Developing an ecologically valid EEG paradigm to parse language subgroups in autism

By

Allison Paige Whitten

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Hearing and Speech Sciences

January 31st, 2019

Nashville, Tennessee

Approved:

James W. Bodfish, Ph.D.

Antje S. Mefferd, Ph.D.

Alexandra P. Key, Ph.D.

Stephen M. Wilson, Ph.D.

Reyna L. Gordon, Ph.D.

Dedicated to the families of children with and without autism spectrum disorder who participated in this research, without whom this work would not have been possible; in particular, to the families of minimally verbal children with autism, who taught a passionate language scientist that what makes all of us truly human is a shared connection that transcends language.

Acknowledgments

First and foremost, I would like to thank my dissertation advisor, Dr. Jim Bodfish, for his invaluable mentorship. It is not lost on me for a moment that I was incredibly lucky to have such an exceptional mentor. Jim has a gift for meeting his students where they are and knowing exactly what they need and how to help them, and I appreciated that at every stage of my Ph.D. his advice and support helped me to grow into an independent researcher and reach my full potential. I will never be able to thank you enough, Jim.

Next, I would like to thank my dissertation committee members. Thank you to Dr. Sasha Key, who taught me so much of what I know about EEG and event-related potentials, and was always willing to answer my questions, no matter how small. Thank you to Dr. Antje Mefferd, who was in the trenches with me designing the stimuli and tasks for this work, and allowed me to grow as a scientist by witnessing how methodically she approaches her research. Thank you to Dr. Stephen Wilson, who brought his incredible language neuroscience expertise to this work, and many thought-provoking discussions as a result (and, his cow jokes aren't so bad either). Finally, thank you to Dr. Reyna Gordon, who contributed EEG expertise beyond event-related potentials, and was a positive influence in my career development through her willingness to chat with me and discuss our shared interests on so many occasions.

I would also like to thank everyone in the Bodfish lab for their helpful insights and discussions over the years: Kathryn Unruh, Robin Shafer, Sahana Nagabhushan Kalburgi, Pumpki Su, Mary Alice Keller, and Kate Slaboch. Additionally, I would like to thank my collaborators on this work, including Dr. Camarata and Dr. Ashmead for their input on my early projects, and Dr. Travis Moore for creating the brilliant R program I used to analyze my data.

Furthermore, I want to thank Jennifer Pilkingon, Renee Ingle, Anna Kulaski, and Brittany Stevenson for allowing me to shadow them during the first few years of the Preschool for Children with Autism in the Bill Wilkerson Pediatric Speech-Language Pathology Clinic. This hands-on experience taught me more about language in children with autism than many articles could, and greatly informed much of this work.

I would also like to thank my family and friends for their support during this journey. Thank you to my parents for their unrelenting encouragement over the years, and especially my mom for introducing me to the world of science and research and fostering my dream to obtain my Ph.D. when I was just a little girl. Thank you to my sister Laura and brother Ben, for preparing me for my Ph.D. by teaching me from a young age that things will not always go my way, but supporting me even when they did not. Also, thank you to Justin, who was there for me during almost the entirety of my Ph.D. and undoubtedly brought out my best work through his diligent roles as my sounding board and my biggest cheerleader.

Finally, thank you to all of my Hearing & Speech colleagues and faculty members for making our department the most supportive environment to pursue a Ph.D. in. I could not imagine better people to get to work with every day, and I will miss all of you.

Table of Contents

Page

Chapter

List of Tables

List of Figures

CHAPTER 1

Introduction

Overview

Language impairments represent a critical aspect of the manifestation of autism spectrum disorder (ASD) throughout development. While the literature base on the presence and nature of language-related deficits in ASD is large, research focusing on the mechanisms underlying language impairments is sparse. Mechanistic, brain-behavior research is needed to guide clinical efforts related to early identification and intervention, especially for the most severely impacted children. Previous studies that have assessed mechanistic accounts have been limited by insufficient characterization of the language abilities of the sample, and the use of synthetic stimuli that lacked ecologically validity. The overall purpose of this dissertation was to create a novel auditory EEG paradigm with ecologically valid stimuli, and apply it to language subgroups in ASD to gain an understanding about potential causes leading to distinct language deficits in ASD. The sections that follow: 1) provide background on the breadth of language impairments in ASD, with a particular focus on the largely ignored subgroup of children with minimally verbal autism, and outline how these impairments have been examined to date using event-related potential (ERPs) measured from EEG, 2) describe a systematic series of experiments using ERPs designed to extend the mechanistic understanding of language impairments in ASD, and 3) detail the implications of the findings of these experiments and identify important areas for future research.

Language Impairment in ASD

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by social communication impairment and the presence of repetitive and restricted behaviors. Deficits in social communication include impaired use of both verbal language and nonverbal gestures to initiate and respond during social interactions. There is a general consensus that pragmatic language, or the social use of language, is universally impaired in ASD (Groen, Zwiers, van der Gaag, & Buitelaar, 2008; Tager-Flusberg, 1981; Wilkinson, 1998). These pragmatic deficits include impaired perception and production of prosody (Paul, Augustyn, Klin, & Volkmar, 2005; Rutherford, Baron-Cohen, & Wheelwright, 2002; Shriberg et al., 2001), use of idiosyncratic language (Loveland, McEvoy, Tunali, & Kelley, 1990; Volden & Lord, 1991), perseveration on specific topics (Baltaxe, 1977), misinterpretation of literal language (Bishop & Adams, 1989), and difficulties with conversational discourse (Capps, Kehres, & Sigman, 1998; Hale & Tager-Flusberg, 2005). Traditionally, structural aspects of language relating to grammar and syntax were thought to be intact in individuals with ASD, however, more recent studies have challenged this claim (Boucher, 2012; Brynskov et al., 2017). Consequently, beyond universal deficits in pragmatic language, research on language in ASD has found significant heterogeneity ranging from individuals who are completely nonverbal to those who are verbally fluent, as well as individuals with a phenotype similar to other language disorders like specific language impairment (Kjelgaard & Tager-Flusberg, 2001).

Such variability in the symptomatology of autistic language deficits has led to a recent attempt to identify language subtypes within ASD (Rapin, Dunn, Allen, Stevens, & Fein, 2009; Tager-Flusberg, 2006). One such subtype, referred to as nonverbal or minimally verbal ASD (MVA), describes individuals with autism who do not acquire functional use of language. Current estimates suggest that this group comprises 25-30% of children with autism (Anderson et al., 2007; Rose, Trembath, Keen, & Paynter, 2016). The amount of spoken language used by minimally verbal individuals with ASD varies from completely mute to use of a few functional words and/or rote phrases, or speech that is entirely echolalic and scripted. Importantly, they may use these words or phrases in only one or two limited contexts (Tager-Flusberg & Kasari, 2013). Kanner's original work in describing eleven cases of ASD included three minimally verbal children whose language abilities were described in the following ways: making noises but saying only one word (Case 3), vocalizations in the form of crying out, humming tunes, with four single words (Case 6), and inarticulate speech expressed in a monotonous singsong tone (Case 7; Kanner, 1943).

Much less is known about the receptive language abilities in MVA because these children likely understand more than they can demonstrate to experimenters. In a comparison of five different receptive language measures, Plesa Skwerer, Jordan, Brukilacchio, and Tager-Flusberg (2016) found that receptive abilities varied significantly among MVA individuals. Depending on the measure used, receptive abilities in their sample of 19 minimally verbal children and adults (5-21 years) ranged from scoring at floor to scoring within normal age equivalents. The majority of receptive measure scales were found to be negatively correlated with the severity scale from the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2012), suggesting that autism symptom severity decreases with increasing receptive abilities. In addition, Rapin et al. (2009) found that children with the lowest functional language scores in a large sample of ASD also presented with low receptive language scores, however, it is unclear whether these children were truly minimally verbal.

The language characteristics of minimally verbal children reviewed above are quite general; ultimately, the expressive and receptive abilities depend crucially upon the definition and measures used in any given study. For example, a study that defines MVA as producing less than 20 words has already limited the scope of the expressive language abilities in their sample, while a study which defines MVA as children without sentence-level expressive language will likely have a wider range of both expressive and receptive abilities. Likewise, a study measuring receptive language using a standardized assessment that asks a minimally verbal child to verbally respond will result in lower scores than use of a receptive measure with no verbal demands. The resulting disparate conclusions about language ability in MVA highlight the need for agreement among researchers on an appropriate definition and measures for use across studies.

Despite variable definitions across studies, there is reason to believe that minimally verbal children differ from another subgroup of children who have acquired the ability to combine words into phrases, often termed the "phrase speech" subgroup (Thurm, Manwaring, Swineford, & Farmer, 2015). This is because the transition from single words to the use of phrases is known to be a distinct stage of language acquisition that usually begins when the child has acquired approximately 50 spoken words (Bates & Dick, 2002; Clark, 2009; Dromi, 1999). Although there have been few longitudinal studies of language development in children with autism, Tager-Flusberg and Calkins (1990) found evidence that children with ASD follow the same trajectory of language acquisition stages as typically developing children. Their results indicate that children with ASD who have moved to the phrase stage have progressed further in their language than those with single words, and greatly widened the scope of their language by being able to use phrases. However, it is important to note the distinction between spontaneously combining words in to phrases and the use of memorized rote phrases that have been documented in minimally verbal children with autism. In the absence of the ability to create novel phrases, rote memorized phrases are often considered "chunks" on par with single word use (Nattinger, 1992). As such, phrase speech autism represents a distinct subgroup from minimally verbal autism, which implies that definitions of minimally verbal autism would be greatly improved by removing arbitrary limits on number of words and instead focusing on the functional demarcation between single words and phrase speech.

To date, the published research literature focusing on minimally verbal children with ASD is severely limited. The primary reason these children are excluded from research is due to the difficulty in testing this population on standard assessments and current experimental paradigms. Consequently, Tager-Flusberg and Kasari(2013) have suggested that research using passive measures like eye-tracking and EEG/MEG which do not require the child to understand verbal directions or make overt responses is urgently needed to test theories of language impairment in MVA.

Applying Theories of Language Impairment to Minimally Verbal Autism

Generally, theories concerning language impairment in ASD have focused on explaining the language deficits of high-functioning individuals. The most commonly researched model has been theory of mind, or the idea that language impairment in ASD is due to an inability to identify the mental states of others (Tager-Flusberg, 1999). However, current tasks assessing theory of mind with narratives and questions are not feasible for use in minimally verbal children, and have produced conflicting results in studies of high-functioning children (Capps et al., 1998; Hadwin, Baron-Cohen, Howlin, & Hill, 1997; Joseph & Tager–Flusberg, 2004; Tager-Flusberg & Sullivan, 1995). Other theories explaining language impairment in ASD include temporal processing deficits and wider temporal binding windows (Boucher, 2001; Kwakye, Foss-Feig, Cascio, Stone, & Wallace, 2010; Stevenson et al., 2014), declarative memory impairment (Boucher, Mayes, &

Bigham, 2008), sensory processing impairment (Marco, Hinkley, Hill, & Nagarajan, 2011; Whitehouse & Bishop, 2008), and decreased social motivation (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012; Tager-Flusberg & Kasari, 2013). While the temporal processing and declarative memory impairment theories typically utilize in-depth memory assessments and psychophysical tasks that are also not practical for use in MVA, the theories of auditory processing impairment and decreased social motivation lend themselves to passive experimental designs that could be used to study language impairment in minimally verbal autism populations, as well as identify possible biomarkers that could be used to detect language impairment in ASD early on.

Social Motivation Theory

Social motivation can be defined as the combination of innate biological factors that interact to create a preference for the social world (Chevallier et al., 2012). When viewed from the perspective of social motivation, language impairment in ASD may result because children with autism fail to orient and attend to social stimuli like human speech, and this in turn constrains their ability to learn language. There is evidence to suggest that this may be the case; studies assessing preference for speech over nonspeech sounds have found that typically developing children show a bias for speech as early as the first few days of life (Vouloumanos & Werker, 2007), while young children with autism show a preference for nonspeech sounds instead (Klin, 1991; Kuhl, Coffey-Corina, Padden, & Dawson, 2005).

In addition to attending to speech, social motivation may play a crucial role in language acquisition by fostering and enhancing the social interactions necessary for language learning (Kuhl, 2007). For example, the ability to respond to joint visual attention with a conversational partner is thought to encourage word learning (Baldwin, 1995), and has been found to continually predict language outcomes in typically developing children (Brooks & Meltzoff, 2005; Markus, Mundy, Morales, Delgado, & Yale, 2000; Morales et al., 2000; Mundy & Gomes, 1998; Tomasello & Farrar, 1986). Many studies have demonstrated impaired joint attention skills in ASD compared to both typically developing children and children with intellectual disabilities (Dawson et al., 2004; Leekam, López, & Moore, 2000; Loveland & Landry, 1986; McArthur & Adamson, 1996; Mundy, Sigman, & Kasari, 1990), and better joint attention skills have been associated with greater language development in children with ASD (Bono, Daley, & Sigman, 2004; Charman, 2003; Kasari, Paparella, Freeman, & Jahromi, 2008). Therefore, decreased social motivation in children with ASD may impair the development of pre-linguistic social interaction skills that contribute to language learning.

For social motivation to promote language acquisition beginning in infancy, it must be inherently rewarding to engage with others. Indeed, Depue and Morrone-Strupinsky (2005) have proposed a neurobehavioral model in which appetitive and consummatory phases of social affiliation are intrinsically rewarding via interactions between dopaminergic and opiate systems. Similarly, in an extensive review on the evolution of social and motivational processes in both human language and birdsong, Syal and Finlay (2011) suggest that language is primarily learned via connections between a vocal cortico-striatal learning pathway and subcortical social-motivation circuitry. In line with these models, the social motivation theory of language impairment in ASD posits that decreased orienting to and engagement with social stimuli may result from decreased neural connections between language-learning regions and dopaminergic circuitry. This is supported by the findings of Abrams et al. (2013), in which children with ASD showed significantly reduced connectivity between a voice-selective region of the superior temporal sulcus (STS) and nodes of the dopaminergic reward circuitry. This study found that children with ASD showed normal connectivity between reward circuitry and the primary auditory cortex, suggesting there may be decreased motivation to attend to speech, but not other sounds. Furthermore, they found that the degree of underconnectivity between the STS and reward processing areas in individual participants with ASD was predictive of more severe communication deficits, suggesting a direct link between abnormal social reward processing and language deficits.

Auditory Processing Theory

One important caveat of the social motivation hypothesis is its assumption that children with autism have the ability to process linguistic information successfully, but they fail to do so only because they are not motivated to allocate neural resources to attending. Another theory, impaired auditory processing, suggests that some children with autism may struggle to learn language due to primary deficits at the level of processing the complex acoustical features of speech that leads to downstream difficulties in learning, storing, and comprehending speech.

The theory of impaired auditory processing posits that deficits occur at some point during the sensory encoding of an auditory stimulus, with the earliest stage of processing beginning in the brainstem. Studies of the auditory brainstem response (ABR) in children with ASD have found mixed results in response to tones (Kwon, Kim, Choe, Ko, & Park, 2007; Tanguay, Edwards, Buchwald, Schwafel, & Allen, 1982; Taylor, Rosenblatt, & Linschoten, 1982; Tharpe et al., 2006) but consistent impaired processing when the tones are presented in more complex paradigms, e.g., faster presentations (Fujikawa-Brooks, Isenberg, Osann, Spence, & Gage, 2010; Källstrand, Olsson, Nehlstedt, Sköld, & Nielzén, 2010) or studies using speech sounds (Russo et al., 2008; Russo, Nicol, Trommer, Zecker, & Kraus, 2009). In both of these studies measuring ABR responses to speech, atypical responses correlated with lower language abilities, providing evidence for impaired auditory processing related to language impairment in ASD.

Following brainstem processing, fine-grained sensory processing takes place in the primary auditory cortex. As mentioned previously, Abrams et al. (2013) found underconnectivity between regions implicated in reward and the voice-selective STS, but found normal connectivity between the primary auditory cortex and these same reward regions. While these results suggest that perhaps auditory processing of speech remains intact and the deficits instead lie in abnormal connections to reward circuitry, the findings of Gervais et al. (2004) suggest that auditory processing of the speech signal is itself abnormal in individuals with ASD. They found that when adults with ASD listen to speech and vocal sounds, they did not activate the voice-selective bilateral STS regions that typically developing adults did. Further evidence for abnormal auditory processing of speech comes from neuroimaging studies finding abnormal hemispheric specialization for speech in children and adults with ASD. The activation pattern observed during processing of speech in typically developing individuals involves stronger activation in language regions of the left hemisphere, yet multiple studies have found right hemisphere dominance in individuals with autism (Boddaert et al., 2004; Eyler, Pierce, & Courchesne, 2012; Flagg, Cardy, Roberts, & Roberts, 2005; Redcay & Courchesne, 2008; Takahashi, Nakahachi, Stickley, Ishitobi, & Kamio, 2016). Notably, two studies have found that children with autism without language impairment do not exhibit this abnormal right-hemisphere dominance (Dawson, Finley, Phillips, & Galpert, 1986; De Fossé et al., 2004), suggesting a potential link between abnormal hemispheric auditory processing and language deficits in ASD.

In line with the theory that auditory processing impairments could be a product of abnormal hemispheric specialization in ASD, Haesen, Boets, and Wagemans (2011) proposed that because the right hemisphere is thought to be more specialized in processing pitch compared to the left hemisphere, this right-hemisphere dominance could explain why so many studies have found enhanced pitch and vowel processing in children and adults with ASD (Bonnel, Mottron, Peretz, Trudel, & Gallun, 2003; Heaton, 2005; Heaton, Hermelin, & Pring, 1998; Järvinen-Pasley & Heaton, 2007; L. Mottron, Peretz, & Ménard, 2000; O'Riordan & Passetti, 2006). However, more recent studies suggest only 10-20% of individuals with ASD show exceptional pitch processing.

Other behavioral studies of auditory processing in ASD have found deficits in the processing of prosody (Golan, Baron-Cohen, & Hill, 2006; Kleinman, Marciano, & Ault, 2001; Rutherford et al., 2002), speech-in-noise (Alcántara, Weisblatt, Moore, & Bolton, 2004; Groen et al., 2009), and hypersensitivity to sounds (Khalfa et al., 2004; see O'Connor, 2012 for a full review). These findings support the hypothesis that a general enhancement of processing low level features like pitch occurs at the expense of processing more complex features which require extracting simultaneous spectral and temporal content (Mottron, Dawson, Soulières, Hubert, & Burack, 2006). The majority of behavioral studies have been restricted to high-functioning children and adults with ASD, however, and thus these findings may not extend to lower-functioning and minimally verbal children who may also show deficits in low-level sensory processing as well.

In summary, impairment in the auditory processing of speech sounds in ASD may range from as early as the brainstem up to abnormal right-hemisphere dominance. In addition, many of these abnormalities have been related to language deficits in ASD, suggesting that impaired auditory processing is an important theory to test in examining causes underlying severe language impairment in minimally verbal children with autism.

Testing Social Motivation & Auditory Processing Theories Using EEG

Event-related potentials (ERPs) measured from EEG have produced a wealth of knowledge on how the brain processes speech sounds. Through decades of research, a variety of ERP components have been identified that index early sensory encoding of the acoustical properties and later higher order cognitive processes related to attention, lexical retrieval, and semantic evaluation of speech. Due to their relative ease of use, ERPs related to speech processing offer a unique opportunity to study language deficits in children and adults with autism spectrum disorder (ASD) who may not tolerate other neuroimaging methods like fMRI and MEG.

Several event-related potentials (ERPs) obtained using an auditory oddball paradigm have been employed to examine deficits at the level of auditory processing and social motivation. This is due to the increased temporal resolution of EEG, which can distinguish between early sensory and later cognitive cortical processes. Event-related potential (ERP) components occurring within 50 – 200 ms after sound onset provide a marker for early auditory sensory processing, while the P3a component occurring around 300 ms provides a measure of attentional capture by salient events, a proxy for social motivation. The auditory oddball paradigm typically involves presentation of a standard sound that is repeated on 80-85% of trials, and a deviant sound presented on 10-15% of trials. The standard and deviant sounds both elicit auditory ERP components which reflect the first stages of cortical auditory processing, while the deviant sound also elicits a P3a component reflective of an attention switch or orienting response to the stimulus.

In ERP studies with children and adolescents, the early auditory components are usually referred to as the P1 and N2 ERPs, often collectively referred to as "obligatory" components because they are robust and reliable. The P1 and N2 are frontocentral components that occur within 200 ms of the auditory stimulus (Ponton, Eggermont, Kwong, & Don, 2000). However, when interstimulus intervals longer than 1 second are used, the adult pattern of N1 and P2 auditory ERPs can be observed in children (Bruneau & Gomot, 1998; Ceponiene et al., 2002; Paetau, Ahonen, Salonen, & Sams, 1995). The N1 reflects encoding of the physical and temporal characteristics of the stimulus, and is thought to be generated in the primary auditory cortex (Risto Näätänen & Picton, 1987). The P2 component is strongest over central sites (approximately 150 – 250 ms), and also indexes physical and temporal characteristics of the stimulus (Crowley & Colrain, 2004), with evidence for neural generators in the primary and secondary auditory cortices (Godey, Schwartz, de Graaf, Chauvel, & Liégeois-Chauvel, 2001). The functional interpretation and source localization of the children's P1 and N2 components are less well defined, but instead together are thought to reflect early acoustical encoding in both primary and secondary auditory cortices before mature development of the adult N1 (Ceponiene et al., 2002).

The early auditory ERP components can be used to test the auditory processing theory of language impairment in children with ASD; support for this theory arises when the auditory ERP components are compared to typically developing individuals and found to be significantly different in either amplitude or latency, reflecting altered amount of neural processing or differing speed of processing, respectively. These differences in the amplitude and latencies between ASD and TD could be found to both speech and nonspeech stimuli, reflecting a global auditory processing deficit, or they may be found to be specific to speech processing. Furthermore, testing differences between the latency to speech and nonspeech stimuli within the ASD group could support deficits in auditory processing if children with ASD process the speech stimuli more slowly than the nonspeech stimuli. However, it is not yet known if typically developing children show latency differences in the auditory ERPs to speech vs. nonspeech processing; if they do show faster processing to speech over nonspeech, it could represent a neural bias to process speech sounds at the earliest stages of auditory processing in TD that may not be present in individuals with ASD.

Following sensory processing reflected by the P1 and N2 components, the P3a component represents an "orienting response" related to involuntary attention switching and initial allocation of attention to a salient stimulus (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2003; Squires, Squires, & Hillyard, 1975). The P3a latency occurs at approximately 300 – 400 ms, and is thought to originate in the anterior cingulate cortex through dopaminergic neurotransmitter activity (Polich & Criado, 2006). This ERP component thus provides a marker for whether a new sound is judged to be salient enough to activate frontal lobe mechanisms and engage further processing; a finding of reduced amplitude or lack of a P3a suggests a deficit in switching attention to a sound. Thus, the P3a can provide evidence for decreased social motivation if children with ASD show decreased P3a amplitude to speech sounds compared to nonspeech sounds.

The Auditory and P3a ERP Components Relation to Language Abilities

In ERP studies of speech processing, the auditory ERPs have been used to study pre-attentive discrimination between different speech sounds in infants, children, and adults (Digeser, Wohlberedt, & Hoppe, 2009; Molfese & Molfese, 1985; Molfese et al., 2002). The P3a has been used to answer questions regarding which language processes require attention vs. automatic processing (X.-D. Wang, Gu, He, Chen, & Chen, 2012; Winkler, Kujala, Alku, & Näätänen, 2003), and which kinds of acoustical speech cues elicit the attention of the listener (Wang, Friedman, Ritter, & Bersick, 2005). In addition to group studies in typical development, one of the most exciting aspects of speech ERP research is the potential to be used as a practical tool to identify children at risk of language disorders early in development. In order to do this, the auditory ERPs must be sensitive enough to measure individual differences that relate to speech and language ability within the typical population.

Several studies have found relationships between the auditory ERPs to speech sounds in infancy and later language performance on standardized tests at 24 months (Benasich et al., 2006), and 3-5 years of age (Choudhury & Benasich, 2011; Molfese & Molfese, 1985, 1997). In addition, Mills, Coffey-Corina, and Neville (1993) found that the spatial distribution of the N2 response can reliably distinguish between words that toddlers comprehend vs. do not comprehend. Last, (Carpenter and Shahin (2013) showed that six-year-old children with N1 and P2 responses that resembled adults performed as well as the adult sample on a phonetic judgement task, while four and five-year-olds showing immature N1 and P2 ERPs performed worse. Their findings suggest that maturational changes in the auditory cortex measured by N1 and P2 provide enhanced speech sound discrimination in children. These studies provide preliminary data to support the use of the auditory ERPs to measure individual differences in language abilities.

The P3a component has been less commonly related to behavioral measures of language functioning in typical development, and more often applied to populations in which attention is thought to be impaired (e.g., attention deficit hyperactivity disorder). However, a few studies suggest that the P3a component could be related to language ability. For example, Getzmann, Falkenstein, and Wascher (2015) found that older adults who performed worse on a forced-word discrimination task showed a significantly longer latency of the P3a component. In addition, Shestakova, Huotilainen, Ceponiene, and Cheour (2003) found that the P3a amplitude was larger to French phonemes in children learning French compared to children who were not learning French, suggesting a role for attention during the language learning process. Interestingly, Liu, Shi, Zhang, Zhao, and Yang (2007) found that children with higher IQs had a significantly shorter latency and larger amplitude of the P3a component to consonant-vowel syllables compared to children with average IQ scores. Given the high correlations between language ability and IQ (Sandel, 1998), this finding combined with the results of other studies reported above suggest further investigation into the relationship between the P3a ERP and language ability is warranted.

Previous Studies Assessing Auditory and P3a ERP Components in Autism Spectrum Disorder

Studies investigating the early auditory and P3a ERPs using an oddball task in children with autism have primarily focused on high-functioning individuals with autism. These studies typically find no differences in the auditory ERP latencies compared to typically developing (TD) children, suggesting normal speed of processing to both speech and nonspeech stimuli (Čeponienė et al., 2003; Gomot, Giard, Adrien, Barthelemy, & Bruneau, 2002a; Kujala et al., 2010; T. Lepistö et al., 2006; Tuulia Lepistö et al., 2005a; Lindström, Lepistö-Paisley, Vanhala, Alén, & Kujala, 2016; Whitehouse & Bishop, 2008). Although, a few studies have found delayed latency to speech (Korpilahti et al., 2007) or shorter latency to nonspeech (Ferri et al., 2003). In contrast to relatively normal latencies, amplitudes of the early auditory ERP responses have often been found to be decreased in children with ASD to speech sounds, indicative of reduced synchronous neural processing (Courchesne, Lincoln, Kilman, & Galambos, 1985; Jansson-Verkasalo et al., 2003; Korpilahti et al., 2007; Lepistö et al., 2005; Whitehouse & Bishop, 2008), with fewer studies finding reduced amplitude to nonspeech stimuli as well (Jansson-Verkasalo et al., 2003; Lepistö et al., 2005).

In regards to the P3a component, multiple studies have found children and adults with autism show reduced P3a amplitudes to speech compared to typically developing samples (Fan &

Cheng, 2014; Kemner, Verbaten, Cuperus, Camfferman, & van Engeland, 1995; Lepistö et al., 2005, 2006, 2007), but no differences or even enhanced P3a amplitudes to nonspeech (Čeponienė et al., 2003; Gomot, Giard, Adrien, Barthelemy, & Bruneau, 2002b; Lepistö et al., 2005, 2006, 2007; Lincoln, Courchesne, Harms, & Allen, 1993; Lindström et al., 2016; Lortie et al., 2017; Whitehouse & Bishop, 2008). Few exceptions to this pattern exist; Whitehouse and Bishop found an enhanced P3a amplitude to speech and decreased P3a amplitude to nonspeech, but this was explained by a crucial paradigm difference in which their deviant speech stimuli always followed a repeated train of nonspeech standard stimuli, and vice versa. The authors consequently suggested that P3a amplitude in children with ASD is driven by a top-down mechanism in which repeated speech stimuli is increasingly tuned out, but repeated nonspeech stimuli is not. In addition, a few studies investigating P3a to tones have found reduced amplitude or delayed P3a responses in low-functioning children with ASD, but without a speech comparison included in these studies, it is unclear whether the P3a amplitude or latency may be further impaired to speech in these children (Ferri et al., 2003; Salmond, Vargha-Khadem, Gadian, de Haan, & Baldeweg, 2007).

The majority of results comparing the P3a component to speech and nonspeech in individuals with ASD support decreased levels of social motivation, or an increased motivation to attend to non-social stimuli in ASD. Together with the results of the early auditory ERP components, these studies suggest evidence for global and speech-specific auditory processing deficits in some individuals with ASD, while deficits in early attentional mechanisms related to social motivation appear to be speech-specific.

The only EEG/ERP study which has included minimally verbal children with ASD used a passive picture-word matching paradigm and found that compared to typically developing children, MVA children showed delayed latency of the auditory P1 to words presented with matching or mismatched pictures. However, since this study did not include verbally fluent children with ASD as a control, it is unclear whether the observed P1 delay in this type of task is specific to MVA children or children with ASD in general. The findings from the previous auditory oddball studies in ASD suggest that passive ERP measures offer a valid way to test the theories of impaired auditory processing and social motivation as mechanisms underlying language impairment in minimally verbal ASD.

Improving the Ecological Validity of Stimuli Used in ERP Speech Processing Studies

Although the use of passive oddball paradigms in ERP research has allowed investigators to study the mechanisms of language impairment in autism spectrum disorder, the conclusions remain limited due to the reduced ecological validity of the experimental stimuli. The majority of previous ERP studies have used synthetic speech stimuli created by a computer rather than natural speech recorded from the human voice. ERP studies that compare speech to nonspeech stimuli have relied on synthetic speech because of the difficulties involved in matching a nonspeech stimulus to the inherent spectro-temporal complexity found in natural speech stimuli. In contrast, the simple nature of synthetic speech allows researchers to easily match the complexity to a simple nonspeech analog, usually a complex tone made up of just three or four sinusoidal waves. Matching the acoustic features and complexity to a nonspeech sound is critical in EEG/ERP research because ERP components arise from synchronized activity from populations of neurons, and thus without matching the stimuli one cannot rule out that any differences are due to larger populations of neurons responding to the more complex stimulus. The use of synthetic speech provides a solution to the matching issue as synthetic speech is most often programmed to include only the most rudimentary aspects of the speech signal (and thus sounds quite robotic).

Recently, researchers have called for the use of naturalistic stimuli in neuroscience experiments of language and auditory processing (Hamilton & Huth, 2018; Theunissen & Elie, 2014). This comes in light of new findings at the level of single neurons whereby responses to simple artificial stimuli cannot predict responses to complex natural sounds (Mizrahi, Shalev, & Nelken, 2014), suggesting that ERP responses from populations of neurons in response to simple synthetic stimuli may not map onto responses from natural complex stimuli. Consequently, the same reason that synthetic speech has been advantageous in ERP research (i.e., less complexity) could also mean that previous auditory ERP results to synthetic speech may not accurately represent neural processing of natural speech. Indeed, comparing ERP studies that have used the same paradigm with synthetic speech vs. natural speech suggest that this is the case.

Using synthetic speech, Sharma, Marsh, and Dorman (2000) and Tremblay and Kraus (2002) found differences in the latencies of the auditory ERP components in response to the synthetic speech syllables. However, Tremblay, Friesen, Martin, and Wright (2003) used a continuum of natural speech syllables and found no differences in latency, suggesting that the auditory cortex may treat synthetic speech differently due to the impoverished frequencies. In addition, when comparing synthetic and natural speech stimuli within the same study, Swink and Stuart (2012) found a faster latency of the early auditory responses to natural than synthetic speech vowels, which is also suggestive of a difference at the earliest level of cortical processing.

Another example of conflicting ERP results comes from the N400 response, in which Hagoort and Brown (2000) found an additional early response occurring around 250 ms to natural speech, while studies using synthetic speech only found a response occurring around 400 ms (Ardal & Donald, 1991; Herning, Jones, & Hunt, 1987). This suggests that natural speech may undergo earlier and more intensive semantic processing. Additionally, Sandridge and Boothroyd

(1996) compared synthetic and natural speech in a commonly used paradigm to test discrimination of speech sounds using the mismatch negativity (MMN). They found that the MMN discrimination response to synthetic speech syllables displayed a more well-defined response with larger amplitudes and a consistent latency compared to natural speech. The MMN to synthetic speech was consistent of synthetic speech responses found in many other studies (see (Näätänen, Paavilainen, Rinne, & Alho, 2007 for a review), while the MMN to natural speech was less robust. This difference in the neural response between natural and synthetic speech most likely has to do with the increased acoustic variability of natural speech. That is, because there is more than one acoustic feature to discriminate at any given time, the discrimination response to natural speech results in a less well-defined ERP response.

The differences outlined above between synthetic and natural speech ERP components are likely related to behavioral findings that suggest synthetic speech is a more difficult stimulus to process across both perception and comprehension tasks. For example, studies have consistently found that the intelligibility of synthetic speech is worse than natural speech on word recognition tasks (Allen, Hunnicutt, Klatt, Armstrong, & Pisoni, 1987; Hawkins & Slater, 1994; Kangas & Allen, 1990; Logan, Greene, & Pisoni, 1989; Nye & Gaitenby, 1973). Even in studies that use relatively higher quality intelligible synthetic speech, comprehension breaks down when task complexity is increased in comparison to performance with natural speech (Drager & Reichle, 2001; Duffy & Pisoni, 1992; Paris, Gilson, Thomas, & Silver, 1995). Furthermore, response rates are increased by 145 ms compared to natural speech (Pisoni, 1981). Importantly, individuals with language impairment or intellectual disability have been found to exhibit increased difficulty with synthetic speech compared to typically developing individuals (see Koul, 2003 for a review), which likely has implications for neuroimaging studies using synthetic speech in these populations.

As a result of these findings, Luce, Feustel, and Pisoni (1983) proposed that processing synthetic speech requires greater cognitive capacity and working memory.

The reason that synthetic speech is more difficult to process may have more to do with what synthetic speech excludes, rather than includes. Indeed, synthetic speech is made up of the "bare bones" of the speech signal – it is often programmed to include only the frequencies that make up each consonant and vowel. This essentially gets rid of the "redundancy" found in the rich spectral and temporal content of natural speech, whereby multiple acoustic features signal the same sound. This is likely not a coincidence; humans communicate in noisy environments frequently, so having more than one cue strongly increases the signal to noise ratio, and the chance that the receiver understands the message. Thus, while it seems that studying only the purely linguistic content of the speech signal using synthetic speech would allow researchers to more directly investigate speech perception, the evidence reported above suggests otherwise. Specifically, because synthetic speech is missing so many of the rich acoustic cues found in natural speech, listeners are forced to use alternate processing strategies that often result in contradictory findings when compared to natural speech.

In order to increase the ecological validity, some research groups have improved the "naturalness" of synthetic speech by starting with a natural utterance and manipulating it to create "semi-synthetic speech" (Alku, Tiitinen, & Näätänen, 1999). While these stimuli do sound significantly more natural and still allow researchers to have full control over the acoustic parameters of the voice, they inherently increase the complexity of the stimuli. Studies that have used these semi-synthetic speech stimuli continued to use complex tones as their matched nonspeech counterpart, and thus their stimuli were no longer matched on complexity.

Consequently, it is imperative to the advancement of ERP speech processing research to include both a naturalistic speech sound and a naturalistic complexity-matched nonspeech sound.

Purpose of the Present Work

The present work had two primary aims: 1) to develop and validate a novel ERP paradigm using naturalistic speech and nonspeech stimuli to study language processing in typical and language-impaired populations; and 2) to apply the novel paradigm to elicit ERP components related to auditory processing (auditory ERP components) and social motivation (P3a component) to language subgroups in ASD to identify differences in their ERP profiles that could relate to potential causes. The remaining chapters are the products of these two aims.

In Chapter 2, a novel ERP oddball paradigm using natural speech matched to an animal vocalization was developed and piloted in a sample of typically developing children. An oddball with synthetic speech and complex tones was also included in this study to compare differences between the auditory and P3a ERP components to synthetic and natural stimuli. In addition, this study provided a framework from which to investigate a novel research question in typical development: whether the early stages of cortical auditory and attentional processing differ for speech compared to nonspeech stimuli, and whether this differs for natural vs. synthetic speech. In Chapter 3, two new oddball tasks with natural speech and nonspeech stimuli were created to assess how the design of the oddball task impacts the auditory and P3a ERP components. This study switched the sound category (speech vs. nonspeech) of the standard and deviant sounds to create congruent (standard and deviant sounds were from the same sound category) and incongruent (standard and deviant sounds differed on the sound category) conditions, and tested a new sample of typically developing children. Finally, Chapter 4 describes the application of an ecologically valid auditory oddball paradigm to three language subgroups in ASD (minimally verbal autism, phrase speech autism, and verbally fluent autism) and a typically developing control group. This study measured the auditory and P3a ERP components elicited by natural speech and nonspeech deviant sounds, and investigated whether distinct ERP profiles indicative of underlying mechanistic differences could be found between language subgroups.

References

Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., … Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, *110*(29), 12060–12065.

Alcántara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107–1114. https://doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x

Alku, P., Tiitinen, H., & Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology*, *110*(8), 1329–1333.

Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., & Pisoni, D. B. (1987). *From text to speech: The MITalk system*. Cambridge University Press. Retrieved from http://dl.acm.org/citation.cfm?id=28587

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., … Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, *75*(4), 594.

Ardal, S., & Donald, M. W. (1991). Long-latency brain responses to incongruent sentences using synthetic speech. *Journal of Psychophysiology*, *5*(4), 315–326.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. *Joint Attention: Its Origins and Role in Development*, 131–158.

