development

One prominent model of speech perception is the *Word Recognition and Phonetic Structure Acquisition model (WRAPSA)* by Jusczyk (1997). This model details a sequence of skills and abilities acquired by infants to explain the development of speech perception in the first year of life. Central to this is the notion that infants innately possess “auditory analyzers” that are capable of encoding most acoustic details from the speech signal. This initial “default” analysis is thought to be broadly tuned and lacking language-specific refinement in early infancy which could explain how infants are capable of detecting sounds of many languages (Werker & Tees,
1984) but fail to recognize subtle but meaningful acoustic differences that adults are capable of distinguishing (Aslin, Pisoni, Hennessy & Perey, 1981).

Balancing the concept of innate abilities with learned ones, WRAPSA proposes a structure of the subconscious mechanisms needed for perceptual learning. This includes an experience-dependent “cue weighting” system and a pattern recognition system that, together, account for the attunement of infants to sounds of their ambient language (Jusczyk, 1995; Kuhl, 2004; Werker & Tees, 1984). The term “weight” reflects the relative amount of perceptual attention directed toward specific acoustic information (Nittrouer, 2002; pp. 718). The cue weighting system is thought to statistically analyze incoming signals, weight the information and develop weighting “schemes” that will be used to emphasize and enhance the detection of features critical for word distinction. These features are then given greater weight in the perceptual system, leading to a refinement of the analyzers and thus shaping future perception.

Werker & Tees’ 1984 Hindi study demonstrated experientially-related refinement by showing that very young infants initially showed sensitivity to both dental /d/ and retroflex /D/ regardless of their ambient language (English or Hindi). By the end of the first year of life however, the sensitivity faded for those whose language (English) considered the contrast to be allophonic but was retained for those whose language (Hindi) recognized it as a meaningful distinction. Complementing the weighting system is the pattern recognition system, which is employed to extracts higher order patterns from inputs and compares them with word templates and exemplars that are stored in the lexicon. This process is strongly related to language experience, as templates are strengthened based on repeated presentations of language input (Jusczyk, 1999; Mattys & Jusczyk, 2001). Once a template is established, subsequent input that
matches an exemplar on critical levels is recognized as the same form or “object”, while those that do not will be perceived as a different form or tagged as lacking meaning.

A limitation of the WRAPSA model is that it does not explicitly address the attentional requirements for each of the systems included. Given the innate nature of input analyzers, they can be considered to act independent of attention as it is suggested that a child does not intentionally attend to these features. Similarly, the cue weighting system also can be conceptualized as acting without attention, as this system facilitates the identification of speech sounds, which can serve to activate or capture attention. Regarding pattern recognition, while Jusczyk does not specifically state that attention is required, he does suggest that pattern detection is greatly enhanced with attentional focus; he also posited that listeners learn to develop attentional routines to facilitate detection of probabilistic information about sound patterns and likely word forms.

Another highly influential model is the *(processing Rich Information from Multidimensional Interactive Representations model (PRIMIR)* by Werker and Curtin (2005). Like WRAPSA, PRIMIR identifies several components or “planes” of perceptual development, including probabilistic learning, similarity detection, and experiential influences. However, unlike WRAPSA, it does not suggest that any of these planes are present at birth. Instead, the model posits that the components needed for effective perception are the result of a statistical learning process, in which biases are established across the different planes based on regularities in language input. Learning initially focuses on general perception of auditory information, and then is refined to establish phonemic representations, with specificity for sounds from the ambient language emerging over time. The general statistical learning mechanisms specified by PRIMIR provide layered simultaneous analysis of acoustic information, prosodic features, segmentation of
the speech signal, syllable extraction, phoneme establishment, and storage of word forms, with each having influence over the evolution of the others by enhancing or damping signal saliency.

The PRIMR model emphasizes the role of attention in facilitating learning processes. Attentional processes are guided by three dynamic filters that are posited to work and evolve together to develop and refine language-specific speech perception abilities for later language learning. The filters include initial biases of the child, the developmental level, and requirements of the specific language task at hand. The filters interact with one another to direct attention to different cues based upon maturation and task-demands.

Similar to WRAPSA and PRIMIR, the Native-Language Magnet Theory (NLM) (Kuhl, 1993; Kuhl & Iverson, 1995), sought to address speech perception development by merging innate abilities and language experience in the formation of perception, through a more distinct framework. The NLM theory postulates that all objects in the perceptual space are analyzed and deemed either relevant or irrelevant to the ambient language experience; attention is not considered to be a necessity for this process. Unlike WRAPSA, which posits that listeners store specific instances of what they hear, NLM suggests that listeners store relevant contrasting objects as phonetic prototypes, each with a specific “neighborhood of influence”. At the center of each phonemic category is the object prototype, which is deemed to be the “best exemplar” and acts as a magnet for other exemplars that closely conform to its relevant features. As experience with language increases, there is a learned refinement of the magnet effect, allowing for more accurate category attraction and specific category development. Non-prototypical objects are not acted upon by the magnet effect and are therefore not perceived as an exemplar of that specific target, creating the potential for the formation of a new phenotypic category as repeated instances occur.
Unlike the NLM model, Best’s Perceptual Assimilation Model (PAM) (1995) conceptualizes speech perception to be a highly active process, requiring the listener to seek out information about vocal tract gestures that form specific phonetic contrasts. This view suggests that perception is a purposeful activity by which the listener must attend to a speech signal to gain information about vocal tract positions in order for effective perception and eventual production of the sound(s). One’s perceptual goal of understanding these gestures focuses attention to auditory objects of interest that will be selected form the speech signal and subjected to further analysis. This model is similar to both Fowler’s Direct Realist Theory of Speech Perception (1986) and Lieberman & Mattingly’s (1986) Motor Speech Theory of Speech Perception insofar as they all suggest that speech events are perceived as being articulatory in nature rather than acoustic or auditory. An implicit assumption of these models is that speech perception is attentionally dependent, as a listener must attend to the utterance’s form in order to approximate the physical gestures.

From this brief review, the complexity of speech perception, and its development is evident. While each of the models proposed a unique set of organizing principles and specific mechanisms for learning, they tended to vary with respect to their specification of attention; some appear to suggest that attention is beneficial though not necessary (e.g., WRAPSA, NLM), while others have more obvious dependency (e.g., PAM). None of these models provide for a comprehensive or quantifiable understanding of the interplay of attention and learning throughout the course of development.

1.3 A Primer on Automaticity

Rooted in cognitive psychology, information processing theory attempts to explain how humans are able to learn and master sophisticated skills by inferring how changes in theoretical mental
structures result in the ability to learn new information and execute new behaviors (Eggen & Kauchack, 2007). Central to information processing models is the notion that in order to master a sophisticated skill, an individual must first master the incremental sub-processes that support it. More specifically, the theory attempts to explain how input from the environment goes through the processes of perception, attention, and storage (Eggen & Kauchack, 2007). Focused on fundamental principles, information processing theory has been applied to the study of learning and behavior in a range of areas, including psychology (Forgas & George, 2001) linguistics (Nyikos & Oxford, 1993; McLaughlin, Rossman & McLeod, 1983), technology (Hann, Hui, Lee & Png, 2007), business management (Roger, Miller & Judge, 1999), education (Mayer, 1996), and literacy (LaBerge & Samuels, 1974).

LaBerge & Samuels (1974) proposed a highly influential information-processing model of complex skill learning focused on reading. They outlined a series of processing steps involving visual, phonological and episodic memory (e.g., combining letters, mapping graphemes to phoneme, transforming phonemes to semantic representations), which combine together to support the transformation of visual stimuli into semantically meaningful information. Assuming that each of the processing steps is the product of learning, they argued that reading fluency depends not only on the accuracy of the steps carried out, but their “automaticity”. A central tenet of their model was that the automation of lower-order operations or skills is necessary to allow for the allocation of time and resources to more sophisticated, higher-order comprehension skills.

For more than three decades, the concept of automaticity has remained a central focus in cognitive psychology, as researchers have worked to operationalize it and apply it to the broader domain of skills learning, beyond reading. Posner and Snyder (1975) highlighted that in order for
a process to be automatic, it should occur with or without intention or awareness. Automatic processes are thought to result from being highly over-practiced and do not, on their own, result in the learning and storage of novel information (Shiffrin and Schneider, 1977). Once automatized, there will be little change with increasing age or changes in mental state; Bargh (1989) went so far as referring to them as “chronically accessible” Viewing automaticity in terms of a memory encoding process, Hasher and Zacks (1978) suggested that the encoding of aspects or attributes of inputs for the process must be automatic; they also reinforced the notion that process must be efficiently performed incidentally or intentionally with little to no improvement made with intention or practice. Importantly, Posner and Snyder, 1975 argued that a truly automatic process should not be vulnerable to interference from simultaneous processes once the process has begun. Such processes require minimal cognitive resources, do not require conscious monitoring (Bargh, 1992) and are not vulnerable to interference from other ongoing operations, unless the concurrent process is competing for input or output channels (Bargh, 1994). As such, an automatic process can actually become a source of interference when not relevant to task performance (e.g., the Stroop Task). Bargh (1994) attempted to summarize the literature by putting forth a set of criteria upon which the automaticity of behaviors could be judged, which includes: awareness, intentionality, efficiency and controllability.

Not surprisingly, the concept of automaticity has emerged in the realm of speech perception - the first step in spoken communication. Understanding how speech perception becomes automatic begins with explaining the underlying perceptual skills necessary for negotiating complex speech signals in a continuous signal (e.g. establishment of phonemes and their boundaries) so that automatic retrieval of word patterns in speech streams can evolve.

1.4 Towards An Automaticity-based Perspective of Speech Perception
While a number of studies claim that speech perception is automatic (Johnson & Ralston, 1994; see Shriﬀen & Schneider, 1984), they do not provide insights into how automaticity develops. As highlighted by the WRAPSA model, which focuses on the ﬁrst years of life, the elements for speech perception may be present at birth, but automatic processing is not. Instead, children are thought to learn to automatically attend to relevant contrasts through experience with the attentional process itself being affected by language learning. Detection of language-speciﬁc contrasts therefore, becomes automatic only after some level of learning has taken place. This is in line with Posner who characterized learning as a highly attention-dependent process in which information must be attended to and prioritized before subsequent memory formation, or learning, can occur.

In recent years, the Automatic Selective Perception (ASP) model of speech perception (Strange, 2010; Strange & Shafer, 2008) has emerged as a solution to the challenges of operationalizing and quantifying interplay of attention and learning in speech perception, and its development. The ASP attempts to explain the reﬁnement of language perception in terms of the reorganization of attention to language-speciﬁc contrasts. Within this framework, speech perception is thought to be a purposeful activity, in which individuals attend to an incoming speech signal in order to perceive intended messages. After extensive experience attending to sounds of an ambient language, individuals weight acoustic information and form Selective Perception Routines (SPRs). Once established, listeners can detect language-speciﬁc sequences, automatically.

According to the ASP model, inexperienced language learners, including children, must purposefully allocate attention to speech signals until SPRs are formed and perception is automatized. It is at this point that attention is no longer required for the detection of language-
specific objects by the SPR. Frequent repeated exposure will strengthen and refine the SPRs leading to over-learning of an object (Crick and Koch, 1990) and eventually to a reciprocal relationship of increased SPR strength and decreased need for attention. SPR “strength”, for the purpose of this study, will refer to the perceptual salience of an auditory signal. Unlike auditory salience, which refers to the size of the physical differences between phonetic contrasts, perceptual salience refers to the physiological and behavioral responses to language-specific contrasts that occur as a function of linguistic experience (see Strange, 2010 for a review). Highly perceptually salient auditory objects attract attention and eventually, do not require focused attention for their perception (Koch, 2004).

Speech perception studies show that both age and language experience contribute to the perceptual salience of phonetic contrasts (Nittroeur & Miller, 1997; Nittroeur, 2002; Shafer, Yu, Datta, 2010; Sundara, Polka, Genesee, 2006). Additionally, Sundara, Polka, and Genesee (2006) found that four year-olds discriminate the English contrast /d-ð/ better than 10-12 month-old children, but poorer than adults and that participants who lacked meaningful experience with the contrast showed poorer discrimination overall regardless of age. Nittroeur and colleagues also provide evidence of developmental changes in perception as they found that children aged 4-8 years-old do not make use of acoustic features the same way adults do when identifying phonemes in a speech signal (Nittroeur and Miller 1997;Nittroeur, 2002). When examining weighting routines during a fricative judgment task, 4-7 year-old children and adults were asked to identify fricatives as being [s] or [ʃ] when followed by an [a] or [u]. An age-related trend was found in which children assigned more weight to the vocalic formant transitions and less to fricative noise when compared to adults. A gradual and reciprocal change in this weighting strategy was found as the age of the children increased, revealing developmental perceptual
changes throughout childhood that have not reached adult-like status by age 7 or 8 (Nittrouer and Miller, 1997; Nittrouer, 2002). It has been suggested that perceptual attention is initially focused on formant transitions that provide information about vocal tract changes. As age and experience increased, the focus then shifts to more subtle acoustic properties that do not involve spectral change, such as silent gaps that specify periods of vocal-tract closure; or periods of stable spectral information that specify place of consonantal constrictions (Nittrouer, 2002).

The notion that language experience, rather than developmental stage alone, shapes perceptual routines is supported by cross-language findings. For example, adult native Japanese speakers were found to have difficulty categorizing and discriminating the English [r-l] distinction, since it is not phonemically contrastive in their native language (Strange, 1995; Strange, 2010). Such difficulties are not present in Japanese infants, who show good perceptual discrimination of this contrast at 6 to 8 months, though intriguingly not at 10 to 12 months of age. Overall, these findings indicate a decline in discriminative sensitivity to this non-native contrast (Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Morio, Menyuk & Best, 1994).

Mounting evidence shows that experiential influences (Best, 1995; Hisagi, Shafer, Strange & Sussman, 2010; Jusczyk, 1999; Mattys & Jusczyk, 2001; Strange, 1995; Starnge & Shafer, 2008; Werk & Tees, 1984) attentional processes (Gomes, Molholm, Ritter, Kurtzberg, Cowan & Vaughan, 2000; Hisagi & Strange, 2011; Krause, 1992; Kraus 1993; Sussman, Ritter & Vaughan, 1998; Hisagi, Shafer, Strange & Sussman, 2010) and maturational factors (Shafer, Yu & Datta, 2011) affect speech perception, leading to increased interest in developing a more complete understanding of speech perception development considering these factors, as well as their interactions. A key challenge to such pursuits is the need for objective measures of speech processing that can be measured without confound throughout development. While behavioral
observations can prove useful in this regard (e.g., comparison of performance outcomes under different attentional conditions), they are not without limitation. In particular, it can be difficult to compare, or equate, behavioral measurements across ages (Bishop, Hardiman, & Barry, 2011) and clinical populations. Additionally, behavioral indices may lack the depth required to capture the complexity of neural phenomena underlying speech processing. Fortunately, electrophysiological tools have emerged as a means of providing objective characterizations of component processes underlying speech processing, and appear to be able to successfully index SPR development.

1.5 Probing the Automaticity of Speech Perception with Event-Related Potentials

Electroencephalography (EEG) is used to non-invasively examine electrical activity of large groups of neurons in the brain. EEG responses are electrical deflections (in the range of microvolts) that occur in response to a stimulus and are detected by surface-electrodes at the scalp. The selective averaging of EEG responses across events of a given type results in event-related potentials (ERPs), which reveal time-locked neural processing (Duncan, Barry, Connolly, Fischer, Michie, Näätänen, Polich, Reinvang, Van Petten, 2009; Klix, Näätänen, & Zimmer 1985; Luck, 2005; Näätänen, 1990).

Of particular relevance to speech perception development is the Mismatch Negativity (MMN) (Näätänen, 1990), an ERP component that is thought to reflect automatic detection of changes in the auditory signal. The MMN is elicited when a frequently occurring auditory *standard* is interrupted by an infrequently occurring *deviant* (Luck, 2005; Näätänen, 1990; Näätänen, Tervaniemi, Sussman, Paavilainen, Winkler, 2001), and is typically described in terms of *latency* and *amplitude*. Latency refers to the period of time between the presentation of the deviant stimulus and the onset of the MMN, while amplitude refers to the peak difference between the standard and deviant stimulus (Amenedo & Escera, 2000). The MMN is commonly
examined using a classic oddball paradigm, in which a repeated sequence of standard stimuli is infrequently interspersed with deviant stimuli. Participants are typically instructed to ignore auditory stimuli and focus on some other activity, such as watching a silent move, reading a book, or performing a visual task (Muller-Grass, Macdonald, Schröger, Sculthorpe, Campbell, 2007; Muller-Gass, Stelmack, Campbell, 2005; Naatanen, Paavilainen, Rinne, & Alho, 2007; Winkler, Karmos, Näätänen, 1996). Central to its utility in the study of automaticity, the MMN is believed to be the product of a pre-attentive, automatic detection system (Naatanen, 1992; Naatanen, 1990; Naatanen, Gaillard, & Mantysalo, 1978; Sculthorpe, Collin, Campbell, 2008).

In considering the significance of the MMN, two competing interpretations have emerged. First, the model adjustment hypothesis postulates that the MMN is an error detection signal that results from a deviation in an established auditory standard. Once regularity in a signal is detected, a predictive model of future inputs emerges. Any unexpected violation of the anticipated standard generates a mismatch response, which triggers online adjustments of the model. In contrast, the adaptation hypothesis argues that the MMN is actually reflects local neural adaptation to the standard stimulus, which causes an attenuation and delay of the obligatory N1 response associated with early auditory processing. It is this N1 differential that the adaptation hypothesis suggests is reflected in the MMN.

While the exact origins of the MMN are not definitively known, the EEG, MEG and fMRI literatures appear to be converging on two principle generators – one located in supratemporal cortices and the other in right prefrontal cortex. The respective role of the two generators can be explained in terms of the model adjustment mode. The temporal lobe based generator is thought to be responsible for the automatic detection of deviations between an incoming auditory stimulus and either memory traces of preceding stimuli, or predictions of future stimuli based upon past.
The higher order prefrontal generator is believed to be responsible for the redirection of attention to an auditory stimulus when the lower-order temporal generator detects a deviation.

Importantly, while the MMN is thought to be independent of attention, the ability to discriminate sound contrasts is not necessarily so. It has been reported that that MMN elicitation is only significant when the standard and deviant stimuli are behaviorally discriminable (Martin, Kurtzberg & Stapells, 1999; Kraus et al. 1996). However, studies have shown that with repeated exposure, difficult to discriminate, unfamiliar auditory stimuli can eventually evoke an MMN (Näätänen, Schroger, Karakas, Tervaniemi, Paavilainen, 1993) suggesting learning or sharpening of the sensory system.

1.6 Mapping the Development of Automatic Speech Perception

The potential value of the MMN for efforts to study the interface between attention and the development of automaticity comes from findings that attentional processes can enhance the MMN (Datta, Shafer, Morr, Kurtzberg, & Schwartz, 2010). For example, a study by Gomes et al. (2000) compared children (ages: 8-12 years old) and adults with respect to attentional modulation of the MMN, using a paradigm in which a standard tone (1000Hz) and three deviants (easy: 1500Hz; medium: 1200 Hz and difficult: 1050Hz) were presented in an oddball design under three conditions of attentional focus (ignore, attend, ignore again). While adults exhibited robust MMN’s for all conditions, with minimal evidence of attentional modulation, the MMN’s in children were significantly modulated by attention for difficult-to-detect deviants. The presence of significant attentional effects in children, but not adults, supports the claim that perception for some auditory contrasts is automatic in adulthood but not yet in childhood. Datta et al. (2010) also provided evidence of attentional modulation of the MMN in children, with peak responses
occurring 50 ms earlier in the attend condition for easy-to-detect vowel contrasts, than in the ignore condition.

In an effort to provide a more comprehensive characterization of speech perception development, Shafer and colleagues examined the mismatch responses (MMRs) in children between 2 months and 11 years old, finding the MMN peak shifted earlier by approximately 11 milliseconds (ms) per year (Morr, Shafer, Kreuzer, Kurtzberg, 2002; Shafer, Morr, Kreuzer, Kurtzberg, 2000). At age 11, adult-like MMN latencies had not been reached, suggesting the protracted nature of speech processing refinement. In addition, a clear MMN was not observed to the more difficult tone contrast before four years of age (although a positive mismatch response was observed). Similarly, MMN was not clearly present until four years of age to a vowel contrast (Shafer, et al., 2010; 2011). Not surprisingly, these various findings have led several studies to consider the potential value of the MMN in identifying and tracking neurodevelopmental disturbances related to language.

While the primary focus of our effort was to map the development of speech perception using the MMN, a number of other ERP components have been shown to be developmentally sensitive and relevant to speech perception. Here, we provide a brief overview of these components; additionally, secondary analyses in the proposed experiments will reference developmental changes seen in these indices, when relevant.

**P100 (P1).** The P100 is an obligatory ERP component that occurs in response to the detection of an auditory stimulus (Näätänen and Picton, 1987), which can be modulated by attention and is developmentally sensitive (Näätänen et al., 1978; Picton and Hillyard, 1974). The component typically manifests as a positive deflection observed between 80 and 140 ms post-stimulus. Evidence for attentional modulation of the P100 comes from studies such as that of
Mangun & Hillyard (1990), who found that when identical stimuli were presented to either an attended or unattended channel, that the P100 in response to the attended channel was larger. From a developmental perspective, the P100 is not detectable early in life. The component initially manifests around 6 months of life, though as the P150, which persists into early childhood (approximately 6 years old). The P100 observed in children is thought to develop in the adult P1 with a steady decrease in latency and amplitude until 20 years of age (Ponton, Eggermont, Kwong and Don, 2000). There is some lack of consensus regarding the developmental phenomenology of the P100, as some have suggested that age-related differences in the waveform may be less related to P1 maturation, and more reflective of amplitude increases in the subsequent N1 (Ponton et al, 2000).

N2b. The N2b, a subcomponent of the N200, is a negative deflection occurring between 180 and 325 milliseconds after presentation of the deviant stimulus (Patel & Azzam, 2005). In contrast to the MMN (a.k.a. the N2a), which is considered to be an index of automatic processing, the N2b is thought to index voluntary processing. This N2b component is elicited when individuals detect a deviation of an infrequent or low-frequency stimulus that is being attended to and believed to reflect detection of mismatch between the actual stimulus and the expected stimulus. The N2b is commonly observed for deviations in auditory, semantic, and orthographic information when the specific dimension is being attended to in a selective attention oddball paradigm. The greater the difference between a standard and deviant stimulus, the larger the N2b response observed. A distinguishing property between the N2b and the MMN, is that the MMN exhibits polarity reversals in the mastoid lead, but the N2b does not (Näätänen et al., 1993). As recently reviewed by Patel and Azzam (2005), age effects for the N2b have been reported throughout childhood into adulthood. For example, in a visual color deviation task, decreases in
response time, error rates and in N2b latency were shown with increasing age, from 7-24 years old (Van Der Stelt, Smulders, 1998); this was interpreted as being suggestive of increasing cognitive and visual discrimination with development. At the other end of the lifecycle, increases in N2b latency are observed in the elderly, possibly reflecting the compromises in attentional control that are characteristic of aging.

**P300:** There are two P300 subcomponents, the P3a and the P3b. The labels P300 or P3 are commonly used to refer to the P3b, while the P3a requires specific designation. The P3a is a positive going waveform occurring around 250-300ms after stimulus presentation, which is thought to reflect an involuntary shift to novel changes in the environment (Polich, 2003). In contrast, the P3b (henceforth referred to as P3) is a large positive-going wave that peaks around 300-500 ms after stimulus presentation. While the P3a is elicited by infrequently occurring deviants in unattended conditions, the P3 is thought to be elicited by infrequently occurring stimuli that are relevant to an intentional/focused task. There is a direct relationship between how improbable the attended target is and the magnitude of the P3 response, with more rare targets eliciting bigger P3 responses (Donchin, 1981; Escera et al., 2000; Gumenyuk, Korzyukov, Alho, Escera, Schroger, Ilmoniemi, Naatanen, 2001.). Morphologically, the P3 tends to be relatively broad waveform, while the P3a is notably more narrow. P3a amplitudes tend to be maximal over frontal/central sites on the scalp, such as Fz or Cz, while P3 amplitudes are generally greater at sites like Pz (Comerchero, Polich, 1999). Of note, the P3a is thought to reflect more than simple deviance detection. Rather, it is thought to index further evaluation of novelty to determine if it is of relevance to task performance. This theory is supported by studies that report increased response times after a stimulus elicited a P3a, suggesting that the deviant triggered an involuntary attention switch, and evaluation of the stimulus to determine the relevance of the deviance for
current task performance (Escera, Alho, Schroger, Winkler, 2000; Woods, 1992). The greater the deviance, the larger the P3a amplitude (Yago et al., 2001). Developmentally, the P3 tends to increase throughout childhood into adulthood, with decreases emerging in advanced aging.