Baltaxe, C. A. M. (1977). Pragmatic Deficits in the Language of Autistic Adolescents. *Journal of Pediatric Psychology*, *2*(4), 176–180. https://doi.org/10.1093/jpepsy/2.4.176

Benasich, A. A., Choudhury, N., Friedman, J. T., Realpe-Bonilla, T., Chojnowska, C., & Gou, Z. (2006). The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. *Neuropsychologia*, *44*(3), 396–411. https://doi.org/10.1016/j.neuropsychologia.2005.06.004

Bishop, D. V. M., & Adams, C. (1989). Conversational characteristics of children with semantic-pragmatic disorder. II: What features lead to a judgement of inappropriacy? *International Journal of Language & Communication Disorders*, *24*(3), 241–263. https://doi.org/10.3109/13682828909019890

Boddaert, N., Chabane, N., Belin, P., Bourgeois, M., Royer, V., Barthelemy, C., … others. (2004). Perception of complex sounds in autism: abnormal auditory cortical processing in children. *American Journal of Psychiatry*, *161*(11), 2117–2120.

Bonnel, A.-M., Mottron, L., Peretz, I., Trudel, M., & Gallun, E. (2003). Enhanced pitch sensitivity in individuals with autism: a signal detection analysis. *Cognitive Neuroscience, Journal Of*, *15*(2), 226–235.

Bono, M. A., Daley, T., & Sigman, M. (2004). Relations among joint attention, amount of intervention and language gain in autism. *Journal of Autism and Developmental Disorders*, *34*(5), 495–505.

Boucher, J. (2001). Time-Parsing and Autism. *Time and Memory: Issues in Philosophy and Psychology*, (1), 111.

Boucher, J. (2012). Research Review: Structural language in autistic spectrum disorder – characteristics and causes. *Journal of Child Psychology and Psychiatry*, *53*(3), 219–233. https://doi.org/10.1111/j.1469-7610.2011.02508.x

Boucher, J., Mayes, A., & Bigham, S. (2008). Memory, language and intellectual ability in low-functioning autism. Retrieved from http://eprints.bournemouth.ac.uk/9558/1/9780521862882c14_p268-290.pdf

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Bruneau, N., & Gomot, M. (1998). Auditory evoked potentials (N1 wave) as indices of cortical development. In *Neuroimaging in child neuropsychiatric disorders* (pp. 113–123). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-95848-9_10

Brynskov, C., Eigsti, I.-M., Jørgensen, M., Lemcke, S., Bohn, O.-S., & Krøjgaard, P. (2017). Syntax and Morphology in Danish-Speaking Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(2), 373–383. https://doi.org/10.1007/s10803-016-2962-7

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism*, *2*(4), 325–344.

Carpenter, A. L., & Shahin, A. J. (2013). Development of the N1-P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neuroscience Letters*, *544*, 56–61. https://doi.org/10.1016/j.neulet.2013.03.041

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech–sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proceedings of the National Academy of Sciences*, *100*(9), 5567–5572.

Ceponiene, R., Yaguchi, K., Shestakova, A., Alku, P., Suominen, K., & Näätänen, R. (2002). Sound complexity and “speechness” effects on pre-attentive auditory discrimination in children. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *43*(3), 199–211.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1430), 315–324.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239.

Choudhury, N., & Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3 and 4 year language and cognitive abilities. *Clinical Neurophysiology*, *122*(2), 320–338. https://doi.org/10.1016/j.clinph.2010.05.035

Courchesne, E., Lincoln, A. J., Kilman, B. A., & Galambos, R. (1985). Event-related brain potential correlates of the processing of novel visual and auditory information in autism. *Journal of Autism and Developmental Disorders*, *15*(1), 55–76. https://doi.org/10.1007/BF01837899

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744.

Dawson, G., Finley, C., Phillips, S., & Galpert, L. (1986). Hemispheric specialization and the language abilities of autistic children. *Child Development*, 1440–1453.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*(2), 271.

De Fossé, L., Hodge, S. M., Makris, N., Kennedy, D. N., Caviness, V. S., McGrath, L., … others. (2004). Language-association cortex asymmetry in autism and specific language impairment. *Annals of Neurology*, *56*(6), 757–766.

Depue, R. A., & Morrone-Strupinsky, J. V. (2005). A neurobehavioral model of affiliative bonding: Implications for conceptualizing a human trait of affiliation. *Behavioral and Brain Sciences*, *28*(3), 313–349.

Digeser, F. M., Wohlberedt, T., & Hoppe, U. (2009). Contribution of spectrotemporal features on auditory event-related potentials elicited by consonant-vowel syllables. *Ear and Hearing*, *30*(6), 704–712.

Drager, K., & Reichle, J. (2001). Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication*, *17*(2), 109–119.

Duffy, S. A., & Pisoni, D. B. (1992). Comprehension of synthetic speech produced by rule: A review and theoretical interpretation. *Language and Speech*, *35*(4), 351–389.

Eyler, L. T., Pierce, K., & Courchesne, E. (2012). A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain*, *135*(3), 949–960.

Fan, Y.-T., & Cheng, Y. (2014). Atypical Mismatch Negativity in Response to Emotional Voices in People with Autism Spectrum Conditions. *PLOS ONE*, *9*(7), e102471. https://doi.org/10.1371/journal.pone.0102471

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*(9), 1671–1680.

Flagg, E. J., Cardy, J. E. O., Roberts, W., & Roberts, T. P. L. (2005). Language lateralization development in children with autism: Insights from the late field magnetoencephalogram. *Neuroscience Letters*, *386*(2), 82–87. https://doi.org/10.1016/j.neulet.2005.05.037

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373.

Fujikawa-Brooks, S., Isenberg, A. L., Osann, K., Spence, M. A., & Gage, N. M. (2010). The effect of rate stress on the auditory brainstem response in autism: a preliminary report. *International Journal of Audiology*, *49*(2), 129–140.

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., … Zilbovicius, M. (2004). Abnormal cortical voice processing in autism. *Nature Neuroscience*, *7*(8), 801–802. https://doi.org/10.1038/nn1291

Getzmann, S., Falkenstein, M., & Wascher, E. (2015). ERP correlates of auditory goal-directed behavior of younger and older adults in a dynamic speech perception task. *Behavioural Brain Research*, *278*, 435–445. https://doi.org/10.1016/j.bbr.2014.10.026

Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., & Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clinical Neurophysiology*, *112*(10), 1850–1859. https://doi.org/10.1016/S1388-2457(01)00636-8

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing Complex Emotion Recognition in Adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *36*(2), 169–183. https://doi.org/10.1007/s10803-005-0057-y

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002a). Hypersensitivity to acoustic change in children with autism: electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(05), 577–584.

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002b). Hypersensitivity to acoustic change in children with autism: electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(05), 577–584.

Groen, W. B., van Orsouw, L., Ter Huurne, N., Swinkels, S., van der Gaag, R.-J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact spectral but abnormal temporal processing of auditory stimuli in autism. *Journal of Autism and Developmental Disorders*, *39*(5), 742–750.

Groen, W. B., Zwiers, M. P., van der Gaag, R.-J., & Buitelaar, J. K. (2008). The phenotype and neural correlates of language in autism: an integrative review. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1416–1425.

Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill, K. (1997). Does teaching theory of mind have an effect on the ability to develop conversation in children with autism? *Journal of Autism and Developmental Disorders*, *27*(5), 519–537.

Haesen, B., Boets, B., & Wagemans, J. (2011). A review of behavioural and electrophysiological studies on auditory processing and speech perception in autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(2), 701–714.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, *38*(11), 1518–1530. https://doi.org/10.1016/S0028-3932(00)00052-X

Hale, C. M., & Tager-Flusberg, H. (2005). Brief Report: The Relationship between Discourse Deficits and Autism Symptomatology. *Journal of Autism and Developmental Disorders*, *35*(4), 519–524. https://doi.org/10.1007/s10803-005-5065-4

Hamilton, L. S., & Huth, A. G. (2018). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *0*(0), 1–10. https://doi.org/10.1080/23273798.2018.1499946

Hawkins, S., & Slater, A. (1994). Spread of CV and v-to-v coarticulation in british English: implications for the intelligibility of synthetic speech. In *ICSLP* (Vol. 94, pp. 57–60). Retrieved from http://www.ling.cam.ac.uk/sarah/docs/hawkins_slater_icslp94.pdf

Heaton, P. (2005). Interval and contour processing in autism. *Journal of Autism and Developmental Disorders*, *35*(6), 787–793.

Heaton, P., Hermelin, B., & Pring, L. (1998). Autism and pitch processing: A precursor for savant musical ability? *Music Perception*, 291–305.

Herning, R. I., Jones, R. T., & Hunt, J. S. (1987). Speech event related potentials reflect linguistic content and processing level. *Brain and Language*, *30*(1), 116–129.

Jansson-Verkasalo, E., Ceponiene, R., Kielinen, M., Suominen, K., Jäntti, V., Linna, S.-L., … Näätänen, R. (2003). Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. *Neuroscience Letters*, *338*(3), 197–200. https://doi.org/10.1016/S0304-3940(02)01405-2

Järvinen-Pasley, A., & Heaton, P. (2007). Evidence for reduced domain-specificity in auditory processing in autism. *Developmental Science*, *10*(6), 786–793. https://doi.org/10.1111/j.1467-7687.2007.00637.x

Joseph, R. M., & Tager–Flusberg, H. (2004). The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology*, *16*(1), 137–155. https://doi.org/10.1017/S095457940404444X

Källstrand, J., Olsson, O., Nehlstedt, S. F., Sköld, M. L., & Nielzén, S. (2010). Abnormal auditory forward masking pattern in the brainstem response of individuals with Asperger syndrome. *Neuropsychiatric Disease and Treatment*, *6*, 289.

Kangas, K. A., & Allen, G. D. (1990). Intelligibility of Synthetic Speech for Normal-Hearing and Hearing-Impaired Listeners. *Journal of Speech and Hearing Disorders*, *55*(4), 751–755. https://doi.org/10.1044/jshd.5504.751

Kanner, L., & others. (1943). Autistic disturbances of affective contact. Retrieved from http://neurodiversity.com/library_kanner_1943.pdf

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*, *76*(1), 125.

Kemner, C., Verbaten, M. N., Cuperus, J. M., Camfferman, G., & van Engeland, H. (1995). Auditory event-related brain potentials in autistic children and three different control groups. *Biological Psychiatry*, *38*(3), 150–165.

Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J.-L., … Collet, L. (2004). Increased perception of loudness in autism. *Hearing Research*, *198*(1), 87–92.

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and Cognitive Processes*, *16*(2–3), 287–308.

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders*, *31*(1), 29–36.

Klin, A. (1991). Young autistic children’s listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42.

Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytky, S., … Moilanen, I. (2007). Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of Autism and Developmental Disorders*, *37*(8), 1539–1549. https://doi.org/10.1007/s10803-006-0271-2

Koul, R. (2003). Synthetic speech perception in individuals with and without disabilities. *Augmentative and Alternative Communication*, *19*(1), 49–58.

Kuhl, P. K. (2007). Is speech learning ‘gated’by the social brain? *Developmental Science*, *10*(1), 110–120.

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental Science*, *8*(1), F1–F12.

Kujala, T., Kuuluvainen, S., Saalasti, S., Jansson-Verkasalo, E., Wendt, L. von, & Lepistö, T. (2010). Speech-feature discrimination in children with Asperger syndrome as determined with the multi-feature mismatch negativity paradigm. *Clinical Neurophysiology*, *121*(9), 1410–1419. https://doi.org/10.1016/j.clinph.2010.03.017

Kwakye, L. D., Foss-Feig, J. H., Cascio, C. J., Stone, W. L., & Wallace, M. T. (2010). Altered auditory and multisensory temporal processing in autism spectrum disorders. *Frontiers in Integrative Neuroscience*, *4*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/pmc3024004/

Kwon, S., Kim, J., Choe, B.-H., Ko, C., & Park, S. (2007). Electrophysiologic assessment of central auditory processing by auditory brainstem responses in children with autism spectrum disorders. *Journal of Korean Medical Science*, *22*(4), 656–659.

Leekam, S. R., López, B., & Moore, C. (2000). Attention and joint attention in preschool children with autism. *Developmental Psychology*, *36*(2), 261.

Lepistö, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Näätänen, R., & Kujala, T. (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical Neurophysiology*, *117*(10), 2161–2171.

Lepistö, Tuulia, Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005a). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1), 147–157.

Lepistö, Tuulia, Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005b). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1), 147–157.

Lepistö, Tuulia, Nieminen-von Wendt, T., von Wendt, L., Näätänen, R., & Kujala, T. (2007). Auditory cortical change detection in adults with Asperger syndrome. *Neuroscience Letters*, *414*(2), 136–140. https://doi.org/10.1016/j.neulet.2006.12.009

Lincoln, A. J., Courchesne, E., Harms, L., & Allen, M. (1993). Contextual probability evaluation in autistic, receptive developmental language disorder, and control children: event-related brain potential evidence. *Journal of Autism and Developmental Disorders*, *23*(1), 37–58.

Lindström, R., Lepistö-Paisley, T., Vanhala, R., Alén, R., & Kujala, T. (2016). Impaired neural discrimination of emotional speech prosody in children with autism spectrum disorder and language impairment. *Neuroscience Letters*. https://doi.org/10.1016/j.neulet.2016.06.016

Liu, T., Shi, J., Zhang, Q., Zhao, D., & Yang, J. (2007). Neural mechanisms of auditory sensory processing in children with high intelligence. *Neuroreport*, *18*(15), 1571–1575.

Logan, J. S., Greene, B. G., & Pisoni, D. B. (1989). Segmental intelligibility of synthetic speech produced by rule. *The Journal of the Acoustical Society of America*, *86*(2), 566–581.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). Autism diagnostic observation schedule, (ADOS-2) modules 1-4. *Los Angeles, California: Western Psychological Services*.

Lortie, M., Proulx-Bégin, L., Saint-Amour, D., Cousineau, D., Théoret, H., & Lepage, J.-F. (2017). Brief Report: Biological Sound Processing in Children with Autistic Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(6), 1904–1909. https://doi.org/10.1007/s10803-017-3093-5

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism and Developmental Disorders*, *16*(3), 335–349.

Loveland, K. A., McEvoy, R. E., Tunali, B., & Kelley, M. L. (1990). Narrative story telling in autism and Down's syndrome. *British Journal of Developmental Psychology*, *8*(1), 9–23.

Luce, P. A., Feustel, T. C., & Pisoni, D. B. (1983). Capacity Demands in Short-Term Memory for Synthetic and .Natural Speech. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *25*(1), 17–32. https://doi.org/10.1177/001872088302500102

Marco, E. J., Hinkley, L. B. N., Hill, S. S., & Nagarajan, S. S. (2011). Sensory processing in autism: a review of neurophysiologic findings. *Pediatric Research*, *69*(5 Pt 2), 48R.

Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, *9*(3), 302–315.

McArthur, D., & Adamson, L. B. (1996). Joint attention in preverbal children: Autism and developmental language disorder. *Journal of Autism and Developmental Disorders*, *26*(5), 481–496.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, *5*(3), 317–334.

Mizrahi, A., Shalev, A., & Nelken, I. (2014). Single neuron and population coding of natural sounds in auditory cortex. *Current Opinion in Neurobiology*, *24*, 103–110.

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development? *Infant Behavior and Development*, *8*(2), 197–211.

Molfese, D. L., & Molfese, V. J. (1997). Discrimination of language skills at five years of age using event-related potentials recorded at birth. *Developmental Neuropsychology*, *13*(2), 135–156.

Molfese, D. L., Molfese, V. J., Key, S., Modglin, A., Kelley, S., & Terrell, S. (2002). Reading and cognitive abilities: Longitudinal studies of brain and behavior changes in young children. *Annals of Dyslexia*, *52*(1), 99.

Morales, M., Mundy, P., Delgado, C. E., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6-through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298.

Mottron, L., Peretz, I., & Ménard, E. (2000). Local and Global Processing of Music in High-functioning Persons with Autism: Beyond Central Coherence? *Journal of Child Psychology and Psychiatry*, *41*(8), 1057–1065. https://doi.org/10.1111/1469-7610.00693

Mottron, Laurent, Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development*, *21*(3), 469–482.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, Risto, & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Nye, P. W., & Gaitenby, J. H. (1973). Consonant intelligibility in synthetic speech and in a natural speech control (modified rhyme test results). *Haskins Laboratories Status Report on Speech Research, SR*, *33*, 77–91.

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience & Biobehavioral Reviews*, *36*(2), 836–854.

O'Riordan, M., & Passetti, F. (2006). Discrimination in autism within different sensory modalities. *Journal of Autism and Developmental Disorders*, *36*(5), 665–675.

Paetau, R., Ahonen, A., Salonen, O., & Sams, M. (1995). Auditory evoked magnetic fields to tones and pseudowords in healthy children and adults. *Journal of Clinical Neurophysiology*, *12*(2), 177–185.

Paris, C. R., Gilson, R. D., Thomas, M. H., & Silver, N. C. (1995). Effect of synthetic voice intelligibility on speech comprehension. *Human Factors*, *37*(2), 335–340.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Pisoni, D. B. (1981). Speeded classification of natural and synthetic speech in a lexical decision task. *The Journal of the Acoustical Society of America*, *70*(S1), S98–S98. https://doi.org/10.1121/1.2019150

Plesa Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*(5), 591–604. https://doi.org/10.1177/1362361315600146

Polich, J. (2003). *Theoretical overview of P3a and P3b*. Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-0294-4_5

Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*(2), 172–185. https://doi.org/10.1016/j.ijpsycho.2005.12.012

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clinical Neurophysiology*, *111*(2), 220–236. https://doi.org/10.1016/S1388-2457(99)00236-9

Rapin, I., Dunn, M. A., Allen, D. A., Stevens, M. C., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology*, *34*(1), 66–84.

Redcay, E., & Courchesne, E. (2008). Deviant functional magnetic resonance imaging patterns of brain activity to speech in 2–3-year-old children with autism spectrum disorder. *Biological Psychiatry*, *64*(7), 589–598.

Rose, V., Trembath, D., Keen, D., & Paynter, J. (2016). The proportion of minimally verbal children with autism spectrum disorder in a community-based early intervention programme. *Journal of Intellectual Disability Research*, *60*(5), 464–477. https://doi.org/10.1111/jir.12284

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, *119*(8), 1720–1731.

Russo, N., Nicol, T., Trommer, B., Zecker, S., & Kraus, N. (2009). Brainstem transcription of speech is disrupted in children with autism spectrum disorders. *Developmental Science*, *12*(4), 557–567.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, *32*(3), 189–194.

Salmond, C. H., Vargha-Khadem, F., Gadian, D. G., de Haan, M., & Baldeweg, T. (2007). Heterogeneity in the Patterns of Neural Abnormality in Autistic Spectrum Disorders: Evidence from ERP and MRI. *Cortex*, *43*(6), 686–699. https://doi.org/10.1016/S0010-9452(08)70498-2

Sandel, L. (1998). The Relationship between Language and Intelligence. Review of Historical Research: Summary# 5. Retrieved from http://eric.ed.gov/?id=ED470647

Sandridge, S. A., & Boothroyd, A. (1996). Using naturally produced speech to elicit the mismatch negativity. *JOURNAL-AMERICAN ACADEMY OF AUDIOLOGY*, *7*, 105–112.

Sharma, A., Marsh, C. M., & Dorman, M. F. (2000). Relationship between N 1 evoked potential morphology and the perception of voicing. *The Journal of the Acoustical Society of America*, *108*(6), 3030–3035.

Shestakova, A., Huotilainen, M., Ceponiene, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *114*(8), 1507–1512.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and Prosody Characteristics of Adolescents and Adults With High-Functioning Autism and Asperger Syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, *34*(3), 691–697.

Syal, S., & Finlay, B. L. (2011). Thinking outside the cortex: social motivation in the evolution and development of language. *Developmental Science*, *14*(2), 417–430.

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders*, *11*(1), 45–56.

Tager-Flusberg, H. (1999). A psychological approach to understanding the social and language impairments in autism. *International Review of Psychiatry*, *11*(4), 325–334. https://doi.org/10.1080/09540269974203

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience Research*, *6*(3), 219–224.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics*, *16*(3), 241–256.

Takahashi, H., Nakahachi, T., Stickley, A., Ishitobi, M., & Kamio, Y. (2016). Relationship between physiological and parent-observed auditory over-responsiveness in children with typical development and those with autism spectrum disorders. *Autism: The International Journal of Research and Practice*, 1362361316680497. https://doi.org/10.1177/1362361316680497

Tanguay, P. E., Edwards, R. M., Buchwald, J., Schwafel, J., & Allen, V. (1982). Auditory Brainstem Evoked Responses in Autistic Children. *Archives of General Psychiatry*, *39*(2), 174–180. https://doi.org/10.1001/archpsyc.1982.04290020040008

Taylor, M. J., Rosenblatt, B., & Linschoten, L. (1982). Auditory Brainstem Response Abnormalities in Autistic Children. *Canadian Journal of Neurological Sciences*, *9*(4), 429–433. https://doi.org/10.1017/S0317167100044346

Tharpe, A. M., Bess, F. H., Sladen, D. P., Schissel, H., Couch, S., & Schery, T. (2006). Auditory characteristics of children with autism. *Ear and Hearing*, *27*(4), 430–441.

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, *15*(6), 355–366. https://doi.org/10.1038/nrn3731

Thurm, A., Manwaring, S. S., Swineford, L., & Farmer, C. (2015). Longitudinal study of symptom severity and language in minimally verbal children with autism. *Journal of Child Psychology and Psychiatry*, *56*(1), 97–104.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, 1454–1463.

Tremblay, K. L., Friesen, L., Martin, B. A., & Wright, R. (2003). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear and Hearing*, *24*(3), 225–232.

Tremblay, Kelly L., & Kraus, N. (2002). Auditory Training Induces Asymmetrical Changes in Cortical Neural Activity. *Journal of Speech, Language, and Hearing Research*, *45*(3), 564–572. https://doi.org/10.1044/1092-4388(2002/045)

Volden, J., & Lord, C. (1991). Neologisms and idiosyncratic language in autistic speakers. *Journal of Autism and Developmental Disorders*, *21*(2), 109–130. https://doi.org/10.1007/BF02284755

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

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

Wang, X.-D., Gu, F., He, K., Chen, L.-H., & Chen, L. (2012). Preattentive extraction of abstract auditory rules in speech sound stream: a mismatch negativity study using lexical tones. *PLoS One*, *7*(1), e30027.

Whitehouse, A. J., & Bishop, D. V. (2008). Do children with autism 'switch off'to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524.

Wilkinson, K. M. (1998). Profiles of language and communication skills in autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *4*(2), 73–79.

Winkler, I., Kujala, T., Alku, P., & Näätänen, R. (2003). Language context and phonetic change detection. *Brain Research. Cognitive Brain Research*, *17*(3), 833–844.

CHAPTER 2

Auditory Event-Related Potentials Index Faster Processing of Natural Speech but Not Synthetic Speech Over Nonspeech Analogs in Children

Introduction

The neural processing of language depends on an initial filtering of the complex acoustic environment in order to extract only the speech signal. This process may be enhanced by a biological preference to orient and attend to human speech over other types of sounds, similar to the way other animals display an early bias to the vocalizations of their own species (Braaten & Reynolds, 1999; Heaton, Miller, & Goodwin, 1978; Marler, 1990; Penna & Meier, 2011). In humans, behavioral findings in infants from sequential looking preference and high-amplitude sucking tasks demonstrate a preference for speech over human nonspeech vocalizations and animal vocalizations that emerges within the first three months of life (Shultz & Vouloumanos, 2010; Vouloumanos, Hauser, Werker, & Martin, 2010). This bias to attend to speech in infants has been shown to predict later vocabulary development (Vouloumanos & Curtin, 2014) and is presumed to contribute to the development of language learning networks in the brain (Kuhl, 2007). However, neural evidence that speech signals are prioritized and processed differently than other nonspeech sounds remains inconsistent.

Previous fMRI studies in adults have found that the upper bank of the bilateral superior temporal sulci (STS) preferentially responds to human speech and nonspeech vocalizations over other environmental sounds (Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Binder et al., 2000; Fecteau, Armony, Joanette, & Belin, 2004) and exhibits stronger activation to human speech than nonspeech vocalizations (Belin et al., 2002). However, more recent studies have found that these same regions show increased activation to artificial nonspeech sounds after training (Leech, Holt, Devlin, & Dick, 2009; Liebenthal et al., 2010). These studies suggest that the bilateral STS activity may reflect auditory expertise with complex sounds rather than being indicative of a unique specialization to the human voice or speech. Other studies using electrocortigraphy (ECoG) have found areas of the superior temporal gyrus specialized for speech (Chan et al., 2013; Mesgarani, Cheung, Johnson, & Chang, 2014), but the degree to which these areas overlap with nonspeech acoustical processing is heavily debated (Peretz, Vuvan, Lagrois, & Armony, 2015), though newer analysis methods may help to distinguish specific regions (Norman-Haignere, Kanwisher, & McDermott, 2015).

The temporal resolution of EEG offers an alternative approach to investigate whether human speech may be processed differently from other sounds in the environment. By comparing the timing of neural responses between speech and nonspeech sounds, these studies can measure whether speech processing is supported by faster, more efficient mechanisms or greater orienting responses to speech. Faster processing of speech would suggest that the signal is prioritized and/or identified earlier in cortical processing, similar to the way that children with musical training display faster auditory processing of a musical sound relative to non-trained children (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006). Faster processing of speech over other types of natural sounds in the environment could confer several advantages for listeners, e.g., faster identification of human voices, faster comprehension of potentially vital communicative messages, and perhaps even a greater ability to segregate the speech stream from other competing sounds (i.e., the cocktail party effect). Similarly, greater orienting responses to speech could be reflective of a biological preference to attend to speech over other types of sounds.

The early sensory processing of auditory stimuli in EEG studies is typically indexed by the N1 and P2 components in adults, and the P1 and N2 components in children. The difference in waveform morphology between the child and adult auditory ERPs is thought to originate from maturational changes in the synaptic pruning of the primary auditory cortex as well as increased connectivity between auditory and frontal lobe areas that is not complete until the age of 12 (Eggermont & Moore, 2012). In children, the P1 and N2 are frontocentral components that occur within 200 ms of the auditory stimulus and are thought to represent encoding of acoustic properties in both primary and secondary auditory cortices (Ponton, Eggermont, Kwong, & Don, 2000). However, when interstimulus intervals longer than 1 second are used, the adult pattern of N1 and P2 auditory ERPs can be observed in young children, presumably because slower presentation rates are less taxing on the immature auditory system (Čeponien, Rinne, & Näätänen, 2002; Paetau, Ahonen, Salonen, & Sams, 1995). At the level of individual participants, these cortical auditory ERPs are seen as robust measures of early auditory processing as they have been found to be highly reliable in both children and adults (Kileny & Kripal, 1987; Räikkönen, Birkás, Horváth, Gervai, & Winkler, 2003; Segalowitz & Barnes, 1993; Tremblay, Friesen, Martin, & Wright, 2003), and predictive of later language performance (Benasich et al., 2006; Choudhury & Benasich, 2011; Molfese & Molfese, 1985, 1997).

While the early auditory ERP components represent early sensory encoding, the P3a component is thought to be an "orienting response" related to involuntary attention switching or initial allocation of attention to a salient stimulus (Friedman, Cycowicz, & Gaeta, 2001; Squires, Squires, & Hillyard, 1975; Polich, 2003; Polich, 2007). The latency of the P3a occurs between approximately 300—400 ms after stimulus onset and has been found to be generated by a fronto-parietal network with involvement of the anterior cingulate cortex (Bledowski, Prvulovic, Goebel, Zanella, & Linden, 2004; Volpe et al., 2007; Wronka, Kaiser, & Coenen, 2012). This ERP component thus provides a marker for whether a new sound is judged to be salient enough to activate frontal lobe mechanisms and engage further cognitive processing.

The majority of previous auditory ERP studies have compared the relative amplitudes of the speech and nonspeech responses. However, a stronger amplitude to one stimulus cannot be used to assess whether speech is processed more efficiently or prioritized in any way because any differences in the amplitude during early sensory processing are likely directly attributed to differences in the acoustic features of the sounds (Woodman, 2010). The few studies that have measured auditory ERP latency differences between speech and nonspeech stimuli in children and adults demonstrate inconsistent findings, showing either faster processing of nonspeech complex tones (Bruder et al., 2011; Ceponiené et al., 2001), or no latency differences between vowels/syllables and nonspeech tones (Eulitz, Diesch, Pantev, Hampson, & Elbert, 1995; Swink & Stuart, 2012; Woods & Elmasian, 1986). Studies comparing the P3a ERP component between speech and nonspeech have found either a larger P3a amplitude for speech compared to nonspeech analogs (Jaramillo et al., 2001; Sorokin, Alku, & Kujala, 2010), or no difference in P3a amplitude between stimuli (Reiche et al., 2013). Only Sorokin et al. (2010) measured P3a latency and found no significant differences between stimuli.

The apparent lack of auditory ERP and P3a ERP evidence for more efficient processing and greater orienting responses to speech could be due to the type of stimuli used. ERP studies comparing speech to nonspeech have relied on synthetic speech because of the difficulties in matching a nonspeech stimulus to the inherent spectro-temporal complexity found in natural speech. Matching the acoustic complexity to a nonspeech sound is critical in EEG/ERP research because ERP components arise from the synchronized activity of populations of neurons. Without matching the stimuli one cannot rule out that any differences are due to larger populations of neurons responding to the more complex stimulus. The simplified nature of a synthetic speech stimulus allows researchers to easily match it on spectro-temporal complexity to a synthetic nonspeech analog. However, use of synthetic speech stimuli and their complexity-matched tones introduces issues of ecologic validity. Synthetic speech is commonly programmed to include only the most rudimentary aspects of the speech signal and thus sounds robotic, and complex tones are not found naturally in the environment. Recently, researchers have called for the use of more naturalistic stimuli in neuroscience experiments of language and auditory processing (Hamilton & Huth, 2018; Theunissen & Elie, 2014). This comes in light of new findings at the level of single neurons whereby responses to simple artificial stimuli do not predict responses to complex natural sounds (Mizrahi, Shalev, & Nelken, 2014). Similarly, ERP responses to simplified synthetic stimuli may not map onto responses from natural complex stimuli. Thus, the results of previous auditory ERP studies that have used synthetic speech may not accurately represent neural processing of natural speech.

The primary challenge in matching human speech stimuli to other nonspeech stimuli is the inherently high spectro-temporal complexity of human speech characterized by peaks of acoustic energy vary significantly in both frequency and amplitude on the order of milliseconds.

Even the most basic speech sound, the vowel, has high spectro-temporal complexity that arises from the quasi-periodic vocal fold vibration as well as the aperiodic, turbulent air that typically escapes with each cycle of vocal fold vibration. This results in a signal in which periods of semi-regular frequency bands (harmonics) occur simultaneously with irregular broadband noise. This precludes matching to many different complex natural sounds that are characterized by only one of these features. For example, musical instruments can produce similar harmonic structures with an initial noise burst but the noise is not sustained across the signal, while nature sounds like wind and rain contain only the broadband noise without harmonic structure. Some research groups have opted to match natural speech to artificial stimuli created by using spectrally-rotated speech or reversed speech. Although this approach is an improvement over the use of purely synthetic stimuli, it contrasts one naturalistic to one synthetic stimulus and so does not overcome concerns with ecologic validity. Other approaches have focused on increasing the "naturalness" of synthetic speech by starting with a natural utterance and manipulating it to create "semi-synthetic speech" (Alku, Tiitinen, & Naatanen, 1999). While these stimuli do sound significantly more natural and allow researchers to have full control over the acoustic parameters of the voice, they inherently increase the complexity of the stimuli. Many studies using these semi-synthetic speech stimuli continue to use complex tones as their matched nonspeech counterpart, and thus their stimuli are no longer matched on complexity.

An alternative solution to match the spectro-temporal complexity of human speech is to use natural nonspeech stimuli that are generated with the same underlying mechanism as human speech. The vocalizations, or "calls" of certain mammals are produced in a human-like fashion; that is, laryngeal vocal fold vibration and filtering via a "vocal tract." Due to the identical sound production mechanisms, the larynges of these mammals (e.g., canines, rabbits, pigs, cows) have been used as animal models for the study of sound production in humans (Alipour & Jaiswal, 2008; Ge, French, Ohno, Zealear, & Rousseau, 2009; Jiang, Raviv, & Hanson, 2001).

The purpose of the present study is to compare the auditory ERPs and P3a ERP component to speech and nonspeech to assess whether speech is processed differently from nonspeech at the level of early auditory processing and involuntary orienting responses in children ages 8—11-years-old, and if this differs for natural vs. synthetic speech stimuli. We created a novel task with ecologically valid natural speech vowels matched to cow vocalizations and used a second task with commonly used synthetic stimuli. The 8-11-year-old age range was selected to obtain a sample of children with similar morphology of their auditory ERPs (Bishop, Hardiman, Uwer, & von Suchodoletz, 2007), and furthermore to be consistent with previous studies using synthetic speech and tone stimuli (Bruder et al., 2011; Ceponiené et al., 2001). We used two separate passive oddball tasks with slow presentation rates for the natural and synthetic stimuli to elicit N1 and P2 auditory ERPs to deviant (infrequent) sounds. The use of passive auditory oddball tasks (i.e., no overt response required from participants) was crucial to our study design in order to test how the brain responds to a novel speech vs. nonspeech sound naturally when no demands are in place. We hypothesized that more efficient processing of speech would be evidenced by faster latencies of the N1, P2, and P3a ERPs to a deviant natural speech vowel sound compared to a deviant cow vocalization, but that consistent with previous studies in children this latency difference would not be present in response to a deviant synthetic speech vowel compared to a deviant complex nonspeech tone (Bruder et al., 2011; Ceponiené et al., 2001). Furthermore, we hypothesized that natural speech sounds would also elicit a greater orienting response reflected by larger P3a amplitudes. These results would support a neural mechanism for the prioritization of speech that begins at the earliest level of cortical auditory processing, and is sustained across later cognitive processing relating to attention.

Methods

Participants

Twenty-two children (15 males) between 8 and 11-years-old (M = 9.5 years, SD = 1.1 years) participated in the study. All children had an IQ within or above the average range (M = 120.9, SD = 8.6, Range = 97 – 134) according to the Kaufman Brief Intelligence Test 2 (KBIT-2; (Kaufman & Kaufman, 2004). At the time of the visit, all children had reported normal hearing and no medical or educational diagnoses. Children were recruited via the Vanderbilt University email distribution service.

Participant Subgroups

Given that children have been found to show mature adult-like N1-P2 morphology of auditory ERP responses when interstimulus intervals over 1 s are used (Ceponiene, Rinne, & Naatanen, 2002; Ceponiene, Cheour, & Näätänen, 1998; Gilley, Sharma, Dorman, & Martin, 2005) as in the present study, we expected all children to show N1-P2 auditory responses to all stimuli. However, during inspection of ERPs from individual participants, it was discovered that although all 22 children displayed N1-P2 responses to the stimuli in the Synthetic Speech Task, children differed in their early sensory ERP components to the stimuli in the Natural Speech Task. While 15 children displayed the expected mature N1-P2 auditory responses, 7 children did not show an N1 response to the natural stimuli. As a result, for the Natural Speech Task only, we sub-grouped these children into the “Mature NST” subgroup (n=15) and “Immature NST Subgroup” (n=7). The difference in waveform morphology between participants can be seen to have an impact on the grand average waveform in Figure 2.4 (discussed later), especially to the natural speech vowel, as averaging together all 22 participants results in what appears to be the absence of any activity until around 100 ms. Children in both the Immature NST and Mature NST subgroups showed similar P3a responses within the same time range, so the full sample was analyzed for the P3a ERP.

Stimuli & Task Design

Stimuli were presented in two passive oddball tasks (Synthetic Speech Task and Natural Speech Task) in which infrequent "deviant" sounds were interspersed within a stream of frequent "standard" sounds. This design allowed for the interpretation of event-related potentials elicited to the rare deviant sound as a measure of how fast the brain processes a new sound in the environment. The order of presenting the Synthetic Speech Task and Natural Speech Task was counterbalanced across participants. To ensure novelty of the deviant sound in both oddball tasks, at least three standards always preceded a deviant sound. Within one trial, six sounds were presented, and the deviant sound occurred randomly in either the 4th, 5th, or 6th position. Across both tasks, sounds were presented with a varying interstimulus interval (ISI) between 1400-2600 ms to prevent habituation to stimulus onset. In addition, 1 s of silence was presented between each trial sequence such that after every six sounds the ISI varied between 2400 – 3600 ms. See Figure 2.1 for the schematic illustration of the first blocks in the Synthetic Speech Task and Natural Speech Task. For both tasks, participants were seated 110 cm from a TV monitor displaying a silent cartoon while auditory stimuli were presented at approximately 70 dB from two loudspeakers located on each side of the monitor. Participants were instructed to watch the cartoon and to ignore the sounds.

All stimuli were edited to have a rise and fall time of 10 ms using a custom-written MATLAB script (The Mathworks, Inc). Further, stimuli were downsampled to a sampling rate of

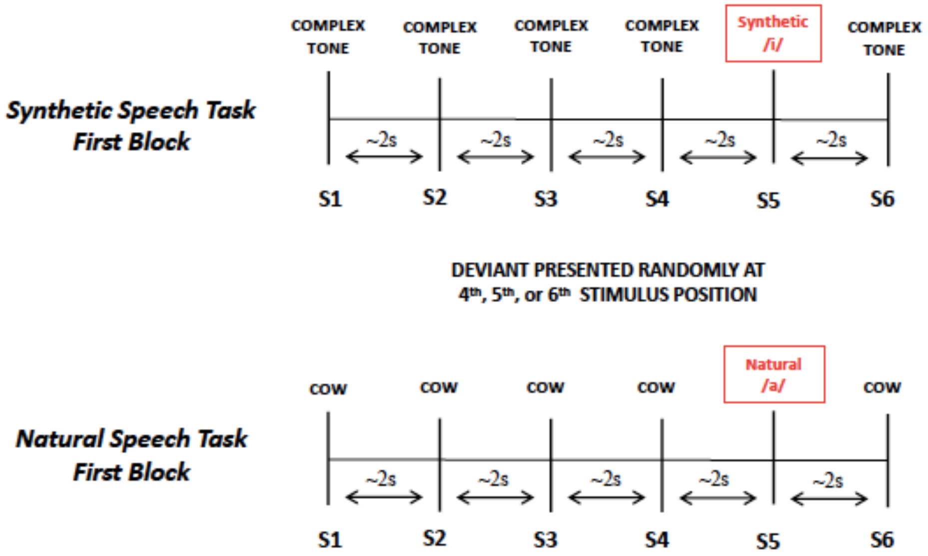


Figure 2.1. Schematic Illustration of Oddball Design. Top: First block of the Synthetic Speech Task, in which a complex tone was presented as the frequent standard sound and the synthetic vowel /i/ was presented as the rare deviant sound. Bottom: First block of the Natural Speech Task, in which a cow sound was presented as the frequent standard sound and the natural vowel /a/ was presented as the rare deviant sound.

44.1 kHz and had a quantization level of 16-bit and were matched on loudness based on their average root mean square (RMS) using Adobe Audition (Adobe Systems, Inc). Table 2.1 presents the acoustic attributes that were used to match stimuli in the Synthetic Speech Task and Natural Speech Task. Figure 2.2 provides the oscillograms, spectrograms, and power spectrum plots of the four deviant stimuli used across tasks.

Synthetic Speech Task

Synthetic speech vowels and complex tones were each presented as standard and deviant sounds in two blocks. As in previous studies (Bruder et al., 2011; Ceponiené et al., 2001; Eulitz et al., 1995; Swink & Stuart, 2012; Woods & Elmasian, 1986), the synthetic speech vowels and tones were matched on overall spectro-temporal complexity by matching on duration (200 ms each), average RMS level (intensity), spectral complexity (four main frequencies/harmonics in each sound), and temporal complexity (rise/fall time, steady-state vowel and tone). The synthetic vowels were created using the Klatt Synthesizer (Klatt, 1980) and the complex tones were created in MATLAB. In the first block, the standard sound was a complex tone composed of four sinusoidal tones (500 Hz, 1000 Hz, 1500 Hz, 2000 Hz) and the deviant sound was a synthetic vowel /i/ generated according to frequency patterns used in previous studies (150 Hz, 400 Hz, 1920 Hz, and 2560 Hz). In the second block, the standard sound was a synthetic speech vowel /a/ generated with frequencies at 150 Hz, 710 Hz, 1100 Hz, and 2640 Hz, and the deviant sound was a complex tone composed of four sinusoidal tones (800 Hz, 1600 Hz, 2400 Hz, 3200 Hz). These stimuli were modeled from a previous study in children (Whitehouse & Bishop, 2008).

Table 2.1 *Acoustic Attributes Used to Match Stimuli in the Synthetic Speech Task and Natural Speech Task.*

Synthetic Speech Task

Block	Stimulus	Harmonics*	Mean Intensity*	Duration*	Jitter	Shimmer	Spectral Entropy
Block 1	Standard: Complex Tone 1	500 Hz, 1000 Hz, 1500 Hz, 2000 Hz	71.26 dB	200 ms	0.00%	0.01%	0.33
	Deviant: Synthetic Vowel /i/	150 Hz, 400 Hz, 1920 Hz , 2560 Hz	71.49 dB	200	0.01%	0.10%	0.51
Block 2	Standard: Synthetic Vowel /a/	150 Hz, 710 Hz, 1100 Hz, 2640 Hz	71.79 dB	200 ms	0.04%	0.21%	0.50
	Deviant: Complex Tone 2	800 Hz, 1600 Hz, 2400 Hz, 3200 Hz	71.26 dB	200 ms	0.00%	0.01%	0.34

Natural Speech Task

Block	Stimulus	F0 Range*	Intensity Range*	Duration*	Jitter	Shimmer	Spectral Entropy
Block 1	Standard: Cow Sound 1	34.17 Hz	11.60 dB	350 ms	0.26%	1.64%	0.75
	Deviant: Natural Vowel /a/	35.75 Hz	11.24 dB	350 ms	0.27%	1.12%	0.69
Block 2	Standard: Natural Vowel /i/	33.63 Hz	10.80 dB	350 ms	0.14%	0.82%	0.77
	Deviant: Cow Sound 2	25.45 Hz	9.86 dB	350 ms	0.12%	1.19%	0.79

Note: The categories marked with a * signify the acoustic attributes that were used to match the stimuli used in each task. Jitter, shimmer, and spectral entropy values are shown to provide measures of the cycle to cycle changes in frequency and amplitude, respectively, and to illustrate differences in the complexity of the synthetic vs. natural stimuli across tasks.

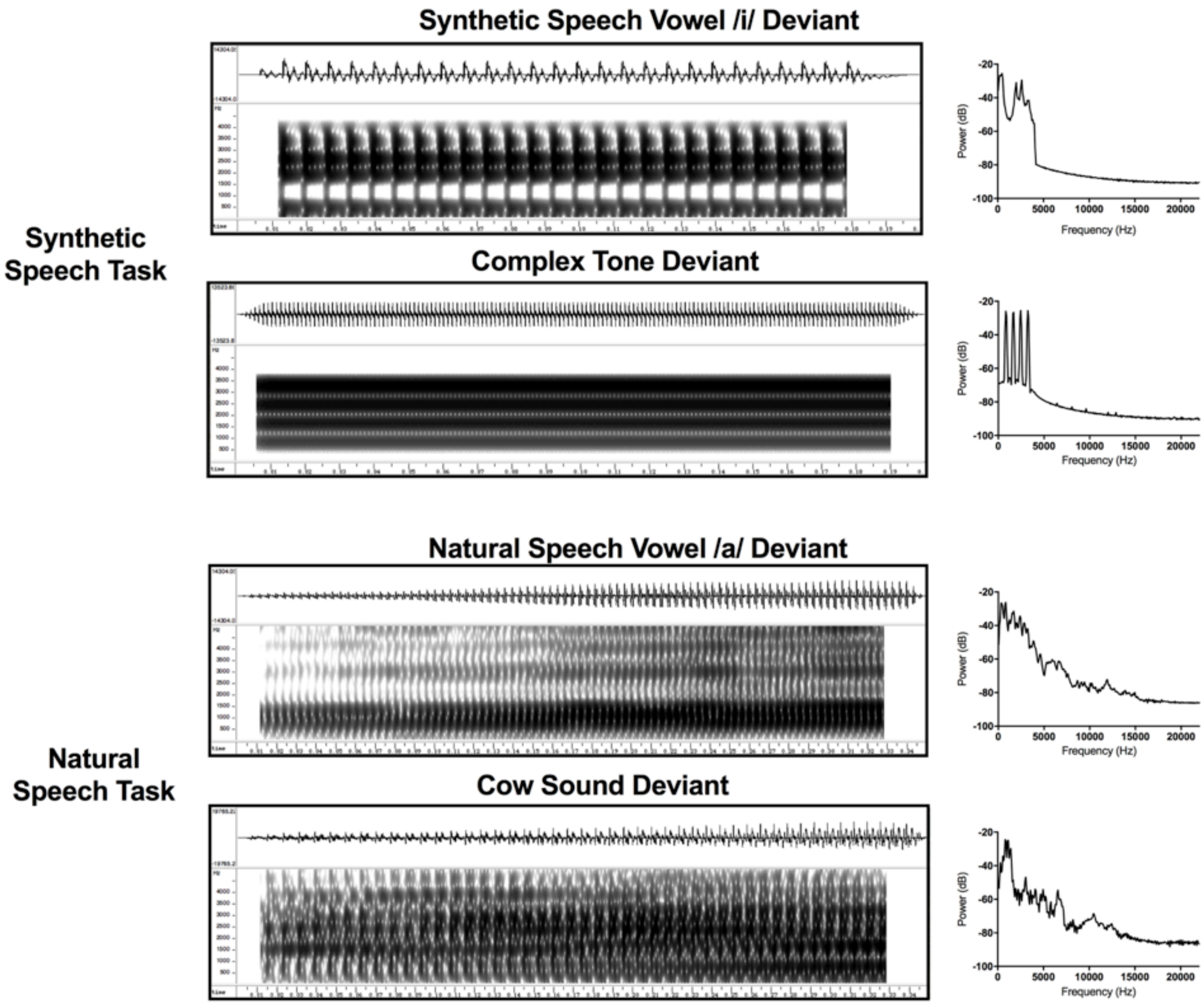


Figure 2.2 Oscillograms, Spectrograms, and Power Spectral Density Plots. The oscillogram, spectrogram, and power spectral density plot are shown side by side for each deviant sound in the Synthetic Speech Task (top) and Natural Speech Task (bottom).

Natural Speech Task

Paralleling the matching process of synthetic stimuli, the goal was to match a natural speech vowel to a natural nonspeech stimulus on overall spectro-temporal complexity and not individual acoustic parameters (e.g., same fundamental frequency or harmonics), as this would lead to stimuli that sounded too similar, especially in the context of a passive task. Animal vocalizations offer an ideal match because of similar mechanisms involved in producing the signals. Given that vocalizations produced by humans and animals are spectrally rich signals with an infinite number of frequencies, these stimuli were not matched in the same way that the synthetic speech and complex tones were matched on the number of frequencies/harmonics in the signal. Instead, we focused our efforts on matching two dynamic spectro-temporal aspects, namely the changes in amplitude and fundamental frequency across the entire stimulus duration.

We chose to use cow vocalizations as our natural nonspeech stimulus for the following reasons. First, the physical production mechanisms of cow vocalizations are virtually identical to those of humans, and the vocal folds of the cow have been used as a model for human vocal folds (Alipour & Jaiswal, 2008). Therefore, the spectral properties of human and cow sounds are comparable with regards to spectral complexity. Second, due to the difference in larynx and vocal tract shape and size between humans and cows, spectral patterns (i.e., fundamental frequency, formant frequencies) sound sufficiently different and give each natural stimulus their unique identity. This results in stimuli that are similar on acoustic attributes but are not confused with one another, such that when we increased the stimulus duration to 350 ms both were easily recognized as a human and a cow rather than ambiguous sounds (piloting these stimuli revealed that 200 ms was too short for the cow sound to be recognized). Third, cow and human vocalizations can comprise a similar frequency range; thus, both are highly relevant for human auditory processing.

Finally, we chose an animal sound that would be familiar to children but not potentially frightening or irritating to listen to (lions roaring, monkeys screeching), or those that may be associated with attachments/emotions that could differ across children (cats vs. dogs).

Matching was achieved by selecting two 350 ms segments from two different cow vocalizations in which the magnitude of change in fundamental frequency (pitch) fell within the range as that of the two human speech stimuli (+/- 10 Hz). Similarly, vocal intensity change across stimulus duration was matched (+/- 1 dB). To verify that the pattern of change in fundamental frequency and intensity were comparable across the matched stimuli dyads, we ran Pearson correlations using the time histories of the extracted acoustic measures (using Praat v 6.0.20, sampling rate 1000 Hz). We found significant correlations with coefficients ranging from 0.937 to 0.995. This matching process allowed us to keep the stimuli natural without any editing, and thus preserved the highest level of ecological validity.

Natural speech vowels and cow sounds were each presented as standard and deviant sounds in two blocks. The natural vowel stimuli were produced with slightly rising pitch and intensity by a female speaker. The stimuli were recorded with a Tascam DR-100 MKII voice recorder in a quiet room. The two natural cow vocalizations were purchased from the Sound-Ideas General HD sound effects collection and were recorded with a Sanken CMs7 microphone. In the first block, the standard sound was a cow sound and the deviant sound was a natural vowel (/a/). In the second block, the standard sound was a natural vowel (/i/) and the deviant sound was a different cow sound.

Praat software (v 6.0.20) was used to analyze acoustic features of the natural stimuli during the matching process (see Table 2.1). To further quantify the spectral complexity of the natural stimuli, we calculated spectral entropy for each sound using the Seewave R package (Sueur et al., 2008). Spectral entropy values closer to 0 indicate simple, spectrally highly predictable stimuli, while values closer to 1 indicate greater spectral complexity (e.g., pure tones have a value around 0). This analysis revealed similar values for spectral entropy: for both speech and nonspeech, spectral entropy was high, with a mean ± SD spectral entropy of .70 ± .069 for speech, and a mean entropy of .79 ± .076 for nonspeech.

EEG Procedures

Neural responses were recorded by Net Station 5.3 software (EGI, Inc., Eugene, OR, USA) and stimulus presentation was controlled by E-Prime 2.0 (PST, Inc., Pittsburgh, PA, USA). Participants wore a hydrocel geodesic EEG net with 128 electrodes (EGI, Inc., Eugene, OR, USA). Electrode impedances were kept below 50 kΩ. The EEG was sampled at 500 Hz and filtered offline using a 0.1 – 30 Hz bandpass filter. During data collection the signal was referenced to Cz, and later re-referenced to the average of all electrodes following artifact rejection.

Data Analysis

ERP epochs were segmented from the continuous EEG data beginning 100 ms before (pre-stimulus baseline) and 800 ms after stimulus onset to the four deviant sounds and all standard sounds that preceded a deviant. The resulting epochs were screened for artifacts using an automated Net Station algorithm followed by manual review. Trials containing eye blinks, lateral eye movements, muscular activity, or electrical noise were excluded from analysis. Individual electrodes with poor signal quality within a trial were corrected using spherical spline interpolation; no more than 10% of electrodes were interpolated on any one trial. The average amount of trials per participant across all sounds was 21 (SD = 5.0). Within each sound condition, the average trial counts were: natural speech (M = 24, SD = 5.6) cow sound (M = 20, SD = 4.0) synthetic speech (M = 21, SD = 5.7), and complex tone (M = 20, SD = 6.1).

Following artifact rejection, the remaining trials for each condition were averaged, re-referenced to the average reference, and baseline-corrected. In order to maximize signal-to-noise ratio, ERPs were measured as the average of a cluster of electrodes rather than a single electrode. The strongest auditory responses were found at a cluster that corresponded to Cz (electrodes 7, 31, 55, 80 106, 129/Cz), while the strongest P3a orienting responses were found at a cluster corresponding to Fz (electrodes 4, 5, 10, 11, 12, 16, 18, 19). The latency and mean amplitude windows were chosen based on the examination of the grand-averaged waveform, resulting in the time windows of 65-115 ms for N1 and 120-200 ms for P2, and 275-400 ms for P3a. Latency was measured as the local peak within the time window. We defined a local peak as the time-point in which at least 10 data points (20 ms) on either side of the peak showed a smaller amplitude. This technique is favorable over simple peak measures (which identify the most positive or negative point within the window) because it is less influenced by noise (Luck, 2014). However, the local peak approach could not be used for the seven children who did not show a clear N1 peak in the Natural Speech Task (the Immature NST subgroup). For these children, the most negative point between 65-115 ms (simple peak) was used for the initial analysis on the full sample.

Latency values to the deviant synthetic speech vowel, complex tone, natural speech vowel, and cow sound at a Cz cluster were compared for N1 and P2 separately in the full sample using a 2 x 2 repeated measures ANOVA where the within-subject factors were stimulus type (synthetic, natural) and sound category (speech, nonspeech). A priori planned simple effects were used to investigate each within-task speech vs. nonspeech comparison (e.g., synthetic speech vs. complex tone, natural speech vs. cow sound) and one exploratory post-hoc across-task test to compare synthetic vs. natural speech.

Given the substantial variability found in N1 morphology between 65-115 ms in individual waveforms in the Natural Speech Task (illustrated by the apparent lack of activity to natural speech until ~100 ms in Figure 2.4, discussed below), latency values between natural speech and the cow sound were analyzed separately in the Immature NST subgroup and Mature NST subgroup using paired t-tests within groups. This was done in order to determine whether the results of the repeated measures ANOVA across all participants were replicated within the subgroups, or if perhaps the maturity of the auditory ERP response led to different patterns in processing natural speech vs. complex nonspeech (i.e., cow sound). Because the Immature NST subgroup (n=7) did not display N1 responses, only P2 measured at the Cz cluster between 120-200 ms was analyzed for this group using paired t-tests. Note that in other paradigms, this response may be similar to the P1 seen in children before the development of N1, however since its timing overlapped directly with the P2 in the Mature NST subgroup's N1-P2 response we will refer to it as P2 for consistency. In the Mature NST Subgroup, N1 (65-115 ms) and P2 (120-200 ms) latencies at the Cz cluster were compared between natural speech and the cow sound using paired t tests.

P3a amplitude was calculated as the mean amplitude between 275-400 ms of the difference wave calculated by subtracting the standard sound from the deviant sound for all four deviant sounds (deviant synthetic complex tone, deviant synthetic /i/, deviant natural cow, deviant natural /a/) from both tasks. P3a latency and mean amplitude values were compared in a 2 x 2 repeated measures ANOVA where the within-subject factors were stimulus type (synthetic, natural) and sound category (speech, nonspeech).

Effect sizes are reported for all significant findings based on the recommendations by Lakens (2013) and calculated using the supplementary tables. We present both within-subjects effect sizes which utilize the correlation between measures and thus result in a power increase associated with within-subjects designs that is best fit to our study, but we also include between-subjects effect sizes in order to generalize to other studies that may compare groups in the future, as well as to facilitate interpretability of both effect sizes given that between-subjects effect sizes are more common. For the repeated measures ANOVA effects, we present both partial eta squared as the within-subjects effect size and generalized eta squared as the between-groups effect size. For simple effects and paired t-tests, we present Cohen's d_z as the within-subjects effect size. For the between-subjects effect size, we applied Hedges correction to Cohen's d_{av} to report Hedges g_{av} that takes into account the positive bias associated with using sample estimates.

Results

Auditory ERPs

The mean and standard deviation of the latency values for N1 and P2 to all stimuli are listed in Table 2.2. Figure 2.3 displays the N1 and P2 ERP results for the Synthetic Speech Task in all participants and Figure 2.4 displays the N1 and P2 ERP results for the Natural Speech Task in all participants. Repeated measures ANOVA for N1 (measured at Cz cluster between 65-115 ms) showed no statistically significant main effect of either stimulus type (synthetic vs. natural, $F(1,21) = .86, p = .36$) or sound category (speech vs. nonspeech, $F(1,21) = .33, p = .57$). However,

Table 2.2 *Mean and SD Values to Auditory ERPs and P3a ERP to All Stimuli Across All Participants and Subgroups.*

All Participants (N=22)

Task	**Stimuli**	**N1 Mean Latency (SD)**	**P2 Mean Latency (SD)**
Synthetic Speech Task	Synthetic Vowel Deviant	96.82 (10.04)	161.27 (15.49)
	Complex Tone Deviant	86.55 (9.99)	149.09 (16.22)
Natural Speech Task	Natural Vowel Deviant	84.64 (14.45)	149.36 (20.51)
	Cow Sound Deviant	93.50 (12.43)	160.45 (15.86)

Immature NST Subgroup (N=7)

Task	**Stimuli**	**N1 Mean Latency (SD)**	**P2 Mean Latency (SD)**
Natural Speech Task	Natural Vowel Deviant	N/A	139.43 (20.26)
	Cow Sound Deviant	N/A	158.86 (13.31)

Mature NST Subgroup (N=15)

Task	**Stimuli**	**N1 Mean Latency (SD)**	**P2 Mean Latency (SD)**
Natural Speech Task	Natural Vowel Deviant	89.60 (13.80)	152.27 (16.71)
	Cow Sound Deviant	95.53 (8.95)	161.20 (17.30)

All Participants (N=22)

Task	**Stimuli**	**P3a Difference Wave Mean Amplitude (SD)**	**P3a Difference Wave Mean Latency (SD)**
Synthetic Speech Task	Synthetic Vowel Deviant	-0.09 (3.15)	346.55 (33.68)
	Complex Tone Deviant	0.44 (4.18)	340.72 (35.87)
Natural Speech Task	Natural Vowel Deviant	-0.04 (3.32)	328.0 (32.53)
	Cow Sound Deviant	1.50 (3.68)	324.73 (35.34)

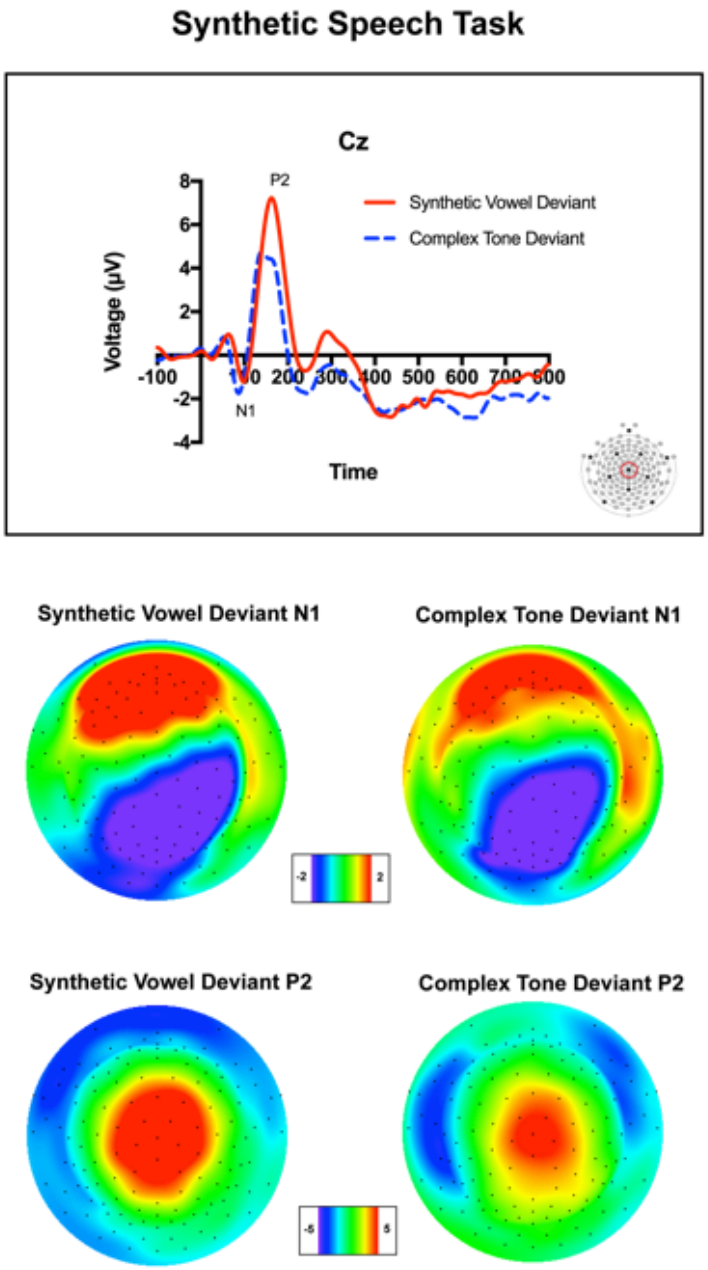


Figure 2.3 Auditory Event-Related Potentials and Topographic Maps to the Deviant Sounds in the Synthetic Speech Task in All Participants (N=22). ERPs are plotted at a Cz cluster of electrodes (7, 31, 55, 80 106, 129/Cz). Topographic maps are plotted at the peak of the N1 and P2 ERPs.

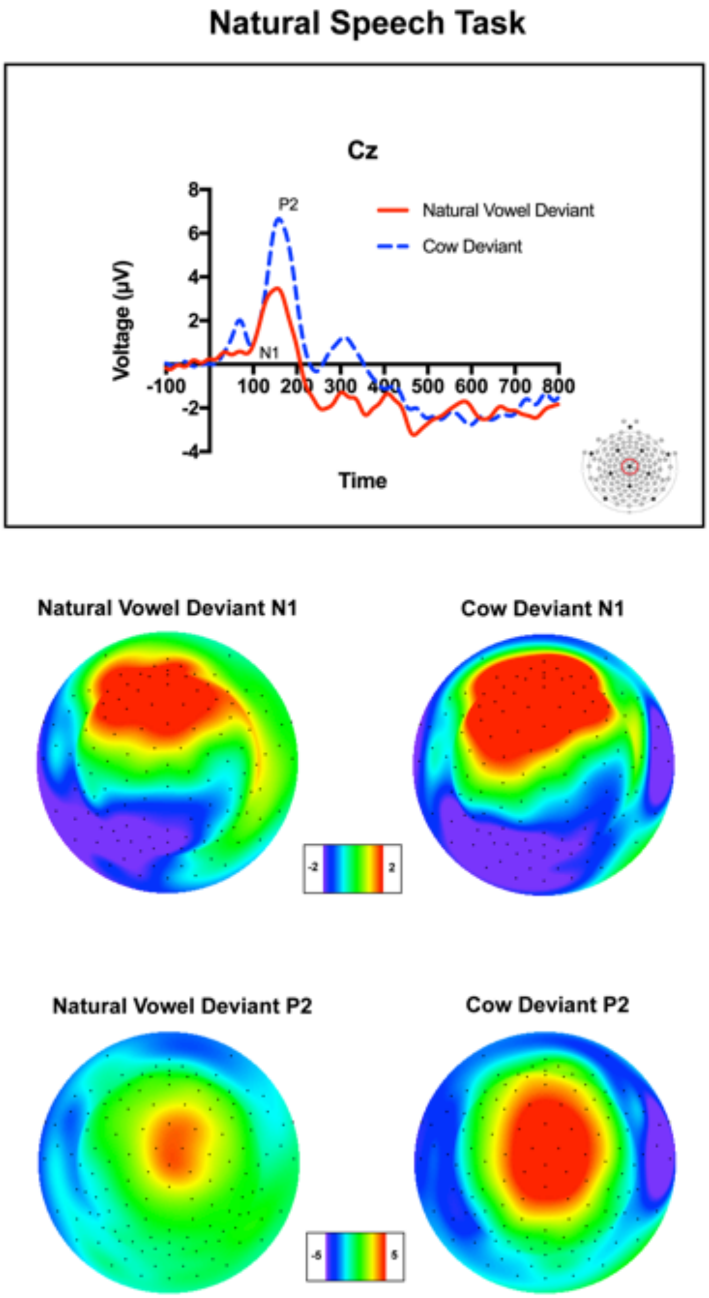


Figure 2.4 Auditory Event-Related Potentials and Topographic Maps for N1 to the Deviant Sounds in the Natural Speech Task in All Participants (N=22). ERPs are plotted at a Cz cluster of electrodes (7, 31, 55, 80 106, 129/Cz). Topographic maps are plotted at the peak of the N1 and P2 ERPs.

the interaction of stimulus type x sound category was significant (*F*(1, 21) = 19.85, *p* < .001, partial eta squared = .486, generalized eta squared = .145). Simple effects on the interaction revealed that N1 latency in the Natural Speech Task was significantly faster to natural speech (M = 84.6 ms, SD = 14.5) compared to the cow sound (M = 93.5 ms, SD = 12.4, *p* = .002, d_z = 0.76, Hedges g_{av} = 0.64). Yet, in the Synthetic Speech task N1 latency was significantly faster to the complex tone (M = 86.6 ms, SD = 10.0) compared to synthetic speech (M = 96.8 ms, SD = 10.0, *p* < .001, d_z = 0.89, Hedges g_{av} = .98). Furthermore, across tasks N1 latency to natural speech was significantly faster than synthetic speech (t(21) *p* = .006, d_z = .65, Hedges g_{av} = .94).

Repeated measures ANOVA for P2 measured at the Cz cluster between 120-200 ms confirmed the same pattern as the N1 findings, in which no significant main effects were found for either stimulus type (*F*(1, 21) = .01, *p* = .94) or sound category (*F*(1, 21) = .07, *p* = .80). However, the stimulus type x sound category interaction was significant (*F*(1, 21) = 20.90, *p* < .001, partial eta squared = .499, generalized eta squared = .108). Again, simple effects revealed that P2 latency in the Natural Speech Task was significantly faster to natural speech (M = 149.4 ms, SD = 20.5) compared to the cow sound (M = 160.5 ms, SD = 15.9, *p* = .002, d_z = .77, Hedges g_{av} = .58), however, in the Synthetic Speech task P2 latency was significantly faster to the complex tone (M = 149.1 ms, SD = 16.2) compared to synthetic speech (M = 161.3 ms, SD = 15.5, *p* = .002, d_z = .75, Hedges g_{av} = .74). In addition, across tasks P2 latency to natural speech was significantly faster than synthetic speech (*p* = .011, d_z = .60, Hedges g_{av} = .63).

Natural Speech Task (NST) Subgroup Analysis

As explained in the Participants section above, children were subgrouped based on their presence of an N1 response to the stimuli in the Natural Speech Task into either the Immature NST or Mature NST Subgroups. Figure 2.5 displays auditory ERP responses in each subgroup to deviant sounds in both the Natural Speech Task and Synthetic Speech Task, demonstrating that while children in the Mature NST Subgroup demonstrated N1 responses in both tasks, children in the Immature NST Subgroup only exhibited N1 responses to the stimuli in the Synthetic Speech Task.

The subgroups did not significantly differ on IQ (*t*(20) = .18 , p = .86) or age (*t*(20) = -.85, p = .41). In addition, data quality measured by number of artifact-free trials included in the analysis could not account for the differences in N1 responses both between subgroups and within participants, as the mean number of trials included did not significantly differ between subgroups for the deviant natural vowel (Immature NST Subgroup M = 23, SD = 6.2, Mature NST Subgroup M = 23.8, SD = 5.2, *t*(20) = -0.32, *p* = .75) or cow sound (Immature NST Subgroup M = 18, SD = 3.6, Mature NST Subgroup M = 20.5, SD = 4.1, *t*(20) = -1.4, *p* = .18). In addition, within the Immature NST Subgroup, a paired t-test found trend-level significance in the average number of trials included between the Natural Speech Task (M=20.5, SD=2.8) and Synthetic Speech Task (M = 18.1, SD = 3.5, *t*(6) = -2.25, *p* = .06), suggesting that children may have been more likely to have more artifact-free trials in the Natural Speech Task than the Synthetic Speech Task, again supporting the argument that the lack of a mature N1 in the Natural Speech Task is not due to poor data quality. Finally, the lack of an N1 response to the natural stimuli in this subgroup is also not

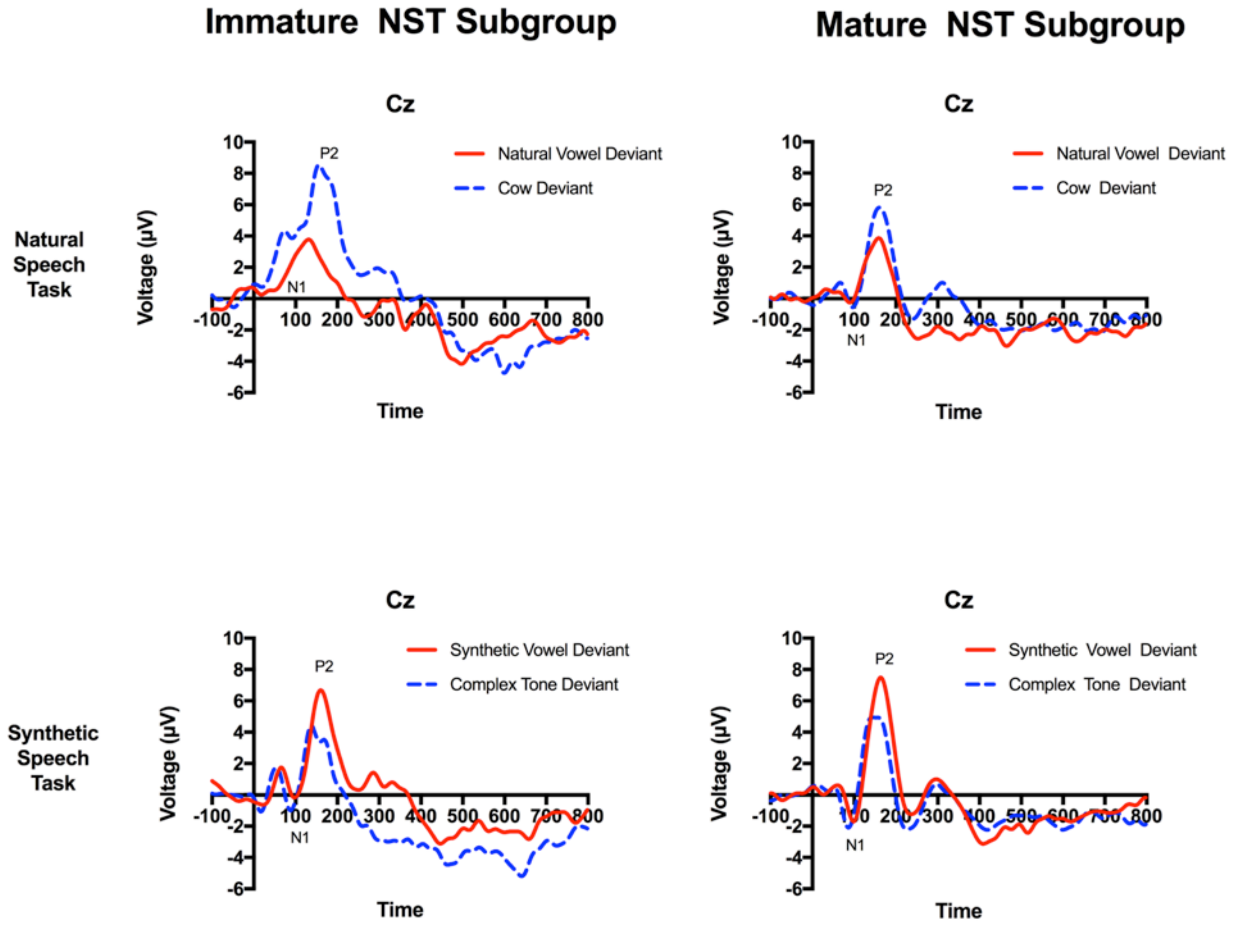


Figure 2.5 Auditory Event-Related Potentials Subgrouped by Immature and Mature Responses to the Natural Speech Task (NST). Auditory ERPs to the deviant sounds in the Synthetic Speech Task are also shown on the bottom for each subgroup to emphasize that both the Immature NST Subgroup and Mature NST Subgroup showed mature N1-P2 auditory ERPs to the deviant sounds in the Synthetic Speech Task; only the Immature NST Subgroup showed immature responses to the deviant sounds in the Natural Speech Task.

likely to be explained by attention differences due to fatigue, as the majority (N=5/7) of these participants were randomly assigned to listen to the Natural Speech Task first.

In both the Immature NST and Mature NST Subgroups, paired t-tests confirmed the results of the repeated measures ANOVA. In the Immature NST Subgroup showing no N1 response, only P2 latency between 120-200 ms was analyzed and found to be significantly faster to natural speech (M = 139.4, SD = 20.3) than the cow sound (M = 158.9, SD =13.3, $t(6) = -2.00$, $p = .046$, $d_z = 0.75$, Hedges $g_{av} = .99$). In the Mature NST Subgroup, natural speech N1 latency (M = 89.6, SD = 13.8) was significantly faster than the cow sound (M = 95.5, SD = 9.0, $t(14) = -2.02$, $p = .03$, $d_z = .52$, Hedges $g_{av} = .48$) and the same was true for P2 (natural speech M = 152.5, SD = 16.7, cow sound M = 161.2, SD = 17.3, $t(14) = -2.64$, $p = .009$, $d_z = .66$, Hedges $g_{av} = .48$).

P3a

Figure 2.6 displays the P3a ERP results (both difference wave and raw condition ERPs) for both the Synthetic Speech Task and Natural Speech Task in all participants. The P3a is reflected by a greater positivity to the deviant sound compared to the standard sound it was paired with, as well as a large positivity in the difference wave; both are presented because only presenting the difference wave can be misleading (Picton et al., 2000). Repeated measures ANOVA for P3a mean amplitude (measured at Fz cluster between 275-400 ms) showed no statistically significant main effect of either stimulus type (synthetic vs. natural, $F(1, 21) = .75$, $p = .40$) or sound category (speech vs. nonspeech, $F(1, 21) = 2.21$, $p = .15$). Additionally, there was no significant stimulus type x sound category interaction (, $F(1, 21) = .32$, $p = .58$).