**LDN.** The late difference negativity (LDN), also referred to as the ‘late MMN’ (Neuhoff, Bruder, Bartling, Warnke, Remschmidt, Müller-Myhsok, Schult-Körne, 2012), is a negative component occurring between 300-550 milliseconds after the onset of rare stimulus (Bishop, Hardiman & Barry, 2011). The LDN tends to be more prominent for speech sounds than non-speech sounds, and tends to be larger in children than adults. Inspired by their findings that the LDN is larger in small deviant stimuli than large deviant stimuli, Bishop et al. (2011) noted that the LDN should not be viewed as simply a later version of the MMN. Instead, they argued that it appears to reflect additional processing demands associated with salient features of a stimulus that are harder to detect; this may also explain why children, who are less experienced with speech perception, exhibit larger LDN responses than adults.

1.7. Remaining Obstacles

While MMN-based approaches to probing automaticity in speech perception have gained significant popularity, there is a continued need for methodological refinement. A number of studies have failed to produce attentional modulation of auditory discrimination as indexed by the MMN, raising concerns about sensitivities to the specific conditions and stimuli employed. For example, Shafer, Morr, Datta, Kurtzberg, & Schwartz (2005) found no evidence of attentional modulation of the MMN in 8-10 year old children in an experiment using [I] and [e] vowel contrast; interestingly, the same laboratory later found evidence of attentional modulation when longer duration versions of these vowels (Datta, et al., 2010) were employed. Disparities in findings across studies suggest the complexity of the process(es) indexed by the MMN. Although
defined based upon stimulus deviance, the MMN is believed to reflect the hierarchical relationship between top-down intentions and bottom-up perceptions (Garrido, Kilner, Stephan, & Friston, 2009; Grimm & Schroger, 2007).

A number of experimental factors have been identified across studies, which may account for some of the variation in findings. In particular, stimulus complexity tends to vary on a variety of dimensions from one laboratory or study to the next (e.g. tones vs. syllables, synthetic vs. natural speech, easy vs. difficult contrasts). Familiarity with the stimuli and contrasts employed can affect the strength of bottom-up processes and thus their susceptibility to top-down modulation. Additionally, task instructions vary substantially, with some studies requiring participants to detect single-stimulus deviants, while others require more complex pattern/rule detection across multiple stimuli.

Most germane to the present work, is the fact that the vast majority of studies make use of a “passive-ignore” condition as the low demand condition in their study of attentional modulation. This can be particularly problematic, as there is no way to ensure that participants are actively allocating their attention away from the stimulus; such a situation can be especially problematic in younger populations, who are generally less prone or able to comply with instructions.

Beyond the challenge of how to optimize MMN-based approaches to ensure reliability and reproducibility across studies, an important gap remains in the study of early to middle adolescence, where the auditory processes indexed by the MMN is thought to reach adult levels. This is reflective of the tendency of studies to compare early and middle childhood to adulthood. Few studies have mapped out the MMN to speech contrasts during adolescence when it is expected to reach adult levels. Those that have, tended to rely on synthetic stimuli and traditional “passive” ignore conditions, raising concerns about their generalizability and completeness.
1.8 Summary

In summary, although encouraging, with respect to validation of automaticity-based models of speech perception, current efforts to map the trajectory of speech perception development via the MMN are incomplete for a number of reasons. First, they tend to make use of suboptimal paradigms that do not allow for careful control or quantification of attentional demands in the manipulations used to probe automaticity; they may not vary attentional demands sufficiently (thereby decreasing sensitivity to age-related changes). Additionally, while a number of MMN-based studies have examined changes in speech perception from early to late childhood, transitions from adolescence to adulthood are yet to be comprehensively examined. An additional limitation of previous studies is that most have used synthetic speech stimuli rather than natural which limits the generalizability of findings as synthetic stimuli lack ecological validity, and are not processed equivalently to natural speech tokens (Nusbaum, Dedina & Pisoni, 1984; Pisoni, 1981).

1.9 The Present Study

The present study includes two experiments that aim to: 1) use MMN to map the development of automatic speech perception for natural stimuli from late childhood into mid-adolescence, when adult levels are expected to be achieved, and 2) refine current methodologies for probing the automaticity of speech perception in child and adolescent populations, using mismatch negativity response.

As later sections will describe in greater detail, experiment 2 builds on the insights obtained from experiment 1 to further refine the manipulation of attentional demands in the ignore condition using the N-back task that is employed to pull attention away from the speech signal.
Additionally, experiment 2 hones in in adolescence, to provide a more thorough examination of the bridge between childhood and adult levels of speech perception.

Chapter 2: Experiment 1

The primary goal of the first study was to examine the maturational trajectory of the MMN to a subtle vowel contrast under carefully controlled attentional conditions. Central to this goal was the inclusion of an attend condition as well as both an active ignore (i.e., ignore the auditory stimuli while mentally solving math equations) and passive ignore (i.e., ignore the stimuli while passively viewing a silent movie) condition to more carefully control for attention. In particular, the active ignore allowed us to ensure that task instructions were followed. The inclusion of two ignore conditions was intended to elucidate a major potential confound in the literature. Additionally, we included a number of improvements as follows: 1) usage of natural speech stimuli to provide greater ecological validity; 2) use of an active ignore task where performance could be quantified and, thus, allow us to assess direction of attentional focus. Given that prior work suggests that the MMN is not adult-like by 11 years of age, we also included an adolescent age range (10-12 years) to determine whether there is mature vowel processing at these ages.

We predicted that adults would show a robust MMN in all conditions and that children would show attentional modulation evidenced by a reduced MMN in the ignore conditions with the most dramatic reduction occurring in the math condition. It was hypothesized that the math condition will pull attention away from the auditory signal to a greater extent than the passive-ignore and that this would be reflected in the MMN.

2.1 Methods

2.1.1 Participants
A total of 34 individuals participated in the first experiment, 8 of whom were children (ages: 6-12) and 11 were adults (ages: 18-40). Among these individuals, 14 were excluded either due to behavioral performance outside of our criteria on the math condition, due to EEG-related artifacts or technical issues in data collection or processing. Two participant groups were established, one consisting of adults (n=11; ages: 18-40) and the other consisting of children (n=8; ages: 6-12). An additional 6 children were excluded from the study because of high artifact, 6 because of poor behavioral accuracy and 10 because of corrupt data files. The child group was composed of an equal number of children ages 6-8 and 10-12 years. Participants were recruited from the greater New York area via flyers and online ads (Craigslist.org & Backpage.org). Prior to participation, we received written consent from each participant’s parent as well as assent from the participant; this procedure is in accordance with The Graduate Center’s IRB requirements and regulations.

_Inclusion/Exclusion Criteria._ All participants were monolingual American-English speakers that resided in monolingual households. Developmental and language history questionnaires, as well as parental interview, were used to exclude any individuals with a history of speech, language, psychiatric or developmental impairments. Additionally, the following formal testing was employed to ensure that language and auditory function were within normal limits: Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4) (Concepts and Following Directions, and Recalling Sentences, Understanding Spoken Paragraphs, and Phonological Awareness), pure-tone audiometric screening (25 dB HL at 500, 1000, 2000, and 4000 Hz). Individuals were excluded if they have any of the following conditions as per caregiver report: emotional or behavioral disturbances, cognitive delay, motor deficits, or neurological signs including seizure disorders or use of seizure medications. Participant histories were negative for hearing
impairment, neurological impairment or psychiatric disorder. A small monetary payment of $25 was provided for participation in the study.

2.1.2. Stimuli Materials

Natural speech sequences /æpə/ and /ɑpə/ ([æ] as in vowel sound in “cap” and [ɑ] as in the vowel sound in “cop”) served as the stimuli. Three tokens of each syllable type, /æpə/ and /ɑpə/, produced by a single male speaker, were used to provide a more naturalistic listening condition. The token durations for /æpə/, measured in milliseconds (ms) are 434ms, 453ms, 433ms with a mean of 440ms and a standard deviation (SD) of 11.27. The /ɑpə/ tokens are 400ms, 380ms, and 420ms, with a mean of 400ms and a SD of 20. The stimuli were presented free field over two loudspeakers, situated approximately 1m from participants at a comfortable listening level of 65 dB SPL with an interstimulus interval (ISI) of 300ms and a stimulus onset asynchrony (SOA) of 800ms.

Natural stimuli were employed because they are more ecologically valid than synthetic speech and provided natural variability across tokens. Native listeners have shown excellent discrimination for these stimulus contrasts but poorer perception has been found for non-native listeners (Shafer, Strange, Ito, Gilichinskaya, Rosas & Kresh, 2011). The order of the stimuli was pseudo-randomized for each subject with presentation adhering to the restriction that deviants must be separated by at least three standards, with an overall probability of 20% deviance and 80%. 136 deviants were delivered for each condition.

2.1.3 Paradigms and Procedures

In the attend condition, a behavioral discrimination task was conducted during which participants were asked to count the deviant /ɑpə/ tokens in six 3-minute blocks. After each
block, the stimuli were stopped and the examiner documented the number of deviants the participant reported hearing. In the math condition, accuracy was recorded to ensure that the participants were engaged in the task and that the task was of an appropriate difficulty level. Math lists were assigned prior to the study based on performance on a brief pre-test. Performance of 75-90% guided selection of the math list to ensure a task that was challenging enough to require attentional focus but was not too difficult as to result in disengagement. Scores higher than 90% on a pre-test were thought to reflect that the list was too easy and thus would not be considered challenging enough to capture attention. In the passive-ignore condition, participants were instructed to attend to a silent movie and ignore the auditory stimuli.

2.1.4. Electrophysiological Recordings

A 65-channel Geodesic net, that included two electrooculargraphy (EOG) electrodes, was placed on the participant’s head. Net electrodes made contact with the scalp via saline soaked sponges. The EEG was amplified using the EGI Geodesic Amplifiers, which included an online bandpass filter (0.1 to 40 Hz filter). NetStation version 4.1.2 was used to record the data at a sampling rate of 250 Hz per channel (continuous mode) for later off-line processing. The continuous EEG was processed off-line, using a low-pass filter of 20 Hz, and segmented into epochs with an analysis time of 200 ms prestimulus baseline to 800 ms poststimulus. The data were then baseline corrected and examined for artifact using Netstation software version 4.1.2. Epochs with excessive artifact (i.e., those with differential average amplitude greater than ±70 uV on more than 5 channels) were excluded from the average. Channels identified as bad on more than 20% of the trials were replaced using a spline interpolation algorithm from surrounding channels. ERP averages were then calculated for each stimulus type (standard, deviant) and baseline corrected using the 100-msec pre-stimulus activity. For each participant, we calculated
difference waveforms for the examination of the MMN, by subtracting the standard waveform from the deviant waveform.

2.1.5. Experimental Design

Participants were seated in a sound-attenuating booth and instructed to listen to a series of natural speech sequences presented in three different attention conditions (passive-ignore, math-ignore, and auditory-attend) in a classic oddball paradigm. The two ignore conditions required participants to ignore auditory stimuli while they either watched a silent video (passive-ignore) or conducted a challenging mathematical task (math-ignore) in which they were required to mentally solve mathematical equations and indicate whether the solution provided was correct via a response box (Yes/No response). The auditory-attend condition required participants to actively attend to the stimuli and mentally count the number of deviant sounds presented. Participants reported the number of deviants at the end of a block. Inclusion of the two separate ignore-conditions enabled the examination of the effects of instructing participants to passively ignore stimuli versus the effects of directing their attentional focus away from the signal to a challenging competing task. EEG recordings were obtained throughout performance of the task.

2.1.6. Data Analysis

Identification of Time Windows. Time windows selected for analysis were identified after visually inspecting the waveforms. Time-points of interest were identified and analysis was conducted to examine them.