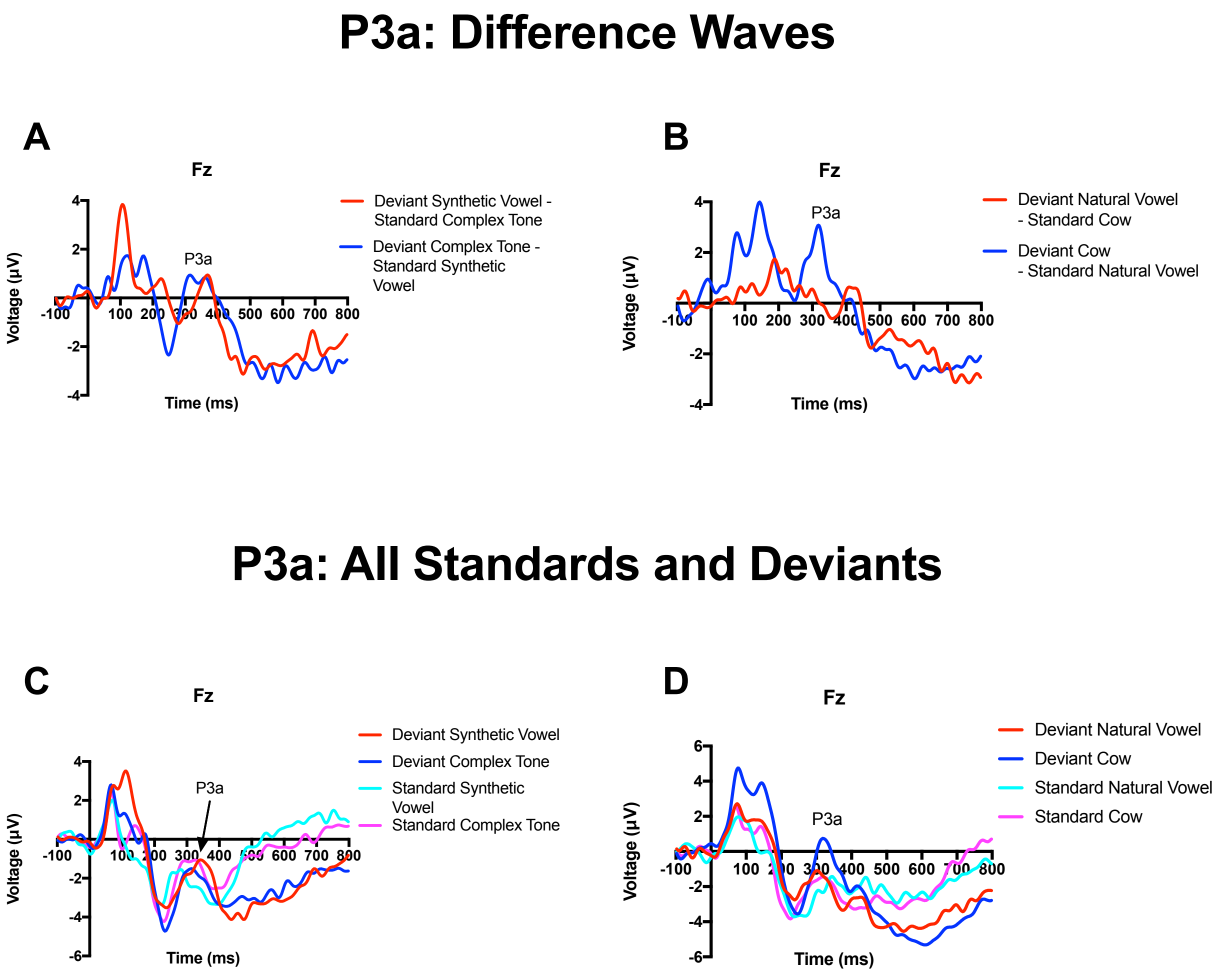


Figure 2.6 P3a ERPs to Deviant Stimuli in the Synthetic Speech Task and Natural Speech Task. A: Difference waves showing the P3a ERPs in the Synthetic Speech Task. B: Difference waves showing the P3a ERPs in the Natural Speech Task. C: All deviant and standard stimuli in the Synthetic Speech Task, presented to display the comparison between the P3a response to deviant and standards in the raw waveforms. D: All deviant and standard stimuli in the Natural Speech Task, presented to display the comparison between the P3a response to deviant and standards in the raw waveforms. The P3a is evidenced when there is both a large positivity in the difference wave (A and B) and greater positivity to the deviant compared to standard sound. Thus, only the Deviant Cow (D, blue line) shows evidence of a true P3a response.

For P3a latency, there was a significant main effect of stimulus type (synthetic vs. natural; $F(1, 21) = 5.74, p = .03$, partial eta squared = .22, generalized eta squared = .19). This was driven by faster P3a latency to natural stimuli (M = 326.36, SD = 33.60) than synthetic stimuli (M = 343.64, SD = 34.51). There was no significant main effect of sound category (speech vs. nonspeech; $F(1, 21) = .58, p = .46$), or significant stimulus type x sound category interaction ($F(1, 21) = .03, p = .87$).

Notably, the results of the P3a standard, deviant, and difference waveforms reveal (Figure 2.6) that only the cow deviant sound elicited evidence of a P3a response with both greater positivity to the deviant compared to the standard natural vowel, and large positivity in the difference wave, although this was not reflected in an interaction between stimulus type and sound category for P3a amplitude. Based on the ERP waveforms, the results of the P3a mean amplitude and latency should be interpreted with caution.

Discussion

The purpose of this study was to investigate whether faster processing and orienting to speech over nonspeech sounds could be evidenced by ERP responses in a sample of typically developing 8–11-year-old children. We hypothesized that previous ERP studies had not found evidence for faster speech processing or greater orienting because of the stimuli used, i.e., synthetic speech and tones that are not as ecologically valid as natural stimuli. We thus used two tasks, a Synthetic Speech Task and a novel Natural Speech Task to examine the auditory ERP and P3a ERP responses in the context of both natural and synthetic speech stimuli. Our results provide neural evidence for more efficient processing of natural speech, but not synthetic speech, that begins in the earliest stages of cortical auditory processing. Furthermore, our results revealed that the maturity of the children’s auditory ERP responses differed across tasks, with all children showing mature N1 responses to the synthetic stimuli but 7 children reverting back to immature child-like responses (no N1) to the more complex natural stimuli. This effect was not driven by age, IQ, or data quality, and thus evidence for a neural bias to process natural speech faster than the nonspeech cow sound was found across participants regardless of the maturity of their auditory ERP responses. In addition, we replicated the finding of previous studies in which auditory ERP latencies to a complex tone were significantly faster than to synthetic speech (Bruder et al., 2011). However, we did not find evidence for a greater orienting response to speech sounds, as in fact we only found evidence for a true P3a response to the natural cow deviant stimulus.

Our finding that children’s early sensory processing was faster to the natural speech vowel in the Natural Speech Task is the first ERP study that we are aware of to support more efficient processing of speech compared to nonspeech stimuli. Given that we did not find faster processing to the synthetic speech vowel in the Synthetic Speech Task, our results suggest that faster processing may be driven by the acoustic properties of the human voice, rather than the linguistic content of the signal. In addition, our finding that the natural speech vowel showed a significantly faster N1 and P2 latency than the synthetic speech vowel in our full sample could be taken as further evidence that the human voice is prioritized in the environment relative to any kind of speech signal. However, this comparison was exploratory and warrants caution in its interpretation because 1) latency differences may be due to differences in complexity of natural speech vs. synthetic speech, and 2) the natural speech vowel deviant was an /a/ and the synthetic speech deviant was an /i/, and thus the latency could be impacted by differences in phonological processing (Grimaldi, Manca, & Di Russo, 2016). A comprehensive comparison of natural vs.

synthetic speech processing was beyond the scope of the present paper, however, a faster latency to natural speech than synthetic speech was also reported by Swink and Stuart (2012).

Our study suggests that different auditory ERP results can arise when using natural vs. synthetic speech matched to nonspeech analogs in a passive oddball task, and this has implications for the use of synthetic speech in neuroimaging research. Several fMRI studies have found differences in more general auditory and language areas of the brain between synthetic and natural speech (Assaneo et al., 2016; Beaucousin et al., 2007; Benson et al., 2001), corroborating neural differences between natural and synthetic speech. Furthermore, many behavioral studies have shown that synthetic speech is actually more difficult to perceive across a variety of tasks (Allen, Hunnicutt, Klatt, Armstrong, & Pisoni, 1987; Drager & Reichle, 2001; Duffy & Pisoni, 1992; Hawkins & Slater, 1994; Kangas & Allen, 1990; Logan, Greene, & Pisoni, 1989; Nye & Gaitenby, 1973; Paris, Gilson, Thomas, & Silver, 1995). One interpretation of these results is that models of speech processing in the brain based on synthetic speech may reflect compensatory processing strategies rather than the true nature of neural speech networks.

In addition to the increased ecological validity of using natural stimuli, our finding that the maturity of children's auditory ERP responses differed between synthetic and natural stimuli adds new insight into the maturation of the auditory system in children. Previous studies using long interstimulus intervals (> 1 s) in children demonstrated that children showed mature N1-P2 (Bruneau & Gomot, 1998; Ceponiene et al., 2002; Paetau et al., 1995); however, these studies used synthetic speech and tone stimuli. We found that even with a longer ISI averaging around 2 s, 7 children who showed mature N1 responses to the synthetic stimuli showed no evidence of an N1 to the natural stimuli. This suggests that the auditory system matures faster for simple sounds. A few cross-sectional ERP studies in children have looked at maturation of the N1 to natural speech and pure tones. Generally these studies have found no differences, or faster maturation of natural speech at temporal sites but not central sites characteristic of the adult fronto-central N1 (Bishop, Hardiman, Uwer, & von Suchodoletz, 2007; Mahajan & McArthur, 2012; Pang & Taylor, 2000), although Bishop et al. (2007) noted substantially more variability in the maturation to natural speech. However, these studies only looked at the group level trajectories, and did not compare whether maturation to speech and tones differed at the individual level as in the present study. Furthermore, because pure tones were not matched to natural speech on either acoustic features or complexity, they could not comment on whether their findings reflected differences in maturation to the speech sound because of the "speech" nature or simply because it was more complex than the pure tone. Our study adds to these findings by matching both synthetic and natural speech to nonspeech counterparts and demonstrating that maturational differences across individuals were due to the complexity of the sound, not the "speechness" of the sound. Because previous studies did not look at individual responses, our finding that seven children reverted back to a child-like immature response to the more complex natural stimuli is a novel contribution to the literature. This result suggests that the maturation of the auditory cortex is not uniform; i.e., the processing of more complex sounds continues to develop concurrently with an already mature auditory processing of simple sounds.

The maturity of the auditory ERP responses was not found to be related to IQ, age, or data quality in our sample, yet this could have been due to not having enough variability in IQ and the need for more participants in each age bin to find age effects. Future studies should investigate which factors are involved in moderating the maturation of complex auditory processing in children, and whether the morphology of the auditory ERPs correlates with mature auditory processing skills. Some evidence exists to support this claim; Kwok et al. (2017) found that the maturity of auditory ERP responses in 7—10-year-old children was significantly related to language abilities, and Carpenter and Shahin (2013) observed that six-year-olds whose N1-P2 responses closely resembled adults' performed as well as the adults behaviorally on a phonetic judgement task.

Our finding that natural speech was processed faster than the cow sound regardless of whether children displayed the mature adult-like N1 waveform morphology suggests that neural prioritization of speech is not related to the maturity of auditory processing. Given that the emergence of the mature N1 is thought to coincide with the structural maturation of intra-cortical connections occurring by age 12 (Eggermont & Moore, 2012), it is assumed that children showing the absence of an N1 ERP response have likely not completed this stage of structural maturation. Accordingly, because faster processing of the natural speech stimulus was found in both the Immature NST an Mature NST ERP Subgroups, faster processing may more likely be driven by the existing thalamo-cortical connections projecting to the auditory cortex that are known to be mature by age five (Moore, 2002). Faster processing of natural speech may be driven by subcortical bottom-up processing pathways. There is also evidence this could be accomplished by an early role from the brainstem in routing important social information that matures during neonatal development (Geva et al., 2017).

Regardless of the precise neural circuitry involved, the question remains: what leads to faster processing of natural speech? Is it due to an innate circuitry that infants are born with, or does it develop due to experience and eventual expertise with listening to the human voice and speech? The finding that a behavioral bias to natural speech over synthetic sounds exists in infants just a few days old could lend support to an innate view (Vouloumanos & Werker, 2007), however, that study used synthetic time-varying sine waves as their nonspeech comparison. In another study that compared human speech to animal vocalizations, representing more natural and complex stimuli similar to the present study, Vouloumanos et al. (2010) found that while neonates showed no preference, by 3 months old infants preferred the human speech suggesting that experience with the human voice led to the preference for speech. Thus, experience may lead to the tuning and prioritization of the speech signal that is similar to the neural selectivity that has been found to develop in other animals as a result of experience with conspecific vocalizations (Amin, Gastpar, & Theunissen, 2013). Our study suggests this process is complete in children by 8—11 years old; future studies could apply our paradigm to infants and toddlers to determine at what age the processing of natural speech becomes faster.

In contrast to the auditory ERP results, we did not find evidence of enhanced orienting to speech over nonspeech supported by the P3a component. Indeed, we only found a true P3a orienting response to the natural cow deviant. Given that other studies have found P3a responses to similar synthetic speech and tone stimuli, it is unclear why our study did not replicate these results. However, it is difficult to say for certain whether our findings are atypical because previous studies primarily only present the difference waves, which often do not reflect the raw standard and deviant waveforms. By only looking at the difference waves, our synthetic speech and tone stimuli also appear to show normal P3a responses. Regardless of whether our synthetic speech and tone stimuli P3a results are similar to previous studies, it is notable that the P3a to the cow deviant was stronger than all other stimuli. We hypothesize that one possibility that could explain our pattern of results is that the standard sound that precedes the deviant sound may have a much larger impact on the orienting response than previously thought. Our reasoning for this relates to the present findings from the auditory ERPs that suggest that natural speech, but not synthetic speech, is prioritized early and processed faster than the other sounds. Given the complex relationship between sensory and attentional processing (Petkov et al., 2004; Woldorff et al., 1993), it would be interesting for a future study to investigate in a controlled design whether the standard sound influences the P3a response. In the present study, the cow deviant was the only stimulus to follow a stream of natural speech sounds, and therefore it's possible that the auditory system is "primed" to orient to any upcoming sound that follows natural speech because the auditory system is automatically more tuned in to the stream of natural speech sounds.

An important limitation of the current study is that we did not systematically control for acoustic differences between stimuli that could lead to early latency differences. Previous studies using simple synthetic stimuli have found that the frequency and rise time of simple synthetic stimuli contribute to N1 latency (Jacobson, Lombardi, Gibbens, Ahmad, & Newman, 1992; Kodera, Hink, Yamada, & Suzuki, 1979; Verkindt, Bertrand, Perrin, Echallier, & Pernier, 1995; Woods, Alain, Covarrubias, & Zaidel, 1993). However, more recent studies using natural vowels and more controlled experimental paradigms suggest that N1 latency does not change solely due to changes in pitch or rise time, and rather is determined by multiple dynamic properties of the stimulus (Grimaldi et al., 2016; Krishnan, Bidelman, Smalt, Ananthakrishnan, & Gandour, 2012). In order to better characterize the speed of processing natural speech, a future study could include child, adult male, and adult female vocalizations that differ in their fundamental frequencies to determine whether natural speech continues to be processed faster than well-matched cow or other animal vocalizations. We hypothesize that acoustic attributes of the human voice (e.g., a quasi-periodic air pressure wave, specific harmonic-to-noise ratio range, specific number and/or frequency range of formants) are already recognized at the level of early auditory ERPs and result in faster processing regardless of speaker pitch, due to the significance of the human voice. Future studies could also add a comparison between nonspeech human vocalizations (e.g., laughs, cries, etc.) and animal vocalizations to determine whether the faster latency is maintained to any human vocalization, or if it is speech-specific.

Conclusion

We created a novel task battery using natural speech to investigate the auditory processing and involuntary orienting of speech vs. nonspeech sounds using ecologically valid stimuli in 8—11-year-old children. Compared to a commonly used oddball task with synthetic stimuli, we found that the natural speech stimulus indexed faster processing than the cow sound regardless of ERP maturity, providing evidence that the brain may be prioritizing processing of the human voice signal rather than speech sounds in general. Similar to previous studies, faster processing was not found in a control task using synthetic speech and complex tones, and instead we replicated findings in which complex tones are processed faster than synthetic speech. The findings of our study suggest that natural speech may be more efficiently processed than synthetic speech, and that the maturity of the auditory ERPs in 8—11-year-old children differs for some children to simple synthetic sounds vs. natural complex sounds, although more research is needed to determine whether the orienting responses to natural speech vs. nonspeech differ. These results have implications for the adoption of naturalistic stimuli in EEG/ERP research, and provide the first ERP evidence for faster processing natural speech than a nonspeech analog.

Given the advantages of faster processing of speech to identify the signal and facilitate faster orienting to and comprehension of communicative messages, the absence of faster processing to speech in some individuals could lead to deficits in language development or social communication skills. The present paradigm could thus be used in future studies to assess whether children with neurodevelopmental conditions associated with atypical language outcomes (e.g. autism, developmental language disorder) do not exhibit faster processing of natural speech.

References

Alipour, F., & Jaiswal, S. (2008). Phonatory characteristics of excised pig, sheep, and cow larynges. *The Journal of the Acoustical Society of America*, *123*(6), 4572–4581. https://doi.org/10.1121/1.2908289

Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., & Pisoni, D. B. (1987). *From text to speech: The MITalk system*. Cambridge University Press. Retrieved from http://dl.acm.org/citation.cfm?id=28587

Amin, N., Gastpar, M., & Theunissen, F. E. (2013). Selective and Efficient Neural Coding of Communication Signals Depends on Early Acoustic and Social Environment. *PLOS ONE*, *8*(4), e61417. https://doi.org/10.1371/journal.pone.0061417

Assaneo, M. F., Sitt, J., Varoquaux, G., Sigman, M., Cohen, L., & Trevisan, M. A. (2016). Exploring the anatomical encoding of voice with a mathematical model of the vocal system. *NeuroImage*, *141*, 31–39. https://doi.org/10.1016/j.neuroimage.2016.07.033

Athena, V., & Suzanne, C. (2014). Foundational Tuning: How Infants' Attention to Speech Predicts Language Development. *Cognitive Science*, *38*(8), 1675–1686. https://doi.org/10.1111/cogs.12128

Beaucousin, V., Lacheret, A., Turbelin, M.-R., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). FMRI Study of Emotional Speech Comprehension. *Cerebral Cortex*, *17*(2), 339–352. https://doi.org/10.1093/cercor/bhj151

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, *13*(1), 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309–312. https://doi.org/10.1038/35002078

Benasich, A. A., Choudhury, N., Friedman, J. T., Realpe-Bonilla, T., Chojnowska, C., & Gou, Z. (2006). The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. *Neuropsychologia*, *44*(3), 396–411. https://doi.org/10.1016/j.neuropsychologia.2005.06.004

Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., & Liberman, A. M. (2001). Parametrically Dissociating Speech and Nonspeech Perception in the Brain Using fMRI. *Brain and Language*, *78*(3), 364–396. https://doi.org/10.1006/brln.2001.2484

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, *10*(5), 512–528. https://doi.org/10.1093/cercor/10.5.512

Bishop, D. V., Hardiman, M., Uwer, R., & von Suchodoletz, W. (2007a). Maturation of the long-latency auditory ERP: step function changes at start and end of adolescence. *Developmental Science*, *10*(5), 565–575. https://doi.org/10.1111/j.1467-7687.2007.00619.x

Bishop, D. V., Hardiman, M., Uwer, R., & von Suchodoletz, W. (2007b). Maturation of the long-latency auditory ERP: step function changes at start and end of adolescence. *Developmental Science*, *10*(5), 565–575. https://doi.org/10.1111/j.1467-7687.2007.00619.x

Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F. E., & Linden, D. E. J. (2004). Attentional systems in target and distractor processing: a combined ERP and fMRI study. *NeuroImage*, *22*(2), 530–540. https://doi.org/10.1016/j.neuroimage.2003.12.034

Braaten, R. F., & Reynolds, K. (1999). Auditory preference for conspecific song in isolation-reared zebra finches. *Animal Behaviour*, *58*(1), 105–111. https://doi.org/10.1006/anbe.1999.1134

Bruder, J., Leppänen, P. H., Bartling, J., Csépe, V., Démonet, J.-F., & Schulte-Körne, G. (2011). An investigation of prototypical and atypical within-category vowels and non-speech analogues on cortical auditory evoked related potentials (AERPs) in 9 year old children. *International Journal of Psychophysiology*, *79*(2), 106–117.

Bruneau, N., & Gomot, M. (1998). Auditory evoked potentials (N1 wave) as indices of cortical development. In *Neuroimaging in child neuropsychiatric disorders* (pp. 113–123). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-95848-9_10

Carpenter, A. L., & Shahin, A. J. (2013). Development of the N1-P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neuroscience Letters*, *544*, 56–61. https://doi.org/10.1016/j.neulet.2013.03.041

Čeponien, R., Rinne, T., & Näätänen, R. (2002). Maturation of cortical sound processing as indexed by event-related potentials. *Clinical Neurophysiology*, *113*(6), 870–882.

Čeponien≐, R., Cheour, M., & Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children: evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(4), 345–354. https://doi.org/10.1016/S0168-5597(97)00081-6

Ceponiené, R., Shestakova, A., Balan, P., Alku, P., Yiaguchi, K., & Naatanen, R. (2001). Children's auditory event-related potentials index sound complexity and "speechness." *International Journal of Neuroscience*, *109*(3–4), 245–260.

Chan, A. M., Dykstra, A. R., Jayaram, V., Leonard, M. K., Travis, K. E., Gygi, B., … Cash, S. S. (2013). Speech-Specific Tuning of Neurons in Human Superior Temporal Gyrus. *Cerebral Cortex*, bht127. https://doi.org/10.1093/cercor/bht127

Choudhury, N., & Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: Prediction to 3 and 4 year language and cognitive abilities. *Clinical Neurophysiology*, *122*(2), 320–338. https://doi.org/10.1016/j.clinph.2010.05.035

Drager, K., & Reichle, J. (2001). Effects of age and divided attention on listeners' comprehension of synthesized speech. *Augmentative and Alternative Communication*, *17*(2), 109–119.

Duffy, S. A., & Pisoni, D. B. (1992). Comprehension of synthetic speech produced by rule: A review and theoretical interpretation. *Language and Speech*, *35*(4), 351–389.

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In *Human auditory development* (pp. 61–105). Springer.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., & Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *Journal of Neuroscience*, *15*(4), 2748–2755.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage*, *23*(3), 840–848. https://doi.org/10.1016/j.neuroimage.2004.09.019

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*(10), 2593–2608. https://doi.org/10.1093/brain/awl247

Ge, P. J., French, L. C., Ohno, T., Zealear, D. L., & Rousseau, B. (2009). Model of Evoked Rabbit Phonation. *Annals of Otology, Rhinology & Laryngology*, *118*(1), 51–55. https://doi.org/10.1177/000348940911800109

Geva, R., Dital, A., Ramon, D., Yarmolovsky, J., Gidron, M., & Kuint, J. (2017). Brainstem as a developmental gateway to social attention. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*. https://doi.org/10.1111/jcpp.12746

Gilley, P. M., Sharma, A., Dorman, M., & Martin, K. (2005). Developmental changes in refractoriness of the cortical auditory evoked potential. *Clinical Neurophysiology*, *116*(3), 648–657. https://doi.org/10.1016/j.clinph.2004.09.009

Grimaldi, M., Manca, A. D., & Di Russo, F. (2016). Electroencephalographical evidence of vowels computation representation in human auditory cortex. *Biolinguistic Investigations on the Language Faculty*, *235*, 79.

Hamilton, L. S., & Huth, A. G. (2018). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *0*(0), 1–10. https://doi.org/10.1080/23273798.2018.1499946

Hawkins, S., & Slater, A. (1994). Spread of CV and v-to-v coarticulation in british English: implications for the intelligibility of synthetic speech. In *ICSLP* (Vol. 94, pp. 57–60). Retrieved from http://www.ling.cam.ac.uk/sarah/docs/hawkins_slater_icslp94.pdf

Heaton, M. B., Miller, D. B., & Goodwin, D. G. (n.d.). Species-specific auditory discrimination in bobtail quail neonates. *Developmental Psychobiology*, *11*(1), 13–21. https://doi.org/10.1002/dev.420110106

Jacobson, G. P., Lombardi, D. M., Gibbens, N. D., Ahmad, B. K., & Newman, C. W. (1992). The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear and Hearing*, *13*(5), 300–306.

Jiang, J. J., Raviv, J. R., & Hanson, D. G. (2001). Comparison of the phonation-related structures among pig, dog, white-tailed deer, and human larynges. *The Annals of Otology, Rhinology, and Laryngology*, *110*(12), 1120–1125. https://doi.org/10.1177/000348940111001207

Kangas, K. A., & Allen, G. D. (1990). Intelligibility of Synthetic Speech for Normal-Hearing and Hearing-Impaired Listeners. *Journal of Speech and Hearing Disorders*, *55*(4), 751–755. https://doi.org/10.1044/jshd.5504.751

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test*. Wiley Online Library. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781118660584.ese1325/summary

Kileny, P. R., & Kripal, J. P. (1987). Test-retest variability of auditory event-related potentials. *Ear and Hearing*, *8*(2), 110–114.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America*, *67*(3), 971–995.

Kodera, K., Hink, R. F., Yamada, O., & Suzuki, J.-I. (1979). Effects of rise time on simultaneously recorded auditory-evoked potentials from the early, middle and late ranges. *Audiology*, *18*(5), 395–402.

Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., & Gandour, J. T. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, *50*(12), 2849–2859.

Kuhl, P. K. (2007). Is speech learning 'gated'by the social brain? *Developmental Science*, *10*(1), 110–120.

Kwok, E. Y. L., Joanisse, M. F., Archibald, L. M. D., Stothers, M. E., Brown, H. M., & Oram Cardy, J. (2017). Maturation in Auditory Event-Related Potentials Explains Variation in Language Ability in Children. *The European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13785

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Leech, R., Holt, L. L., Devlin, J. T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *The Journal of Neuroscience: The Official*

Journal of the Society for Neuroscience, *29*(16), 5234–5239. https://doi.org/10.1523/JNEUROSCI.5758-08.2009

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex (New York, N.Y.: 1991)*, *20*(12), 2958–2970. https://doi.org/10.1093/cercor/bhq045

Logan, J. S., Greene, B. G., & Pisoni, D. B. (1989). Segmental intelligibility of synthetic speech produced by rule. *The Journal of the Acoustical Society of America*, *86*(2), 566–581.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press. Retrieved from https://books.google.com/books?hl=en&lr=&id=y4-uAwAAQBAJ&oi=fnd&pg=PR5&dq=An+Introduction+to+the+Event-Related+Potential+Technique&ots=jA36D2RPIz&sig=nOoNKrH1EFySoIxVZBtSlUAD0s8

Mahajan, Y., & McArthur, G. (2012). Maturation of auditory event-related potentials across adolescence. *Hearing Research*, *294*(1–2), 82–94. https://doi.org/10.1016/j.heares.2012.10.005

Marler, P. (1990). Song Learning: The Interface between Behaviour and Neuroethology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *329*(1253), 109–114. https://doi.org/10.1098/rstb.1990.0155

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Mizrahi, A., Shalev, A., & Nelken, I. (2014). Single neuron and population coding of natural sounds in auditory cortex. *Current Opinion in Neurobiology*, *24*, 103–110.

Molfese, D. L., & Molfese, V. J. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development? *Infant Behavior and Development*, *8*(2), 197–211.

Molfese, D. L., & Molfese, V. J. (1997). Discrimination of language skills at five years of age using event-related potentials recorded at birth. *Developmental Neuropsychology*, *13*(2), 135–156.

Moore, J. K. (2002). Maturation of human auditory cortex: implications for speech perception. *Annals of Otology, Rhinology & Laryngology*, *111*(5_suppl), 7–10.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281–1296.

Nye, P. W., & Gaitenby, J. H. (1973). Consonant intelligibility in synthetic speech and in a natural speech control (modified rhyme test results). *Haskins Laboratories Status Report on Speech Research, SR*, *33*, 77–91.

Paetau, R., Ahonen, A., Salonen, O., & Sams, M. (1995). Auditory evoked magnetic fields to tones and pseudowords in healthy children and adults. *Journal of Clinical Neurophysiology*, *12*(2), 177–185.

Pang, E. W., & Taylor, M. J. (2000). Tracking the development of the N1 from age 3 to adulthood: an examination of speech and non-speech stimuli. *Clinical Neurophysiology*, *111*(3), 388–397.

Paris, C. R., Gilson, R. D., Thomas, M. H., & Silver, N. C. (1995). Effect of synthetic voice intelligibility on speech comprehension. *Human Factors*, *37*(2), 335–340.

Penna, M., & Meier, A. (n.d.). Vocal Strategies in Confronting Interfering Sounds by a Frog from the Southern Temperate Forest, Batrachyla antartandica. *Ethology*, *117*(12), 1147–1157. https://doi.org/10.1111/j.1439-0310.2011.01973.x

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Phil. Trans. R. Soc. B*, *370*(1664), 20140090. https://doi.org/10.1098/rstb.2014.0090

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, *7*(6), 658–663. https://doi.org/10.1038/nn1256

Picton, T. w., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a., Johnson, R., … Taylor, M. j. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(02), 127–152. https://doi.org/null

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: evidence from multi-channel evoked potentials. *Clinical Neurophysiology*, *111*(2), 220–236. https://doi.org/10.1016/S1388-2457(99)00236-9

Räikkönen, K., Birkás, E., Horváth, J., Gervai, J., & Winkler, I. (2003). Test-retest reliability of auditory ERP components in healthy 6-year-old children. *NeuroReport*, *14*(16), 2121.

Segalowitz, S. J., & Barnes, K. L. (1993). The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology*, *30*(5), 451–459.

Shultz, S., & Vouloumanos, A. (2010). Three-Month-Olds Prefer Speech to Other Naturally Occurring Signals. *Language Learning and Development*, *6*(4), 241–257. https://doi.org/10.1080/15475440903507830

Swink, S., & Stuart, A. (2012). Auditory long latency responses to tonal and speech stimuli.(Report). *Journal of Speech, Language, and Hearing Research*, *55*(2), 447.

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience*, *15*(6), 355–366. https://doi.org/10.1038/nrn3731

Tremblay, K. L., Friesen, L., Martin, B. A., & Wright, R. (2003). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear and Hearing*, *24*(3), 225–232.

Verkindt, C., Bertrand, O., Perrin, F., Echallier, J.-F., & Pernier, J. (1995). Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *96*(2), 143–156.

Volpe, U., Mucci, A., Bucci, P., Merlotti, E., Galderisi, S., & Maj, M. (2007). The cortical generators of P3a and P3b: A LORETA study. *Brain Research Bulletin*, *73*(4), 220–230. https://doi.org/10.1016/j.brainresbull.2007.03.003

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, *81*(2), 517–527.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

Whitehouse, A. J., & Bishop, D. V. (2008). Do children with autism 'switch off'to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722–8726. https://doi.org/10.1073/pnas.90.18.8722

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8), 2031–2046. https://doi.org/10.3758/BF03196680

Woods, D. L., Alain, C., Covarrubias, D., & Zaidel, O. (1993). Frequency-related differences in the speed of human auditory processing. *Hearing Research*, *66*(1), 46–52.

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

Wronka, E., Kaiser, J., & Coenen, A. M. L. (2012). Neural generators of the auditory evoked potential components P3a and P3b. *Acta Neurobiologiae Experimentalis*, *72*(1), 51–64.

CHAPTER 3

Isolating the Methodological Features of the Auditory Oddball Task that Reliably Elicit the P3a Orienting Response to Naturalistic Sounds in Children

Introduction

The ability to rapidly switch attention towards a novel salient sound offers a crucial advantage in natural environments. Research using event-related potentials (ERPs) measured from the EEG signal has identified the P3a ERP component as a neural marker of an involuntary attention switch, often termed the "orienting response." The P3a is thought to reflect the process by which neural resources are allocated towards the evaluation of a salient stimulus (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2003; Squires, Squires, & Hillyard, 1975). The latency of the P3a occurs between approximately 300—400 ms after stimulus onset and has been found to be generated by a fronto-parietal network with involvement of the anterior cingulate cortex (Bledowski, Prvulovic, Goebel, Zanella, & Linden, 2004; Volpe et al., 2007; Wronka, Kaiser, & Coenen, 2012). Because the P3a is elicited automatically to a salient stimulus even in passive tasks, it has been extensively studied with a variety of sound stimuli in both typical studies of attention in children and adults (Chobert, Marie, François, Schön, & Besson, 2011; Rogenmoser, Elmer, & Jäncke, 2014; Seppänen, Pesonen, & Tervaniemi, 2012; Shestakova, Huotilainen, Ceponiene, & Cheour, 2003; Wang, Friedman, Ritter, & Bersick, 2005; Wronka, Kaiser, & Coenen, 2008) and numerous clinical conditions including language disorders, dyslexia, ADHD, psychosis, and even comatose patients (Fischer, Dailler, & Morlet, 2008; Hämäläinen, Leppänen, Guttorm, & Lyytinen, 2008; Hermens et al., 2010; Lincoln, Courchesne, Harms, & Allen, 1993; Yang et al., 2015). In contrast, studies assessing the factors that influence the P3a amplitude and latency are much more rare, and thus interpretations of findings across studies varying in task features and stimuli are limited.

The most common paradigm used to elicit the P3a response is the oddball task, in which a frequent standard sound is presented on 80-90% of trials and an infrequent deviant sound is presented on only 10-20% of trials. If the deviant sound is deemed salient enough to warrant an attention switch, the P3a ERP is demonstrated by a fronto-central positivity on the scalp. P3a amplitude and latency have been found to be modulated by the amount of acoustic difference between the standard and deviant sound, i.e., larger amplitudes are found to a deviant sounds that are more easily discriminated from the standard sound (Berti, Roeber, & Schröger, 2004; Comerchero & Polich, 1998; Katayama & Polich, 1998; Wronka et al., 2012). In addition, P3a amplitude is also modulated by task demands, such that although the P3a is found even when participants are working on a task and ignoring the sounds (Alho, Escera, Díaz, Yago, & Serra, 1997; Schröger & Wolff, 1998), amplitude is decreased when participants are focused on a more difficult task compared to an easier task (Berti & Schröger, 2003; Harmony et al., 2000). These studies demonstrate that the P3a component is not an automatic response to any novel sound, but instead is influenced by features of the task and standard sound. However, because all previous studies have only measured responses to pure tones that differed on frequency, one factor that has not been studied is whether the type of sound presented can differentially affect the P3a, both in terms of the sound category of the deviant or the category of the standard sound preceding it. To date, there has been little research on how these methodological manipulations within the auditory oddball paradigm impact the auditory P3a response.

One methodological feature of the auditory oddball paradigm that may have an impact on the P3a orienting response is a comparison between speech vs. nonspeech deviant stimuli. Speech represents an important sound signal in the environment for humans due to its essential role in communication, and thus it may be adaptive for orienting responses to be stronger and faster to speech in the environment over other types of sounds. Behaviorally, this has been supported by faster detection of novel speech sounds compared to nonspeech in an oddball task (Vouloumanos, Kiehl, Werker, & Liddle, 2001). In an ERP task, if speech sounds are more salient, then deviant speech sounds should elicit a larger P3a amplitude compared to deviant nonspeech sounds.

A second methodological feature of the auditory oddball paradigm that may have an impact on the P3a orienting response is the nature and extent of the similarity between the frequent standard stimuli and the infrequent deviant stimuli. Standard stimuli could be either congruent with the deviant stimulus (e.g. a standard speech stimulus and a deviant speech stimulus where the standard and deviant differ in pitch), or incongruent with the deviant stimulus (e.g. a standard nonspeech stimulus and a deviant speech stimulus). Given that for humans speech is likely to be more heavily monitored in our environments, then an auditory stream of standard speech sounds may also impact whether a deviant sound elicits an attention switch, regardless of what the upcoming deviant sound is. It is often assumed that the formation of auditory streams is an automatic bottom-up process (Sussman, Horváth, Winkler, & Orr, 2007), although it can be influenced by top-down processing as well as experience (Snyder, Holder, Weintraub, Carter, & Alain, 2009; Sussman et al., 2007; Zuijen, Sussman, Winkler, Näätänen, & Tervaniemi, 2004). Therefore, it's possible that the orienting response may be "primed" in the context of a speech stream compared to a nonspeech stream. For example, Whitehouse and Bishop (2008) found that children with autism only exhibited a P3a response to stimuli that followed a stream of standard nonspeech sounds, but not a stream of standard speech sounds. The authors suggested that this could be due to decreased motivation to attend to a stream of speech that becomes tuned out over time via top-down processing.

Very few studies have compared the P3a response between speech and nonspeech stimuli in typically developing samples. Jaramillo et al. (2001) and Sorokin, Alku, and Kujala (2010) both found larger P3a amplitude for speech stimuli, while Reiche et al. (2013) found no differences in P3a between speech and nonspeech. However, Reiche et al. (2013) did not present a stream of standard sounds, and both Jaramillo et al. (2001) and Sorokin et al. (2010) presented speech and nonspeech conditions in separate blocks, so it remains unknown whether the category of the stream of standard sounds impacts the orienting response to a deviant sound. Additionally, one limitation of these studies is that they used tone stimuli as the nonspeech stimuli, which may not reflect a P3a orienting response to an ecologically valid nonspeech stimulus.

In the present study, we aimed to assess the importance of sound category (speech vs. nonspeech) of both the standard and deviant sound on the P3a orienting response using ecologically valid speech and nonspeech stimuli in children aged 8—14-years-old. Given the complex relationship between sensory and attentional processing (Petkov et al., 2004; Woldorff et al., 1993), we also measured the P1 and N2 auditory ERP components that precede the P3a ERP. The P1 and N2 are thought to index cortical encoding of acoustic properties of sound using the same generators as adults in both primary and secondary auditory cortices (Eggermont & Moore, 2012; Ponton, Eggermont, Khosla, Kwong, & Don, 2002). We hypothesized that the importance of the speech signal would lead to 1) faster P1 and N2 latency to speech deviants compared to nonspeech deviants; 2) faster latency and increased P3a amplitude to speech deviants compared to nonspeech; and 3) increased P3a amplitude to a nonspeech deviant that followed a stream of speech standards compared to a nonspeech deviant following a stream of nonspeech standards.