Site Selection for Analysis. Six frontocentral sites were selected for analysis. Frontocentral sites were selected because research has consistently demonstrated the largest MMNs at these locations. These sites included Geodesic site 13 (F3) and site 17 (C3), site 62 (F4) and site 54 (C4) and two midline sites, site 4 (FZ) and the vertex (CZ).
Statistical Analysis. The primary goal of our analysis was to determine differences between children and adults with respect to the impact of attentional condition on the MMN. In order to accomplish this, at each electrode site, we made use of a 3-way ANOVA (implemented in SPSS) that consisted of the following factors: group (children, adult), attention condition (attend, passive-ignore, passive-math), and time (five 20-second intervals between 100-200ms). Given potential age- and attention-related variations in the time characteristics of the MMN, we repeated our analyses for two additional time intervals (200-300ms, 300-400ms). We made use of a standard p < 0.05 alpha criteria for significance, and employed the Greenhouse Geisser correction to account for potential violations of sphericity. The Helmert contrast implemented in SPSS was used for post-hoc comparisons; this contrast compares the first level of the factor the remaining factors, then the second level to the remaining factors and so on (e.g., for attentional condition: attend vs, [ignore-math, ignore-passive], ignore-math vs. ignore-passive).

2.2 RESULTS

Prior to testing for age- and attention-related differences in the MMN, we first inspected the mean responses to standard and deviant trials to ensure the appropriateness of our measurements. Figures 1a and 1b, show the grand mean ERP waveforms for standard and deviant trials for the two age groups, respectively. Figure 2a and 2b display the subtraction waveforms at the six sites selected for statistical analysis for children. Figures 2c and 2d display the subtraction waveforms at the six sites selected for statistical analysis for adults. Across all sites, ANOVA revealed a significant main effect of time in one or more of the time intervals examined (100-180; 200-280; 300-380).
Of note, C4 was characterized by a later response than the other sites (F3, FZ, F4, C3, CZ), only showing a significant effect of time in the 300-380 millisecond time-range. Significant interactions observed in each of the three time windows are reported below.

Early interval 100-200 ms

In the 100-200 millisecond time window, a three-way interaction of group X condition X time was significant at Fz (F(8, 128)=2.72; p < 0.04) and C4 (F(8, 128)= 2.613; p < 0.044). In order to further investigate the group X condition X time at Fz, we first visually inspected our findings and carried out post-hoc analyses. Consistent with the presence of a 3-way interaction, the magnitude of attentional modulation attend vs. ignore [passive, math] appeared to increase with time in children, but not adults. The helmert post-hoc contrast supported this notion. Specifically, the magnitude of group-differences in attentional modulation was significantly smaller at earlier intervals, ms 100 and 200, relative to later time-intervals (100 ms: F(4,64)= 7.32, p < .016; 200 ms: F(4,64)=5.26; p < 0.036). No other post-hocs for this interaction achieved significance.

Next, we visually inspected the three-way interaction at C4, finding that children, but not adults, showed evidence of attentional modulation of the MMN. Specifically, adults show comparable MMN amplitude for the three attentional conditions, from 100-200 ms; in contrast, children exhibited no MMN for the attend condition and MMN negativity for the two ignore conditions. Post-hocs only supported the presence of a group difference in the MMN in the comparison of the two ignore conditions at 120 ms versus later time intervals (F(1,16)=6.076; p < 0.025), and 160 ms versus 180 msec (F(1,16)= 4.328; p < 0.054).

Beyond the two 3-way interactions noted above, a condition X time interaction was found at C4 (F(8, 128)= 8.392; p< 0.00001) with a trend noted at F4 (F(8,128)= 2.279; p < 0.085). A main
effect of time was found at Fz (F(4,64) =6.50; p< 0.006), F4 (F(4,64)=7.443 ; p < 0.001) and Cz (F(4,64)=3.176 ; p< 0.055). Post hoc showed difference between 100 and later (F=5.553; p<.032) and 160-180 (F=3.870; p< .067).

Middle interval 200-300 ms

A three-way interaction of condition X group X time was significant at F4 (F(8,128)= 4.881; p < 0.002). Visual inspection of the three-way interaction at F4 and post-hoc tests suggested that group differences in the impact of attentional modulation changed with time. Adults appeared to have an MMN response in the 200-300 ms interval only for the attend condition, while children had an MMN response to both the attend and ignore conditions (albeit somewhat smaller for the ignore condition). Helmert-based post hoc tests confirmed the presence of significant group differences in the impact of attentional modulation (attend vs. ignore) on the MMN, with significant effects being found in the comparison at each time interval (200 vs. later [F(1,16)=8.197; p < .011], 220 vs. later, F(1,16)=14.097; p < .002), 240 vs. later [F(1,16)= 7.045; p < .017] and 260 compared to 280 [F(1,16)= 4.118; p < .059]). The two ignore conditions did not differ significantly.

A two-way interaction of condition X time was found at C4 (F(8,128) =4.979; p < 0.007), which was the only site that did not reveal a main effect of time (F3(4,64) [F(4,64)= 9.124; p < 0.001], Fz [F(4,64)= 17.422; p < 0.00001], F4 [F(4,64)=11.425; p <0.001], C3 [F(4,64)= 8.923; p < 0.002], and Cz [F(4,64)=5.244 ; p < 0.023]). For the condition X time interaction at C4 (F(8, 127) = 4.979; p < 0.007), a less prominent response was noted for the math condition than the ignore-passive and attend conditions. In support of this observation, post hocs showed differences between the two ignore conditions at 200 versus later (F(1,16)=20.605; p < 0.00001), 220 vs later
Late interval, 300-400 ms

A condition X time interaction was found for C3 (F(8, 128)=3.565; p < 0.022), Cz (F(8, 128)=3.897; p < 0.015) and C4 (F(8,128)=4.032; p < 0.006). A significant time X group interaction was found for C3 (F(4, 64)= 4.523; p < 0.022) and a main effect of time was found for Fz (F(4, 64)= 6.94; p < 0.007), F4(F(4, 64)= 4.392; p < 0.019), and C4(F(4, 64)= 4.270; p < 0.012).

Regarding the condition X time interaction, inspection of C3 suggested that the MMN in the attend condition was returning to baseline (i.e., becoming less negative) more rapidly than the two ignore conditions. This observation was supported by post-hocs comparing the attend vs. ignore conditions at 300 vs. later (F(1,16)= 10.433; p < .005), 320 vs. later (F(1,16)= 7.452; p < .015) and 360 vs. 380 (F(1,16)=4.198; p < .057). Cz showed differences between attend and ignore conditions at the first (F(1,16)=10.817; p < .005) and second (F(1,16)=5.984; p < .026) time intervals, with the math condition being more negative in the early, but not at later intervals.

Finally, the significant time X group interaction appeared to reflect the children having continued negativity in this time window, while the adults had more subtle variation with time. Consistent with this observation, post hoc contrasts for 300 vs. later (F(1,16)=6.531; p < .021) and 360 vs. 380 (F(1,16)=6.483; p < .022) were significant.
Figures from Experiment 1.

Figure 1a. Children’s Grand Means to standards and deviants averaged across all conditions at C3, Cz, C4, F3, Fz, and F4.
Figure 1b. Adult Grand Means to standards and deviants averaged across all conditions at C3, Cz, C4, F3, Fz, and F4.
Figure 2a. Children’s subtraction waves (deviant minus standard) at C3, Cz & C4 for the attend, passive-ignore and math-ignore conditions.
Figure 2b. Children’s subtraction waves (deviant minus standard) at F3, Fz & F4 for the attend, passive-ignore and math-ignore conditions.
Figure 2c. Adult subtraction waves (deviant minus standard) at C3, Cz & C4 for the attend, passive-ignore and math-ignore conditions.
2.3 DISCUSSION

The primary goal of experiment 1 was to examine the development of automaticity in speech perception, as indexed by the MMN, using a cross-sectional sample of children ages 6-14 year old children, with an adult sample (18-40 years) included as a reference. Here, we reported preliminary results from the first phase of data collection, in which we verified our ability to
produce attentional modulation of the MMN in children; these findings are in contrast to adults, in which attentional modulation was minimal. As will be discussed later, we attempted to optimize the ignore condition to allow us to better examine pre-attentive processing. However, our findings suggest a more complex picture than expected, as will be discussed below. Examination of age-related changes in the child group was limited by sample size at the present time, but will be addressed in the next phase of data acquisition.

Overall, our findings from the first experiment were consistent with a larger body of interest suggesting that by adulthood, SPR development achieves a level of automaticity and thus attentional factors show little effect on modulation of the MMN to these native language speech sounds. This is in contrast to children, where attentional manipulations (e.g., attend vs. ignore) of the MMN are known to occur to some auditory contrasts. Prior work in early childhood populations (e.g., ages 4-5 years old) have emphasized that under certain conditions, a positive mismatch response (pMMR) can occur prior to the negative response, and is thought to attenuate by ages 6-7 years (Shafer et al., 2010). Interestingly, the present study, the C4 site exhibited evidence of a pMMR in the attend condition for children that appears to overlap and attenuate the MMN response. Given that the mean age is substantially higher than 7 years, this result is somewhat surprising, and may reflect differences related to the stimuli employed in the present work (e.g., natural speech). They also suggest a possible uncoupling of neural systems during this condition given that the other sites did not show this response. One other possibility is that the deviation from the expected results may reflect the sample size; future data collection will allow us to rule out this possibility.

The determination of optimal test conditions for the examination of age-related variation in automaticity of speech perception via the MMN remains an open challenge. Experiment 1
attempted to address this by varying the cognitive demands of the distracting condition employed as a reference in the assessment of MMN attentional modulation. Both distraction conditions employed (ignore-passive, ignore-math) did successfully achieve attentional modulation, and help to differentiate children and adults. However, the two ignore conditions did not differ as hypothesized. We had expected the ignore-math condition to be more attentionally demanding, and thus produce greater attentional modulation than the ignore passive. Instead, we found a more complex picture, with MMN modulation of the ignore-math being either being equivalent to that with the ignore-passive (e.g., adult Fz), or trending towards being smaller (e.g., adult Cz), depending on the site being examined. These findings suggest against simplistic conceptualizations of task difficulty, and highlight that the neural systems modulated by a specific task may vary depending upon the nature of the task. Future efforts attempting to optimize attentional modulation in children through the inclusion of an active task may benefit from more systematic examinations of task difficulty via parametric manipulations of demands in a single task, or comparison of tasks differing in one or more well-prescribed domains. Regarding passive tasks, future work may benefit from systematic variation of stimulus variables that can impact engagement.

Characterization of late-stage SPR development via attentional modulation of the MMN was a primary goal of the present work. Based upon prior work, we suspect that the SPRs may approach adult levels at some time after 11 years old, since the auditory system shows increased maturity after this age (Morr et al., 2002; Shafer et al., 2000). Anticipating that the process is complete by the early high school years, we set the upper bound of our target recruitment goal to 14 years of age. In the second experiment we recruited older children to investigate this issue.
2.4 LIMITATIONS OF THE FIRST EXPERIMENT AND CONSIDERATIONS

A notable limitation of experiment 1 is the small number of participants included after exclusion of participants based upon performance during the math task. Math task performance parameters were included in the participant inclusion criteria as a means of ensuring that the condition was attentionally demanding. Specifically, we set a minimum accuracy criteria to guarantee that participants were compliant with task demands. Additionally, we removed participants in whom responses were deemed to be excessively rapid, even when accuracy was appropriate, as this could allow for significant gaps in time between trials, which would represent low attention periods. This decision proved to be overly-exclusionary as the range of time in which participants took to respond varied more than initially anticipated in the study construction, and participant exclusion rates for the present work were exceedingly high. In the next phase of the present project, we make use of behavioral observation as an alternative criterion for excluding participants based upon attentional engagement.

The MMN holds significant potential for application in the study of clinical populations characterized by impairments in speech processing and possible identification of clinically meaningful biomarkers. Examples of successful applications to date include identification of differences in auditory processing in children with specific-language impairment (Shafer, et al., 2011), autism (Dunn, Gomes & Gravel, 2008), dyslexia (Bonte, Poelmans, & Blomert, 2007; Bruder, Leppanen, Bartling, Csepe, Demonet, Schulte-Korne, 2011; Noordenbos et al, 2012; 2013), language impairments (Ahmmed, Clarke, & Adams, 2008; Shafer et al. 2005), attention deficit (Gomes, Duff, Flores, & Halperin, 2013; Halperin, & Schulz, 2006), and psychiatric disorders (Todd, Harms, Schall, & Michie, 2013). It is exciting to think of the clinical applications of electrophysiological measures for elucidation of neural differences underlying
speech and language disorders, but we must temper expectations. A key prerequisite for this pursuit is the careful mapping of MMN to a particular auditory contrast across development in a large-scale typically developing sample, as well as clinical populations. This is necessary to obtain both, a more complete neurodevelopmental understanding of the phenomena, and when and how disease processes can impact it.

In summary, experiment 1 presented preliminary findings have the potential to expand upon the existing literature regarding developmental changes in attentional modulation of the MMN to difficult speech contrasts. They highlight the need for continued optimization of paradigm design and greater consideration of the nature of mental operations associated with the ignore conditions used to isolate pre-attentive processing, especially when attempting to use more naturalistic stimuli and environment. Experiment 2 attempted to provide a more detailed characterization of late-stage development of speech perception using the MMN measure.

**Chapter 3: Experiment 2**

Building on insights gained from experiment 1, we conducted a second experiment that aimed to: 1) examine attention modulation of speech discrimination using different tasks than used for experiment 1, and 2) examine neural correlates of speech discrimination across the adolescent and teenage years. While the findings of experiment 1 suggested the value of moving beyond the traditional “passive” ignore condition for the MMN, the math condition employed proved to be problematic for participants. The broad range of processes that differ between the passive and math ignore conditions made it difficult to interpret the nature of differences in findings between the two conditions. Additionally, while robust differences were detected
between child and adult populations, the trajectory of maturation after age 11 years into adulthood is unknown.

The second experiment attempted to overcome the limitations encountered in the preliminary work by including a cognitively demanding “ignore task” for which successful performance in child and adolescent populations is known to be feasible, and a specific construct can be manipulated to examine the impact on speech perception at the level measured by MMN. Specifically, we made use of the N-back task, a gold standard test of working memory (WM) (Kane & Engle, 2002). Comparison of the 2-back and 0-back conditions of this task was intended to allow for the isolation of working memory function, with the 0-back serving as a vigilance task and the 2-back requiring continued updating of information in working memory for successful task performance. An additional benefit of the 2-back condition is that working memory maintenance processes continue between trials, ensuring that a cognitive load remains present. A number of prior studies have demonstrated the ability of school age children and adolescents (e.g., Vuontela et al.) to perform the N-back task and its variants. It is important, however, to recognize that we chose these tasks to lead to different levels of engagement of attention on a visual task so that the participants would have fewer resources available to process the auditory information.

Germane to the present work, a number of prior studies have demonstrated the feasibility of using the n-back task and other working memory manipulations to see whether attentional modulation affects speech perception of the target contrasts (Fan, Hsu, Cheng, 2013; Lv, Wang, Qiu, Feng, Wei, 2010; Otten, Claude, Picton, 2000; Ruhnau, Wetzel, Widmann, Schroger, 2010; van Rhijn, Roeber, O’Shea, 2013). An initial examination of the impact of a visual N-back on the auditory MMN in adults (Takegata et al., 2005) failed to find evidence of modulation, though it did find changes in late-orienting ERPs components. In contrast, a study of adults by Lv, Wang,
Qiu, Feng, Tu, Wei (2010) found that parametric variation of visual WM load successfully modulated the MMN produced by environmental sounds (standard stimulus: tones); with increasing WM load (recalling 3 digits vs. 7 digits) they found enhanced MMN amplitude to deviant/novel sounds in the high-load condition, showing enhanced change detection processes occurring during the more demanding WM task. These findings are similar to findings in our pilot work, where a subset of electrode sites showed trends towards enhanced MMN amplitude in the cognitively demanding math condition, relative to the passive condition. Interestingly, Lv and colleagues reported an attenuation of the Novelty-P3 as load demands increased. This finding suggests reduced orienting to the deviant sound when the participants were engaged in a high demand task. Finally, a study of 9-10 year olds by Ruhnau, Wetzel, Widdman & Shroger (2010) examined the MMN responses during an auditory-visual task in which the auditory stimuli was presented during visual 0-back and 2-back tasks. Standard auditory stimuli (sinusoidal tones) were occasionally replaced with novel (environmental) sounds, which elicited MMN responses. Age-related differences in the MMN to unexpected novel sounds were noted, though did not interact with memory load.

The primary goal of the present experiment is to examine the late maturational stages of speech perception development, using the MMN elicited to naturalistic stimuli (i.e., a relatively difficult vowel contrast) as our index. Central to this goal, we will make use of a more refined paradigm than the first experiment to probe neural discrimination of the vowel contrasts. These new tasks allow for greater control and quantification of cognitive load through inclusion of the 0- and 2-back in our ignore conditions. We focused on the examination of children from 10-16 years of age to provide a more comprehensive characterization of the latter stages of speech perception than what has been reported in the literature to date.
**Hypotheses.** Based upon the existing literature and our prior work, including experiment 1, we hypothesized that: 1) adults would show robust MMN, without significant attentional modulation in all conditions. One or more of the ignore conditions are expected to show some level of attentional modulation to differentiate children and adults. With regard to the two ignore conditions, we expected the 2-back ignore to be more attentionally demanding, and thus produce greater attentional modulation than the 0-back condition. 2) children 10-16.9 would show greater modulation of the MMN relative to adults; among the child group, we predicted a dimensional relationship between modulation and age, with the older ages (15+) approaching adult-like levels of modulation.

**3.1 Methods**

**3.1.1 Participants**

Participant included individuals from 10-25 years. These fell into the following age ranges: (10 n=2; 11 n=1; 12 n=4; 13 n=3; 14 n=1; 15 n=3; 16 n=3; 18 n=3; 19 n=2; 2- n=2; 21 n=2; 22 n=3; 23 n=1; 24 n=2; 25 n=2). Recruitment procedures and consent obtainment were the same as those used in Experiment 1, in accordance with The Graduate Center’s IRB requirements and regulations.

**Inclusion/Exclusion Criteria.** All participants were monolingual, native New Yorkers, speakers of American-English who resided in monolingual households at the time of participation.

Developmental and language history questionnaires, as well as parental interview, were used to exclude individuals with a history of speech, language, psychiatric or developmental impairments. Additionally, the following formal testing was employed to ensure that language and auditory function were within normal limits: Clinical Evaluation of Language Fundamentals, 4th edition.
(CELF-4) (Subtests: Understanding Spoken Paragraphs, Phonological Awareness subtests and Number Repetition – Forward Backward Digit Span (Semel, Wiig, & Secord, 2003), pure-tone audiometric screening (20dB or 25dB SPL at 500, 1000, 2000, and 4000 Hz) in a sound-attenuated booth using a Welch Allyn Audioscope (Model 92680). Hearing screenings were conducted at 20dB SPL for all participants and adjusted to 25dB SPL if they could not pass initially. No participants were excluded solely for requiring the pure tone screening to be adjusted at 25dB SPL, provided they satisfied all other inclusion requirements. Individuals were excluded if they have any of the following conditions as per preliminary evaluation, caregiver report or behavioral observations: emotional or behavioral disturbances, cognitive delay, motor deficits, neurological signs including seizure disorders or use of seizure medications. Participant histories were negative for hearing impairment, neurological impairment or psychiatric disorder. A small monetary payment of $12 per hour was provided for participation in the study.

3.1.2 Stimulus Materials

Natural speech sequences /æpə/ and /ɑpə/ ([æ] as in vowel sound in “cap” and [ɑ] as in the vowel sound in “cop”) were employed as the stimuli. Three tokens of each syllable type, /æpə/ and /ɑpə, produced by a single male speaker, were used to induce categorical perception. The token durations for /æpə/, measured in milliseconds (ms) are 434ms, 453ms, 433ms with a mean of 440ms and a standard deviation (SD) of 11.27. The /ɑpə/ tokens are 400ms, 380ms, and 420ms, with a mean of 400ms and a SD of 20. The stimuli were presented free field over two loudspeakers, situated approximately 1m from participants at a comfortable listening level of 65 dB SPL with an interstimulus interval (ISI) of 300ms and a stimulus onset asynchrony (SOA) of 800ms.

The same natural stimuli used in experiment one were employed in this experiment as they
are more ecologically valid than synthetic speech and provide natural variability across tokens. The order of the stimuli were pseudo-randomized for each subject with presentation adhering to the restriction that deviants must be separated by at least three standards, with an overall probability of 20% deviants and 80% standards.

3.1.3 Experimental Procedures

Participants were seated in a sound-attenuating booth while they were presented with a series of natural speech sound sequences using a classic oddball design. The auditory stimuli were continuously presented in six 3-minute blocks. Attentional conditions included the auditory-attend, 0-back-ignore (low demand), and 2-back-ignore (high demand). During the auditory attend blocks, participants were required to actively attend to the stimuli and mentally count the number of deviants sounds presented and to report the number of deviants at the end of each block. During the two ignore blocks (0-back, 2-back), participants were required to ignore the auditory stimuli while performing the classic N-back task. More specifically, participants were presented a sequence of centrally located, black, capital letters on a computer display (display duration = 800 milliseconds) and required to determine whether each letter: a.) matches a target letter specified before task performance (e.g., ‘M’) (0-back condition), or b) matches the letter presented 2 letters back (e.g. for the sequence ‘H-J-I-K-O-K-O’, ‘K’ and ‘O’ would be match responses). Letter stimuli in the two N-back blocks will be presented at a rate of one per 4 seconds, and participants will be required to press a button on a standard response box each time they detect a matching stimulus; no response is required for non-matching trials. EEG recordings will be obtained throughout the experimental session (expected recording time: 18 minutes + breaks).
Preliminary data from our pilot study revealed that 18 minutes of recording was adequate for ensuring a satisfactory number of good trials to be collected withoutfatiguing the which led to degraded compliance (i.e., task performance) and data quality (e.g., increased eye blink artifacts, motion). Inclusion of the two separate ignore-conditions allowed for the examination of the MMN to ignored sounds under differing levels of task difficulty. The presentation of the conditions were randomized and counterbalanced to prevent order effects.

3.1.4. Electrophysiological Recordings

A 65-channel Geodesic net by Electrical Geodesics, Inc. (EGI), which includes two electrooculargraphy (EOG) electrodes, were placed on the participant’s head. Net electrodes made contact with the scalp via saline soaked sponges. The EEG was amplified using the EGI Geodesic Amplifiers, and included an online bandpass filter (0.1 to 40 Hz filter). NetStation version 4.1.2 was used to record the data at a sampling rate of 250 Hz per channel (continuous mode) for later off-line processing. The continuous EEG was recorded for later, off-line processing, using a low-pass filter of 20 Hz, and segmented into epochs with a time-window that extended from 200 msec before the stimulus onset to 800 msec after stimulus onset. The data were baseline corrected and examined for artifact using Netstation software version 4.1.2. Epochs with excessive artifact (i.e., those with differential average amplitude greater than ±100 uV on more than 5 channels) were excluded from the average. Channels identified as bad on more than 20% of the trials were replaced using a spline interpolation algorithm from surrounding channels. ERP averages were then calculated for each stimulus type (standard, deviant) and baseline corrected using the 100-msec pre-stimulus activity. For each participant, we calculated difference waveforms for the examination of the MMN, by subtracting the standard waveform from the deviant waveform.
3.1.5 Statistical Analysis

A more objective analysis strategy was used in this second experiment (the Appendix includes the identical planned comparisons to the first experiment to show that the outcome is the same). First, to help identify time periods in which event-related activity was strongest, we calculated Global Field Power (GFP) (Lehmann and Skrandies, 1980). The GFP is calculated as the standard deviation of the mean amplitude of the full array of electrodes at any given time point. This calculation reveals the timepoints where the greatest variance occurred; in particular, dipolar activity is highlighted when electrodes record activity over both poles. Consistent with prior work, we used this measure, to guide the selection of time windows for analysis (Shafer et al., 2007). GFP was plotted in conjunction with Fz and the mastoids to select time windows of interest. The mastoid leads were included because the MMN inverts with frontal-central sites. This process resulted in two time-periods of interest: 100-400ms and 450-650ms. Figure 7 (below) illustrates GFP and the mastoids for each of the three conditions. The first time-period appeared to coincide with MMN-related phenomena, and to a lesser degree P3 (for the attend condition). The second window appeared to reflect the LDN – the late-occurring counterpart of the MMN.