Methods

Participants

Sixteen typically developing children between the ages of 8 and 14 years were recruited to participate in this study (M = 9.9 years, SD = 1.6). Children were recruited through the Vanderbilt University email distribution service. Children were excluded from the TD group for any of the following reasons: medical or educational diagnoses, presence of language impairment, or their first language was not English. Normal hearing was confirmed in all children using a brief pure tone audiometry screener based on the guidelines from the American Academy of Audiology (Bright et al., 2011) with pure tones at 1000 Hz, 2000 Hz, and 4000 Hz. All participants were found to have normal IQ (M = 96.4, SD = 10.7) as assessed by the abbreviated IQ measure from the Stanford Binet-5 (Roid, 2003).

Stimuli & Task Design

Speech and nonspeech stimuli were presented in two passive oddball tasks in which infrequent "deviant" sounds were presented within a stream of frequent "standard" sounds. In the Congruent Oddball Task, the sound category of the standard and deviant sounds in each block was the same, such that a standard and deviant speech sound were presented in one block and a standard and deviant cow sound were presented in another block. In the Incongruent Oddball Task, speech and nonspeech stimuli were interspersed together within the same block, such that in one block speech was the standard sound and nonspeech was the deviant sound, and vise versa in the other block. The standard speech and nonspeech sounds were kept the same across tasks. In contrast, the deviant sounds changed across tasks to create the congruent vs. incongruent conditions. This design allowed for the direct comparison of deviant type (speech, nonspeech) as well as standard type (speech, nonspeech) across tasks.

To ensure novelty of the deviant sounds in both oddball tasks, at least three standards always preceded a deviant sound. Within one trial, six sounds were presented, and the deviant sound occurred randomly in either the 4th, 5th, or 6th position. Across both tasks, sounds were presented with a varying interstimulus interval (ISI) between 700-900 ms to prevent habituation to stimulus onset. For both tasks, participants were seated 120 cm from a TV monitor displaying a silent movie (Wall-E) while auditory stimuli were presented at approximately 70 dB from two loudspeakers located on each side of the monitor. All sound stimuli were 350 ms in duration and edited to have a rise and fall time of 10 ms using a custom-written MATLAB script (The Mathworks, Inc). Stimuli were downsampled to a sampling rate of 44.1 kHz with a quantization level of 16-bit. All stimuli were matched on loudness based on their average root mean square (RMS) using Adobe Audition (Adobe Systems, Inc).

The speech and nonspeech stimuli were adapted from a previous study in our lab (Whitten et al., under review). This study created a novel oddball paradigm using natural speech and nonspeech stimuli by matching vowel stimuli to cow vocalizations. Cow vocalizations provided a unique opportunity to match a natural ecologically valid nonspeech stimulus to the spectro-temporal complexity of speech. This was possible because cow vocalizations are generated with the same underlying mechanisms as human speech, i.e., laryngeal vocal fold vibration and filtering via a “vocal tract.” The process of matching the vowel and cow sound stimuli has been described in detail in Whitten et al. (2019). Briefly, stimuli were matched on overall complexity by matching two dynamic spectro-temporal aspects, namely by matching the changes in amplitude within 1 dB and the changes in fundamental frequency within 10 Hz across the entire stimulus duration. Pearson correlations on acoustic measures between stimuli resulted in significant correlations with coefficients ranging from 0.937 to 0.995.

Our previous task presented two different pairs of natural speech and nonspeech stimuli interspersed as standard and deviant sounds, similar to the design of the Incongruent Oddball Task in the present study. Therefore, it was necessary to adapt the stimuli to create stimuli that could be used in both the Congruent Oddball Task and an Incongruent Oddball Task. Previous studies presented either a vowel or syllable as the standard sound with a novel vowel/syllable as the deviant, and a standard complex tone with a different deviant complex tone. Thus, a logical extension in the present study would have been to create a speech block with two natural vowels and a nonspeech block with two cow sounds. However, we reasoned that the change from one vowel to a different vowel may be more salient than the change from one cow sound to another cow sound and consequently could bias the P3a attentional response. This is because changing from one vowel to another vowel is likely more meaningful given the importance of discriminating different phonemes in language processing. Furthermore, it would be difficult to match the acoustic difference between two vowels and two cow vocalizations. While humans vary only fundamental frequency and formants between two vowels, we found through acoustic analysis of our previous dataset that cow vocalizations which vary in formants also vary the amount of noise and harshness in the signal.

Congruent Oddball Task

To ensure that the acoustic difference between standards and deviants did not differ across speech and nonspeech blocks, and that one deviant stimulus did not potentially convey more meaning than the other in the Congruent Oddball Task (i.e., vowel change vs. cow change), we decided to use an equivalent pitch change across blocks. Specifically, to create the stimuli for the Congruent Oddball Task, we edited the stimuli from our previous study using Adobe Audition to create vowel and cow sound stimuli that increased in mean pitch (F0) by 60% (7.2 semitones) from the standard to deviant sound. This pitch change represented the largest difference between speech and nonspeech stimuli that still maintained the natural quality of the original stimuli (i.e., no distortions, still recognizable as cow sound and /i/ vowel spoken by female). For the Congruent Oddball Task, this resulted in a congruent speech block that presented a standard low pitch /i/ (mean F0: 224.4 Hz) and deviant high pitch /i/ (mean F0: 359.2 Hz), and a congruent nonspeech block that presented a standard low pitch cow (mean F0: 87.5 Hz) and a deviant high pitch cow (mean F0: 140.4 Hz).

Incongruent Oddball Task

In the Incongruent Oddball Task, it was crucial to use the same standard sounds as the Congruent Oddball Task, and intersperse the speech and nonspeech deviant sounds. However, it would not be possible to keep the same standard sounds if we wanted to also equate the speech and nonspeech stimuli on pitch within each block. Furthermore, because the cow stimuli have a lower pitch than the female speaker, this would result in one block where the pitch rose from standard to deviant, and one block where the pitch fell. Therefore, in order to equate the acoustic difference between standards and deviants across blocks within the Incongruent Oddball Task, the deviant and standard sounds were simply swapped across blocks. For example, in the incongruent speech block, the standard sound was the low pitch /i/ vowel and the deviant sound was the low pitch cow sound. In the incongruent nonspeech block it was the opposite, in which the standard sound was the low pitch cow sound and the deviant sound was the low pitch /i/ vowel.

The Congruent Oddball Task was presented first in all participants. This was a result of concurrent data collection across a variety of severity levels of children with autism spectrum disorder in which we wanted to ensure that all children received the same task first because the majority would not be able to complete the full task battery. However, within each task the block order was counterbalanced across participant. See Figure 3.1 for a schematic illustration of both tasks.

Data Analysis

Neural responses were recorded by Net Station 5.4 software (EGI, Inc., Eugene, OR, USA) and stimulus presentation was controlled by E-Prime 2.0 (PST, Inc., Pittsburgh, PA, USA). Participants wore high-density hydrocel geodesic EEG nets with 128 electrodes (EGI, Inc., Eugene, OR, USA). Electrode impedances were kept below 50 kΩ. The EEG was sampled at 500 Hz and filtered offline using a 0.1 – 30 Hz bandpass filter. During data collection the signal was referenced to Cz, and later re-referenced to the average of all electrodes following artifact rejection.

ERP epochs were segmented from the continuous EEG data beginning 100 ms before (pre-stimulus baseline) and 600 ms after stimulus onset to all standard sounds that preceded a deviant sound, and both deviant sounds. The resulting epochs were screened for artifacts using an automated Net Station algorithm followed by manual review. Trials containing eye blinks, lateral eye movements, muscular activity, or electrical noise were excluded from analysis. Individual

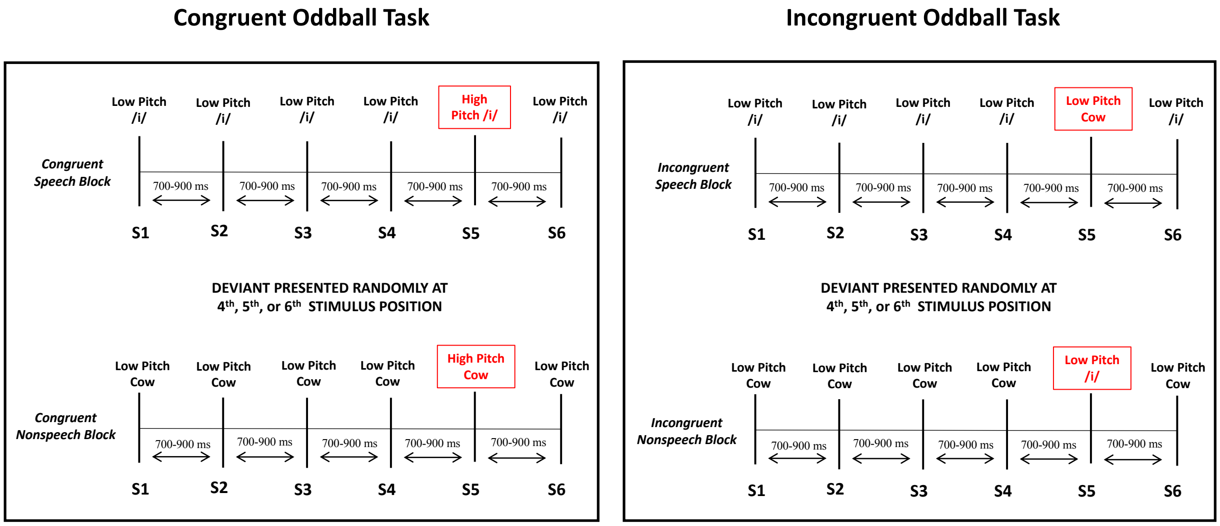


Figure 3.1 Design of Congruent and Incongruent Oddball Tasks.

electrodes with poor signal quality within a trial were corrected using spherical spline interpolation; no more than 10% of electrodes were interpolated on any one trial.

Following artifact rejection, the remaining trials for each condition were averaged, re-referenced to the average reference, and baseline-corrected. In order to maximize signal-to-noise ratio, ERP amplitudes and latencies were measured as the average of a cluster of electrodes rather than a single electrode. The strongest auditory P1 and N2 responses were found at a cluster that corresponded to Fz (electrodes 4, 5, 10, 11, 12, 16, 18, 19). P1 and N2 latency and amplitude windows were chosen based on the examination of the grand-averaged waveform and further confirmed within each individual participant's ERPs to ensure that each individual displayed the ERP of interest within the time window. This resulted in the time windows of 50-150 ms for P1 and 200-300 ms for N2. Because such wide windows were necessary to encompass P1 and N2 ERPs in every subject, amplitude was measured as peak amplitude rather than mean amplitude as this would lead to averaging over the entire window that is not representative of the peak of interest. Peak latency and peak amplitude were measured at the local peak within the time window. We defined a local peak as the time-point in which at least 10 data points (20 ms) on either side of the peak were reduced in amplitude. This technique is favorable over simple peak measures (which identify the most positive or negative point within the window) because it is less influenced by noise (Luck, 2014).

The strongest P3a response across participant groups was again found at a cluster corresponding to Fz (electrodes 4, 5, 10, 11, 12, 16, 18, 19). The P3a time window was chosen based on the grand-average waveforms of all groups and selected as 300-400 ms. P3a amplitude was calculated as the mean amplitude between 300-400 ms of the difference wave calculated by subtracting the standard sound from the deviant sound for all four deviant sounds (congruent speech deviant, congruent nonspeech deviant, incongruent speech deviant, incongruent nonspeech deviant) from both tasks.

P1 and N2 peak latency values to the four deviant speech and nonspeech stimuli from both tasks (congruent speech deviant, congruent nonspeech deviant, incongruent speech deviant, incongruent nonspeech deviant) were compared in a 2 x 2 x 2 repeated measures ANOVA where the within-subject factors were deviant type (speech, nonspeech), standard type (speech, nonspeech) and ERP component (P1, N2). ERP Component was included for latency to test whether P1 and N2 latency may be impacted differentially by deviant and standard type, however a significant main effect of ERP latency was ignored due to being measured at different time points. Because P1 and N2 have opposite polarity, peak amplitude values were analyzed separately for each component in a 2 x 2 repeated measures ANOVA where the within-subject factors were deviant type (speech, nonspeech) and standard type (speech, nonspeech). P3a latency and mean amplitude values were compared in a 2 x 2 repeated measures ANOVA where the within-subjects factors were deviant type (speech, nonspeech) and standard type (speech, nonspeech). Estimated marginal means (EMM) and standard errors (SE) are presented for all significant effects.

Results

Sensory Responses: P1 and N2

For P1 and N2 latency, there was no significant main effect of deviant type (speech vs. nonspeech; *F*(1, 15) = 3.26, *p* = .09, partial eta squared = .18) and no significant main effect of standard type (speech, nonspeech; *F*(1, 15) = 4.11, *p* = .06, partial eta squared = .22). However, there was a significant deviant type x standard type interaction effect (*F*(1, 15) = 6.26, *p* = .02, partial eta squared = .30). Post hoc tests on the interaction revealed that this was due to the speech deviant latency remaining stable regardless of whether it followed a speech or nonspeech standard sound, while this was not true for the nonspeech deviants. Specifically, the nonspeech deviant sound showed faster P1 and N2 latency when it was preceded by a speech standard sound (EMM = 158.81, SE = 4.95) than a nonspeech standard sound (EMM = 178.44, SE = 4.32, *p* = .004). P1 and N2 latency to the speech deviant following a speech standard sound (EMM = 160.31 ms, SE = 3.89 ms) was similar to the speech deviant following a nonspeech standard sound (EMM = 158.94, SE = 4.47), and thus the slowed latency of the nonspeech deviant following a nonspeech standard was significantly slower than the speech deviant following a nonspeech standard (*p* = .01). The deviant type x standard type interaction is shown in Figure 3.2.

For both P1 and N2 amplitude, there was a significant main effect of deviant type (P1: *F*(1, 15) = 4.68, *p* = .047, partial eta squared = .24, N2: *F*(1, 15) = 8.17, *p* = .01, partial eta squared = .35). P1 amplitude was significantly larger to nonspeech deviants across tasks (EMM = 3.12, SE = 0.44) than to speech deviants (EMM = 1.70, SE = 0.52), while N2 amplitude was significantly larger to speech deviants (EMM = -4.93, SE = 0.62) than nonspeech deviants (EMM = -3.52, SE = 0.68). There was no main effect of standard type on P1 amplitude (*F*(1, 15) = 0.24, *p* = .63, partial eta squared = .02) or N2 amplitude (*F*(1, 15) = 1.83, *p* = .20, partial eta squared = .11). There was also no significant deviant type x standard type interaction for P1 amplitude (*F*(1, 15) = 0.04, *p* = .84, partial eta squared = .003) or N2 amplitude (*F*(1, 15) = 0.55, *p* = .47, partial eta squared = .04).

Orienting Response: P3a. P3a latency showed no main effect of deviant type (*F*(1, 15) = 0.07, *p* = .79, partial eta squared = .005), but a significant main effect of standard type (*F*(1, 15) = 8.72, *p* = .01, partial eta squared = .37). This was driven by a significantly faster P3a latency for

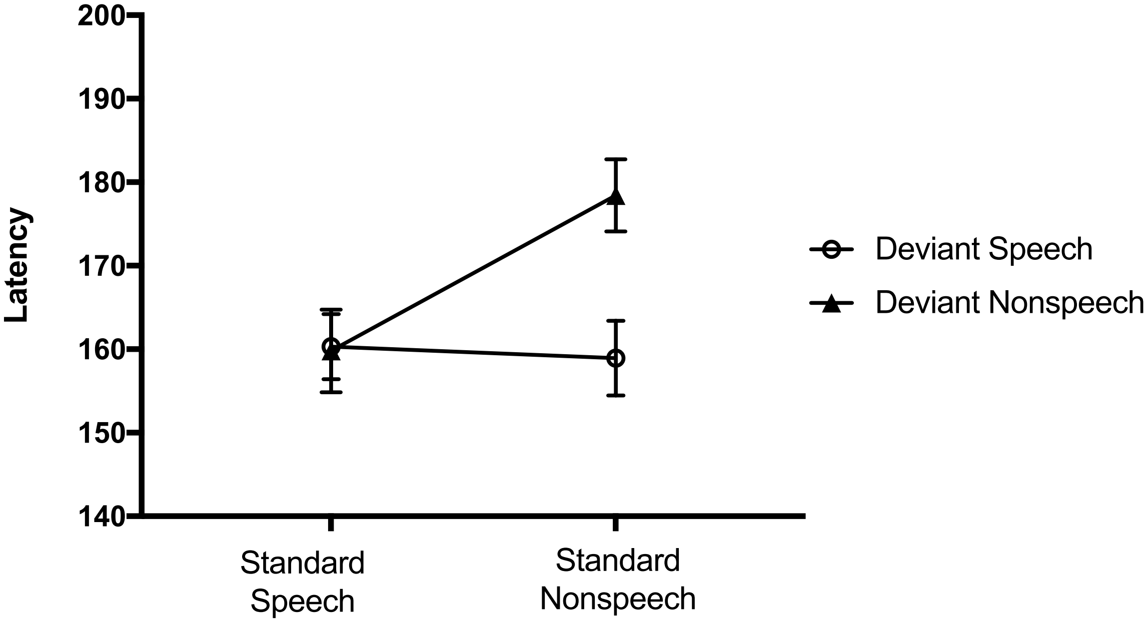


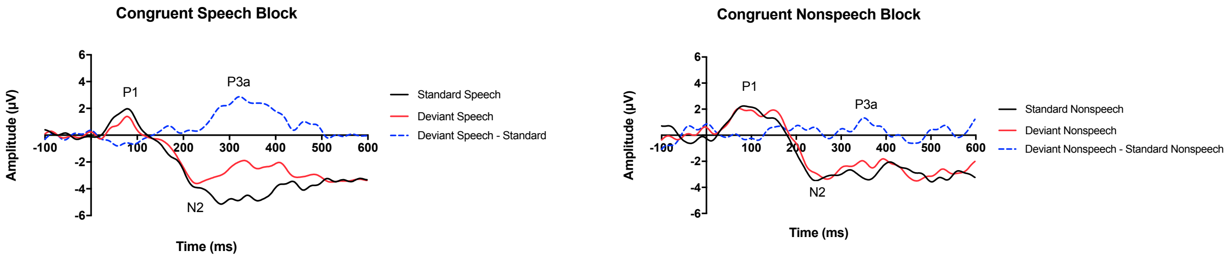
Figure 3.2 Standard Type x Deviant Type Interaction for P1 and N2 Latency. The type of standard sound presented (speech, nonspeech) is displayed on the X axis, while the Y axis shows the average of the P1 and N2 latencies to the four deviant sounds. The open circles reflect the deviant speech sounds presented after standard speech (Congruent Oddball Task) and standard nonspeech (Incongruent Oddball Task). Similarly, the dark triangles reflect the deviant nonspeech sounds presented after standard speech (Incongruent Oddball Task) and standard nonspeech (Congruent Oddball Task).

deviants following standard speech sounds (EMM = 336.69, SE = 6.57) compared to standard nonspeech sounds (EMM = 363.31, SE = 5.01, *p* = .01). There was no significant deviant type x standard type interaction (*F*(1, 15) = 2.14, *p* = .16, partial eta squared = .13).

For P3a mean amplitude, there was no significant main effect of deviant type (*F*(1, 15) = 2.55, *p* = .13, partial eta squared = .15). However, similar to P3a latency there was a significant main effect of standard type (*F*(1, 15) = 5.80, *p* = .03, partial eta squared = .28), in which deviant sounds that followed standard speech sounds showed significantly larger P3a amplitude (EMM = 3.25, SE = 0.67) than deviants that followed standard nonspeech sounds (EMM = 0.09, SE = 0.87, *p* = .03). There was also no significant deviant type x standard type interaction for P3a amplitude (*F*(1, 15) = 0.11, *p* = .75, partial eta squared = .007).

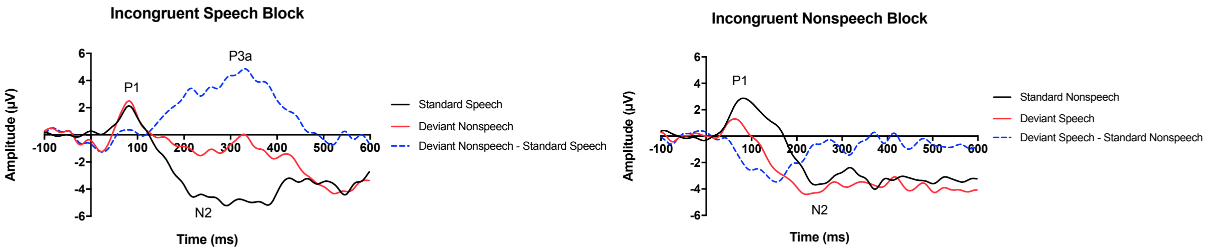
Figure 3.3 presents the P1, N2, and P3a results for all four blocks across oddball tasks. The P3a is reflected by a greater positivity to the deviant sound compared to the standard sound it was paired with, as well as a large positivity in the difference wave; both are presented because only presenting the difference wave can be misleading (Picton et al., 2000). In addition, Figure 3.4 presents the deviant stimuli that followed the standard speech sounds vs. standard nonspeech sounds in order to display the effect of the standard sounds.

Congruent Oddball Task


Congruent Speech Block

Congruent Nonspeech Block

Incongruent Oddball Task


Incongruent Speech Block

Incongruent Nonspeech Block

Figure 3.3 P1, N2, and P3a Results in the Congruent and Incongruent Oddball Tasks. P1 and N2 are labeled to the standard and deviant sounds. The P3a is reflected by a greater positivity to the deviant sound compared to the standard sound it was paired with, as well as a large positivity in the difference wave (dashed line). No P3a is seen in the incongruent nonspeech block.

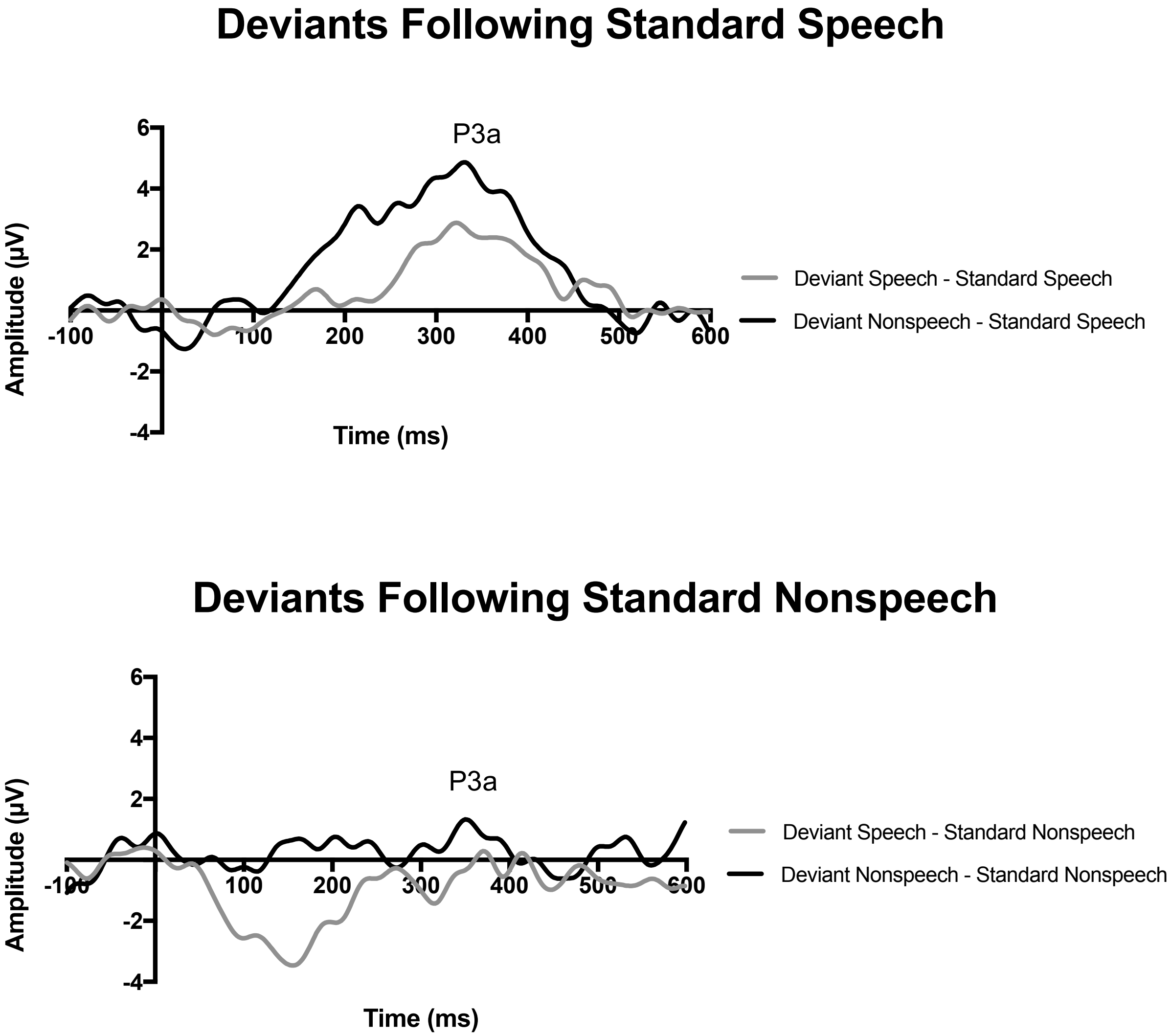


Figure 3.4 Deviant P3a Responses Grouped by Standard Sound. Top: The difference waves to the deviant speech and deviant nonspeech sounds that followed standard speech are presented, gray: congruent speech block, black: incongruent speech block. Bottom: The difference waves to the deviant speech and deviant nonspeech sounds that followed standard nonspeech are presented, gray: incongruent nonspeech block, black: congruent nonspeech block

Discussion

The goal of the present study was to examine whether the type of standard and deviant sounds used in an oddball paradigm affects the P3a orienting response, and whether this may start during the earliest stages of cortical auditory processing (P1 and N2) in 8—14-year-old children. The motivation for this study was based on the hypothesis that the relative importance of the speech signal may lead to earlier and greater auditory processing and/or orienting responses compared to nonspeech sounds, both when speech is a deviant sound and when it is the standard sound. We tested this hypothesis by using two tasks. In the Congruent Oddball Task, a congruent speech block and congruent nonspeech block were presented with standard and deviant sounds that were the same category in each block (speech or nonspeech). In the Incongruent Oddball Task, the same standard sounds from the Congruent Oddball Task were presented, but in the incongruent speech block the deviant sound was the low pitch cow, and in the incongruent nonspeech block the deviant sound was the low pitch vowel. This design allowed us to compare across tasks in a Standard Type x Deviant Type repeated measures ANOVA for P1, N2, and P3a amplitude and latency to the four deviant sounds (congruent speech deviant, congruent nonspeech deviant, incongruent speech deviant, incongruent nonspeech deviant).

We found that P3a latency was significantly faster to both deviant types (speech or nonspeech) that followed a standard speech sound, and similarly P3a amplitude was significantly larger to deviants that followed a standard speech sound. In fact, Figures 3.3 and 3.4 show little or no evidence of a P3a response to either the speech or nonspeech deviant that followed a standard nonspeech stream (evidenced by a greater positivity to the deviant sound compared to the standard sound). This finding was likely related to early cortical auditory processing, as we found that while P1 and N2 amplitude did not change based on the standard sound, P1 and N2 speed of processing showed a significant interaction between the deviant type and standard type. While auditory processing speed to speech deviants remained stable, the nonspeech deviant received a “boost” in latency only after following a standard stream of speech sounds, but was slower following standard nonspeech sounds.

Our results emphasize the importance of the sound context for both early sensory and attentional processing, and suggest that a stream of speech sounds may prime the auditory system to “tune in” at some level to encode the acoustic features of any upcoming stimulus faster (P1, N2), as well as orient to it faster or at all (P3a). This is consistent with evidence that speech processing engages a more efficient network than nonspeech; Dehaene-Lambertz et al. (2005) found that when participants perceived sinewave analogues as speech, there was a significantly faster discrimination response in “speech mode” than when the same participants perceived the same signals as nonspeech. This is likely related to expertise with the speech signal that results in more efficient processing, but it could also be due to the extremely fast fluctuations of speech (e.g., only ~10 ms to discriminate differences in voice onset time) that necessitates a system that is primed to be ready for the next meaningful acoustic change. This may be accomplished through enhanced cortical tracking of speech sounds relative to other sounds. Golumbic et al. (2013) found that increasing attention modulated the amount of cortical tracking to speech streams, but even the ignored speech streams were still represented in lower levels of auditory processing, which could be serving as a way to ensure that at some level speech is monitored and the system is “primed” to be ready to quickly encode upcoming messages.

The enhanced sensory and attentional responses following standard speech compared to nonspeech sounds also suggest a more nuanced model of novelty sound orientation. It has traditionally been assumed that even in the context of an ignored stream, any sound that deviates strongly enough from a repeating sound will engage bottom-up sensory working memory processes that then prompt a top-down P3a attentional response (Polich, 1989, 2003, 2007). Conversely, we found that speech and nonspeech stimuli that differ by exactly the same amount (7.2 semitones) do not always produce the same P3a response, which suggests that streams of speech sounds may differentially engage top-down attentional resources before the evaluation of a deviant sound. The effect of speech vs. nonspeech streams has implications for studies comparing speech or human vocalizations to other environmental stimuli, which have often interspersed sounds together, similar to our Incongruent Oddball Task (Bidet-Caulet et al., 2017; Bruneau & Gomot, 1998; Lucia, Clarke, & Murray, 2010; Rogier, Roux, Belin, Bonnet-Brilhault, & Bruneau, 2010). Indeed, our findings could relate to the stronger right-hemisphere response to sounds that followed the human voice found by Renvall et al. (2012).

One limitation of the present study is that we did not counterbalance the order of the Congruent and Incongruent Oddball Tasks, and thus every participant received the Incongruent Oddball Task second. However, it is unlikely that potential fatigue to the second task can explain the findings. This is because we did counterbalance the order of the incongruent speech and nonspeech block across participants, and therefore if participants were too fatigued then we should not have seen a strong P3a response in the incongruent speech condition.

A second limitation is that we did not measure the mismatch negativity component (MMN) that indicates discrimination between the standard and deviant sound before attentional orienting (Näätänen, Paavilainen, Rinne, & Alho, 2007). We were unable to measure this component due to the requirement of a large number of trials (>150), and we aimed to create a short passive task that could be used in clinical studies. Therefore, it will be important for a future study to investigate whether the lack of a P3a response to deviants following nonspeech standards is related to difficulty discriminating; however, because the blocks in our study were matched acoustically, discrimination would not have been more difficult. Thus, if the MMN is reduced or absent in these blocks, it is likely also reflective of top-down inhibition that starts as early as sensory processing, and could suggest that the MMN is also modulated by the type of standard sound.

In conclusion, we found that the sound context impacts the orienting response, and our results highlight the importance of a repeating speech sound compared to an ecologically valid nonspeech sound. Future work could include children younger than 8 years in order to assess whether the impact of standard and deviant sounds on auditory processing and attention develops over time as language experience increases, or if it is present as early as infancy. Similarly, this paradigm could be used in children with language disorders and children with autism to determine if atypical standard and deviant speech sound processing is associated with language impairments.

References

Alho, K., Escera, C., Díaz, R., Yago, E., & Serra, J. M. (1997). Effects of involuntary auditory attention on visual task performance and brain activity. *NeuroReport*, *8*(15), 3233.

Berti, S., Roeber, U., & Schröger, E. (2004). Bottom-Up Influences on Working Memory: Behavioral and Electrophysiological Distraction Varies with Distractor Strength. *Experimental Psychology*, *51*(4), 249–257. https://doi.org/10.1027/1618-3169.51.4.249

Berti, S., & Schröger, E. (2003). Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. *European Journal of Neuroscience*, *17*(5), 1119–1122. https://doi.org/10.1046/j.1460-9568.2003.02527.x

Bidet-Caulet, A., Latinus, M., Roux, S., Malvy, J., Bonnet-Brilhault, F., & Bruneau, N. (2017). Atypical sound discrimination in children with ASD as indicated by cortical ERPs. *Journal of Neurodevelopmental Disorders*, *9*(1), 13.

Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F. E., & Linden, D. E. J. (2004). Attentional systems in target and distractor processing: a combined ERP and fMRI study. *NeuroImage*, *22*(2), 530–540. https://doi.org/10.1016/j.neuroimage.2003.12.034

Bright, K., Greeley, C. O., Eichwald, J., Loveland, C. O., & Tanner, G. (2011). American Academy of Audiology Childhood Hearing Screening Guidelines.

Bruneau, N., & Gomot, M. (1998). Auditory evoked potentials (N1 wave) as indices of cortical development. In *Neuroimaging in child neuropsychiatric disorders* (pp. 113–123). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-95848-9_10

Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced Passive and Active Processing of Syllables in Musician Children. *Journal of Cognitive Neuroscience*, *23*(12), 3874–3887. https://doi.org/10.1162/jocn_a_00088

Comerchero, M. D., & Polich, J. (1998). P3a, perceptual distinctiveness, and stimulus modality. *Cognitive Brain Research*, *7*(1), 41–48. https://doi.org/10.1016/S0926-6410(98)00009-3

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*(1), 21–33. https://doi.org/10.1016/j.neuroimage.2004.09.039

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In *Human auditory development* (pp. 61–105). Springer.

Fischer, C., Dailler, F., & Morlet, D. (2008). Novelty P3 elicited by the subject's own name in comatose patients. *Clinical Neurophysiology*, *119*(10), 2224–2230. https://doi.org/10.1016/j.clinph.2008.03.035

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373.

Hämäläinen, J. A., Leppänen, P. H. T., Guttorm, T. K., & Lyytinen, H. (2008). Event-related potentials to pitch and rise time change in children with reading disabilities and typically reading children. *Clinical Neurophysiology*, *119*(1), 100–115. https://doi.org/10.1016/j.clinph.2007.09.064

Harmony, T., Bernal, J., Fernández, T., Silva-Pereyra, J., Fernández-Bouzas, A., Marosi, E., … Reyes, A. (2000). Primary task demands modulate P3a amplitude. *Cognitive Brain Research*, *9*(1), 53–60. https://doi.org/10.1016/S0926-6410(99)00044-0

Hermens, D. F., Ward, P. B., Hodge, M. A. R., Kaur, M., Naismith, S. L., & Hickie, I. B. (2010). Impaired MMN/P3a complex in first-episode psychosis: Cognitive and psychosocial associations. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *34*(6), 822–829. https://doi.org/10.1016/j.pnpbp.2010.03.019

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M., & Alho, K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Cognitive Brain Research*, *12*(3), 459–466.

Katayama, J., & Polich, J. (1998). Stimulus context determines P3a and P3b. *Psychophysiology*, *35*(1), 23–33. https://doi.org/10.1111/1469-8986.3510023

Lakatos, P., Barczak, A., Neymotin, S. A., McGinnis, T., Ross, D., Javitt, D. C., & O'Connell, M. N. (2016). Global dynamics of selective attention and its lapses in primary auditory cortex. *Nature Neuroscience*, *19*(12), 1707–1717. https://doi.org/10.1038/nn.4386

Lincoln, A. J., Courchesne, E., Harms, L., & Allen, M. (1993). Contextual probability evaluation in autistic, receptive developmental language disorder, and control children: event-related brain potential evidence. *Journal of Autism and Developmental Disorders*, *23*(1), 37–58.

Lucia, M. D., Clarke, S., & Murray, M. M. (2010). A Temporal Hierarchy for Conspecific Vocalization Discrimination in Humans. *The Journal of Neuroscience*, *30*(33), 11210–11221. https://doi.org/10.1523/JNEUROSCI.2239-10.2010

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press. Retrieved from https://books.google.com/books?hl=en&lr=&id=y4-uAwAAQBAJ&oi=fnd&pg=PR5&dq=An+Introduction+to+the+Event-Related+Potential+Technique&ots=jA36D2RPIz&sig=nOoNKrH1EFySoIxVZBtSlUAD0s8

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, *7*(6), 658–663. https://doi.org/10.1038/nn1256

Picton, T. w., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a., Johnson, R., … Taylor, M. j. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(02), 127–152. https://doi.org/null

Polich, J. (1989). P300 from a passive auditory paradigm. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *74*(4), 312–320. https://doi.org/10.1016/0168-5597(89)90061-0

Polich, J. (2003). *Theoretical overview of P3a and P3b*. Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-0294-4_5

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology*, *113*(3), 407–420. https://doi.org/10.1016/S1388-2457(01)00733-7

Reiche, M., Hartwigsen, G., Widmann, A., Saur, D., Schröger, E., & Bendixen, A. (2013). Involuntary attentional capture by speech and non-speech deviations: A combined behavioral–event-related potential study. *Brain Research*, *1490*, 153–160. https://doi.org/10.1016/j.brainres.2012.10.055

Renvall, H., Staeren, N., Siep, N., Esposito, F., Jensen, O., & Formisano, E. (2012). Of cats and women: Temporal dynamics in the right temporoparietal cortex reflect auditory categorical processing of vocalizations. *NeuroImage*, *62*(3), 1877–1883. https://doi.org/10.1016/j.neuroimage.2012.06.010

Rogenmoser, L., Elmer, S., & Jäncke, L. (2014). Absolute Pitch: Evidence for Early Cognitive Facilitation during Passive Listening as Revealed by Reduced P3a Amplitudes. *Journal of Cognitive Neuroscience*, *27*(3), 623–637. https://doi.org/10.1162/jocn_a_00708

Rogier, O., Roux, S., Belin, P., Bonnet-Brilhault, F., & Bruneau, N. (2010). An electrophysiological correlate of voice processing in 4- to 5-year-old children. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *75*(1), 44–47. https://doi.org/10.1016/j.ijpsycho.2009.10.013

Roid, G. H. (2003). Stanford–Binet Intelligence Scale, Fifth Edition Itasca, IL: Riverside Publishing.

Schröger, E., & Wolff, C. (1998). Behavioral and electrophysiological effects of task-irrelevant sound change: a new distraction paradigm. *Cognitive Brain Research*, *7*(1), 71–87. https://doi.org/10.1016/S0926-6410(98)00013-5

Seppänen, M., Pesonen, A.-K., & Tervaniemi, M. (2012). Music training enhances the rapid plasticity of P3a/P3b event-related brain potentials for unattended and attended target sounds. *Attention, Perception, & Psychophysics*, *74*(3), 600–612. https://doi.org/10.3758/s13414-011-0257-9

Shestakova, A., Huotilainen, M., Ceponiene, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *114*(8), 1507–1512.

Snyder, J. S., Holder, W. T., Weintraub, D. M., Carter, O. L., & Alain, C. (2009). Effects of prior stimulus and prior perception on neural correlates of auditory stream segregation. *Psychophysiology*, *46*(6), 1208–1215. https://doi.org/10.1111/j.1469-8986.2009.00870.x

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

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.

Sussman, E. S., Horváth, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streams. *Perception & Psychophysics*, *69*(1), 136–152. https://doi.org/10.3758/BF03194460

Volpe, U., Mucci, A., Bucci, P., Merlotti, E., Galderisi, S., & Maj, M. (2007). The cortical generators of P3a and P3b: A LORETA study. *Brain Research Bulletin*, *73*(4), 220–230. https://doi.org/10.1016/j.brainresbull.2007.03.003

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of Sounds in the Auditory Stream: Event-Related fMRI Evidence for Differential Activation to Speech and Nonspeech. *Journal of Cognitive Neuroscience*, *13*(7), 994–1005. https://doi.org/10.1162/089892901753165890

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

Whitehouse, A. J., & Bishop, D. V. (2008). Do children with autism 'switch off'to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524.

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences*, *90*(18), 8722–8726. https://doi.org/10.1073/pnas.90.18.8722

Wronka, E., Kaiser, J., & Coenen, A. M. (2008). The auditory P3 from passive and active three-stimulus oddball paradigm. *Acta Neurobiologiae Experimentalis*, *68*(3), 362–372.

Wronka, E., Kaiser, J., & Coenen, A. M. L. (2012). Neural generators of the auditory evoked potential components P3a and P3b. *Acta Neurobiologiae Experimentalis*, *72*(1), 51–64.

Yang, M.-T., Hsu, C.-H., Yeh, P.-W., Lee, W.-T., Liang, J.-S., Fu, W.-M., & Lee, C.-Y. (2015). Attention deficits revealed by passive auditory change detection for pure tones and lexical tones in ADHD children. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00470

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Schroeder, C. E. (2013). Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a “Cocktail Party.” *Neuron*, *77*(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Zuijen, T. L. van, Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2004). Grouping of Sequential Sounds—An Event-Related Potential Study Comparing Musicians and Nonmusicians. *Journal of Cognitive Neuroscience*, *16*(2), 331–338. https://doi.org/10.1162/089892904322984607

CHAPTER 4

Distinct Event-Related Potential Profiles Characterize Language Subgroups in Autism

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by social communication impairment and the presence of repetitive and restricted behaviors. Deficits in social communication include impaired use of both verbal language and nonverbal gestures to initiate and respond during social interactions. There is a general consensus that pragmatic language, or the social use of language, is universally impaired in ASD (Groen, Zwiers, van der Gaag, & Buitelaar, 2008; Tager-Flusberg, 1981; Wilkinson, 1998). These pragmatic deficits include impaired perception and production of prosody (Paul, Augustyn, Klin, & Volkmar, 2005; Rutherford, Baron-Cohen, & Wheelwright, 2002; Shriberg et al., 2001), use of idiosyncratic language (Loveland, McEvoy, Tunali, & Kelley, 1990; Volden & Lord, 1991), perseveration on specific topics (Baltaxe, 1977), misinterpretation of literal language (D. V. M. Bishop & Adams, 1989), and difficulties with conversational discourse (Capps, Kehres, & Sigman, 1998; Hale & Tager-Flusberg, 2005). Traditionally, structural aspects of language relating to grammar and syntax were thought to be intact in individuals with ASD, however, more recent studies have challenged this claim and provide support for a variety of grammatical language deficits in children with ASD (Boucher, 2012; Brynskov et al., 2017).

Despite widespread evidence for language impairments in ASD, the mechanisms underlying these deficits remain poorly understood. One explanation that has been used to describe a variety of autistic behaviors is the social motivation theory. Social motivation can be defined as the combination of innate biological factors that interact to create a preference for the social world (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012). In relation to language impairment, the social motivation theory posits that children with autism are not motivated to orient and attend to social stimuli like human speech, and this in turn constrains their ability to learn language. There is evidence to suggest that this may be the case; while studies assessing preference for speech over nonspeech sounds have found that typically developing children prefer to listen to speech as early as the first few days of life (Vouloumanos & Werker, 2007), studies in children with autism find a preference for nonspeech sounds instead (Klin, 1991; Kuhl, Coffey-Corina, Padden, & Dawson, 2005). In addition, Abrams et al. (2013) found that children with ASD showed significantly reduced connectivity between the superior temporal sulcus (STS), a region implicated in voice and speech processing, and nodes of the dopaminergic reward circuitry. Yet, they found normal connectivity between reward circuitry and the primary auditory cortex, suggesting there may be decreased motivation to attend to speech but not other sounds. Furthermore, the degree of underconnectivity between the STS and reward processing areas in individual participants with ASD was predictive of more severe communication deficits, suggesting a direct link between abnormal social reward processing and language deficits.

One important caveat of the social motivation hypothesis is its assumption that children with autism have the ability to process linguistic information successfully, but they fail to do so only because they are not motivated to allocate neural resources to attending. This explanation does not take into account any sensory processing deficits that could lead to downstream difficulties in learning, storing, and comprehending speech. In contrast, the auditory processing theory suggests that deficits in the sensory processing of sound leads to language impairments.

This theory is supported by behavioral studies in ASD that have found deficits in the processing of prosody (Golan, Baron-Cohen, & Hill, 2006; Kleinman, Marciano, & Ault, 2001; Rutherford et al., 2002), speech-in-noise (Alcántara, Weisblatt, Moore, & Bolton, 2004; Groen et al., 2009), and hypersensitivity to sounds (Khalfa et al., 2004; see O'Connor, 2012 for a full review). Furthermore, there is evidence that some individuals with ASD may have an enhanced ability to process simple acoustic features like pitch at the expense of more complex processing like those involved in speech, which require extracting simultaneous spectral and temporal content (Mottron, Dawson, Soulières, Hubert, & Burack, 2006).

An optimal way to study auditory processing and social motivation as potential factors underlying language impairment in ASD is to measure their neural correlates in the brain. The temporal resolution of EEG allows researchers to measure event-related potentials (ERPs) that can distinguish between early sensory and later cognitive cortical processes. In children, the P1 and N2 are the dominant auditory ERP responses to sound, with P1 peaking around 100 ms and N2 peaking around 250 ms in frontocentral locations after stimulus onset (Ceponiene, Rinne, & Näätänen, 2002; Sussman, Steinschneider, Gumenyuk, Grushko, & Lawson, 2008), although children show mature adult-like N1 and P2 ERP responses with interstimulus intervals longer than 1 second (Čeponien, Rinne, & Näätänen, 2002; Gilley, Sharma, Dorman, & Martin, 2005; Paetau, Ahonen, Salonen, & Sams, 1995). The children's P1 and N2 are thought to index cortical encoding of acoustic properties of sound using the same generators as adults in both primary and secondary auditory cortices (Eggermont & Moore, 2012; Ponton, Eggermont, Khosla, Kwong, & Don, 2002). As such, these ERP components can be used to test the auditory processing theory of language impairment in children with ASD. Following sensory processing reflected by the P1 and N2 components, the P3a component represents an "orienting response" related to involuntary attention switching and initial allocation of attention to a salient stimulus (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2003; Squires, Squires, & Hillyard, 1975). The P3a latency occurs between approximately 300 – 400 ms, and is thought to originate in the anterior cingulate cortex through dopaminergic neurotransmitter activity (Polich & Criado, 2006), although recent studies also suggest a primary role of acetylcholine as well (Caldenhove, Borghans, Blokland, & Sambeth, 2017). The P3a can thus be used as a proxy for social motivation to detect whether social stimuli are deemed salient enough to warrant an attention switch.

Several previous studies have compared the early auditory (P1, N2) and P3a ERP components to speech and nonspeech sounds using a passive oddball task in children with autism. The passive oddball task presents a string of standard sounds that are repeated on 80-85% of trials, and a deviant sound presented on 10-15% of trials. The standard and deviant sounds both elicit the auditory ERP components reflecting the first stages of cortical auditory processing, while the deviant sound also elicits the P3a component if the stimulus is salient enough to trigger an attention switch or orienting response to the stimulus. The majority of previous studies have presented vowel or consonant-vowel stimuli as the speech stimuli in one block, and complex tones as the nonspeech stimuli in another block. These studies typically find no differences in the auditory ERP latencies compared to typically developing (TD) children, suggesting normal speed of processing to both speech and nonspeech stimuli (Ceponiene et al., 2003; Gomot, Giard, Adrien, Barthelemy, & Bruneau, 2002a; Kujala et al., 2010; Lepisto et al., 2006; Lepisto et al., 2005a; Lindstrom, Lepisto-Paisley, Vanhala, Alen, & Kujala, 2016; Whitehouse & Bishop, 2008). Although, a few studies have found delayed latency to speech (Korpilahti et al., 2007) or shorter latency to nonspeech (Ferri et al., 2003). In contrast to relatively normal latencies, amplitudes of the early auditory ERP responses have often been found to be decreased in children with ASD to speech sounds, indicative of reduced synchronous neural processing (Courchesne, Lincoln, Kilman, & Galambos, 1985; Jansson-Verkasalo et al., 2003; Korpilahti et al., 2007; Lepisto et al., 2005; Whitehouse & Bishop, 2008), with fewer studies finding reduced amplitude to nonspeech stimuli (Bidet-Caulet et al., 2017; Jansson-Verkasalo et al., 2003; Lepistö et al., 2005). Furthermore, Whitehouse and Bishop (2008) found reduced P1 and N2 ERP amplitudes to speech in a passive condition that disappeared when the children were required to pay attention to speech in an active condition, suggesting that the common finding of reduced P1 and N2 amplitude to speech could be driven by a top-down attentional mechanism in which repeated speech stimuli are tuned out.

Studies measuring the P3a component to speech and nonspeech sounds have found that children and adults with autism have reduced P3a amplitudes to speech compared to typically developing samples (Fan & Cheng, 2014; Lepisto et al., 2006; Lepisto et al., 2005b; Lepisto, Nieminen-von Wendt, von Wendt, Näätänen, & Kujala, 2007), but no differences or even enhanced P3a amplitudes to nonspeech (Ceponiene et al., 2003; Gomot, Giard, Adrien, Barthelemy, & Bruneau, 2002b; Lepisto et al., 2006; Lepistö et al., 2005a, 2007; Lincoln, Courchesne, Harms, & Allen, 1993; Whitehouse & Bishop, 2008). Few exceptions to this pattern exist; Fan and Cheng (2014) did not find a P3a to their synthetic nonspeech stimuli in either TD or ASD children, and Whitehouse and Bishop (2008) found an enhanced P3a amplitude to speech coupled with a decreased P3a amplitude to nonspeech, but this was explained by a crucial paradigm difference in which their deviant speech stimuli always followed a repeated train of nonspeech standard stimuli, and vice versa. Similar to their P1-N2 findings in the active condition, the authors consequently suggested that decreased P3a amplitude in children with ASD is mediated by a top-down mechanism in which repeated speech stimuli are increasingly tuned out, but repeated nonspeech stimuli are not. In addition, a few studies investigating P3a responses only to tones have found reduced amplitude or delayed P3a responses in low-functioning children with ASD, but without a speech comparison included in these studies, it is unclear whether the P3a amplitude or latency may be further impaired to speech in these children (Ferri et al., 2003; Salmond, Vargha-Khadem, Gadian, de Haan, & Baldeweg, 2007).

The majority of the results measuring the P1, N2, and P3a responses in ASD suggest that decreased social motivation is the primary deficit, and this leads to the observed attenuation of P1, N2, and P3a to speech via top-down processing. However, there are two factors which limit conclusions about whether auditory processing or social motivation contribute to language impairments in ASD. The first is that previous studies have rarely characterized the language abilities of their sample or investigated the relationship between ERPs and individual differences in language abilities, and it may be that different language profiles show differential relationships to auditory processing or attentional deficits. Furthermore, language abilities in children with ASD range from severe impairments in nonverbal/minimally verbal individuals with a couple words up to mild impairments in verbally fluent individuals with only a few pragmatic deficits. The few studies which have measured their sample's language abilities have included a variety of severity levels. Because event-related potentials are averaged at the group level, such heterogeneity may be obscuring true deficits or strengths that are present in only some children within the sample.

The second factor confounding the interpretation of previous studies is that the type of stimuli used are not ecologically valid, and therefore the findings may not translate to neural processing in the real world. All but one of the previous studies have used either 1) robotic-sounding synthetic speech stimuli that are matched on complexity to synthetic tones, both of which lack ecological validity, or 2) semi-synthetic speech stimuli or natural speech stimuli that are ecologically valid, but are not matched on complexity to an ecologically valid nonspeech stimulus.

Matching on complexity is crucial because ERP components arise from the synchronized activity of populations of neurons, and without matching one cannot rule out that any differences are due to larger populations of neurons responding to the more complex stimulus (Bardy, Van Dun, & Dillon, 2015). The simplified nature of a synthetic speech stimulus allows researchers to easily match it on spectro-temporal complexity to a synthetic nonspeech tone. However, this simplicity could lead to misleading results in populations with ASD, especially given that behavioral studies have found that sensory processing deficits in ASD are often found in response to more complex rather than simple stimuli (Mottron, Dawson, Soulières, Hubert, & Burack, 2006).

Recently, Bidet-Caulet et al. (2017) became the only study thus far that has presented ecologically valid and complexity-matched speech and nonspeech stimuli to children with autism. However, they used many different vocal and environmental sounds rather than matching pairs of stimuli. Averaging ERPs to sound stimuli across many acoustic dimensions results in flattened auditory ERP peaks, which could be why they did not find P1 or N2 differences between ASD and TD to speech stimuli. Additionally, this study did not measure the P3a component, and therefore no study thus far has been able to measure in an ecologically valid way whether auditory processing or attentional processing linked to social motivation may contribute to different types of language impairments in children with ASD.

The purpose of the present study was to expand upon previous studies measuring the P1, N2, and P3a ERP components to speech and nonspeech in children with ASD and typically developing children ages 8—14-years-old by 1) by subgrouping children with ASD based on their language abilities, including children with the most severe language impairment (minimally verbal autism; MVA), intermediate language impairment (phrase speech autism; PSA), and mild language impairment (verbally fluent autism; VFA), and 2) by using novel ecologically valid natural speech stimuli matched on complexity to natural nonspeech stimuli. Our lab previously validated a passive oddball task using natural speech vowels matched on complexity to cow vocalizations in typically developing children (Whitten, Mefferd, Key, & Bodfish, under review). This study demonstrated that typically developing children show faster auditory processing of natural speech vowels compared to cow sounds, providing evidence for more efficient processing of speech in typically developing children. The present study adapted the stimuli from our previous paradigm to create a new oddball task and apply it to language subgroups of children with ASD. We hypothesized that 1) children in the MVA and PSA language subgroups would exhibit increased latency and decreased amplitude of the P1 and N2 ERP components compared to VFA and TD children; 2) all ASD language subgroups would exhibit decreased P3a amplitude to speech, but increased P3a amplitude to the nonspeech cow in comparison to TD; and 3) that only the TD children would exhibit faster processing of the P1 and N2 components to the speech deviant compared to the nonspeech cow deviant.

Methods

Participants

Typically developing children and children with a diagnosis of autism spectrum disorder between the ages of 8 and 14 years were recruited to participate in this study. The 8—14-year-old age range was selected to be consistent with previous studies (Ceponiene et al., 2003; Jansson-Verkasalo et al., 2003; Lincoln et al., 1993; Whitehouse & Bishop, 2008) as well as to obtain a sample of truly minimally verbal children rather than “pre-verbal” children, given that children who are minimally verbal typically do not become verbal beyond the age of seven years (Pickett, Pullara, O'Grady, & Gordon, 2009). See Table 4.1 for demographic features of the sample.

Typically developing children (TD)

TD children were recruited through the Vanderbilt University email distribution service. Children were excluded from the TD group for any of the following reasons: medical or educational diagnoses, a family member with a diagnosis of autism spectrum disorder, presence of language impairment, first language was not English, they scored above the cut-off for ASD on the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003), or they failed the hearing screen during their lab visit (see below). This resulted in a final sample of 14 typically developing children (2 females) with a mean age of 9.71 years (SD: 1.49) and mean IQ of 95.07 (SD: 10.50) as measured by the Stanford Binet Intelligence Scales (Roid, 2003).

Children with autism spectrum disorder (ASD)

Children with a diagnosis of autism spectrum disorder were recruited through a variety of resources at Vanderbilt University and the greater Nashville community, including the Vanderbilt Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt Kennedy Center StudyFinder, Autism Tennessee, the Vanderbilt University ResearchMatch.org database, and flyers put up in Vanderbilt University Medical Center clinics. Participants were required to have a diagnosis of ASD from a licensed clinician. All diagnoses were confirmed during the lab visit through evaluations by trained study personnel using the gold standard Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012). Children with ASD were excluded if they had comorbid psychiatric diagnoses (one child: schizoaffective disorder) or brain abnormalities (one child: cortical dysplasia), or if English was not their first language (one child).

Table 4.1 *Demographic Features by Subgroup*

Group	**Age**	**IQ**	**Sex**	**Language Ability***
Minimally Verbal ASD (MVA; n = 8)	11.38 (2.62)	48.75 (3.88)	5 males, 3 females	Single words (n = 8)
Phrase Speech ASD (PSA; n =4)	9.25 (0.96)	67.75 (11.59)	4 males, 0 females	Three and four-word phrases (n=4)
Verbally Fluent ASD (VFA; n=14)	8.93 (1.33)	97.00 (21.57)	11 males, 3 females	Complex Sentences (n=14)
Typically Developing (TD; n=14)	9.71 (1.49)	95.07 (10.50)	12 males, 2 females	Complex Sentences (n=14)

*Language ability at the level of single words, phrases, or complex sentences was measured from natural language sample obtained during ADOS for all ASD participants, and measured during SB-5 for TD participants.

ASD language subgroups. Children with ASD were subgrouped based on their language abilities observed during the hour-long ADOS-2 assessment. The ADOS-2 provides a semi-structured play-based assessment that has often been used to obtain a natural language sample to identify minimally verbal autism (Bal, Katz, Bishop, & Krasileva, 2016; Gordon et al., 2011; Kasari et al., 2014; Mucchetti, 2013; Thurm, Manwaring, Swineford, & Farmer, 2015). The benefit of using the ADOS to obtain a natural language sample is that it is standardized across children and research studies such that the children are exposed to the same toys, presses, and tasks. Children with ASD were placed in the minimally verbal group if they were administered the ADOS Module 1 and received a score on item A1 of 2 (n = 3; "recognizable single words or word approximations only; must use at least 5 different words during the ADOS-2 evaluation"), 3 (n = 4; "at least one word or word approximations but fewer than five words used during ADOS-2 evaluation"), or 4 (n = 1; "No spontaneous use of words or word approximations"). Children with ASD were placed in the phrase speech group if they were administered Module 2 and received a score on Item A1 of 0 (n = 3; "Non-echoed phrase speech of three or more words per utterance; some grammatical markings, such as plurals or tense"), or 1 (n = 1; "Speech is primarily two- or three-word utterances, with minimal or no grammatical markings"). Children were placed in the verbally fluent group if they were administered Module 3 and received a score on Item A1 of 0 (n = 13; "Uses sentences in a largely correct fashion (must use some complex speech), or 1 (n = 1; "Some relatively complex speech (occasional utterances with two or more clauses), but with recurrent grammatical errors not associated with use of dialect"). In order to obtain a parent report measure of the amount of words that the minimally verbal children were able to produce, 6/8 parents completed the MacArthur Bates Communication Developmental Inventory (MCDI; Bates et al., 2004). Out of 396 words total in the MCDI, MVA parents reported a range of 0-363 words (M = 187.33, SD = 158.58).

Three children were excluded from the minimally verbal group due to noncompliance during the EEG task: two children would not tolerate the EEG net, and one child produced too much movement artifact resulting in unusable ERP data. This resulted in a sample of eight children in the minimally verbal autism group (3 females) with a mean age of 11.38 years (SD: 2.62) and a mean IQ of 48.75 (SD: 3.88). The phrase speech group included 4 participants (0 females) with a mean IQ of 67.75 (SD: 11.59) and mean age of 9.25 years (SD: 0.96). The verbally fluent sample included 14 participants (3 females) with a mean IQ of 97.00 (SD: 21.57) and mean age of 8.93 years (SD: 1.33). MVA participants had significantly lower IQ scores than VFA and TD participants ($p < .001$), however this did not reach significance for MVA compared to PSA ($p = .05$). PSA participants had significantly lower IQs than VFA and TD participants ($p < .01$). VFA and TD participants did not significantly differ on IQ ($p = .77$).

Stimuli & Task Design

In the oddball task, speech and nonspeech stimuli were presented in two blocks (speech and nonspeech) in which infrequent "deviant" sounds were presented within a stream of frequent "standard" sounds. To ensure novelty of the deviant sound, at least three standards always preceded a deviant sound. Within one trial, six sounds were presented, and the deviant sound occurred randomly in either the 4th, 5th, or 6th position. Sounds were presented with a varying interstimulus interval (ISI) between 700-900 ms to prevent habituation to stimulus onset. In addition to the oddball task, one additional task with a simple stimulus was presented. In the pure tone task, a 1000 Hz pure tone created using Adobe Audition (Adobe Systems, Inc) was presented

50 times with an ISI of 6-9 seconds. Given that the natural stimuli in the oddball task are highly complex auditory stimuli, this task was added to allow us to test whether auditory ERP responses in language subgroups of children with ASD are also impaired to a simple sound stimulus.

All sound stimuli were 350 ms in duration and edited to have a rise and fall time of 10 ms using a custom-written MATLAB script (The Mathworks, Inc). Stimuli were downsampled to a sampling rate of 44.1 kHz with a quantization level of 16-bit. All stimuli were matched on loudness based on their average root mean square (RMS) using Adobe Audition. Participants were seated 120 cm from a TV monitor displaying a silent movie while auditory stimuli were presented at approximately 70 dB from two loudspeakers located on each side of the monitor. Verbally fluent ASD and typically developing children watched the silent movie Wall-E, while MVA and PSA children were allowed to choose their own video to increase compliance. During both the oddball task and pure tone task, participants were instructed to watch the movie and ignore the sounds.

The speech and nonspeech stimuli were adapted from a previous study in our lab (Whitten et al., under review). This study created a novel oddball paradigm using natural speech and nonspeech stimuli by matching vowel stimuli to cow vocalizations. Cow vocalizations provided a unique opportunity to match a natural ecologically valid nonspeech stimulus to the spectro-temporal complexity of speech. This was possible because cow vocalizations are generated with the same underlying mechanisms as human speech, i.e., laryngeal vocal fold vibration and filtering via a "vocal tract." The process of matching the vowel and cow sound stimuli has been described in detail in Whitten et al. (2019). Briefly, stimuli were matched on overall complexity by matching two dynamic spectro-temporal aspects, namely by matching the changes in amplitude within 1 dB and the changes in fundamental frequency within 10 Hz across the entire stimulus duration.

Pearson correlations on acoustic measures between stimuli resulted in significant correlations with coefficients ranging from 0.937 to 0.995.

Our previous task presented natural speech and cow sound standard and deviant sounds interspersed together within the same block. In the present study, in order to relate our findings to the majority of previous studies in ASD, we modified these stimuli to create a new task in which speech and nonspeech standard and deviant stimuli are presented in separate blocks. Previous oddball studies in ASD have used two different vowel sounds in the speech block and two different complex tones in the nonspeech block. Thus, a logical extension in the present study would have been to create a speech block with two natural vowels and a nonspeech block with two cow sounds. However, we reasoned that the change from one vowel to a different vowel may be more salient than the change from one cow sound to another cow sound and thus could bias the P3a attentional response. This is because changing from one vowel to another vowel is likely more meaningful given the importance of discriminating different phonemes in language processing. Furthermore, it would be difficult to match the acoustic difference between two vowels and two cow vocalizations because while humans vary only fundamental frequency and formants between two vowels, we found through acoustic analysis of our previous dataset that cow vocalizations which vary in formants also vary the amount of noise and harshness in the signal.

To ensure that the acoustic difference and salience between standards and deviants did not differ across speech and nonspeech blocks, we decided to use an equivalent pitch change across blocks. Specifically, we edited the stimuli from our previous study using Adobe Audition to create vowel and cow sound stimuli that increased in mean pitch (F0) by 60% (7.2 semitones) from the standard to deviant sound. This pitch change represented the largest difference between speech and nonspeech stimuli that still maintained the natural quality of the original stimuli (i.e., no distortions, still recognizable as cow sound and /i/ vowel spoken by female). This resulted in a speech block that presented a standard low pitch /i/ (mean F0: 224.4 Hz) and deviant high pitch /i/ (mean F0: 359.2 Hz), and a nonspeech block that presented a standard low pitch cow (mean F0: 87.5 Hz) and a deviant high pitch cow (mean F0: 140.4 Hz). As in our previous study, the vowels and cow sounds still rose in pitch over the length of the stimulus. Adobe Audition's pitch editor did change the pitch range from our previous study particularly for the low pitch stimuli; however, it changed them by the same amount such that the standard sounds remained matched across blocks and the deviant sounds remained matched across blocks, and thus the acoustic difference between standard and deviants within blocks was also the same. Praat software (v 6.0.20) was used to measure acoustic features of the stimuli, while spectral entropy for each sound was calculated using the Seewave R package (Sueur et al., 2008). Table 4.2 presents acoustic features of all sounds, and Figure 4.1 displays a schematic of the oddball and pure tone tasks.

Pilot data revealed that some minimally verbal and phrase speech ASD participants may not be able to comply with both tasks. Therefore, task order was not counterbalanced to ensure that all minimally verbal participants received the oddball task. However, the order of speech and nonspeech blocks within the oddball task was counterbalanced across all participants. All minimally verbal (n=8) and phrase speech (n=4) ASD participants completed the oddball task with usable data, and only one minimally verbal participant was not able to complete the pure tone task (n=7). All VFA (n=14) and TD (n=14) participants completed both tasks, except for one VFA participant whose data were excluded from the pure tone task due to a recording issue during data collection.

Table 4.2 *Acoustic Features of Oddball Task Stimuli*

Acoustics of Oddball Task Stimuli

Block	**Stimulus**	**Mean F0**	**F0 Range***	**Mean Intensity**	**Intensity Range***	**Duration***	**Spectral Entropy**
Speech Block	Standard: Low Pitch Vowel /i/	224.4 Hz	24.7 Hz	73.1 dB	10.3 dB	350 ms	0.798
	Deviant: High Pitch Vowel /i/	359.2 Hz	38.2 Hz	73.1 dB	9.8 dB	350 ms	0.761
Nonspeech Block	Standard: Low Pitch Cow	87.5 Hz	23.8 Hz	73.5 dB	11.8 dB	350 ms	0.762
	Deviant: High Pitch Cow	140.4 Hz	36.1 Hz	73.5 dB	11.1 dB	350 ms	0.759

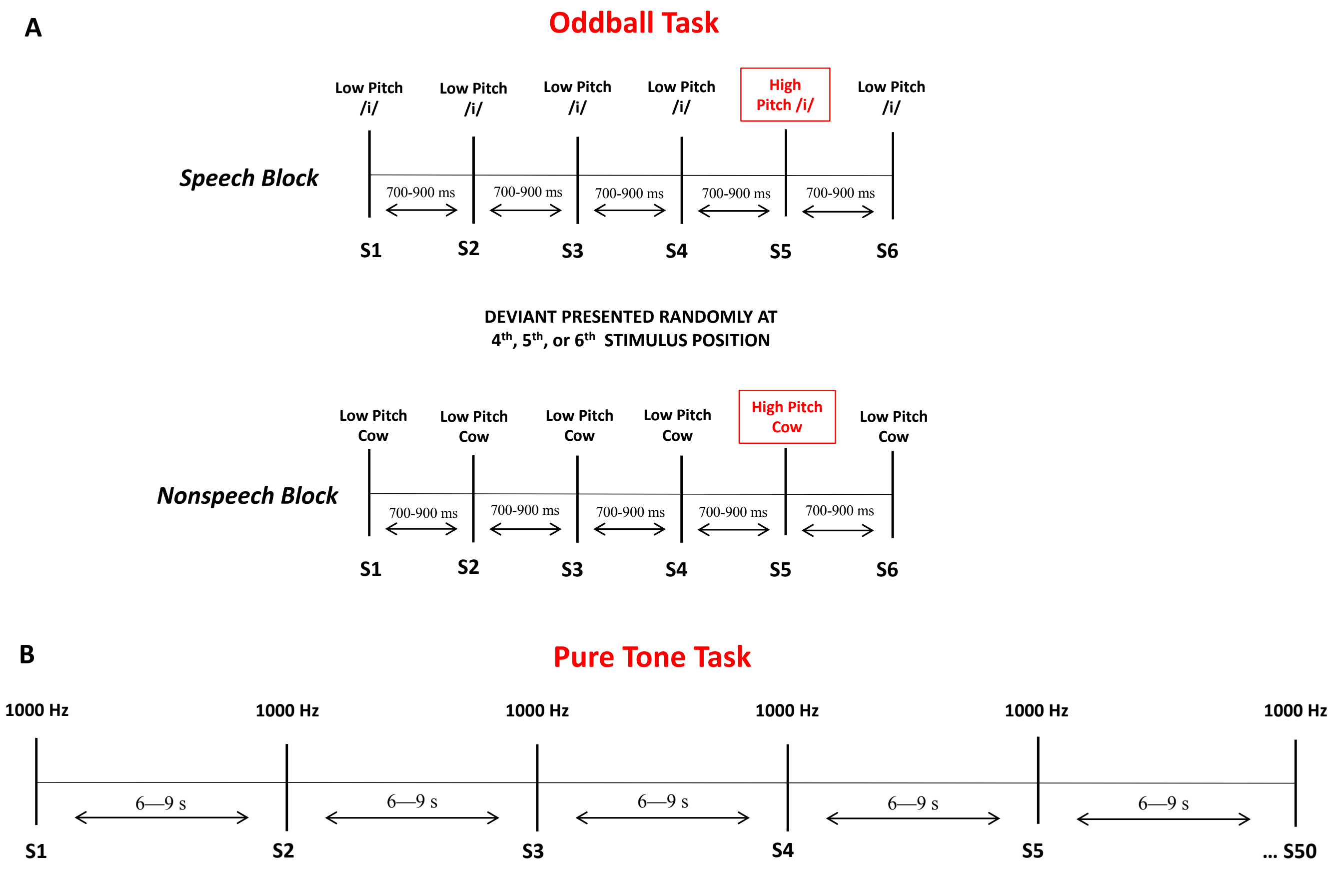


Figure 4.1 Oddball Task and Pure Tone Task Schematic. A) In the oddball task, one speech block and one nonspeech block were presented to each participant with block order counterbalanced across participants. In the speech block, the standard repeating sound was a low pitch /i/ and the infrequent deviant sound was a high pitch /i/. In the nonspeech block, the standard repeating sound was a low pitch cow and the infrequent deviant was a high pitch cow. All stimuli were presented in trial sequences of six stimuli in which the standard sound was always presented in the first three positions and the deviant sound was randomly presented in the 4th, 5th, or 6th position. B) In the pure tone task, 1000 Hz pure tones were presented every 6—9 seconds for a total of 50 trials.

Measures

Autism Diagnostic Observation Schedule (ADOS-2)

The ADOS is considered the gold standard assessment tool for diagnosing ASD in a semi-structured play setting. It has also been used frequently to obtain a natural language sample in ASD and to identify minimally verbal autism (Bal et al., 2016; Gordon et al., 2011; Kasari et al., 2014; Mucchetti, 2013; Thurm et al., 2015). In the present study, the ADOS-2 was used both to confirm ASD diagnosis as well as to subgroup children with ASD into language subgroups based on their natural language sample.

Stanford Binet-5 Abbreviated IQ Test

The Stanford Binet is an IQ test developed to measure developmental or intellectual disabilities in children. The five factors being tested are knowledge, quantitative reasoning, visual-spatial processing, working memory, and fluid reasoning. The present study used the abbreviated version to obtain a quick estimate of each child's developmental level.

Pure Tone Audiometry Screener

All children who were able to comply were given a brief behavioral audiometry test to assess for normal hearing using pure tones, based on the guidelines from the American Academy of Audiology (2011) at 1000 Hz, 2000 Hz, and 4000 Hz. All participants in the VFA and TD groups passed the screener. In contrast, one child in the PSA group and zero children in the MVA group were able to comply. For these children, parents confirmed that they did not have any concerns about the child's hearing, and that the child had had a normal hearing test in the past few years.

Neural responses were recorded by Net Station 5.4 software (EGI, Inc., Eugene, OR, USA) and stimulus presentation was controlled by E-Prime 2.0 (PST, Inc., Pittsburgh, PA, USA).

Participants wore high-density hydrocel geodesic EEG nets with 128 electrodes (EGI, Inc., Eugene, OR, USA). Electrode impedances were kept below 50 kΩ. The EEG was sampled at 500 Hz and filtered offline using a 0.1 – 30 Hz bandpass filter. During data collection the signal was referenced to Cz, and later re-referenced to the average of all electrodes following artifact rejection.

Data Analysis

ERP epochs were segmented from the continuous EEG data beginning 100 ms before (pre-stimulus baseline) and 600 ms after stimulus onset to all standard sounds that preceded a deviant sound, deviant sounds, and the pure tone stimulus. The resulting epochs were screened for artifacts using an automated Net Station algorithm followed by manual review. Trials containing eye blinks, lateral eye movements, muscular activity, or electrical noise were excluded from analysis. Individual electrodes with poor signal quality within a trial were corrected using spherical spline interpolation; no more than 10% of electrodes were interpolated on any one trial. The average amount of trials in each condition (standard speech, standard nonspeech, deviant speech, deviant nonspeech, pure tone) across groups was 25.5 trials; significant differences in trial counts were found only to the deviant speech sound between MVA (M = 19.9, SD = 8.2) and TD (M = 29.5, SD = 7.6, $p = .04$, Scheffe correction for multiple comparisons).

Following artifact rejection, the remaining trials for each condition were averaged, re-referenced to the average reference, and baseline-corrected. In order to maximize signal-to-noise ratio, ERP amplitudes and latencies were measured as the average of a cluster of electrodes rather than a single electrode. Electrode clusters were chosen based on the strongest response in the grand average topographic maps during the selected time window. Time windows for amplitude and latency of ERP components were chosen based on the examination of the grand average waveform, and confirmed within each individual participant's ERP to ensure that each participant displayed the ERP of interest within the time window. For the auditory ERPs in the pure tone task and the oddball task, peak latency and peak amplitude were measured at the local peak within the time window. We defined a local peak as the time-point in which at least 10 data points (20 ms) on either side of the peak were reduced in amplitude. This technique is favorable over simple peak measures (which identify the most positive or negative point within the window) because it is less influenced by noise (Luck, 2014). For the P3a component in the oddball task, amplitude was measured as the mean amplitude within the time window of interest.

As mentioned previously, when interstimulus intervals longer than 1 second are used as in our pure tone task, the adult pattern of N1 and P2 auditory ERPs is observed in children. As a result, N1 and P2 ERPs were measured in this task. N1 peak latency and amplitude was measured between 80—140 ms, and P2 peak latency and amplitude was measured between 145—245 ms. The strongest responses for N1 and P2 at these windows was found at a cluster of electrodes corresponding to Cz (7, 31, 55, 80 106, 129/Cz).

In the oddball task, P1 amplitude and latency was measured between 50-150 ms and 200-300 ms for N2. In these time windows, the strongest P1 and N2 auditory responses in all groups were found at a cluster that corresponded to Fz (electrodes 4, 5, 10, 11, 12, 16, 18, 19). Because such wide windows were necessary to encompass P1 and N2 ERPs in every subject, amplitude was measured as peak amplitude rather than mean amplitude as this would lead to averaging over the entire window that would not be representative of the peak of interest.

The strongest P3a response to the deviant sounds in the oddball task across participant groups was found at a cluster corresponding to Fz (electrodes 4, 5, 10, 11, 12, 16, 18, 19). The P3a time window was chosen based on the grand-average waveforms of all groups and selected as 300-

400 ms. P3a amplitude was calculated as the mean amplitude between 300-400 ms of the difference wave, calculated by subtracting the standard sound from the deviant sound for both deviant sounds in the oddball task. Difference waves were used to calculate the mean P3a amplitude in each group in order to control for differences in response to the standard sounds. However, because just looking at the difference waves can often be misleading with regards to a true P3a response, we also present figures showing the standard and deviant sounds within each group (where greater positivity to the deviant in the 300-400 ms range represents the P3a orienting response). Visual inspection of individual P3a ERP difference waves revealed that many participants did not display a peak during this time window, and thus peak latency was not measured.