Next, to reduce the 65 electrode channels to a small set, we performed spatial principal components analysis (sPCA) on two time-periods (100-400; 450-600). PCA maximizes variance (Dien, 2010) and allowed identification of sites that covaried, and thus, could be collapsed (averaged). A covariance matrix served as the input. The components that served as the output were examined to allow selection of a set of sites that showed a strong relationship in both analyses. The PCA was carried out for all participants together and for each group (adults, children) separately, to confirm that the first two principal components showed consistent factor
loadings for both the children and adults (See Figures 3 and 4 for first two principal components from 100-400 time windows, and 5 and 6 for 450-600 time windows). The first two components accounted for a large amount of variance (over 90%): in the attend, zero-back and two-back conditions, respectively (100-400ms: 94% accounted for by first two components [attend], 93% [0-back], 90% [2-back]; 450-600ms: 96%, 76%, 81%). The set of electrodes with the highest weightings of the same sign were selected for averaging. We identified seven regions of interest: anterior pole (6, 10, 11 & 14), left anterior (15,16, 20, 21 ), vertex (4,5,9, vertex, 17, 18), central posterior (29, 33, 34, 42, 41), right posterior (42, 46, 47), right inferior (48, 51, 44, 45), left inferior (31, 35, 56). The data were reduced in this way, rather than by back-projecting the PCA factor loadings on to the data because this strategy allowed selected the set of sites that was most consistent across the analyses of the two age-groups and the three conditions. In addition, we wanted to compare left, right, anterior, posterior, superior, inferior regions, and the PCA analysis did not always separate the regions in the manner.
Figure 3. Electrode sites that loaded heavily on the first component from the PCA for 100-400. The top row shows the sites from the analysis with all participants, in the attend, 0-back and 2-back conditions, and the middle and bottom row shows children and adults, respectively. Sites in red had similar loadings and sites in blue had similar loadings, but of the opposite sign to those in red.
Figure 4. Electrode sites that loaded heavily on the second component from the PCA for 100-400. The top row shows the sites from the analysis with all participants, in the attend, 0-back and 2-back conditions, and the middle and bottom row shows children and adults, respectively. Sites in red had similar loadings and sites in blue had similar loadings, but of the opposite sign to those in red.
Figure 5. Electrode sites that loaded heavily on the first component from the PCA for 450-600. The top row shows the sites from the analysis with all participants, in the attend, 0-back and 2-back conditions, and the middle and bottom row shows children and adults, respectively. Sites in red had similar loadings and sites in blue had similar loadings, but of the opposite sign to those in red.
Figure 6. Electrode sites that loaded heavily on the second component from the PCA for 450-600. The top row shows the sites from the analysis with all participants, in the attend, 0-back and 2-back conditions, and the middle and bottom row shows children and adults, respectively. Sites in red had similar loadings and sites in blue had similar loadings, but of the opposite sign to those in red.

Having identified optimal time-periods (100-400ms, 450-600ms) for examination, and 7 regions of interest (anterior pole, left anterior, vertex, central posterior, right posterior, right inferior, left inferior), for each region of interest, we next calculated the mean amplitude response for each time-period. At each lead, prior to calculating the mean amplitude for each time-window (100-400, 450-600), we refined the time-windows to only include those timepoints in which a significant difference between standard and deviant trials was noted for one or more of the trial
types. This was accomplished by carrying out one-sample t-tests on the difference waveform (deviant - standard) at 20 msec intervals for each of the trial types at a given lead, which allowed us to identify those timepoints in which a non-zero amplitude ($p < 0.05$) was present for one or more trial types in each lead.

To directly compare the groups and conditions, two-factor repeated measures ANOVAs, in which group (child, adult) was treated as a between-subject factor, and condition (attend, 0-back, 2-back) was used as a within-subject factor were carried out. We made use of a standard $p < 0.05$ alpha criteria for significance, and employed the Greenhouse Geisser correction to account for potential violations of sphericity. For post-hoc comparisons, we made use of the Helmert contrast implemented in SPSS to compare the levels of attentional condition (i.e., [0- + 2-back] vs. attend, 2-back vs. 0-back).

### 3.2 RESULTS

<table>
<thead>
<tr>
<th>Table A. Summary of results from experiments 1 and 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
</tr>
<tr>
<td>Age groups 6-12 years</td>
</tr>
<tr>
<td>18-40 years</td>
</tr>
<tr>
<td>Fz &amp; Cz (Vertex) 100-200 msec time window</td>
</tr>
<tr>
<td>• Increasing attentional demands decreased the MMN to the vowel contrast in children more than adults</td>
</tr>
<tr>
<td>E4 200-300 msec time window</td>
</tr>
<tr>
<td>• Increasing attentional demands decreased the MMN to the vowel contrast in children more than adults</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
</tr>
<tr>
<td>Age Groups 10-16 years</td>
</tr>
<tr>
<td>18-25 years</td>
</tr>
<tr>
<td>Left Inferior Region 100-400 msec time window</td>
</tr>
<tr>
<td>• Increasing attentional demands decreased the MMN in children more than adults.</td>
</tr>
<tr>
<td>• A significant negative correlations between amplitude and age was found in the attend condition and a positive correlation between amplitude and age was found in the 2-Back condition, confirming the distinction.</td>
</tr>
<tr>
<td>Vertex</td>
</tr>
<tr>
<td>• A trend toward a larger MMN in children than adults was noted, but no significant evidence of attentional modulation (i.e., no effect of, or interaction with condition).</td>
</tr>
</tbody>
</table>
Behavioral Findings

Behavioral performance was good for all conditions. In the attend condition, participants achieved an accuracy of 94.0% overall for the counting of auditory deviants, with each participant achieving > 80% accuracy for each block. Counting accuracy was significantly correlated with age ($r = 0.4; p < 0.01$). For the 0-back and 2-back conditions, accuracies were 98.9% and 92.6% respectively, with each participant achieving >80% in each block. No significant correlations were found with age.

Visualizing Obligatory Components

Before examining the MMN, we first looked at the waveforms to the standard stimuli in the attend condition to verify the presence of the obligatory components (P1-N1-P2-N2); these components provide an important index of the functioning of the auditory system and quality of the data obtained (Ponton and Eggermont, 2001, Kraus, 1999). Figure 7 illustrates the presence of the obligatory components in all age groups; additionally, it captures the expected age-related changes in the N2, which decreases with age (Czigler and Balazs, 2005). Age-related decreases in the N2 were made more evident when the child group was divided into two groups (10-13.5; 13.6-16 years). This figure indicated good signal to noise ratio in our data,
Figure 7a. Waveform to standard stimuli in attend condition at Cz. The obligator peaks P1-N1-P2-N2 are labeled. The P2 and N2 are of greater amplitude for the youngest age group compared to adults. The 13-16 year olds show intermediate N2 amplitudes.

*Age-related changes in the GFP*

Figure 7b shows plots of the GFP, Fz and mastoids, upon which we selected the 100-400 and 450-600 time windows (as described in the methods). For each of the newly defined time periods (100-400, 450-600), as well as the larger epoch (-200-800), we found age-related decreases in the mean GFP for all three task (See Figure 8). In contrast, differences in GFP across task types (i.e., attend vs. n-back, 0-back vs. 2-back) did not vary with age.
Figure 7b. Plot of GFP, Mastoids and Fz with selected time-windows 100msec [time point 74] to ~400msec [time point 149] and ~450msec [time point 62] to 600msec [time point 199]
Figure 8. Correlation between GFP and age under each of the attentional conditions (attend, 0-back, 2-back).

Comparison of Time-Window Responses for PCA-based Leads

Following the division of our sample into two groups (child, adult), we first focused on the comparison of children versus adults. Figures 9, 10 and 11 depict the grand mean ERPs and standard error of the mean for the PCA-based regions of interest under the attend, 0-back and 2-back conditions, respectively.
Figure 9. Child vs. adult responses for attend condition (time [ms] on the x-axis and amplitude [µV] on the y-axis).
Figure 10. Child vs. adult responses for 0-back condition (time [ms] on the x-axis and amplitude [µV] on the y-axis).
Two factor ANOVA for mean amplitude in the 100–400 ms time period showed significant main effects of condition, but not age for the anterior pole ($F(2, 76)=9.037; p<.0001$), central posterior ($F(2, 76)=11.228, p<.0001$) and right posterior ($F(2, 76)=6.556, p<.003$) regions; no condition by age-group interactions were detected. The anterior pole exhibited a negative response and the posterior regions exhibiting positive responses. For the anterior pole, post-hoc comparison revealed a more significant negative response in the attend condition than the two n-back ignore conditions (anterior pole: $F=24.474(1, 38), p<.0001$); the n-back conditions did not significantly differ. For the two posterior regions, post-hoc comparison found that the attend condition produced more positive responses than the two n-back conditions (central posterior: $F(1, 38)=...$
11.228, p < 0.001; right posterior: F(1, 38) = 6.556, p < 0.003); the n-back conditions did not significantly differ.

Two regions of interest were notable for age-related variation. First, the vertex exhibited a trend towards greater negativity for children than adults (F(1, 38)=3.082, p<.087). Second, the left inferior region of interest exhibited a significant condition * group interaction (F(2, 76)=6.603, p<.002).

Visual examination indicated that the left inferior response in this time window was more positive for children in the attend condition, and more positive for adults in the 2-back condition. The MMN manifests at left inferior regions as a positive-going waveform, not a negative wave; the vertex response was negative wave, consistent with MMN. Figure 12 provides a clear illustration of attentional modulation in the child group with a reduction in MMN amplitude for the ignore conditions relative to the attend condition, particularly for the 2-back task. For the adult group, the apparent reduced MMN amplitude in the attend condition (relative to the ignore conditions) may reflects overlap and attenuation by the P3b. Post-hoc comparisons confirmed that age-related differences in the response differed between conditions (attend vs. 2-back: F(1, 38) = 13.141; p < .001).

Figure 12. Condition by age-group interaction for mean amplitude in the 100-400 window at left inferior.
A two factor ANOVA for mean amplitude within the 450-600 windows found significant main effects of condition, but not age-group, for the anterior pole (F(2, 76)=3.610, p<.037), left anterior (F(2, 76)=3.907, p<.027), vertex (F(2, 76)=5.990, p<.004), central posterior (F(2, 76)=21.77, p<.000) and right posterior (F(2, 76)=15.820, p<.00001) regions of interest. The anterior regions exhibited a negative response and the vertex and posterior regions exhibiting positive responses in these time ranges. For the anterior regions, post-hoc comparison revealed significantly more negative responses in the attend than ignore conditions (anterior pole: F(1, 38)=5.693; p<.022; left anterior: F(1, 38)=3.907), with no differences being noted between the two n-back conditions. For the vertex and posterior regions, post-hoc comparison revealed significantly more positive responses for attend conditions than ignore conditions (p<.027, vertex: F(1, 38)=5.990, p<.004, central posterior: F(1, 38)=.080, p<.779, right posterior: F(1, 38)=26.874, p<.0001); no significant differences were noted between the 0-back and 2-back conditions.

To examine age-related effects for the P3 phenomena, we carried out one additional ANOVA, focused on the 300-400 time-period at the RP region. This revealed a significant condition by age-group interaction (F(2, 76)=3.892, p < 0.025). To follow-up this interaction, the three conditions were examined separately. A significant group differences was only observed for the Attend condition (F(1, 38)=6.798, p < 0.013). The adults and children did not differ for either of the n-back ignore conditions. Examination of Figure 13 shows a more positive P3 response in the Attend condition for adults than children.
Figure 13. Child vs. adult responses for attend condition at RP (time [ms] on the x-axis and amplitude [µV] on the y-axis).

**Correlations**

We examined the relationship between amplitude for each region and age. Correlations were found for the left inferior (LI), anterior pole (and left anterior (LA) regions.

For the left inferior region, where the MMN was seen as increased positivity to the deviant, the mean amplitude for the 100-400 window was negatively correlated with age for the attend condition. \( r = -0.38, p < 0.016 \). Specifically, the younger had a more positive response than the older participants. In contrast, the left inferior amplitude was positively correlated with age in the 2-back condition \( r = 0.38, p < 0.016 \). In this case, the younger individuals had a more negative response than older participants. These relationships are shown in Figure 14.

For the anterior pole region in the 2-back condition, the amplitude was negatively correlated with age. Specifically, the amplitude was more negative with increasing age in both the 100-400 ms interval \( r = -0.38, p < 0.016 \) and the 450-600 ms interval \( r = -.31, p < 0.044 \) windows; the zero-back condition did not show a correlation \( r = -0.11 \).

For the left anterior (LA) region in the late interval (450-600 ms), the mean amplitude was positively correlated with age in the 2-back condition \( r = 0.32, p < 0.044 \); specifically, the LA
response became more positive with increasing age. The LA region did not show any age-relationship in the 100-400 window.