Peak latency and peak amplitude values to the deviant speech and nonspeech stimuli in the oddball task were compared for P1 and N2 measured at the Fz cluster separately in the full sample using a 2 x 4 mixed model repeated measures ANOVA where the within-subject factor was sound category (speech, nonspeech) and the between-subject factor was group (MVA, PSA, VFA, TD). For P1 and N2, a priori planned comparisons were used to investigate latency differences to the speech vs. nonspeech deviant sounds within each group to test whether speech was processed significantly faster than nonspeech in any group. In addition, a priori planned comparisons were used to test for significant differences between MVA, PSA, VFA, and TD groups on P1 and N2 latency and amplitude to the speech deviant only to test whether MVA and PSA groups differed from VFA or TD on early auditory speech processing. P3a mean amplitude values measured from the difference waves (deviant – standard sounds) at the Fz cluster were compared using a 2 x 4 mixed model repeated measures ANOVA where the within-subject factors were sound category (speech, nonspeech) and the between-subject factor was group (MVA, PSA, VFA, TD). Peak latency and peak amplitude values to the 1000 Hz pure tone presented during the Pure Tone Task were compared for N1 and P2 in the full sample using separate one-way ANOVAs where the independent factor was group (MVA, PSA, VFA, TD) and the dependent variables were N1 latency, N1 amplitude, P2 latency, and P2 amplitude.

Due to the differences in sample sizes across groups, Type III sum of squares was used to correct for unbalanced designs based on the results of the simulation study by Landsheer and van den Wittenboer (2015). Estimated marginal means (EMM) and standard errors (SE) predicted from the models are presented. When testing pairwise comparisons, the Benjamini-Hochberg FDR procedure (Benjamini & Hochberg, 1995) with a false discovery rate of 0.20 was used to control for multiple comparisons. Only results that remained significant after correction are reported. Effect sizes are reported based on the recommendations by Lakens (2013) and calculated using the supplementary tables accompanying the paper. We present within-subjects effect sizes which utilize the correlation between measures and thus result in a power increase associated with within-subjects designs that is best fit to our study, but we also include between-subjects effect sizes in order to generalize to other studies that may compare groups in the future, as well as to facilitate interpretability of both effect sizes given that between-subjects effect sizes are more common. For the repeated measures ANOVA main effects and interaction effects, we present both partial eta squared as the within-subjects effect size and generalized eta squared as the between-groups effect size. For pairwise comparisons, we present Cohen's d_z as the within-subjects effect size. For a comparable between-subjects effect size in the within-subjects contrasts, we applied Hedges correction to Cohen's d_{av} to report Hedges g_{av} that takes into account the positive bias associated with using sample estimates. For between-subjects paired comparisons, both Cohen's d and

Hedges g_s are reported. For the Pure Tone Task analysis which contained only a between-groups factor, partial eta squared is presented.

Results

Auditory ERPs

The natural speech and nonspeech stimuli in the oddball task elicited P1-N2 auditory ERP responses in all groups, displayed in Figure 4.2. However, visible morphology differences are seen the grand-averages between groups, with MVA and PSA groups showing less well-defined peaks than VFA and TD groups. Figure 4.3 shows the deviant speech and nonspeech stimuli separately within each group in to highlight these differences. This was not due purely to lower sample sizes in the MVA and PSA compared to VFA and TD groups, as randomly choosing n = 4 and n = 7 individuals from the VFA and TD groups resulted in similar grand average waveforms to the full n = 14 sample. Notably, the VFA and TD groups exhibited less well-defined P1 and N2 peaks than is typically found to synthetic stimuli in previous studies. Less well-defined peaks to natural stimuli was found previously in our study in typically developing children (Whitten et al., under review) as well as Bishop, Hardiman, Uwer, and von Suchodoletz (2007). In contrast, the simple stimulus (1000 Hz pure tone) in the pure tone task elicited well-defined N1-P2 ERP responses in all groups.

N1-P2: Pure Tone Task

In response to the simple pure tone stimulus (1000 Hz), there was no significant difference

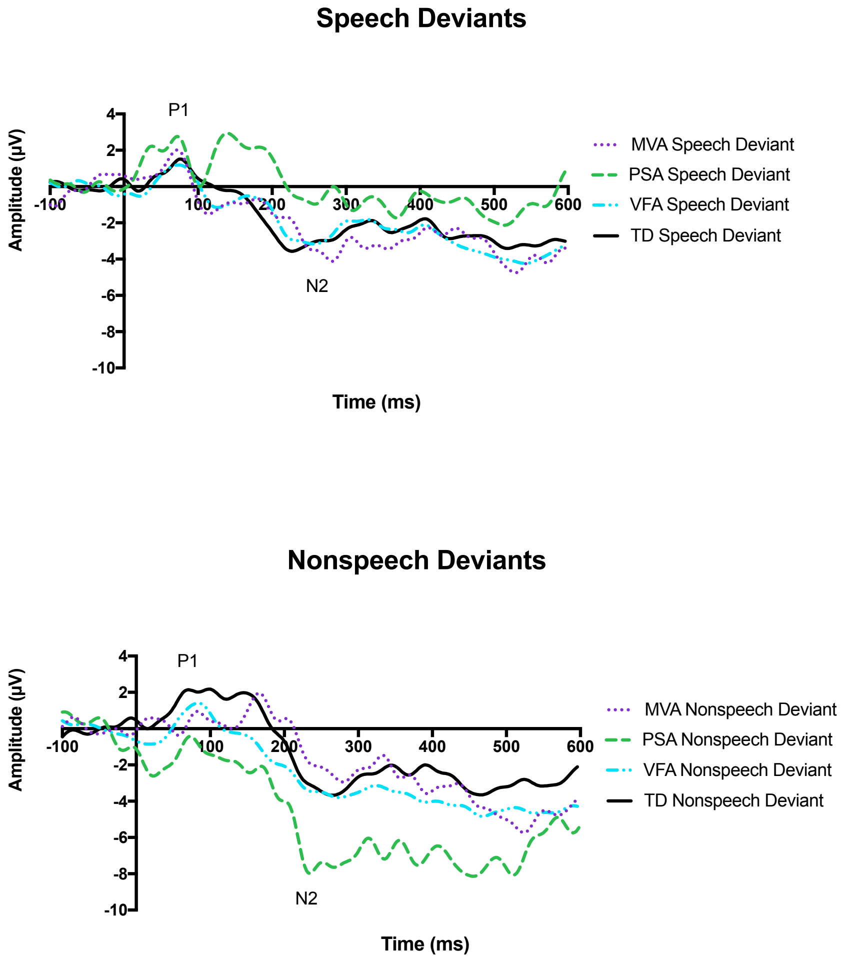


Figure 4.2 P1 and N2 Auditory ERPs to Speech and Nonspeech Deviants Across Groups. Top: P1 and N2 auditory ERPs to the speech deviant (high pitch /i/) during the speech block in all groups. Bottom: P1 and N2 auditory ERPs to the nonspeech deviant (high pitch cow sound) during the nonspeech block in all groups. P1 peak latency and amplitudes were measured as the most positive local peak between 50 and 150 ms, and N2 peak latency and amplitudes were measured as the most negative local peak between 200 and 300 ms.

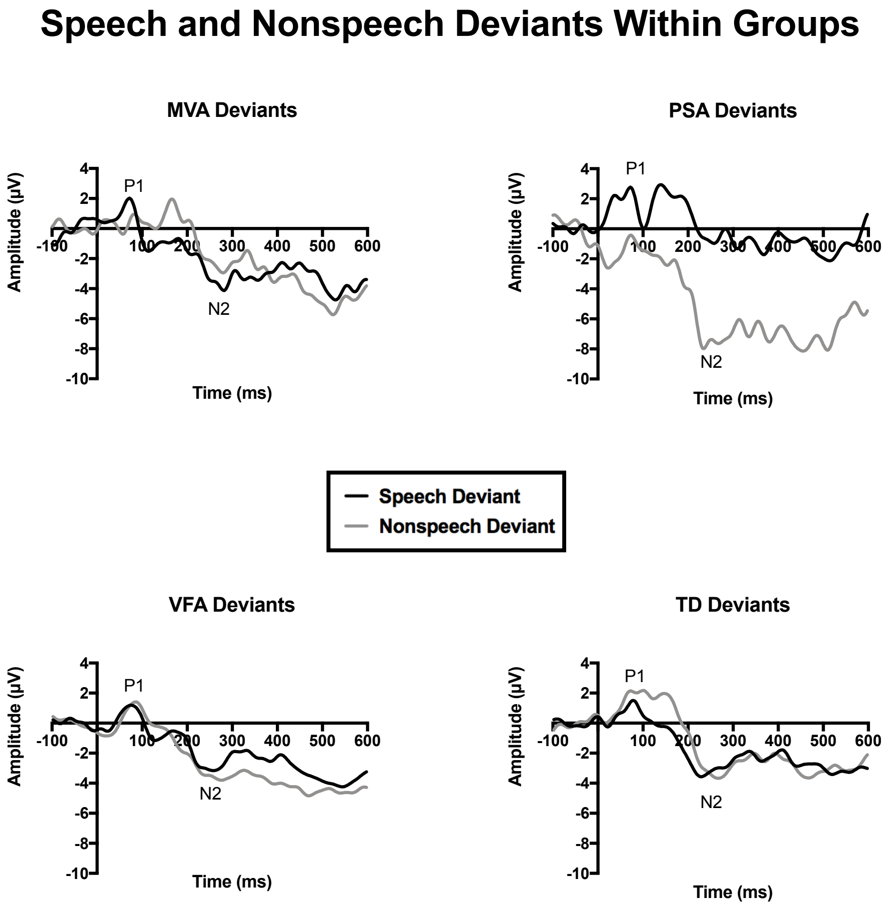


Figure 4.3 P1 and N2 Auditory ERPs to Speech and Nonspeech Deviants Within Groups. P1 peak latency and amplitudes were measured as the most positive local peak between 50 and 150 ms, and N2 peak latency and amplitudes were measured as the most negative local peak between 200 and 300 ms.

between groups on N1 latency or amplitude measured at the Cz cluster between 80 and 140 ms (N1 latency: *F*(3, 34) = .88, *p* = .463, partial eta squared = .07; N1 amplitude: *F*(3, 34) = .10, *p* = .96, partial eta squared = .01). Similarly, there was no significant difference between groups to P2 latency or amplitude measured at the Cz cluster between 145 and 245 ms (P2 latency: *F*(3, 34) = .19, *p* = .90, partial eta squared = .02; P2 amplitude: *F*(3, 34) = 1.05, *p* = .38, partial eta squared = .09). Figure 4.4 displays N1 and P2 responses to the 1000 Hz pure tone in all groups.

P1-N2: Oddball Task

For P1 latency measured at the Fz cluster between 50-150 ms, there was no significant main effect of sound category (speech vs. nonspeech), (*F*(1, 36) = .56 *p* = .46, partial eta squared = .02, generalized eta squared = .01) and no significant main effect of group (MVA, PSA, VFA, TD), *F*(3, 36) = 1.33, *p* = .28, partial eta squared = .10, generalized eta squared = .05). Additionally, there was no significant interaction effect of sound category x group (*F*(3, 36) = 1.09, *p* = .37, partial eta squared = .08, generalized eta squared = .04). Planned comparisons for P1 latency within each group revealed significantly faster latency to speech than nonspeech only in the TD group (speech EMM = 83.29, SE = 6.56, nonspeech EMM = 104.43, SE = 7.34, *p* = .04, $d_z = 0.74$, $g_{av} = 0.76$). Planned comparisons for P1 latency to speech comparing the language-impaired groups to VFA and TD found no significant results.

There was no significant main effect of sound category for P1 amplitude measured at the Fz cluster between 50-150 ms (*F*(1, 36) = 2.41, *p* = .13, partial eta squared = .06, generalized eta squared = .03). There was also no main effect of group (*F*(3, 36) = .50, *p* = .68, partial eta squared = .04, generalized eta squared = .02). However, there was a significant sound category x group

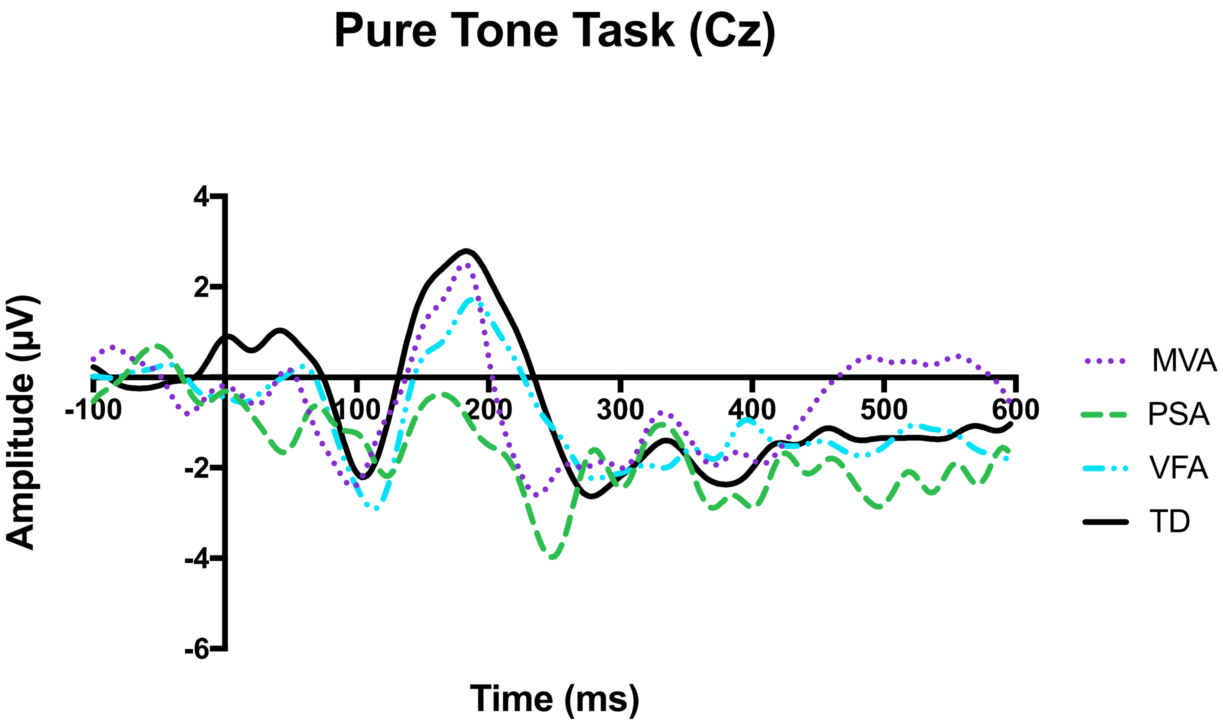


Figure 4.4 N1 and P2 Auditory ERPs to the Pure Tone Task in All Groups. N1 peak latency and amplitude was measured as the most negative local peak between 80 and 140 ms, and P2 amplitude and latency was measured as the most positive local peak between 145 and 245 ms.

interaction (*F*(3, 36) = 3.53, *p* = .024, partial eta squared = .23, generalized eta squared = .11). Simple effects revealed that this was due to significantly larger P1 amplitude to speech (EMM = 3.80, SE = 1.15) compared to nonspeech (EMM= -0.04, SE = .96) in the PSA group only (*p* = .009, $d_z = 1.48$, $g_{av} = 2.00$), while MVA, VFA, and TD showed no significant differences between speech and nonspeech P1 amplitudes. Planned comparisons for P1 amplitude to speech comparing the language-impaired groups to VFA and TD found no significant results.

No significant main effect of sound category was found for N2 latency measured at the Fz cluster between 200-300 ms (*F*(1, 36) = 2.01, *p* = .17, partial eta squared = .05, generalized eta squared = .02) and no significant main effect of group, *F*(3, 36) = 1.38, *p* = .27, partial eta squared = .10, generalized eta squared = .06). Additionally, the sound category x group interaction did not reach significance (*F*(1, 36) = 2.40, *p* = .08, partial eta squared = .17, generalized eta squared = .08). Planned comparisons for N2 latency within each group revealed significantly faster latency to speech than nonspeech again in the TD group (speech EMM = 235.00, SE = 6.76, nonspeech EMM = 258.00, SE = 5.99, *p* = .008, $d_z = 0.66$, $g_{av} = 0.88$), and also in the VFA group (speech EMM = 245.43, SE = 6.76, nonspeech EMM = 263.57, SE = 5.99, *p* = .03, $d_z = 0.73$, $g_{av} = 0.79$). Planned comparisons for N2 latency to speech comparing the language-impaired groups to VFA and TD found significantly slower latency to speech in the MVA group (EMM = 267.50, SE = 8.94) compared to the TD group (*p* = .006, d = 1.36, $g_s = 1.24$), and almost reached significance in MVA compared to the VFA group (*p* = .057, d = 1.05, $g_s = .96$). Given the moderate effect size of the sound category x group interaction, performing post hoc comparisons for N2 latency to nonspeech across groups revealed no significant differences.

For N2 amplitude measured at the Fz cluster between 200-300 ms, there was a significant main effect of sound category, ($F(1, 36) = 4.95$, $p = .03$, partial eta squared = .12, generalized eta squared = .05) in which N2 amplitude to speech was significantly smaller (EMM = -3.84, SE = .50) than nonspeech (EMM = -5.27, SE = .59). There was no significant main effect of group, $F(3, 36) = 0.32$, $p = .81$, partial eta squared = .03, generalized eta squared = .0006). However, the sound category x group interaction was also significant ($F(3, 36) = 5.07$, $p = .005$, partial eta squared = .30, generalized eta squared = .13), suggesting caution in interpreting the sound category main effect across all groups. Simple effects tests on the interaction revealed that the PSA group was driving the main effect in which their N2 amplitude to nonspeech (EMM = -8.64, SE = 1.63) was significantly larger than speech (EMM = -1.78, SE = 1.37, $p < .001$, $d_z = 1.86$, $g_{av} = 1.84$), while this pattern did not hold in MVA, VFA, or TD groups (p = .32, .90, .996, respectively). The observed N2 amplitude in PSA was unexpectedly large (EMM = -8.64, SE = 1.63), and although this may be impacted by one PSA participant with a peak amplitude of -12.4, all other PSA participants also showed large N2 peak amplitudes as well with values of -5.6, -8.8, and -7.7. Planned comparisons for N2 amplitude to speech comparing the language-impaired groups to VFA and TD groups found no significant results.

P3a

Oddball Task

There was a significant main effect of sound category (speech vs. nonspeech) for P3a amplitude measured between 300-400 ms ($F(1, 36) = 13.57$, $p = .001$, partial eta squared = .27, generalized eta squared = .16), due to a larger P3a amplitude to speech (EMM = 1.69, SE = .58) than nonspeech (EMM = -1.50, SE = .62) across groups. In addition, there was a significant main effect of group (*F*(3, 36) = 3.61, *p* = .02, partial eta squared = .23, generalized eta squared = .13), resulting from a significantly smaller P3a amplitude in MVA (EMM = -1.78, SE = .82) across both speech and nonspeech compared to TD (EMM = 1.50, SE = .62, *p* = .003, d = .98, g_s = .90). However, there was also a significant interaction between sound category and group (*F*(3, 36) = 2.89, *p* = .049, partial eta squared = .19, generalized eta squared = .11), suggesting that the main effects alone do not accurately represent the results. Simple effects tests on the interaction demonstrated that a significantly larger P3a amplitude to speech over nonspeech was found only in the PSA and VFA groups (PSA: speech EMM = 4.61, SE = 1.62, nonspeech EMM = -2.97, SE = 1.72, *p* = .003, d_z = 1.80, g_{av} = 2.25; VFA: speech EMM = 1.65, SE = .87, nonspeech EMM = -1.99, SE = 0.92, *p* = .008, d_z = .89, g_{av} = 1.11). Larger P3a amplitude to speech also trended toward significance in the TD group (speech EMM = 2.60, SE = .87, nonspeech EMM = .40, SE = .92, *p* = .096, d_z = .42, g_{av} = .72), but little evidence for larger amplitude to speech was found in the MVA group (*p* = .70, d_z = .12, g_{av} = .15). Furthermore, the interaction revealed that the MVA group showed a significantly lower P3a amplitude to speech (EMM = -2.11, SE = 1.15) than all other groups (PSA: *p* = .002, d = 1.78, g_s = 1.50; VFA: *p* = .01, d = 1.10, g_s = 1.01, TD: *p* = .002, d = 1.54, g_s =1.42).

Figure 4.5 displays the difference waves in each group for the speech block, calculated by subtracting the standard speech sound from the deviant speech sound. Similarly, Figure 4.6 displays the difference waves in each group for the nonspeech block, calculated by subtracting the

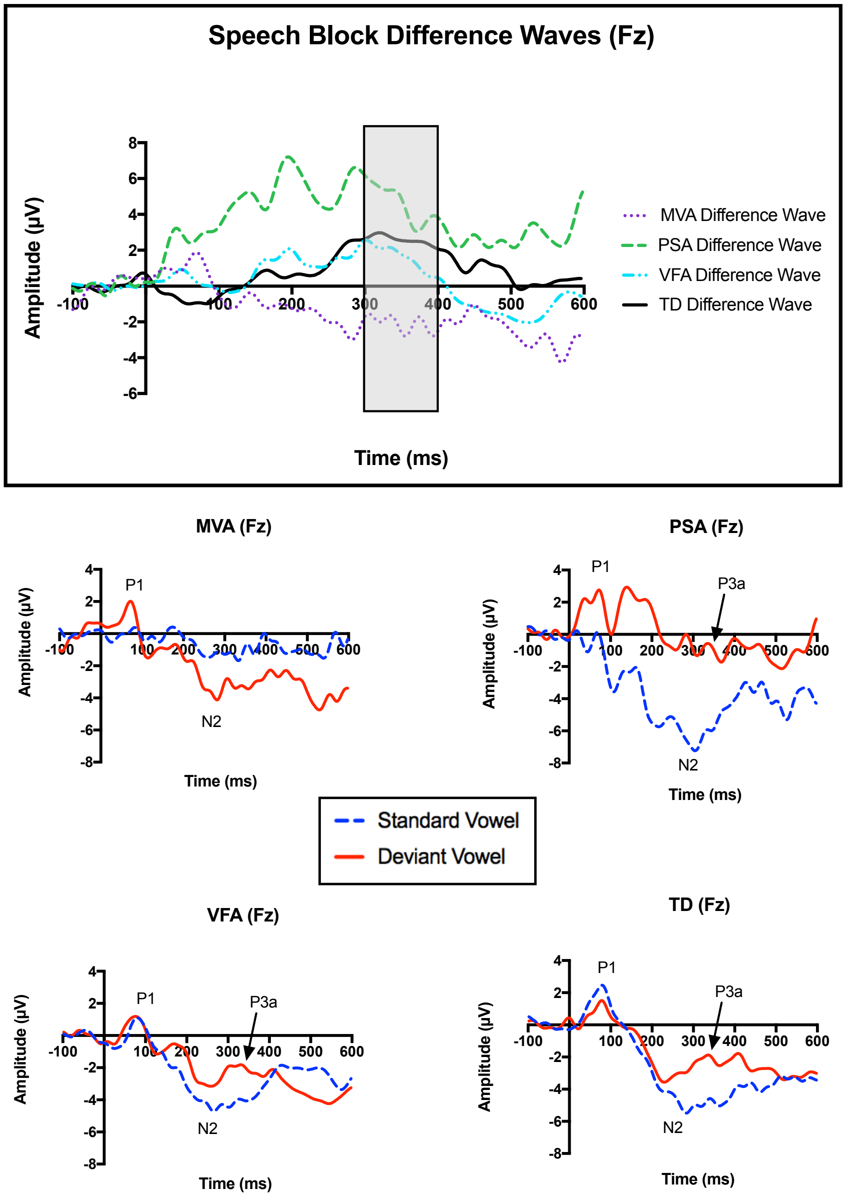


Figure 4.5 P3a ERP Responses to the Speech Deviant in All Groups. Top: Difference waves from the speech block (subtracting the standard speech sound from the deviant speech sound) in each group. The shaded portion represents the P3a time window, measured as the mean amplitude between 300—400 ms. Bottom: ERP responses to the standard and deviant speech sounds within each group. P3a responses to the deviant demonstrated by stronger positivity to the deviant speech sound between 300—400 ms are labeled in the PSA, VFA, and TD groups (no evidence for P3a is found in the MVA group).

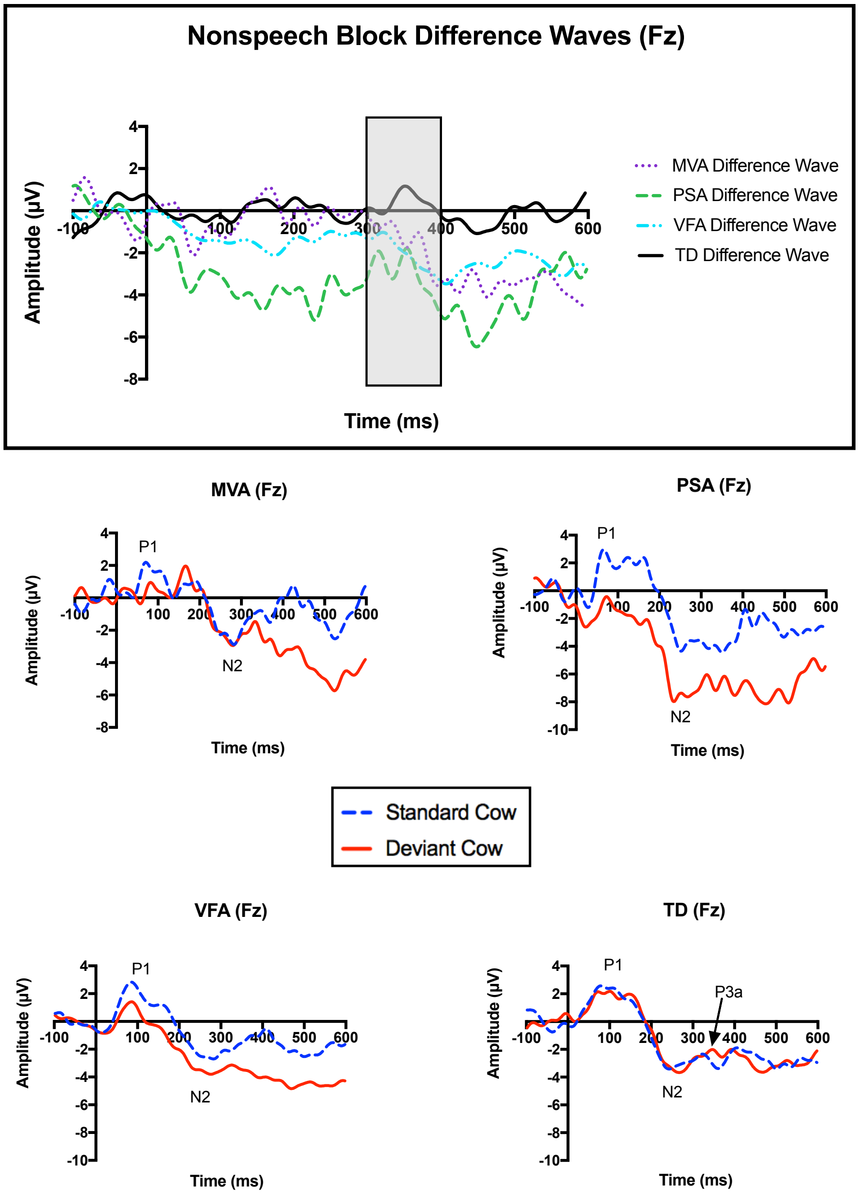


Figure 4.6 P3a ERP Responses to the Nonspeech Deviant in All Groups. Top: Difference waves from the nonspeech block (subtracting the standard nonspeech sound from the deviant nonspeech sound) in each group. The shaded portion represents the P3a time window, measured as the mean amplitude between 300—400 ms. Bottom: ERP responses to the standard and deviant nonspeech sounds within each group. A potential P3a responses to the deviant demonstrated by stronger positivity to the deviant speech sound between 300—400 ms is labeled only in TD group (no evidence for P3a is found in the ASD groups).

standard nonspeech sound from the deviant nonspeech sound. Because difference waves are often misleading (Picton et al., 2000), Figures 4.5 and 4.6 also include the raw ERPs to the standard and deviant sound within each group to better illustrate the P3a effect within each group. It can be seen from Figure 4.6 that the TD group was the only group showing an orienting response to the nonspeech deviant (greater positivity in the 300-400 ms range to the deviant speech), and only the MVA group does not show an orienting response to the speech deviant (no evidence of greater positivity in the 300-400 ms range to the deviant speech sound).

Discussion

In this study, we sought to expand upon previous ERP studies in autism spectrum disorder that have investigated the role of auditory processing and social motivation by 1) subgrouping children on language abilities and 2) using novel ecologically valid natural speech and nonspeech stimuli. We subgrouped children with ASD aged 8—14-years-old into three language subgroups: minimally verbal autism (MVA: single words), phrase speech autism (PSA: three- and four-word phrases), and verbally fluent autism (VFA: fluent sentences), and included a typically developing control group (fluent sentences). Our findings confirmed that children with ASD with different language abilities show distinctive auditory processing (P1, N2) and social motivation/attention (P3a) ERP profiles. Specifically, we propose below that our results suggest that 1) the MVA group shows a “double hit” with deficits in both auditory processing and social motivation, 2) the PSA group shows only abnormal auditory processing, and 3) the VFA group shows no abnormal processing and generally follows the same pattern of responses as the TD group. These findings suggest that subgrouping language abilities in this way highlighted meaningful differences between groups that has likely been washed out in previous studies that have collapsed children with varying language abilities into one sample. However, our finding that no ASD subgroup displayed a P3a component to the nonspeech stimulus differs from most previous studies. This is likely due to our novel use of natural nonspeech stimuli, which illustrates that prior results to synthetic stimuli may not map onto ecologically valid stimuli, and further suggests that children with ASD do not treat all nonspeech stimuli equally.

The results of the P1 and N2 auditory ERPs to the oddball task stimuli support auditory processing impairments in both the MVA and PSA groups, but with different types of processing deficits in each group. Specifically, the MVA group appears to show slowed processing of speech, while the PSA group shows abnormalities in the strength of auditory responses to speech vs. nonspeech. In the MVA group, slower processing was demonstrated by significantly slower N2 latency to speech in the MVA group compared to the TD group, which almost reached significance compared to the VFA group as well (p = .057, similar effect sizes of 1.36 and 1.05 in TD and VFA contrasts, respectively). Interestingly, post hoc comparisons did not reveal evidence for delayed processing of nonspeech in MVA, suggesting it may be a speech-specific deficit, but this was not confirmed by a significant interaction for N2 latency (interaction p = .08). The PSA group did not demonstrate slowed processing, but instead they were the only group to demonstrate significant auditory ERP amplitude differences between speech and nonspeech, with a significantly larger P1 to speech and a significantly larger N2 to nonspeech. Because larger ERP amplitudes result from larger populations of synchronized neurons, this suggests abnormally strong responses for P1 to speech and N2 to nonspeech. Although amplitude differences are often reflective of normal differences arising from neural populations responding to different acoustic features of the stimuli (Antinoro, Skinner, & Jones, 1969; Wunderlich & Cone-Wesson, 2001), the finding that no other group showed this pattern indicates it is more likely representative of abnormal synchronization of cortical auditory neurons in the PSA group to speech vs. nonspeech at different points in time. Notably, the MVA and PSA groups did not show evidence of abnormal auditory processing to the simple pure tone stimulus, suggesting that their auditory processing deficits arise only to complex spectro-temporal sounds.

Measuring P1 latency to speech and nonspeech separately within each group indicated that only the TD group displayed a significantly P1 faster latency to the speech deviant. This suggests that the TD group may prioritize incoming speech sounds over other types of sounds due to the importance of the speech signal, and that children with ASD may not share this neural processing strategy; however, a significant interaction would be needed to make this claim for certain. The sound category x group interaction for N2 latency was marginally significant (p = .08), in which the TD group followed the same pattern of faster latency to speech than nonspeech. Contrary to our hypothesis, the VFA group also displayed faster N2 latency to speech than nonspeech, but given that they did not show this for P1 latency, this finding suggests a slowed recognition of the speech signal that perhaps becomes processed more efficiently than nonspeech during processing in secondary or belt auditory cortex areas.

The P3a results demonstrated that the MVA group displayed a significantly smaller P3a amplitude to speech than every other group, suggesting that these children have speech processing deficits at the level of both early auditory processing and attention. Figure 4.5 further demonstrates that the MVA group was the only group that did not show any P3a response to the deviant speech sound. In addition, the finding that the PSA and VFA groups showed a significantly larger P3a response to speech than nonspeech conflicts with previous studies which have found larger P3a amplitudes to nonspeech in ASD samples (Ceponiene et al., 2003; Gomot et al., 2002b; T. Lepisto et al., 2006; Lepisto et al., 2005a, 2007; Lincoln et al., 1993; Lindstrom et al., 2016; Whitehouse

& Bishop, 2008). Indeed, Figure 4.6 illustrates that the only group to show evidence for a P3a orienting response to nonspeech was the TD group.

The lack of a P3a orienting response to the nonspeech stimulus in the ASD groups conflicts with the majority of previous studies. The TD group showed only weak evidence for a P3a in the nonspeech condition, although this likely contributed to the finding that the P3a was significantly stronger to speech in PSA and VFA groups but did not reach significance in the TD group. It's unclear whether this was because the children in the present study did not discriminate the difference between the two cow sounds and therefore did not orient to it, or if they did discriminate but did not judge it as salient enough to warrant an attention switch. Evidence of discrimination is often measured by the mismatch negativity component (MMN), however, this was not possible in the present study because this component requires upwards of 150 trials which would not have been feasible for the language-impaired ASD subgroups. However, given that we controlled the acoustic difference across speech and nonspeech blocks with a pitch increase of 60%, it is unlikely that the presence of the P3a to speech and absence to nonspeech was due to difficulties in discrimination. The amount of acoustic difference between our study and previous studies also cannot explain why we did not find a P3a to nonspeech in the ASD groups, as previous studies finding a P3a have employed a more difficult discrimination task with only a 10% increase in pitch (Ceponiene et al., 2003).

Instead, the discrepant findings are likely due to the use of natural nonspeech stimuli in our task, and therefore imply that the consensus within the literature that children with autism orient more strongly to nonspeech sounds does not apply to all nonspeech sounds. It may be that children with autism find synthetic nonspeech more salient than natural nonspeech sounds. The findings of (Lortie et al., 2017) suggest that this preference is likely due to the natural quality of rather than greater complexity. They compared two types of nonspeech sounds, natural biological (a finger click and mouth suction sound) and a synthetic sound matched on complexity, and found that children with autism showed larger P3a responses to the synthetic nonspeech sound. These results also align with behavioral findings that children with autism preferred listening to complex synthetic sounds over natural speech (Kuhl et al., 2005). However, Fan and Cheng (2014) also did not find a P3a response to synthetic nonspeech in adults with autism, so future research should aim to disentangle whether P3a responses to synthetic vs. natural nonspeech stimuli may differ in children and adults with ASD. Nonetheless, our results highlight that P3a results from studies using synthetic stimuli may not map onto results using natural stimuli and this has implications for the continued use of synthetic stimuli in ASD event-related potential studies.

Together, the auditory ERP and P3a results suggest a model in which distinct profiles of auditory processing and social motivation characterize language subgroups in ASD. The most severe group, children who are minimally verbal, exhibit delayed auditory processing and decreased social motivation to speech. Slowed processing of speech could contribute to difficulty acquiring full use of language due to the incredibly fast fluctuations in speech, which are on the order of just ~30-50 ms for recognizing segmental features including formant transitions and voice onset time. Cunningham, Nicol, Zecker, and Kraus (2000) found that N2 latency significantly predicted performance on a behavioral auditory processing task in children with learning disabilities, which may be relevant to our MVA findings here. Decreased social motivation to orient to speech sounds could further contribute to language impairment in MVA children by leading to decreased experience with the speech signal, or decreased opportunities for shared interaction that create crucial moments for language learning (Dawson et al., 2004). Furthermore, these auditory processing and social motivation processes could interact to create an unfavorable environment to learn language. There is increasing evidence that sensory perception is impacted by our predictions (Panichello, Cheung, & Bar, 2013; Summerfield & Egner, 2009), and thus decreased experience could lead to diminished predictions about the upcoming speech sounds, and further contribute to auditory sensory processing difficulties. It is important to note that orienting to speech in this group may have started out normal earlier in development and decreased over time as language skills did not progress. Future work could use the oddball task in the present study in a sample of younger minimally verbal children with ASD; however, because language outcomes in minimally verbal ASD toddlers are variable up until the age of five (Thurm et al., 2015; but possibly even later, see: Wodka, Mathy, & Kalb, 2013), it would be ideal to do a longitudinal design to confirm whether these children remain minimally verbal.

In contrast to MVA children, children with ASD who have acquired functional use of phrases but have not acquired fluent sentences (here termed phrase speech autism; PSA) appear to have intact social motivation to attend to speech, but abnormal neural synchrony across populations of neurons reflected by both increased and decreased amplitudes to speech and nonspeech during auditory processing. Recently, Yu, Wang, Huang, Wu, and Zhang (2018) measured inter-trial phase coherence as an index of neural synchrony and found that the amount of phase synchrony in the theta band was a significant predictor of both P1 and N2 amplitudes in children with autism spectrum disorder. This work suggests that the amount of phase synchrony to speech and nonspeech stimuli may be driving these amplitude differences, although it is unclear why phase synchrony would be enhanced to speech in the P1 time range and enhanced to nonspeech in the N2 time range. One possibility is that a stronger P1 to the speech vowel is reflective of increased neural synchrony as a compensatory mechanism to prioritize speech signals, while stronger N2 to the nonspeech cow sound could represent delayed in-depth acoustic processing following the decreased P1 response.

Our finding that the verbally fluent autism group showed similar ERPs to the TD group suggests that their mild language impairments are likely related to later cognitive processes not measured in this study. It also lends support to the need to subgroup children based on language abilities in auditory ERP studies, as in a collapsed design these children that do not differ from TD would contribute to lower power to find true differences in language-impaired children.

One limitation of the present study is the small sample size of the language-impaired ASD groups (MVA n=8, PSA n=4). This likely led to reduced power to find significant group main effects and interactions. However, small sample sizes are direct result of the population of interest; minimally verbal children make up 20-30% of children with autism (Anderson et al., 2007), and the percentage of children with autism with phrase speech level language has not been well quantified. Future studies could aim to pool samples across multiple research sites in order to gain larger sample sizes in each subgroup. Another limitation is that the differences found between groups could be driven by differences in intellectual ability rather than language ability. We did not control for IQ due to the inappropriate use of IQ as a covariate in designs such as ours in which lower IQ is a known characteristic of the sample rather than random sampling error (Dennis et al., 2009; Miller & Chapman, 2001). In order to directly test the impact of IQ, a future study could include an additional control group of children with intellectual disabilities who have acquired language abilities above single words and phrase speech. Finally, the morphological differences in the P1 and N2 grand average waveforms to complex natural stimuli across language subgroups suggest that our ERP analysis may have missed meaningful differences between groups. Specifically, the less well-defined and/or bifurcated peaks found in the MVA and PSA groups could result from increased variability of the neural responses across trials in these children. This could be due to differences in the amount of neural synchrony (reflected in amplitude) across trials, or stable neural synchrony but variations in the timing of the response across trials; either result would lead to a decreased signal-to-noise ratio in the grand average waveform. One way to analyze what led to less well-defined peaks in future work would be to perform single-trial analysis. Several studies using single-trial analysis suggest that children with ASD show increased variability in their single trial ERP responses (Dinstein et al., 2012; Milne, 2011), and it may be that children with ASD and more severe language impairments show increased variability compared to verbally fluent children that contributes to their language deficits.

In conclusion, we found that using ERPs to measure auditory processing and involuntary attention as a proxy for social motivation to ecologically valid stimuli yielded significant differences in neural processing between language subgroups of children with ASD. Our results have implications for designing new treatments for children with ASD based on their auditory processing and social motivation/attentional processing ERP profiles, with minimally verbal children requiring interventions at both the sensory processing and social motivation levels. In contrast, children at the phrase speech level appear to require only auditory processing interventions, and verbally fluent individuals appear to have deficits caused by higher-order cognitive processes not explored in the present study. Given the current lack of successful interventions especially for minimally verbal children with ASD (Brignell et al., 2018), the present study contributes to an urgent need to understand mechanisms causing severe language impairment in these children as well as children with more moderate language impairments.

References

Abrams, D. A., Lynch, C. J., Cheng, K. M., Phillips, J., Supekar, K., Ryali, S., … Menon, V. (2013). Underconnectivity between voice-selective cortex and reward circuitry in children with autism. *Proceedings of the National Academy of Sciences*, *110*(29), 12060–12065.

Alcántara, J. I., Weisblatt, E. J. L., Moore, B. C. J., & Bolton, P. F. (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *Journal of Child Psychology and Psychiatry*, *45*(6), 1107–1114. https://doi.org/10.1111/j.1469-7610.2004.t01-1-00303.x

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., … Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, *75*(4), 594.

Antinoro, F., Skinner, P. H., & Jones, J. J. (1969). Relation between Sound Intensity and Amplitude of the AER at Different Stimulus Frequencies. *The Journal of the Acoustical Society of America*, *46*(6B), 1433–1436. https://doi.org/10.1121/1.1911881

Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of minimally verbal across instruments: evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, n/a-n/a. https://doi.org/10.1111/jcpp.12609

Baltaxe, C. A. M. (1977). Pragmatic Deficits in the Language of Autistic Adolescents. *Journal of Pediatric Psychology*, *2*(4), 176–180. https://doi.org/10.1093/jpepsy/2.4.176

Bardy, F., Van Dun, B., & Dillon, H. (2015). Bigger Is Better: Increasing Cortical Auditory Response Amplitude Via Stimulus Spectral Complexity. *Ear and Hearing*, *36*(6), 677–687. https://doi.org/10.1097/AUD.0000000000000183

Bates, E., Dale, P. S., Fenson, L., Hartung, J. P., Pethick, S., Reilly, J. S., … Thal, D. (2004). *MacArthur Communicative Development Inventory*. Retrieved October 11th.

Bidet-Caulet, A., Latinus, M., Roux, S., Malvy, J., Bonnet-Brilhault, F., & Bruneau, N. (2017). Atypical sound discrimination in children with ASD as indicated by cortical ERPs. *Journal of Neurodevelopmental Disorders*, *9*(1), 13.

Bishop, D. V., Hardiman, M., Uwer, R., & von Suchodoletz, W. (2007). Maturation of the long-latency auditory ERP: step function changes at start and end of adolescence. *Developmental Science*, *10*(5), 565–575. https://doi.org/10.1111/j.1467-7687.2007.00619.x

Bishop, D. V. M., & Adams, C. (1989). Conversational characteristics of children with semantic-pragmatic disorder. II: What features lead to a judgement of inappropriacy? *International*

Journal of Language & Communication Disorders, *24*(3), 241–263. https://doi.org/10.3109/13682828909019890

Boucher, J. (2012). Research Review: Structural language in autistic spectrum disorder – characteristics and causes. *Journal of Child Psychology and Psychiatry*, *53*(3), 219–233. https://doi.org/10.1111/j.1469-7610.2011.02508.x

Bright, K., Greeley, C. O., Eichwald, J., Loveland, C. O., & Tanner, G. (2011). American Academy of Audiology Childhood Hearing Screening Guidelines.

Brignell, A., Chenausky, K. V., Song, H., Zhu, J., Suo, C., & Morgan, A. T. (2018). Communication interventions for autism spectrum disorder in minimally verbal children. *Cochrane Database of Systematic Reviews*, (11). https://doi.org/10.1002/14651858.CD012324.pub2

Brynskov, C., Eigsti, I.-M., Jørgensen, M., Lemcke, S., Bohn, O.-S., & Krøjgaard, P. (2017). Syntax and Morphology in Danish-Speaking Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(2), 373–383. https://doi.org/10.1007/s10803-016-2962-7

Caldenhove, S., Borghans, L. G. J. M., Blokland, A., & Sambeth, A. (2017). Role of acetylcholine and serotonin in novelty processing using an oddball paradigm. *Behavioural Brain Research*, *331*, 199–204. https://doi.org/10.1016/j.bbr.2017.05.031

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism*, *2*(4), 325–344.

Čeponien, R., Rinne, T., & Näätänen, R. (2002). Maturation of cortical sound processing as indexed by event-related potentials. *Clinical Neurophysiology*, *113*(6), 870–882.

Čeponienė, R., Lepistö, T., Shestakova, A., Vanhala, R., Alku, P., Näätänen, R., & Yaguchi, K. (2003). Speech–sound-selective auditory impairment in children with autism: they can perceive but do not attend. *Proceedings of the National Academy of Sciences*, *100*(9), 5567–5572.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239.

Courchesne, E., Lincoln, A. J., Kilman, B. A., & Galambos, R. (1985). Event-related brain potential correlates of the processing of novel visual and auditory information in autism. *Journal of Autism and Developmental Disorders*, *15*(1), 55–76. https://doi.org/10.1007/BF01837899

Cunningham, J., Nicol, T., Zecker, S., & Kraus, and N. (2000). Speech-Evoked Neurophysiologic Responses in Children with Learning Problems: Development and Behavioral Correlates of Perception. *Ear and Hearing*, *21*(6), 554.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*(2), 271.

Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, *15*(3), 331–343. https://doi.org/10.1017/S1355617709090481

Dinstein, I., Heeger, D. J., Lorenzi, L., Minshew, N. J., Malach, R., & Behrmann, M. (2012). Unreliable Evoked Responses in Autism. *Neuron*, *75*(6), 981–991. https://doi.org/10.1016/j.neuron.2012.07.026

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In *Human auditory development* (pp. 61–105). Springer.

Fan, Y.-T., & Cheng, Y. (2014). Atypical Mismatch Negativity in Response to Emotional Voices in People with Autism Spectrum Conditions. *PLOS ONE*, *9*(7), e102471. https://doi.org/10.1371/journal.pone.0102471

Ferri, R., Elia, M., Agarwal, N., Lanuzza, B., Musumeci, S. A., & Pennisi, G. (2003). The mismatch negativity and the P3a components of the auditory event-related potentials in autistic low-functioning subjects. *Clinical Neurophysiology*, *114*(9), 1671–1680.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373.

Gilley, P. M., Sharma, A., Dorman, M., & Martin, K. (2005). Developmental changes in refractoriness of the cortical auditory evoked potential. *Clinical Neurophysiology*, *116*(3), 648–657. https://doi.org/10.1016/j.clinph.2004.09.009

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing Complex Emotion Recognition in Adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *36*(2), 169–183. https://doi.org/10.1007/s10803-005-0057-y

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002a). Hypersensitivity to acoustic change in children with autism: electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(05), 577–584.

Gomot, M., Giard, M.-H., Adrien, J.-L., Barthelemy, C., & Bruneau, N. (2002b). Hypersensitivity to acoustic change in children with autism: electrophysiological evidence of left frontal cortex dysfunctioning. *Psychophysiology*, *39*(05), 577–584.

Gordon, K., Pasco, G., McElduff, F., Wade, A., Howlin, P., & Charman, T. (2011). A communication-based intervention for nonverbal children with autism: What changes?

Who benefits? *Journal of Consulting and Clinical Psychology*, *79*(4), 447–457. https://doi.org/10.1037/a0024379

Groen, W. B., van Orsouw, L., Ter Huurne, N., Swinkels, S., van der Gaag, R.-J., Buitelaar, J. K., & Zwiers, M. P. (2009). Intact spectral but abnormal temporal processing of auditory stimuli in autism. *Journal of Autism and Developmental Disorders*, *39*(5), 742–750.

Groen, W. B., Zwiers, M. P., van der Gaag, R.-J., & Buitelaar, J. K. (2008). The phenotype and neural correlates of language in autism: an integrative review. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1416–1425.

Hale, C. M., & Tager-Flusberg, H. (2005). Brief Report: The Relationship between Discourse Deficits and Autism Symptomatology. *Journal of Autism and Developmental Disorders*, *35*(4), 519–524. https://doi.org/10.1007/s10803-005-5065-4

Jansson-Verkasalo, E., Ceponiene, R., Kielinen, M., Suominen, K., Jäntti, V., Linna, S.-L., … Näätänen, R. (2003). Deficient auditory processing in children with Asperger Syndrome, as indexed by event-related potentials. *Neuroscience Letters*, *338*(3), 197–200. https://doi.org/10.1016/S0304-3940(02)01405-2

Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P., Landa, R., … Almirall, D. (2014). Communication Interventions for Minimally Verbal Children With Autism: A Sequential Multiple Assignment Randomized Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(6), 635–646. https://doi.org/10.1016/j.jaac.2014.01.019

Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J.-L., … Collet, L. (2004). Increased perception of loudness in autism. *Hearing Research*, *198*(1), 87–92.

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced theory of mind in high-functioning adults with autism. *Journal of Autism and Developmental Disorders*, *31*(1), 29–36.

Klin, A. (1991). Young autistic children’s listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42.

Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytky, S., … Moilanen, I. (2007). Processing of affective speech prosody is impaired in Asperger syndrome. *Journal of Autism and Developmental Disorders*, *37*(8), 1539–1549. https://doi.org/10.1007/s10803-006-0271-2

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental Science*, *8*(1), F1–F12.

Kujala, T., Kuuluvainen, S., Saalasti, S., Jansson-Verkasalo, E., Wendt, L. von, & Lepistö, T. (2010). Speech-feature discrimination in children with Asperger syndrome as determined with the multi-feature mismatch negativity paradigm. *Clinical Neurophysiology*, *121*(9), 1410–1419. https://doi.org/10.1016/j.clinph.2010.03.017

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Lepistö, T., Silokallio, S., Nieminen-von Wendt, T., Alku, P., Näätänen, R., & Kujala, T. (2006). Auditory perception and attention as reflected by the brain event-related potentials in children with Asperger syndrome. *Clinical Neurophysiology*, *117*(10), 2161–2171.

Lepistö, Tuulia, Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005a). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1), 147–157.

Lepistö, Tuulia, Kujala, T., Vanhala, R., Alku, P., Huotilainen, M., & Näätänen, R. (2005b). The discrimination of and orienting to speech and non-speech sounds in children with autism. *Brain Research*, *1066*(1), 147–157.

Lepistö, Tuulia, Nieminen-von Wendt, T., von Wendt, L., Näätänen, R., & Kujala, T. (2007). Auditory cortical change detection in adults with Asperger syndrome. *Neuroscience Letters*, *414*(2), 136–140. https://doi.org/10.1016/j.neulet.2006.12.009

Lincoln, A. J., Courchesne, E., Harms, L., & Allen, M. (1993). Contextual probability evaluation in autistic, receptive developmental language disorder, and control children: event-related brain potential evidence. *Journal of Autism and Developmental Disorders*, *23*(1), 37–58.

Lindström, R., Lepistö-Paisley, T., Vanhala, R., Alén, R., & Kujala, T. (2016). Impaired neural discrimination of emotional speech prosody in children with autism spectrum disorder and language impairment. *Neuroscience Letters*. https://doi.org/10.1016/j.neulet.2016.06.016

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). Autism diagnostic observation schedule, (ADOS-2) modules 1-4. *Los Angeles, California: Western Psychological Services*.

Lortie, M., Proulx-Bégin, L., Saint-Amour, D., Cousineau, D., Théoret, H., & Lepage, J.-F. (2017). Brief Report: Biological Sound Processing in Children with Autistic Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(6), 1904–1909. https://doi.org/10.1007/s10803-017-3093-5

Loveland, K. A., McEvoy, R. E., Tunali, B., & Kelley, M. L. (1990). Narrative story telling in autism and Down's syndrome. *British Journal of Developmental Psychology*, *8*(1), 9–23.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press. Retrieved from https://books.google.com/books?hl=en&lr=&id=y4-uAwAAQBAJ&oi=fnd&pg=PR5&dq=An+Introduction+to+the+Event-Related+Potential+Technique&ots=jA36D2RPIz&sig=nOoNKrH1EFySoIxVZBtSlUAD0s8

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48. https://doi.org/10.1037/0021-843X.110.1.40

Milne, E. (2011). Increased Intra-Participant Variability in Children with Autistic Spectrum Disorders: Evidence from Single-Trial Analysis of Evoked EEG. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00051

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced Perceptual Functioning in Autism: An Update, and Eight Principles of Autistic Perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43. https://doi.org/10.1007/s10803-005-0040-7

Mucchetti, C. A. (2013). Adapted shared reading at school for minimally verbal students with autism. *Autism*, *17*(3), 358–372. https://doi.org/10.1177/1362361312470495

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: a review. *Neuroscience & Biobehavioral Reviews*, *36*(2), 836–854.

Paetau, R., Ahonen, A., Salonen, O., & Sams, M. (1995). Auditory evoked magnetic fields to tones and pseudowords in healthy children and adults. *Journal of Clinical Neurophysiology*, *12*(2), 177–185.

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive Feedback and Conscious Visual Experience. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00620

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Pickett, E., Pullara, O., O'Grady, J., & Gordon, B. (2009). Speech acquisition in older nonverbal individuals with autism: a review of features, methods, and prognosis. *Cognitive and Behavioral Neurology*, *22*(1), 1–21.

Picton, T. w., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a., Johnson, R., … Taylor, M. j. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology*, *37*(02), 127–152. https://doi.org/null

Polich, J. (2003). *Theoretical overview of P3a and P3b*. Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-0294-4_5

Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*(2), 172–185. https://doi.org/10.1016/j.ijpsycho.2005.12.012

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology*, *113*(3), 407–420. https://doi.org/10.1016/S1388-2457(01)00733-7

Roid, G. H. (2003). Stanford–Binet Intelligence Scale, Fifth Edition Itasca, IL: Riverside Publishing.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders*, *32*(3), 189–194.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Salmond, C. H., Vargha-Khadem, F., Gadian, D. G., de Haan, M., & Baldeweg, T. (2007). Heterogeneity in the Patterns of Neural Abnormality in Autistic Spectrum Disorders: Evidence from ERP and MRI. *Cortex*, *43*(6), 686–699. https://doi.org/10.1016/S0010-9452(08)70498-2

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and Prosody Characteristics of Adolescents and Adults With High-Functioning Autism and Asperger Syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., & Lawson, K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research*, *236*(1), 61–79. https://doi.org/10.1016/j.heares.2007.12.001

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders*, *11*(1), 45–56.

Thurm, A., Manwaring, S. S., Swineford, L., & Farmer, C. (2015). Longitudinal study of symptom severity and language in minimally verbal children with autism. *Journal of Child Psychology and Psychiatry*, *56*(1), 97–104.

Volden, J., & Lord, C. (1991). Neologisms and idiosyncratic language in autistic speakers. *Journal of Autism and Developmental Disorders*, *21*(2), 109–130. https://doi.org/10.1007/BF02284755

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

Whitehouse, A. J., & Bishop, D. V. (2008). Do children with autism ‘switch off’to speech sounds? An investigation using event-related potentials. *Developmental Science*, *11*(4), 516–524.

Wilkinson, K. M. (1998). Profiles of language and communication skills in autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *4*(2), 73–79.

Wodka, E. L., Mathy, P., & Kalb, L. (2013). Predictors of Phrase and Fluent Speech in Children With Autism and Severe Language Delay. *Pediatrics*, *131*(4), e1128–e1134. https://doi.org/10.1542/peds.2012-2221

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

Yu, L., Wang, S., Huang, D., Wu, X., & Zhang, Y. (2018). Role of inter-trial phase coherence in atypical auditory evoked potentials to speech and nonspeech stimuli in children with autism. *Clinical Neurophysiology*, *129*(7), 1374–1382. https://doi.org/10.1016/j.clinph.2018.04.599

CHAPTER 5

Discussion

Summary of the Work

The goal of the present work was to develop a passive auditory EEG paradigm using ecologically valid stimuli that could be used to parse language subgroups in children with autism from severe to mild language impairments. This goal was motivated in large part by a vast gap in the autism field in which children with the most severe language impairments, minimally verbal autism (MVA), have been largely neglected from research (Tager-Flusberg & Kasari, 2013). This group makes up almost 30% of individuals with autism spectrum disorder (Anderson et al., 2007), and yet a recent review found little evidence for any current successful language interventions in these individuals (Brignell et al., 2018). The primary reason for their exclusion from research is due to noncompliance with research tasks, which necessitates the development of novel passive measures with a focus on paradigms that can provide mechanistic explanations about causes of language impairment. However, despite widespread evidence for a broad array language impairments in ASD (Boucher, 2012; Groen, Zwiers, van der Gaag, & Buitelaar, 2008), research into the mechanistic underpinnings of a variety of language impairments remains scarce within the autism population as a whole. Therefore, the present work describes a series of EEG studies conducted to develop a short passive paradigm designed to test two prominent theories of language impairment in children with autism across varying language abilities: 1) auditory processing theory and 2) social motivation theory.

The auditory processing theory posits that some children with autism may struggle to learn language due to primary deficits at the level of processing the complex acoustical features of speech that leads to downstream difficulties in learning, storing, and comprehending speech. In contrast, the social motivation theory posits that children with autism are not motivated to orient and attend to social stimuli like human speech, and this in turn constrains their ability to learn language. Event-related potentials (ERPs) measured from EEG were used to distinguish early auditory processing from involuntary attentional factors related to social motivation. To test the theory of impaired auditory processing, the integrity of the early auditory ERP components were assessed, reflective of cortical encoding of sound (Eggermont & Moore, 2012; Ponton, Eggermont, Khosla, Kwong, & Don, 2002). To test the theory of decreased social motivation, the P3a component was assessed to a social stimulus (speech) which reflects an "orienting response" related to involuntary attention switching and initial allocation of attention to a salient stimulus (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2003; Squires, Squires, & Hillyard, 1975).

Previous studies in autism that have measured the P1, N2, and P3a ERP components were limited by two primary factors that prevented a mechanistic understanding of language impairments. First, previous studies either did not characterize the language abilities of their samples or included children with a variety of severity levels that obscured the group-average event-related potentials. Second, the majority of these studies did not use complexity-matched ecologically valid natural speech and nonspeech stimuli, and therefore the findings may not be related to deficits that children face when processing language in the real world.

The present work improved upon these limitations by developing a novel task with ecologically valid natural speech and nonspeech stimuli and subgrouping children with autism based on their functional language abilities into three groups (minimally verbal autism, MVA;

phrase speech autism, PSA; verbally fluent autism, VFA). Developing the task required matching a natural nonspeech stimulus to the spectro-temporal complexity of a natural speech stimulus. This was accomplished through the use of cow vocalizations, which have been used as model for human vocal production due their similarities (Alipour & Jaiswal, 2008).

Study 1 describes the first version of the task that created two pairs of a natural speech vowels and cow sounds presented in an oddball paradigm in typically developing children. Importantly, this study also included a synthetic task with commonly used stimuli from previous studies in order to compare results between synthetic and natural ecologically valid stimuli. Yet, this task did not elicit the P3a component to natural or synthetic speech in typically developing children, which motivated the design of Study 2. Two new oddball designs were contrasted in Study 2 in order to investigate the impact of speech vs. nonspeech standard and deviant sounds on the P3a component. Finally, the results of the first two studies were used to select a novel ecologically valid EEG paradigm, and this paradigm was applied to examine differences between language subgroups in children with autism in Study 3.

Main Findings

Study 1 revealed that a direct comparison between synthetic and natural stimuli resulted in faster auditory processing of the natural speech stimuli only, and furthermore established for the first time that auditory processing reflected by ERPs in children matures at a different rate for complex natural stimuli. These results provide evidence against continuing the use of synthetic stimuli in neuroimaging research, while also highlighting that the speech signal is prioritized regardless of cortical auditory maturity. Study 2 compared natural speech vs. nonspeech cow sounds in the standard (frequent) vs. deviant (infrequent) positions in an oddball task and found that both auditory processing and involuntary orienting are directly impacted by the type of standard sound. Specifically, a stream of standard speech sounds appeared to prime the auditory system to “tune in” to encode the acoustic features of any upcoming stimulus faster (P1, N2), as well as orient to it faster or at all (P3a). Overall, the results of Study 1 and Study 2 highlight the salience of the speech signal for typically developing children and demonstrate that speech processing engages an efficient network beginning at the earliest levels of cortical processing, consistent with the results of Dehaene-Lambertz et al. (2005).

Study 3 applied the Congruent Oddball Task from Study 2 to three language subgroups (MVA, PSA, VFA) in autism and a control group of typically developing children. This study demonstrated that the auditory processing ERPs (P1, N2) and involuntary orienting (P3a) show distinct profiles of impairments in children with ASD with varying levels of language impairments, and therefore provide evidence for different mechanistic explanations that lead to differences in functional language abilities. Specifically, minimally verbal ASD children with the most severe language impairment exhibited both delayed auditory processing and decreased social motivation to attend to speech. ASD children who had acquired functional phrases but not fluent sentences (PSA group) exhibited auditory processing deficits but intact social motivation leading to a normal orienting response to speech. Finally, children with ASD who are verbally fluent with no language impairments (VFA group) showed similar ERPs to the typically developing control group, characterized by intact auditory processing of speech signals and intact social motivation to attend to speech.

The present body of work demonstrates the unique nature of speech processing that can be revealed when ecologically valid stimuli are used in both typically developing children and children with autism and varying functional language abilities. There is increasing evidence that the heterogeneity of autism spectrum disorder presentations can be subgrouped using various factors that detect meaningful differences between subgroups (Anderson, Montazeri, & de Bildt, 2015; Green et al., 2015; Lai, Lombardo, Auyeung, Chakrabarti, & Baron-Cohen, 2015; Montgomery et al., 2016; Whitten, Unruh, Shafer, & Bodfish, 2018). The present work represents the first auditory EEG paradigm to validate that language subgroups in ASD do exist and can be meaningfully distinguished by separate profiles of auditory processing and social motivation. Thus, the answer to the overarching question that guided the present work – whether auditory processing or social motivation contribute to language impairments in children with ASD – is that it depends on the subgroup one is studying. These findings provide a crucial first step in understanding the mechanisms that lead to vastly different language outcomes, and paves the way for novel interventions that target the unique mechanism associated with each language subgroup.

Limitations

The primary limitation of the present work is the small sample sizes of the ASD language subgroups, with only n = 8 and n = 4 in the MVA and PSA subgroups, respectively. Unfortunately, this is an inherent problem in studies like ours because these populations represent a smaller percentage of children with ASD. To help overcome this limitation in the future, research groups could work together across sites to pool samples to gain larger sample sizes in each subgroup. Another limitation is that the ERP differences found between language subgroups could be driven by differences in intellectual ability rather than language ability, given that our groups also differed on IQ. This may be improved by the use of an IQ measure that allows for the calculation of a separate nonverbal IQ, as our IQ results showed that the MVA and PSA subgroups were similar on nonverbal raw scores. However, in order to directly test the impact of IQ, a future study could also include an additional control group of children with intellectual disabilities who have acquired language abilities above single words and phrase speech.

One additional limitation of the present study is that we were unable to confirm normal hearing in many of the children in the MVA and PSA language subgroups, which may have contributed to the observed auditory processing deficits in these children. Unfortunately, conducting and interpreting hearing screens in children with autism remains challenging, as even verbally fluent children with ASD can show severe sensory sensitivities that often interfere with testing. Although the pure tone audiometry screener used in the present study is considered the gold standard for testing hearing, it requires compliance from the participant and thus the results of this screen are likely less valid for ASD subgroups in this study with more severe impairments (PSA and MVA). Other measures that do not require a response like tympanograms and otoacoustic emissions can evaluate the function of the cochlea and middle ear systems (Hall & Swanepoel, 2009), however, these measures can also be very difficult to obtain clean results in children with autism because they require input into the ear canal and require the child to sit incredibly still. Furthermore, without a full behavioral screen, neither of these measures can confirm normal hearing on their own; they can only suggest that the function of the tympanic membrane and cochlea are within normal limits, but not that auditory signals are successfully utilized within the brain to substantiate normal hearing.

Future Directions

Study 1 established that cortical auditory processing of natural speech is more efficient than nonspeech in typically developing 8—11-year-olds, a finding that was also replicated in Study 3 with typically developing 8—14-year-olds. These results confirm that speech is prioritized in early auditory processing by age eight, but it remains an open question at what age this process is complete and how this processing advantage develops. Event-related potentials have been studied extensively in infants (Haan, 2013), and therefore a direct extension of the present findings would be to expand the age range down to assess natural speech vs. nonspeech efficiency in infants as young as a few days old. We hypothesize that experience listening to and encoding the specific parameters of the human voice interacts with intrinsic social motivation that leads to the development of efficient speech processing mechanisms within the first year of life. Study 3 suggested that children with autism and more severe language impairments (MVA, PSA subgroups) may not develop faster processing of speech compared to nonspeech. This would also be an interesting finding to follow up on to confirm whether this is the case, and whether it may be related to decreased social motivation early in life.

A second future direction would be to extend the present paradigm to younger children with autism to determine whether the distinct auditory processing and social motivation profiles found between MVA, PSA, and VFA subgroups reflect the cause of language impairments or the consequence. In other words, the observed processing deficits in MVA and PSA children may reflect causal mechanisms or they could arise as a result of the language impairment itself over time. For example, MVA children showing a lack of social motivation to attend to speech in later childhood could result instead from auditory processing deficits with the speech signal. This could then lead to abnormal attention later in life as speech becomes a less relevant signal to attend to over time. Similarly, PSA children who showed normal social motivation to attend to speech may have started out by not attending to speech early on, but then developed top-down orienting responses as a compensatory mechanism. The older age range used in the present study was crucial to our design as the first study to examine differences in language subgroups. This is because we had to ensure that the language subgroups represented a stable phenotype. Indeed, it has been found that language trajectories are highly variable in children with autism, with 70% of children who are minimally verbal as toddlers achieving phrase level speech at or after age four (Wodka, Mathy, & Kalb, 2013). Therefore, children cannot be reliably subgrouped into minimally verbal, phrase speech, and verbally fluent speech subgroups in younger children. However, this offers an opportunity for the present paradigm to be used in younger children when subgroups may be able to be created on the basis of different ERP profiles. Future studies could thus assess whether the same profiles of auditory processing and involuntary attention observed here are found in younger children, and whether they could potentially predict differential language outcomes over time.

Another future direction relates to the less well-defined and/or bifurcated peaks found in the MVA and PSA groups in Study 3, which could result from increased variability of the neural responses across trials in these children. The use of single-trial ERP analysis in a future study (ideally with more trials than the present study) would be a logical next step to investigate individual variability in the MVA and PSA groups. Several studies using single-trial analysis suggest that children with ASD show increased variability in their single trial ERP responses (Dinstein et al., 2012; Milne, 2011). It may be that children with ASD and more severe language impairments show increased variability compared to verbally fluent children, especially in light of studies that have not found greater neural variability in high-functioning (indicating verbally fluent) samples (Coskun et al., 2009). Greater variability across single trials may be consistent with the neural noise hypothesis in autism, which suggests that much of autism symptomology may arise from abnormal widespread neural processing that is characterized by increased noise within and across networks (Rubenstein & Merzenich, 2003; Simmons et al., 2007).

Finally, we suggest two additional EEG analysis methods that could be applied to further elucidate the neural processing differences between language subgroups in ASD. Time-frequency analysis could be applied to assess whether neural oscillations do not entrain as well to speech in MVA and PSA subgroups compared to VFA children. There is evidence that this is the case in some children with autism (Jochaut et al., 2015), and enhanced entrainment has been associated with increased comprehension of speech within typical development (Peelle, Gross, & Davis, 2013). In addition, there is increasing evidence that the "state" the brain is in contributes to how an upcoming stimulus is perceived (Barry, de Pascalis, Hodder, Clarke, & Johnstone, 2003; Britz, Díaz Hernàndez, Ro, & Michel, 2014). Therefore, EEG microstate analysis could be applied to measure the pre-stimulus brain states in ASD language subgroups to determine whether this has an impact on auditory processing and orienting responses in language subgroups in autism.

Conclusion

The implications of the results of the three studies in this dissertation can be summarized by three interrelated themes: 1) the human-generated natural speech signal is a unique and highly salient sound for typically developing children, with complex interactions between sensory and attentional neural processing supporting a highly efficient speech processing network; 2) natural speech processing in children with autism shows a graded effect, with the most severely language-impaired children showing deficits in both auditory processing and social motivation, moderately language-impaired children showing deficits only in auditory processing, and mildly language-impaired children showing similar responses as a typical control group; and 3) the use of natural ecologically valid stimuli is a requirement in order to demonstrate these effects. The significance of this work is indicated in both basic and applied research, as the resulting novel naturalistic oddball task can be applied to answer questions in typical development and a variety of language-related clinical disorders. If the present results are independently replicated, then the ecologically valid EEG paradigm developed here may be capable of identifying distinct mechanisms that contribute to the diverse language outcomes that exist in autism. A more nuanced mechanistic understanding of language impairment in autism is likely to contribute to efforts to develop more focused and effective interventions. Thus, the EEG paradigm developed here may have the potential to guide intervention research for the subset of children with autism with the most severe deficits, who arguably have the greatest need for effective interventions to improve their language outcomes.

References

Alipour, F., & Jaiswal, S. (2008). Phonatory characteristics of excised pig, sheep, and cow larynges. *The Journal of the Acoustical Society of America*, *123*(6), 4572–4581. https://doi.org/10.1121/1.2908289

Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., … Pickles, A. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *Journal of Consulting and Clinical Psychology*, *75*(4), 594.

Anderson, G. M., Montazeri, F., & de Bildt, A. (2015). Network Approach to Autistic Traits: Group and Subgroup Analyses of ADOS Item Scores. *Journal of Autism and Developmental Disorders*, *45*(10), 3115–3132.

Barry, R. J., de Pascalis, V., Hodder, D., Clarke, A. R., & Johnstone, S. J. (2003). Preferred EEG brain states at stimulus onset in a fixed interstimulus interval auditory oddball task, and their effects on ERP components. *International Journal of Psychophysiology*, *47*(3), 187–198. https://doi.org/10.1016/S0167-8760(02)00151-4

Boucher, J. (2012). Research Review: Structural language in autistic spectrum disorder – characteristics and causes. *Journal of Child Psychology and Psychiatry*, *53*(3), 219–233. https://doi.org/10.1111/j.1469-7610.2011.02508.x

Brignell, A., Chenausky, K. V., Song, H., Zhu, J., Suo, C., & Morgan, A. T. (2018). Communication interventions for autism spectrum disorder in minimally verbal children. *Cochrane Database of Systematic Reviews*, (11). https://doi.org/10.1002/14651858.CD012324.pub2

Britz, J., Díaz Hernàndez, L., Ro, T., & Michel, C. M. (2014). EEG-microstate dependent emergence of perceptual awareness. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00163

Coskun, M. A., Varghese, L., Reddoch, S., Castillo, E. M., Pearson, D. A., Loveland, K. A., … Sheth, B. R. (2009). Increased response variability in autistic brains? *NeuroReport*, *20*(17), 1543. https://doi.org/10.1097/WNR.0b013e32833246b5

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*(1), 21–33. https://doi.org/10.1016/j.neuroimage.2004.09.039

Dinstein, I., Heeger, D. J., Lorenzi, L., Minshew, N. J., Malach, R., & Behrmann, M. (2012). Unreliable Evoked Responses in Autism. *Neuron*, *75*(6), 981–991. https://doi.org/10.1016/j.neuron.2012.07.026

DiStefano, C., Dickinson, A., Baker, E., & Spurling Jeste, S. (2019). EEG data collection in children with ASD: The role of state in data quality and spectral power. *Research in Autism Spectrum Disorders*, *57*, 132–144. https://doi.org/10.1016/j.rasd.2018.10.001

Eggermont, J. J., & Moore, J. K. (2012). Morphological and functional development of the auditory nervous system. In *Human auditory development* (pp. 61–105). Springer.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, *25*(4), 355–373.

Green, S. A., Hernandez, L., Tottenham, N., Krasileva, K., Bookheimer, S. Y., & Dapretto, M. (2015). Neurobiology of Sensory Overresponsivity in Youth With Autism Spectrum Disorders. *JAMA Psychiatry*, *72*(8), 778–786. https://doi.org/10.1001/jamapsychiatry.2015.0737

Groen, W. B., Zwiers, M. P., van der Gaag, R.-J., & Buitelaar, J. K. (2008). The phenotype and neural correlates of language in autism: an integrative review. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1416–1425.

Haan, M. de. (2013). *Infant EEG and Event-Related Potentials*. Psychology Press. https://doi.org/10.4324/9780203759660

Hall, J. W., & Swanepoel, D. W. (2009). *Objective Assessment of Hearing*. Plural Publishing.

Jochaut, D., Lehongre, K., Saitovitch, A., Devauchelle, A.-D., Olasagasti, I., Chabane, N., … Giraud, A.-L. (2015). Atypical coordination of cortical oscillations in response to speech in autism. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00171

Lai, M.-C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/Gender Differences and Autism: Setting the Scene for Future Research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 11–24. https://doi.org/10.1016/j.jaac.2014.10.003

Milne, E. (2011). Increased Intra-Participant Variability in Children with Autistic Spectrum Disorders: Evidence from Single-Trial Analysis of Evoked EEG. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00051

Montgomery, C. B., Allison, C., Lai, M.-C., Cassidy, S., Langdon, P. E., & Baron-Cohen, S. (2016). Do Adults with High Functioning Autism or Asperger Syndrome Differ in Empathy and Emotion Recognition? *Journal of Autism and Developmental Disorders*, *46*(6), 1931–1940. https://doi.org/10.1007/s10803-016-2698-4

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Polich, J. (2003). *Theoretical overview of P3a and P3b*. Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-0294-4_5

Ponton, C., Eggermont, J. J., Khosla, D., Kwong, B., & Don, M. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clinical Neurophysiology*, *113*(3), 407–420. https://doi.org/10.1016/S1388-2457(01)00733-7

Rubenstein, J. L. R., & Merzenich, M. M. (2003). Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes, Brain, and Behavior*, *2*(5), 255–267.

Simmons, D. R., McKay, L., McAleer, P., Toal, E., Robertson, A., & Pollick, F. E. (2007). Neural noise and autism spectrum disorders. *Perception*, *36*, 119–120.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally Verbal School-Aged Children with Autism Spectrum Disorder: The Neglected End of the Spectrum. *Autism Research*, *6*(6), 468–478. https://doi.org/10.1002/aur.1329

Whitten, A., Unruh, K. E., Shafer, R. L., & Bodfish, J. W. (2018). Subgrouping Autism Based on Symptom Severity Leads to Differences in the Degree of Convergence Between Core Feature Domains. *Journal of Autism and Developmental Disorders*, *48*(6), 1908–1919. https://doi.org/10.1007/s10803-017-3451-3

Wodka, E. L., Mathy, P., & Kalb, L. (2013). Predictors of Phrase and Fluent Speech in Children With Autism and Severe Language Delay. *Pediatrics*, *131*(4), e1128–e1134. https://doi.org/10.1542/peds.2012-2221