Figure 14. Mean amplitude for the left inferior lead for the attend and 2-back conditions in the 100-400 time window.

4. DISCUSSION

The goal of the second experiment was to examine whether adults and children showed differences in neural discrimination of speech when attention was focused away from the speech information. We hypothesized that the maturation of speech processing is an ongoing process that continues into late adolescence. Consistent with prior work, the deviants produced MMN response with a frontocentral distribution, which is characteristic of adults (Martin et al., 2003). Negative deflections were noted in the anterior pole, left anterior and vertex regions of interest; comparison of our responses in the various leads against the mastoids confirmed the responses in these leads were the MMN. The left and right inferior leads exhibited positive deflections, which were consistent with the presence of an MMN in these areas as well. Regarding age-related changes in MMN responses, the anterior pole regions showed a response pattern that was consistent with the MMN. Specifically, the response was more negative to the deviant than the
standard. The anterior pole region demonstrated the predicted pattern of results, with age-related increases in negativity in the 2-back condition, but not the 0-back condition. This is consistent with our prediction that greater attention-related modulation of the MMN would be observed in children than adults.

For the left inferior region of interest, we found that for children, the response was more positive than that found for adults in the attend condition, and less positive than the response for adults for the more cognitively demanding (2-back) ignore condition. Given that the MMN produces a positive response at inferior sites, this pattern suggests that children had a larger MMN for the attend condition than adults. While this was unexpected, it is not without precedence as others have shown larger MMNs in children compared to adults (Kraus, McGee, Sharma, Carrell & Nicol, 1992). Additionally, Martin et al. (2003) demonstrated larger MMN responses in lateral-temporal areas for children than adults. Consistent with our categorical analyses of age, analyses treating age as a continuous variable for the left inferior region supported these findings, with response magnitude (indicative of the MMN) being negatively associated with age in the attend condition and positively associated with age in the 2-back condition.

While our primary predictions focused on the MMN, other components also showed age-related changes in neural indices of speech discrimination. The anterior pole region revealed age-related increases in negativity in response to deviants in the 450-600 time-window. This late negative component, termed the late discriminative negativity (LDN) has been reported in a number of other studies of speech discrimination (Čeponienė, Cheour, & Naatanen, 1998; Cheour, Korpilahti, Kraus, Holopainen, & Lang, 2001; Datta, et al, 2010; Hill, McArthur, & Bishop, 2004; Korpilahti, Lang, & Aaltonen, 1995; Shafer et al., 2005). The LDN is typically elicited by deviance in oddball paradigms, and follows the MMN, peaking around 400 msec
(Čeponienė, et al., 1998). The component is more commonly elicited in children than adults, with a progressive decrease in amplitude noted as a function of age (Cheour, Korpilahti, Martynova, & Lang, 2001). The literature is yet to provide clear interpretation of the LDN. One possibility is that the LDN reflects the reorienting of attention back to a task following an involuntary attention switch (Schroger & Wolf, 1998; Wetzel, Widmann, Berti & Schroger, 2006). Alternately, the LDN may reflect higher-order or deeper processing of deviance following initial detection of sensory event, which is evidenced by the completion of the MMN (Čeponienė, Lepisto, Soineinen, Aronen, Alku & Naatanen, 2004). Despite the uncertainty of the significance of the LDN, it does serve to index speech discrimination and reflects age-related changes during development (Cheour, et al., 2001). Some studies have suggested the LDN may actually be a more sensitive index of age-related changes than the MMN, thus accentuating its potential value (Cheour, Čeponienė, Lepanen, Tervaniemi, Naatanen, 2001). Though such assertions need to be balanced with other considerations, such as findings that the MMN appears to be more sensitive to more subtle acoustic differences than the LDN (Čeponienė, et al. 2002).

Second, the right posterior region exhibited age-related increases in response magnitude within the 300-400 msec window for the attend condition. This increase in positivity is consistent with the P3b/P300, which is associated with the orienting and allocation of attentional resources in accord with task demands (e.g., Polich et al 1990).

Our findings regarding the P3b are not necessarily surprising, as progressive, age-related developmental increases in amplitude have been reported between 4 and 25 years of age (Polich et al 1990; Satterfield, Schell, Nicholas, Atterfield, Freece, 1990). While our work primarily focused on P3b amplitude measure, it is worth noting that P3b latency findings have also been
reported as reflecting the speed of attentional allocation (Houlihan, Stelmack & Campbell, 1998; Polich, 2007) and higher-level cognitive performance (Pelosi, Holly, Slade, Hayward, Barret & Blumhardt, 1992). However, the developmental literature has provided somewhat more of a complicated picture for the P3b. While some have suggested progressive decreases in P3b latency from ages 4 to 20 (Polich, Ladish & Burns, 1990; Courchesne, 1990), others have suggested that latency changes level off around 16 years (Fuchigami et al, 1993), or even begin to increase in late adolescence/early adulthood (Friedman, Pitnam, Ritter, Hamberger & Berman 1992). Visual inspection of our findings suggested some potential evidence of shorter latencies in adults, but these findings were not clear and are beyond the scope of our current examination. Findings of age-related changes for the P3b in the present work suggest maturational changes in top-down processes, as well as changes in bottom-up speech discrimination.

4.1 General Discussion

The overarching goal of the present study was to map the developmental time course of speech discrimination at the neural level indexed by MMN with the expectation that speech discrimination will become more automatic with age. In each of the experiments, three conditions (one attend and two ignore) were included to assess the automaticity of deviance detection of a vowel phoneme contrast; automaticity in speech discrimination would be seen as no difference in MMN under different amounts of attention. The first experiment focused on children aged 6-12 years and included a math task designed to draw attention away from the speech stimuli, in a more readily objective manner than the traditional passive ignore condition. This task was successful in showing age-related differences in attentional modulation of the MMN, with showing significantly larger MMN responses than adults; results from our math ignore condition were generally comparable to those obtained with our traditional passive ignore condition.
However, a limitation of the math task was that it proved to be challenging for a number of participants; this led to significant loss of participants, when using a strict criterion for behavioral performance on the math task. Additionally, it is difficult to quantify any differences in how attention is engaged between the passive and math ignore conditions from a behavioral condition.

The second experiment focused on children 10-16 years of age and used the 0- and 2-back tasks as distractor conditions because they are better-controlled, and allowed for a more straightforward quantification of performance. The zero-back task was somewhat similar to the traditional passive task in that it engaged attention in terms of vigilance. In other words, it successfully shifts attention away from the sounds using a minimally demanding task. In contrast, the 2-back task required selective attention to successfully perform the working memory task. Our results suggest that the 2-back task did, in fact, direct attention to the visual modality, as indicated by significant decreases in MMN response in a number of electrode sites, regardless of age. While as expected the 2-back condition produced greater modulation of the MMN response in children than adults, it was surprising that the MMN in the attend condition was larger in children.

We also observed positive response over posterior sites in the attend condition that was consistent with the P3b (Polich, 2007) and expected for this task. This P3b also showed an age effect, seen as increased positive with increasing age. In the later time frame for the ignore conditions (passive and math in experiment 1 and 2-back in experiment 2) we saw age-related increases in the late negativity in the anterior pole region. This late negativity may indicate reorienting to the stimulus change. Altogether, these results suggest that the development of speech processing continues into adolescence. Below we discuss these findings in relation to our hypotheses and the literature.
4.2 Maturation of Speech Discrimination

Developmental models of speech processing, such as WRAPSA and PRIMIR, tend to focus on the early years of life and often imply that speech processing is fully developed by grade school. However, there is some evidence that speech development continues across the grade school years (Nittrouer, & Miller, 1997; Shafer, et al., 2010). We hypothesized that the maturation of is an ongoing process that continues into late adolescence. In the first study we tested children from 6 to 12 years old and found that neural measures of discrimination differed between adults and children. Our findings extend those of Shafer et al. (2010) by demonstrating that age-related differences remain when using naturalistic stimuli and more cognitive demanding ignore conditions in the MMN paradigm. In the second experiment we wanted to determine whether older children (10 to 16 years) showed adult-like responses; this later developmental period is only beginning to receive more significant attention in electrophysiological studies of speech perception. Using a notably different experimental paradigm for studying speech perception, our findings proved to be highly consistent with those of Bishop et al. (2011). Their findings indicated that speech discrimination is not fully immature in the grade school years. Of note, while Bishop argued that more sophisticated time-frequency analytic measures (e.g., phase synchronization) are needed for the detection of age-related differences in the MMN. In our study, we were able to see difference in MMN amplitude without these types of analyses, possibly because of differences in the stimuli (e.g., tones and syllables vs. VCV) and/or attentional manipulations employed (Bishop et al. only included ignore condition). It is important to note that our findings emphasize the importance of looking at the trajectory of developmental phenomena into middle to late adolescence is consistent with larger trends in the developmental neuroscience.
literature (Di Martino, Fair, Satterthwaite, Kelly, Castellanos, Thomason, Craddock, Luna, Leventhal, Zuo, Milham, 2014).

4.3 Assessing Automaticity

The main goal of this study was to examine whether adults and children showed differences in neural discrimination of speech when attention was focused away from the speech information. Different tasks were used to probe attentional modulation of the MMN in experiments 1 and 2, with the goal of finding ignore conditions that more fully engage a participant’s attention to allow examination of automaticity. Strange’s (2011) Automatic Selective Perception model argues that first language speech perception is highly automatic in adults, and thus, attention to or away from the English vowel contrasts should not have an effect on adults. In contrast, as demonstrated in experiment 1, the MMN associated with speech discrimination in children 6-12 years of age remains dependent on the availability of attentional resources. Findings of age-related modulation of the MMN were surprisingly similar for the two ignore conditions employed in experiment 1 (passively viewing a video, performing a math task while being presented VCV contrasts); this was unexpected given the seemingly large differences in the attentional demands of the two conditions. This led us to consider a number of factors that made it difficult to directly compare the demands of the two ignore tasks, including the lack of objective behavioral indices of performance in the traditional passive task, dramatic differences in visual, attentional and computational demands, and the lack of time-limits in the math task.

In Experiment 2, the passive and math task were eliminated and replaced with two n-back task conditions (0-back, 2-back). Unlike the tasks in experiment 1, the n-back tasks had a steady presentation rate, were time-sensitive, and required a response. The 0-back task requires participants to attend to a minimally-complex, button-push vigilance task. The 2-back increases
cognitive demands through the introduction of a working memory component. As intended, our refined MMN paradigm successfully demonstrated our ability to modulate the MMN in adolescents. In addition to replicating our findings, given our somewhat surprising pattern of results for the attend condition (i.e., larger MMN in children than adult in the left inferior region), future work may consider inclusion of a broader developmental age range (e.g., 6-21) to gain further insights into the trajectories of the modulation of the MMN. Additionally, inclusion of a larger number of gradations for attentional demands may be considered. In sum, across the two experiments, we have found that manipulation of the MMN using tasks of different difficulty does not appear to be as straightforward as we hypothesized; in several leads, MMN amplitudes did not appear to be directly related to task demands (e.g., math and passive ignore conditions did not differ).

4.4 Nature of the Speech Stimuli

A key question that arises in the interpretation of our results is their generalizability. In response to concerns about the tendency of prior work to rely on relatively simplistic tones or syllables, the present work used vowel contrasts that were natural and therefore had greater complexity. One question is whether this goes far enough? Future work may benefit from examination of use of hard-to-detect natural speech contrasts, which are later to develop. For example, consonant contrasts, such as /z/ and /ð/ are perceptually more difficult, and thus may show larger age-related effects at later stages of development. Even before such pursuits, it may be optimal to take a group of 10-16 year olds and compare MMN responses and attentional modulation across a range of stimuli, differing in the detectability of speech contrasts. This would allow determination of optimal stimuli for mapping the later stages of speech perception development.
A few key limitations of the present work should be noted. First, we employed a smaller number of trials per condition than commonly used in the MMN literature because a higher number of trials in a child population would likely be too long for children to tolerate; fewer trials leads to less statistical power and necessitates a larger sample size of participants. Thus, we cannot rule out the possibility that more subtle effects could emerge if larger number of trials had been used. Second, we made use of a cross-sectional design. Longitudinal designs are superior at mapping the trajectories of developmental processes, because they decrease susceptibility to individual differences that are not related to age confounding the result. Although desirable, such studies also present difficulties, including the long time-span necessary to complete the study and the challenge of attrition.

5. Conclusion

Refinements in the MMN paradigm were successful in replicating prior age-related differences in MMN modulation between children and adults, and demonstrating more complex patterns of age-related change into adolescence. Additionally, the P3b phenomena showed age-related changes extending into adolescence. This finding suggested the importance of considering maturational changes in both top-down and bottom-up processes related to speech discrimination. Future work would benefit from further examination of speech perception into the teen years. A greater focus on speech contrasts known to be later to acquire, such as consonant place distinctions. Such consonant contrasts are often difficulties for children with language and reading deficits (Bradlow and Kraus, 1996).

In summary, the present work has a number of notable implications. In particular, the findings of our work highlight the need for: 1) continued examination of speech perception into late childhood and adolescence, 2) greater attention to the stimuli and tasks employed in
developmental studies using the MMN paradigm, and 3) consideration of MMN in relation to other ERP components, such as the obligatories and P3b. The present work indicates the value of examining teenagers to understand deficits in the age-range. Finally, the present work and future work that will be derived from it may have far reaching implications for the classroom, where greater awareness of the ongoing nature of development for speech processing can be used to optimize classroom environments and the delivery of instruction, especially for clinical populations.
6. Supplementary Figures

Figure 15 Children’s MMN waveforms at C3, C4 and Cz in Attend, 0-Back and 2-Back conditions
Figure 16 Children’s MMN waveforms at F3, F4 and Fz in Attend, 0-Back and 2-Back conditions
Figure 17 Adult’s MMN waveforms at C3, C4 and Cz in Attend, 0-Back and 2-Back conditions
Figure 18 Adult’s MMN waveforms at F3, F4 and Fz in Attend, 0-Back and 2-Back conditions
Figure 19 Children’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend condition.

Figure 20 Adult’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend condition.
Figure 21 Children’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend condition

Figure 22 Adult’s MMN at Fz, Left Mastoid and Right Mastoid in the Attend
Figure 23 Children’s MMN at Fz, Left Mastoid and Right Mastoid in the 2-Back condition

Figure 24 Adult’s MMN at Fz, Left Mastoid and Right Mastoid in the 2-Back condition
Figure 25  Grand Means Standards and Deviants in the Attend condition
Figure 26 Grand Means Standards and Deviants in the 0-Back condition
Figure 27 Grand Means Standards and Deviants in the 2-Back condition
7. APPENDIX 1

Primary Analysis with Age Treated as a Dimensional Variable

Given the stated goal to examine the trajectory of development for speech perception, we first carried out a statistical analysis of our data from a dimensional perspective with age as a covariate in our 2-factor (condition, time) repeated-measures ANOVA model. While our analyses treated age as a dimensional variable, we still felt visualization by age-groupings to be helpful for providing the reader with a richer appreciation of the data; as such, we provide plots of the mean event-related responses for three groupings of participants based on the lower-, middle and upper-terciles for our sample (10-13.6, 13.7-16, 18-25).

Similar to experiment 1, for each of the six a priori leads of interest (Fz, F3, F4, Cz, C3, C4), we first inspected the mean responses to standard and deviant trials to ensure the appropriateness of our measurements before testing for age- and attention-related differences in the MMN. Grand mean ERP waveforms for standard and deviant trials are depicted for the attend (Figures 28 and 29), 0-back (Figures 30 and 31) and 2-back (Figures 32 and 33). Figure 34 shows the difference waves in the attend condition by age group (10-13.6 years, 13.6-16 years, 18-25 years) at F3, C3, Fz, Cz, F4 and C4. Figure 35 portrays the mean event-related responses for each of the conditions (attend, 0-back, 2-back), averaged across all participants, for each of the primary leads of interest (F3, C3, Fz, Cz, F4, C4). Figures 36, 37 and 38 display difference waves by conditions for each age group, by site. Additionally, in Figure 39 we provide plots of the mastoid lead responses against the leads of interest (across all participants), to facilitate confirmation of our judgments regarding MMN identification by the reader.
As will be described in greater detail in the following sections, repeated measures ANOVA revealed a significant main effect of time in one or more of the time intervals examined (100-180; 200-280; 300-380). Fz was characterized by a later response than the other sites (F3, F4, Cz, C3, C4), only showing a significant effect of time in the 200-300 millisecond time-range. Significant interactions observed in each of the three time windows are reported below.

Figure 28. Standard vs. Deviant Responses for Attend condition plotted with standard error (red = deviant, blue = standard) with error bars at C3, C4 and Cz
Figure 29. Standard vs. Deviant Responses for Attend condition plotted with standard error (red = deviant, blue = standard) with error bars at F3, F4 and Fz
Figure 30. Standard vs. Deviant Responses for 0-Back condition plotted with standard error (red = deviant, blue = standard) with error bars at C3, F C4 and Cz
Figure 31. Standard vs. Deviant Responses for 0-Back condition plotted with standard error (red = deviant, blue = standard) with error bars at F3, F4 and Fz.
Figure 32. Standard vs. Deviant Responses for 2-Back condition plotted with standard error (red = deviant, blue = standard) with error bars at C3, C4 and Cz
Figure 33. Standard vs. Deviant Responses for 2-Back condition plotted with standard error (red = deviant, blue = standard) with error bars at F3, F4 and Fz
Figure 34. Difference waves in the attend condition by age group at F3, C3, Fz, Cz, F4 and C4
Figure 35 Grand Means difference waves in three attentional conditions at F3, C3, Fz, Cz, F4 a and C4
Figure 36. Difference waves in three attentional conditions at Fz and Cz
Figure 37. Difference waves in three attentional conditions at F3 and C3
Figure 38. Difference waves in three attentional conditions at F4 and C4
Early interval 100-200 ms

In the 100-200 millisecond time window, which is expected to isolate the MMN, a significant time X age interaction was noted at C3 (F= 4.034; p < 0.023); post-hoc
examination revealed a significant positive correlation between age and response amplitude at timepoints 140-159 (r = 0.51; p < 0.0004), 160-179 (r = 0.46; p < 0.001), and 180-199 (r = 0.40; p < 0.005), suggesting decreased MMN size (i.e., less negative amplitude) with age. A condition X time interaction was noted at C3 (F= 3.638; p < 0.012) with marginal effects at F4 (F=1.96; p < 0.096); in both cases, visual inspection suggested that the MMN for the attend condition was more negative than the two n-back ignore conditions (see Figure 40 for C3); post-hoc comparison confirmed that the linear effect of time differed between attend and N-back conditions (F=4.463; p < 0.041), and the linear (and quadratic) effects of time differed between 0-back and 2-back (F=5.765; p M 0.021). No significant condition X time X age interactions were noted. Main effects for time were found at F3 (F =3.371; p < 0.031), Fz (F=3.309 ; p < 0.04) and C3 (F=8.179 ; p < 0.001).

Figure 40  Condition * Time at C3 in the 100-200 time window.
Middle interval 200-300 ms

In the 200-300 millisecond time window, which tended to isolate the N2b and initial rise of the P3 components, an interaction of time x age was significant at F4 (F=3.237; p < 0.049) with marginal significance noted at Fz (F=3.05; p < 0.061); post-hoc correlation of the individual timepoints with age for F4, revealed positive relationships between age and response amplitude (i.e., less negative MMN amplitude with age), but none reaching significance. Only marginal effects were noted for condition x age x time at F3 (F=2.348; p < 0.062) and F4 (F=2.183; p < 0.067) and for condition x time at F4 (F=2.204; p < 0.064). Main effects for time were noted at F4 (F= 8.6; p < 0.001), Fz (F=7.982; p < 0.001), and C3 (F=3.506; p < 0.03).

Late interval 300-400 ms

In the 300-400 millisecond time window, a two-way interaction of condition X time was significant at C4 (F=3.794; p < 0.007) and Cz (F=5.069; p < 0.002). Visual inspection (see Figure 41) suggested these findings reflected the presence of a P3b for the attend condition; post-hoc comparison confirmed that the linear effect of time differed between attend and N-back conditions at C4 (F = 11.207; p < 0.002) and Cz (F=9.072; p < 0.005); no significant difference was revealed in the response for the two n-back trial types. A time x age interaction was significant at Cz (F=5.69; p < 0.01) with marginal significance at Fz (F=3.088; p < 0.052), C3 (F=2.838; p < 0.066) and C4 (F=2.044; p < 0.094). For Cz, post-hoc examination revealed a significant positive correlation between age and P3b response amplitude timepoints 300-319 (r = 0.39; p < 0.006) and 320-339
(r = 0.37; p < 0.01). Main effects for time were found at Fz (F= 6.446; p < 0.003), C3 (F=7.28; p < 0.001), C4 (F=4.059; p < 0.015) and Cz (F=12.797; p < 0.0001).

Figure 41. Condition * Time at C4 and Cz in the 300-400 time window.

3.2.4 Primary Analyses with age treated as a categorical variable

Given the limited age-related findings in our dimensional analyses, we repeated our primary analyses treating age as a categorical variable with two levels: adult (ages: 18-40) and child (ages 10-16.9). See Figures 42-47 for visualizations of waveforms for children and adults in each condition at C3, C4, Cz, F3, F4 and Fz.
Figure 42 Children and Adult MMN waveforms in Attend condition at C3, C4 and Cz

Figure 43 Children and Adult MMN waveforms in Attend condition at F3, F4 and Fz
Figure 44 Children and Adult MMN waveforms in 0-Back condition at C3, C4 and Cz

Figure 45 Children and Adult MMN waveforms in 0-Back condition at F3, F4 and Fz
Figure 46 Children and Adult MMN waveforms in 2-Back condition at C3, C4 and Cz

Figure 47 Children and Adult MMN waveforms in 2-Back condition at F3, F4 and Fz
As will be described in greater detail in the following sections, repeated measures ANOVA (within-subject factors: condition, time; between-subject factor: age-group) revealed a significant main effect of time in one or more of the time intervals examined (100-180; 200-280; 300-380; see Table 1). C4 and Cz were characterized by later responses than the other sites (F3, F4, Cz, C3), with Cz starting to show a significant effect in the 200-300 time range and Cz only showing a significant effect of time in the 300-400 millisecond time-range, which is characteristic of the P3. Significant interactions observed in each of the three time windows are reported below as well.

*Early interval 100-200 ms*

In the 100-200 millisecond time window, which is expected to isolate the MMN, significant time X condition interactions were noted at F3 (F = 3.928; p < 0.003) and C3 (F = 4.88, p < 0.002). For each of these leads, post-hoc comparisons revealed a more negative event-related response for the attend condition than the two ignore conditions. Specifically, for F3, attend vs. ignore differed with respect to linear effect of time (F = 10.1, p < 0.006), while the two ignore conditions did not differ. Similarly, for C3, attend vs. ignore differed with respect to a cubic effect of time (F = 11.179, p < 0.002). See Figure 48
In the 200-300 millisecond time window, which is also expected to isolate the MMN, a significant time X age-group interaction was noted in Fz (F = 5.525, p < 0.0001), with children showing a more negative response than adults (See Figure 49). Significant condition X time interactions were noted in Cz (F = 4.175, p < 0.005), C3 (F = 2.634, p < 0.032) and C4 (F = 3.023, p < 0.025) (See Figure 50). A significant condition * time * age-group interaction was noted at F3 (F = 4.2, p < 0.003) (See Figure 51). Visual inspection revealed a surprising ordering of responses for children where the 0-back had the weakest MMN response for children and attend the strongest. For adults, the MMN for attend was greatest, and the 2-back was smallest.
Figure 49. Time * Group (child, adult) at Fz.

Figure 50. Condition * Time at C3 and C4.
Figure 51. Condition * Time * Age-group at F3.

**Late interval 300-400 ms**

Finally, in the 300-400 millisecond time window, which is expected to isolate the P3, a significant condition * time interactions was noted in Fz (F=3.258; p < 0.001), Cz (F = 15.357, p < 0.0001), C3 (F = 7.762, p < 0.0001) and C4 (F = 8.979, p < 0.0001) See Figure 52,. As would be expected for the P3b, in each of these leads, post-hoc comparisons revealed a significantly greater response for the attend condition than the ignore conditions, which did not differ from one another.
Figure 52. Condition * Time * Age-group at Fz, Cz, C3, C4.
8. Table 2. Repeated measure ANOVAs for: 100-200ms, 200-300ms, 300-400ms time windows, labeled A, B and C, respectively. *Note: 0= p < 0.0001 in significance column.
References


